REVSTAT – Statistical Journal
Volume 14, Number 1, February 2016, 49–73

HIERARCHICAL DYNAMIC BETA MODEL

Authors: Cibele Queiroz Da-Silva
– Departmento de Estatística, Universidade de Brasília, Brasília, Brazil
cibeleqs@unb.br

Helio S. Migon
– Departmento de Estatística, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
migon@im.ufrj.br

Received: August 2013 Revised: August 2014 Accepted: September 2014

Abstract:

- We develop a hierarchical dynamic Bayesian beta model for modelling 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 no only over time but also across the series. Our methodology was applied to both real and simulated data. The real data sets used in this article include 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 March 2012. A new parametrization of the precision parameter makes 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 a more meaningful interpretation in terms of the magnitude of the scale.

Key-Words:

- *dynamic models; beta distribution; hierarchical models; Bayesian analysis.*

AMS Subject Classification:

- 91B84, 62M10.

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 situations 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 GLM's when both mean and dispersion are to be modeled, 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 the 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 modelling 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 whose range is described by the standard unit interval (0,1). The proposed frequentist βARMA 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 modelling 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.

Even though it is possible to individually fit time series that share common features, gains are obtained when those series are analyzed jointly (Gamerman and Migon, 1993). Naturally, by disregarding existing common features shared by a given set of time series (e.g. trends, seasonal behavior, etc) one could end up with poorer analyses and forecasts.

We also would like to stress the fact that Cargnoni *et al.* (1997) and Gamerman and Migon (1993) do not deal with the situation of fitting the *dispersion process*, a feature that we introduce in our present model formulation. Thus, in this paper we address the issues of formulating a hierarchical dynamic beta model that allows dealing with a set of related time series, each one following related beta models that may present time-evolving mean and precision parameters.

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.

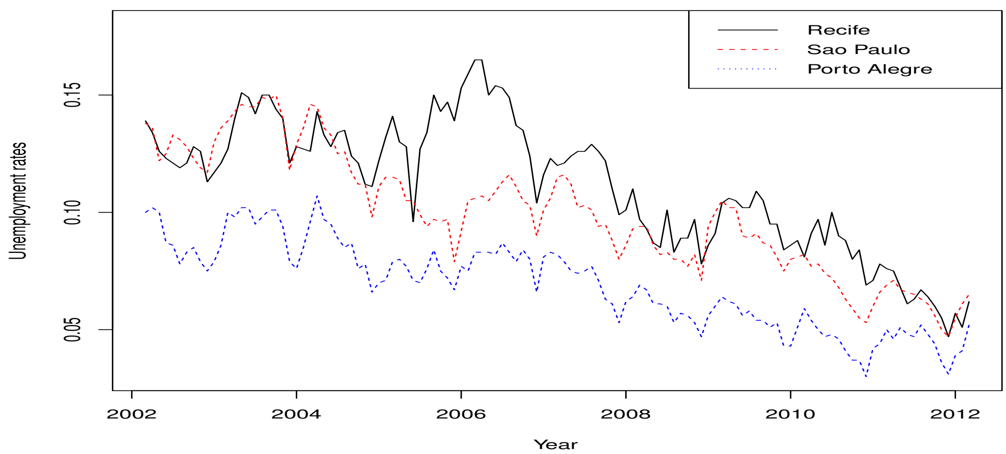


Figure 1: Observed unemployment rates in the cities of Recife, São Paulo and Porto Alegre — Brazil.

This article is organized as follows. In Section 2 we introduce the hierarchical dynamic beta model. 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 modelling a set of I time series of rates or proportions, y_{it}, $i = 1, ..., I$, which share certain characteristics which allows us to treat them in the class of the hierarchical models.

Da-Silva *et al.* (2011) 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

$$p\big(y_{it} \mid \mu_{it}, \zeta_{it}\big) \;=\; \frac{\Gamma(\zeta_{it})}{\Gamma(\zeta_{it}\,\mu_{it})\,\Gamma\big(\zeta_{it}(1-\mu_{it})\big)}\; y_{it}^{\zeta_{it}\mu_{it}-1}(1-y_{it})^{\zeta_{it}(1-\mu_{it})-1}\,, \tag{2.1}$$

and we have $E(y_{it} \mid \mu_{it}, \zeta_{it}) = \mu_{it}$ and $V(y_{it} \mid \mu_{it}, \zeta_{it}) = \mu_{it}(1-\mu_{it})/(1+\zeta_{it})$, with $0 \le \mu_{it} \le 1$ and $\zeta_{it} > 0$, $t = 1, ..., N$ and $i = 1, ..., I$.

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 (2.1), let $\phi_{it} = \frac{1}{1+\zeta_{it}}$ so that $\zeta_{it} = \frac{1-\phi_{it}}{\phi_{it}}$. Thus, $0 < \phi_{it} < 1$, and the *observation equation* of the model is now written as

Observation equation: Let

$$p\big(y_{it} \mid \mu_{it}, \phi_{it}\big) \;=\; \frac{y_{it}^{\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1}\,(1-y_{it})^{(1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)-1}}{B\left(\mu_{it}\left(\frac{1-\phi_{it}}{\phi_{it}}\right),\,(1-\mu_{it})\left(\frac{1-\phi_{it}}{\phi_{it}}\right)\right)}\;, \tag{2.2}$$

with $i = 1, ..., I$, $t = 1, ..., N$ and $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$, be the *observation equation* of the dynamic model. Let $y = (y_1, ..., y_N)$ with $y_t = (y_{1t}, ..., y_{It})'$, $t = 1, ..., N$.

The model structure is such that we have I time series in study, in which $(y_{it} \mid \mu_{it}, \phi_{it})$ is independent of $(y_{jt} \mid \mu_{jt}, \phi_{jt})$ for $i \ne j$. Equation (2.2) incorporates heterogeneity in the precision parameter that may occur both over time or across the series.

Other components which are essential in the description of our hierarchical dynamic beta model include

(**i**) the definition of real transformations applied to μ_{it} and ϕ_{it}, allowing the use of some simplifying Gaussian properties;

(ii) the description of **structural equations** represented in terms of linear models relating the transformed parameters and the latent states and

(iii) the representation of the **system equation** of the dynamic model in which the state parameters are related to surrogate observation equations described by the structural equations.

In order to describe the structural equation, two link function, $h_1(\cdot)$ and $h_2(\cdot)$, associated to, respectively, the *mean process* and the *dispersion process*, should be defined. These are real valued transformations and are useful in the model construction since some of the nice properties of the Gaussian dynamic linear models (DLM's) follow from that.

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. Let $\eta_t = (\eta_{1t}, ..., \eta_{It})'$, i.e., η_t is a $2I\times 1$ vector of *structural parameters* for all the I series at time t with $\eta_{it} = (\eta_{1it}, \eta_{2it})'$, thus $\eta = (\eta_1, ..., \eta_N)$.

Now, y_{it} is parametrized by η_{it}, i.e., $(y_{it}\,|\,\eta_{it}) \sim \text{Beta}(y_{it}\,|\,\eta_{it})$.

Structural equations: Let

$$\eta_t = F_t\theta_t + v_t\ , \qquad v_t \sim N(\mathbf{0}, V)\ , \tag{2.3}$$

with $\eta_{it} = F_{it}\theta_t + v_{it}$ be the *structural equation* in our model formulation. The error term v_{it} in the structural equation is assumed to follow a Gaussian distribution with zero mean vector and covariance matrix V_i, i.e., $v_{it} \sim N(\mathbf{0}, V_i)$, with $t = 1, ..., N$ and $i \in \{1, ..., I\}$.

In equation (2.3) the term θ_t, representing the *state parameter* of the dynamic model at time t, is a real valued s-dimensional vector of latent states. Besides, $F_t = (F_{1t}, ..., F_{It})'$ is the $2I\times s$ design matrix for all the I series at time t, $v_t = (v_{1t}, ..., v_{It})'$ is the $2I\times 1$ vector of errors for the structural equations and $V = \text{block-diag}(V_1, ..., V_I)$ is a $(2I\times 2I)$ block diagonal matrix.

System Equation: Let

$$\theta_t = H_t\theta_{t-1} + w_t\ , \qquad w_t \sim (\mathbf{0}, W)\ , \tag{2.4}$$

with $t = 1, ..., N$, be the *system equation* of the dynamic model.

The error term w_t in the system equation is assumed to follow a Gaussian distribution with zero mean vector and covariance matrix W, i.e., $w_t \sim N(\mathbf{0}, W)$, with $t = 1, ..., N$. Additionally, we assume that the error terms w_t and v_{it} are all mutually independent.

The s-dimensional covariance matrix W (for the s-dimensional vector of latent states, θ_t), is assumed to be block-diagonal including k blocks, with $k \leq s$.

Those blocks are associated to the effects included in the latent states. Thus, $W = \text{block-diag}(W_1, ..., W_k)$. The matrix H_t is a specified $s \times s$ state evolution matrix.

The hierarchical dynamic beta model (HDBM) requires the specification of a $(2I \times 2I)$ covariance matrix V in the structural equations and another covariance matrix W for the state vector. That might become complicated for large matrix dimensions. In many applications it may be sufficient to model simpler dependences, in particular to allow individual random effects. That is why in our proposed model both V and W are block-diagonal matrices.

Notice that equations (2.3) and (2.4) 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 a substantial simplification in the posterior computations of the parameters η and θ, as described in Cargnoni *et al.* (1997).

3. MODELLING THE LATENT COMPONENTS OF THE HDBM

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, $i = 1, ..., I$ and $t = 1, ..., N$. We take the logit transformation of both μ_{it} and ϕ_{it} and, to η_{1it} and η_{2it}, 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:

$$\begin{aligned} \eta_{1it} &= \log\left(\frac{\mu_{it}}{1-\mu_{it}}\right) = F_{i1t}\theta_t + v_{i1t} , \qquad v_{i1t} \sim N(0, V_{i1}) , \\ \eta_{2it} &= \log\left(\frac{\phi_{it}}{1-\phi_{it}}\right) = F_{i2t}\theta_t + v_{i2t} , \qquad v_{i2t} \sim N(0, V_{i2}) , \end{aligned} \tag{3.1}$$

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

In equation (3.1) 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

$$\eta_{1it} = \beta_t + \lambda_{t0} + \gamma_{it} + v_{i1t} . \tag{3.2}$$

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

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

$$\begin{aligned} \beta_t &= \beta_{t-1} + \delta_{t-1} + w_{\beta_t} , \\ \delta_t &= \delta_{t-1} + w_{\delta_t} , \end{aligned} \tag{3.3}$$

Free-form Seasonal effects:

$$\begin{aligned} \lambda_{tr} &= \lambda_{t-1,r+1} + w_{tr} \ , \qquad r = 0, ..., p-2 \ , \\ \lambda_{t,p-1} &= \lambda_{t-1,0} + w_{t,p-1} \ , \end{aligned} \tag{3.4}$$

First-order polynomial effects for the offset term:

$$\gamma_{it} = \gamma_{i,t-1} + w_{\gamma_{it}} \ , \tag{3.5}$$

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 the size of the seasonal cycle as p.
- γ_{it} is an offset parameter representing deviations of the observed rate in region i at time t with respect to the average β_t;
- v_{i1t} represents the region i series-specific stochastic deviation.

In equation (3.1) 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

$$\eta_{2it} = \psi_t + \alpha_{it} + v_{i2t} \ . \tag{3.6}$$

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

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

$$\begin{aligned} \psi_t &= \psi_{t-1} + \xi_{t-1} + w_{\psi_t} \ , \\ \xi_t &= \xi_{t-1} + w_{\xi_t} \ , \end{aligned} \tag{3.7}$$

First-order polynomial effects for the offset term:

$$\alpha_{it} = \alpha_{i,t-1} + w_{\alpha_{it}} \ , \tag{3.8}$$

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 observed rate in region i at time t with respect to the average ψ_t;
- v_{i2t} represents the region i series-specific stochastic deviation.

Identifiability restrictions:

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

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

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

That is,

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

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

The state vector θ_t for generic-sized p cycles is represented by

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

Consider the following design matrices:

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

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 (3.3). Such model is called a *linear growth model* and it includes a time-varying slope 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 $\theta_t = J\theta_{t-1} + (w_{\beta_t}, w_{\delta_t})'$.

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 $h = 1, ..., p$, 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 quarterly data y_t, $t = 1, 2, ...$, so that when y_{t-1} refers to the first quarter of the year, y_t refers to the second one.

Due to 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 to formulate the desired quarterly seasonal pattern.

Considering $I = 3$ sub-populations or regions, the design matrices associated to the hierarchical beta dynamic model given by expressions (2.3), (2.4) and (3.1) to (3.8) are given by

$$H = \text{block-diag}\big(J, P, J, \mathbf{I}_2, \mathbf{I}_2\big) , \tag{3.9}$$

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

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

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

The incorporation of seasonal components in the model can also be done by using the Fourier Representation Theorem (see Pole, West and Harrison, 1994, pp. 49) in which any cyclical function of period p defined by a set of p effects $\psi_1, ..., \psi_p$, can be expressed as a linear combination of sine and cosine terms. Let $\omega = 2\pi/p$, then there exist $(p-1)$ real numbers $a_1, ..., a_h$; $b_1, ..., b_{h-1}$ such that, for $j = 1, ..., p$,

$$\psi_j = a_h \cos(\pi j) + \sum_{r=1}^{h-1} \big[a_r \cos(\omega r j) + b_r \sin(\omega r j)\big] , \tag{3.10}$$

where $p = 2h$ if p is even, and $p = 2h-1$ with $a_h = 0$ if p is odd. The Fourier coefficients a_r and b_r are known quantities and we usually set $a_h = 0$. Thus equation (3.10) can be written as $\psi_j = \sum_{r=1}^{h} S_r(j)$, where

$$S_r(j) = a_r \cos(\omega r j) + b_r \sin(\omega r j) = A_r \cos(\omega r j + \gamma_r) ,$$
$$A_r = (a_r^2 + b_r^2)^{1/2} \quad \text{and} \quad \gamma_r = \arctan(-b_r/a_r) .$$

The terms $S_r(j)$ is called the r-th harmonic. A_r, ωr and γ_r describe, respectively, the *amplitude*, the *frequency* and the *phase* of $S_r(j)$.

For seasonal cycles of even size p (say quarters), we replace matrix P by G where $G = \text{block-diag}\big(J_2(1,\omega), J_2(1,2\omega), ..., J_2(1,(p/2-1)), -1\big)$, with $G^p = G$ and

$$J_2(1,\omega) = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}.$$

For a second-order polynomial trend two harmonic model, the design matrices are given by

$$G = \text{block-diag}\big(J_2(1,\omega), J_2(1,2\omega)\big) ,$$

$$H = \text{block-diag}\big(J, G, J, \mathbf{I}_2, \mathbf{I}_2\big) , \tag{3.11}$$

$$F_1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

$$F_2 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$F_3 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & -1 & -1 \end{pmatrix}.$$

3.1. 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 $i = 1, ..., I$ and $t = 1, ..., N$.

(2) For $i = 1, ..., I$ and $t = 1, ..., N$ we simulate n (say, $n = 1{,}000$) samples from a beta distribution $\text{Beta}\Big(\mu_{it}\Big(\frac{1-\phi_{it}}{\phi_{it}}\Big),\, (1-\mu_{it})\,\Big(\frac{1-\phi_{it}}{\phi_{it}}\Big)\Big)$ and then we take the average value of those draws.

(3) For the confidence bands we repeat steps (1) and (2) for calculating the 2.5% and 97.5% percentiles of the posterior distribution of η_{it}.

The k-step-ahead forecasts for the states are obtained by the repeated application of the system equation (see expression (2.3)), that is,

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

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,

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

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 $R_t(0) = C_t$. Therefore the "future" θ_t values are obtained by successively sampling from the system equation followed by the evaluation of the structural equation (see expression (2.4)). The forecast rates are then obtained by running steps (1) to (3) given above.

4. BAYESIAN ANALYSIS

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

It is more convenient to work with the precision matrices instead of with the covariances matrices. Let $\Phi_{0i} = V_i^{-1}$, $i = 1, ..., I$, $\Phi_l = W_l^{-1}$, $l = 1, ..., k$, $\Phi_0 = \text{block-diag}(\Phi_{01}, ..., \Phi_{0I})$ and $\Phi = \text{block-diag}(\Phi_1, ..., \Phi_k)$. Suppose that Φ_{0i}, $i = 1, ..., I$, follow independent Wishart distributions such that $\Phi_{0i} \sim W(\nu_{0i}, S_{0i})$, where S_{0i} is a symmetric positive definite matrix of dimensions $p_i \times p_i$. Similarly, $\Phi_l \sim W(\varsigma_l, Z_l)$, follows independent prior distributions for $l = 1, ..., k$, where Z_l is a symmetric positive definite matrix of dimensions $q_l \times q_l$.

The joint posterior distribution is given by

$$\begin{aligned} p\big(\eta, \theta, \Phi_0, \Phi \mid y\big) \propto & \left[\prod_{t=1}^{N} \left(\prod_{i=1}^{I} \text{Beta}(y_{it} \,|\, \eta_{it})\, N\big(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}\big)\right) N\big(\theta_t; H_t\theta_{t-1}, \Phi^{-1}\big)\right] \\ & \times N(\theta_0; m_0, C_0) \prod_{i=1}^{I} W(\Phi_{0i}; \nu_{0i}, S_{0i}) \prod_{l=1}^{k} W(\Phi_l; \varsigma_l, Z_l) \ . \end{aligned} \tag{4.1}$$

The Markov chain Monte Carlo (MCMC) procedure used for the inferential processes involves sampling from the full conditional posteriors $p(\eta \mid \theta, \Phi_0, \Phi, y)$, $p(\theta \mid \eta, \Phi_0, \Phi, y)$ and $p(\Phi_0, \Phi \mid \eta, \theta, y)$.

4.1. Sampling from $p\big(\theta \mid \eta, \Phi_0, \Phi, y\big)$

As mentioned before, the equations (2.3) and (2.4) 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

$$
\begin{aligned}
p\big(\theta \mid \eta, \Phi_0, \Phi\big) &= p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_t \mid \theta_{t+1}, ..., \theta_N, \eta, \Phi_0, \Phi\big) \\
&= p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_t \mid \theta_{t+1}, \eta, \Phi_0, \Phi\big) \\
&\propto p\big(\theta_N \mid \eta, \Phi_0, \Phi\big) \prod_{t=0}^{N-1} p\big(\theta_{t+1} \mid \theta_t, \eta, \Phi_0, \Phi\big)\, p\big(\theta_t \mid \eta, \Phi_0, \Phi\big)\ . \qquad (4.2)
\end{aligned}
$$

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 allows us to draw a sample $\theta_N, \theta_{N-1}, ..., \theta_0$ from $p(\theta \mid \eta, \Phi_0, \Phi)$ as follows:

(1) Filtering

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

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

(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 $t = N-1, ..., 0$, 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

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

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

Given θ, Φ_0 and Φ, the η_{it}'s are mutually independent. 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

$$
p\big(\eta_{it} \mid \theta_t, \Phi_{0i}, \Phi, y_{it}\big) \propto p\big(y_{it} \mid \eta_{it}\big)\, p\big(\eta_{it} \mid \theta_t, \Phi_{0i}\big)\ . \qquad (4.3)
$$

The second term on the right-hand side of the full conditional (4.3) 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 (2.2), such that $\eta_{1it} = h_1(\mu_{it})$ and $\eta_{2it} = h_2(\phi_{it})$.

Since the distribution $p(\eta_{it} \mid \theta_t, \Phi_{i0}, y_{it})$ does not have a closed form, it is necessary to use the Metropolis–Hastings algorithm (M-H) (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 M-H random-walk with symmetric normal proposal for η_{it}:

(**a**) Draw $\eta^*_{1it} \sim q_1(\eta^{m-1}_{1it}, \eta^*_{1it}) \overset{d}{=} N(\eta^{m-1}_{1it}, \Phi^{-1}_{1i0})$ and $\eta^*_{2it} \sim q_2(\eta^{m-1}_{2it}, \eta^*_{2it}) \overset{d}{=} N(\eta^{m-1}_{2it}, \Phi^{-1}_{2i0})$.

(**b**) Calculate the acceptance probability $\alpha(\eta^{m-1}_{it}, \eta^*_{it}) = \min\{1, R_{\eta_{it}}\}$, where

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

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

(**c**) Set

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

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

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

$$\begin{aligned} p\big(\Phi_l \mid \eta, \theta, \Phi_0, y\big) &\propto \left[\prod_{t=1}^{N}\prod_{m=1}^{k} |\Phi_m|^{1/2} \exp\left\{-\frac{1}{2}(\theta_t - H_t\theta_{t-1})^T\, \Phi(\theta_t - H_t\theta_{t-1})\right\}\right] \\ &\quad \times |\Phi_l|^{\varsigma_l-(p_l+1)/2} \exp\{-\operatorname{tr}(Z_l\Phi_l)\} \\ &\propto |\Phi_l|^{N/2+\varsigma_l-(p_l+1)/2} \exp\left\{-\operatorname{tr}\left(\frac{1}{2}\sum_{t=1}^{N} ZZ_{ll,t}\Phi_l\right) - \operatorname{tr}(Z_l\Phi_l)\right\} \qquad (4.4) \\ &\propto |\Phi_l|^{N/2+\varsigma_l-(p_l+1)/2} \exp\left\{-\operatorname{tr}\left(\left(\frac{1}{2}\,ZZ_{l\cdot} + Z_l\right)\Phi_l\right)\right\}, \end{aligned}$$

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

$$\big(\Phi_l \mid \eta, \theta, \Phi_0, y\big) \sim \text{Wishart}\left(\frac{N}{2} + \varsigma_l,\ \frac{1}{2}ZZ_{l\cdot} + Z_l\right), \qquad l = 1, ..., k\ .$$

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

$$
\begin{aligned}
p\big(\Phi_{i0} \mid \eta, \theta, \Phi_0, y\big) \;&\propto\; \left[\prod_{t=1}^{N} N\big(\eta_{it}; F_{it}\theta_t, \Phi_{0i}^{-1}\big)\right] W(\Phi_{0i}; \nu_{0i}, S_{0i}) \\
&\propto\; \left[\prod_{t=1}^{N} |\Phi_{0i}|^{1/2}\, \exp\left\{-\frac{1}{2}\,(\eta_{it} - F_{it}\theta_t)^T\, \Phi_{0i}(\eta_{it} - F_{it}\theta_t)\right\}\right] \\
&\qquad \times\, |\Phi_{0i}|^{\nu_{0i}-(p_{0i}+1)/2}\, \exp\big\{-\operatorname{tr}(S_{0i}\Phi_{0i})\big\} \\
&\propto\; |\Phi_{0i}|^{N/2+\nu_{0i}-(p_{0i}+1)/2}\, \exp\left\{-\operatorname{tr}\left(\left(\frac{1}{2}SS_{\eta_i} + S_{0i}\right)\Phi_{0i}\right)\right\},
\end{aligned}
\tag{4.5}
$$

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

$$
\big(\Phi_{0i} \mid \eta, \theta, \Phi, y\big) \;\sim\; \text{Wishart}\left(\frac{N}{2} + \nu_{0i},\, \frac{1}{2}SS_{\eta_i} + S_{0i}\right), \qquad i = 1, ..., I\;.
$$

4.4. The case of static dispersion parameters

It is also possible to describe a beta hierarchical model such that the dispersion process does not vary with time, i.e., the precision parameters are static. In such case, the vector η_t on the left-hand side of the structural equation (2.3) will only include the term related to the mean process and $\eta_{it} = \eta_{1it} = h_1(\mu_{it})$. However, we can still associate a link function to the precision parameters, and we will denote it by $\eta_{2i} = h_2(\phi_i)$, $i = 1, ..., I$.

The observation equation for such case is then $(y_{it} \mid \eta_{it}, \eta_{2i}) \sim \text{Beta}(y_{it} \mid \eta_{it}, \eta_{2i})$, $i = 1, ..., I$. The Bayesian analysis for such situation can be adapted from the one we just described in the previous sections.

The MCMC developments for η_{it} are largely the same described in Section 4.2, but now they will be conditioned upon the current values of η_{2i}. Additionally, for a given prior distribution for η_{2i}, the corresponding full conditional distribution is

$$
p\big(\eta_{2i} \mid \eta, \theta, \Phi_0, \Phi, y\big) \;\propto\; \left[\prod_{t=1}^{N} \text{Beta}(y_{it} \mid \eta_{it}, \eta_{2i})\right] p(\eta_{2i})\;. \tag{4.6}
$$

In this work the prior $p(\eta_{2i})$ was set to be a Gaussian distribution, with parameters chosen as the average mean and average variance of the initial estimated values of η_{2i}, $i = 1, ..., I$, that were obtained from separate MCMC runs for each of the individual time series.

5. A SIMULATION STUDY

We applied the model described in Section 3 to simulated data in which we considered $N = 72$ time points (say, six years), $I = 3$ sub-populations and cycles of size $p = 4$.

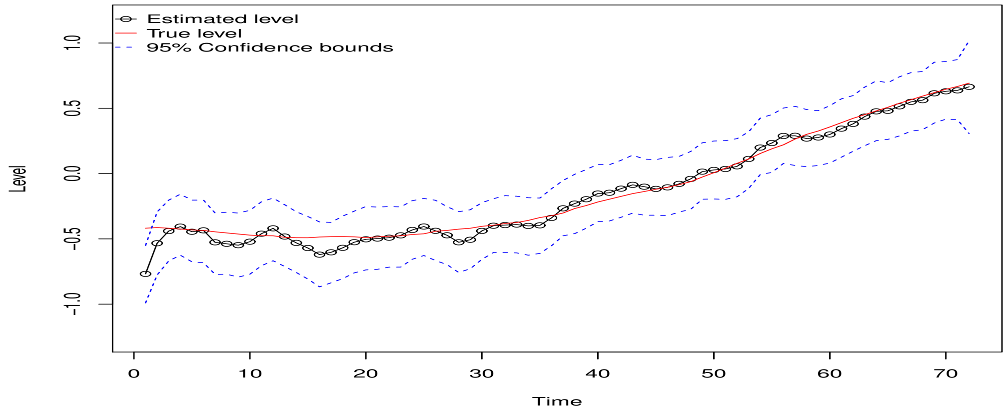
In order to obtain initial values for the MCMC procedure, we estimated the parameters involved by running separate DLM models (described by equations (2.3) and (2.4)) for each of the sub-populations. All the routines were written using the `R` language (`http://www.r-project.org/`). We also made extensive 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 $\log\left(\frac{\tilde{\sigma}^2_{it}}{1-\tilde{\sigma}^2_{it}}\right)$ with $\tilde{\sigma}^2_{it} = \mathrm{var}(y_i)/(y_{it}(1-y_{it}))$ (see properties of expression (2.1)).

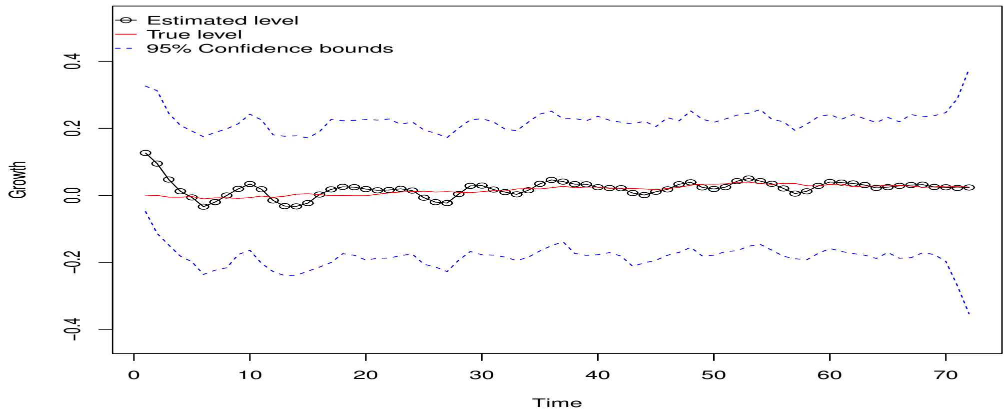
For the simulated data we considered a hierarchical dynamic beta model in which a second-order polynomial trend seasonal effects were fitted to the parameters related to the mean, μ_{it}, and a second-order polynomial effects was fitted to the parameters related to the precision, ϕ_{it}. We run chains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

Figure 2 shows the true values (in red) used in the simulations, the estimated values of the parameters involved in expressions (3.2), and the respective confidence bands for the main effects of level, growth and seasonality. Figure 3 shows the estimated proportions for each of the sub-populations and their corresponding confidence bands. As we can observe all the effects and probabilities are well estimated.

(a)



(b)



(c)

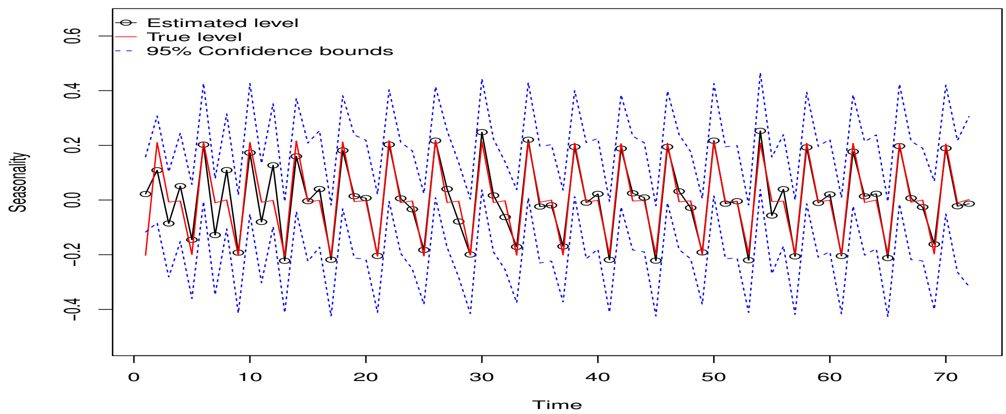
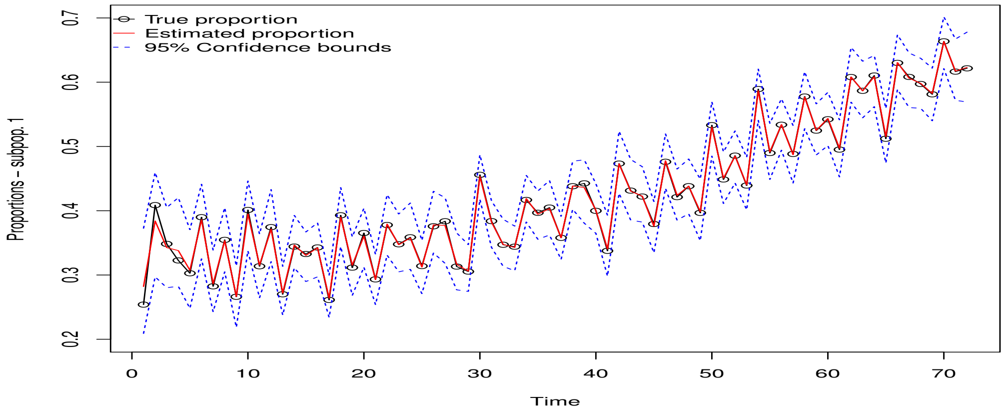
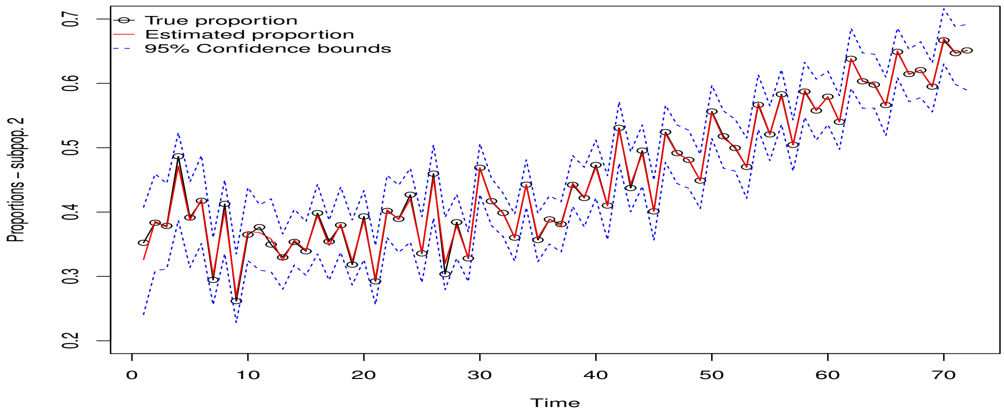


Figure 2: Simulated data — estimated values and 95% credibility bounds for the components of η_{1it}: (a) Level (β_t), (b) Growth (δ_t), (c) Seasonality (λ_t).

(a)



(b)



(c)

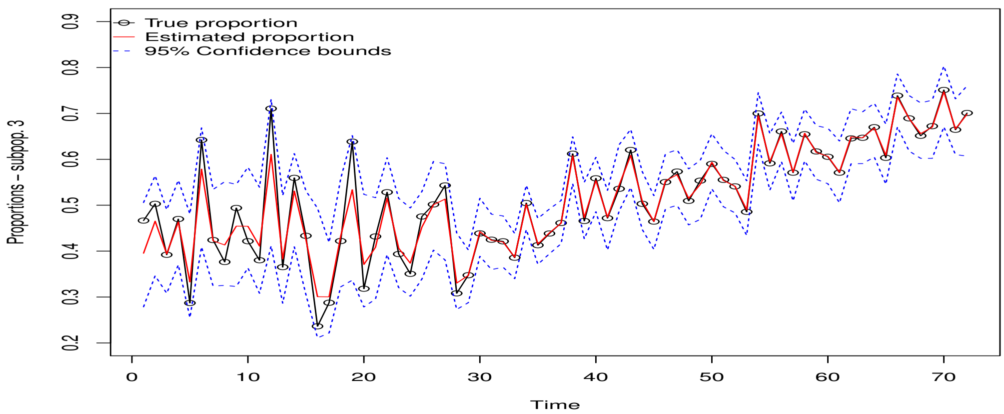


Figure 3: Simulated data — estimated proportions and 95% credibility bounds for the three sub-populations.

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 the Introduction (see Section 1). The three sub-populations in our the analysis are Recife, São Paulo and Porto Alegre, i.e., $I = 3$. We analyze monthly unemployment rates (MUR) based on PME data in the period from March 2002 to July 2011 ($N = 113$ observations). As a procedure for checking the performance of the model, forecast rates are also provided. We used MUR data for the months of August 2011 to March 2012.

The Brazilian Institute of Geography and Statistics acknowledges the necessity of incorporating seasonal components in any analyses based on MUR data (`http://www.ibge.gov.br`). In fact the MURs are affected by yearly cycles caused by factors such as climatic changes, Christmas festivities and school vacations.

To the *mean process*, we considered a hierarchical dynamic beta model in which a second-order polynomial trend seasonal effects model (with cycles of $p = 12$ months) was fitted to the parameters related to the mean, μ_{it} (see expressions (3.1) to (3.5)). The model structure was similar to the one used in Da-Silva *et al.* (2011), and we opted for using a free-form seasonal effects model (see equation (3.9)) instead of using harmonic analysis via Fourier representation (see equation (3.10)), just to be consistent with Da-Silva *et al.* (2011).

To the *dispersion process* we used two models: **Model 1** describes a static hierarchical model with respect to the precision parameter (see Section 4.4). **Model 2** adds dynamics to the precision parameter of the beta model. For that purpose we use a second-order polynomial trend effects model (see expressions (3.6) to (3.8)).

Considering Model 2 and the parameters of both mean and dispersion processes, the design matrices H and $F = (F_1', F_2', F_3')'$ are the ones defined by expression (3.9). For Model 1 those matrices are:

$$H = \text{block-diag}(J, P, \mathbf{I}_2) ,$$
$$F_1 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 1 & 0 \end{pmatrix} , \qquad F_2 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & 0 & 1 \end{pmatrix} \qquad \text{and}$$
$$F_3 = \begin{pmatrix} 1 & 0 & 1 & \mathbf{0}_{1\times(p-2)} & -1 & -1 \end{pmatrix} .$$

We used the mean absolute deviation (MAD), the mean square error (MSE) and the Deviance Information Criterion (DIC) (Spiegelhalter *et al.*, 2002) to compare the forecasting accuracies of Model 1 and Model 2. The MAD and MSE are defined, respectively, by the following formulae: $MAD = \frac{1}{n}\sum_{t=1}^{n} |e_t|$ and $MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2$, where $e_t = Y_t - E(Y_t \,|\, y_{1:t-1})$ (see Section 4.1 for details).

According to Spiegelhalter *et al.* (2002), the DIC is a measure of fit based on a trade-off between the fit of the data to the model and the corresponding complexity of the model: DIC = goodness of fit + complexity. The fit for model M_i is measured in terms of the posterior distribution of the deviance statistic, $D(\theta_i) = -2\log L(Y \,|\, \theta_i)$, while complexity is measured by an estimate of the effective number of parameters:

$$d_i \;=\; \overline{D}_i - D(\bar{\theta}_i) \;=\; E\Big(D(\theta_i \mid Y, M_i) - D(E(\theta_i \mid Y, M_i)\Big)\,,$$

i.e., the posterior mean deviance minus deviance evaluated at the posterior mean of the parameters. The DIC is defined as:

$$DIC(M_i) \;=\; D(\bar{\theta}_i) + 2\,d_i \;.$$

The DIC generalizes the AIC (Akaike, 1973) in the sense that it explicitly applies to non nested non IID problems. Besides that, DIC can be approximated via MCMC samples from the posterior density. For the reasons exposed so far, in this work we used DIC instead of either AIC or BIC (Schwarz, 1978).

Models with smaller DIC are better supported by the data. The DIC is a positive number, in general. However, it can be negative but such occurrence does not pose any difficulty in terms of model comparison, since the focus is in the difference between two values and not in the DIC value itself.

In general, the d_i component is a positive value. However, it can be negative in cases where the likelihood function is not log-concave, when there is a conflict between the prior and the likelihood or when the posterior distribution of the parameters is too skewed or symmetric and multi-modal, so that the posterior mean/median are poor measures of central tendency. In those cases, the use of the posterior mode can be a fair alternative.

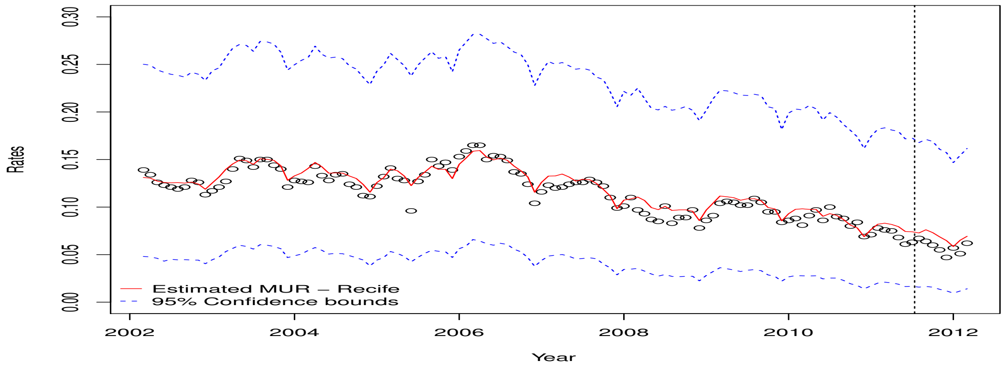
Table 1 displays the MAD, MSE, the effective number of parameters, d_i, and DIC values for Model 1 and Model 2. As we observe, the total MAD and total MSE values for Model 2 are somewhat smaller than those values for Model 1. However, the DIC for Model 2 is much smaller than the DIC for Model 1, giving strong indication that Model 2 provides a superior fit compared to Model 1. Additionally, the effective number of parameters, d_i, for both, Model 1 and Model 2, are positive. Thus besides of the limitations of the DIC, in our applications it seems to be performing properly.

Table 1: MSE, MAD, d_i and DIC values for Models 1 and 2.

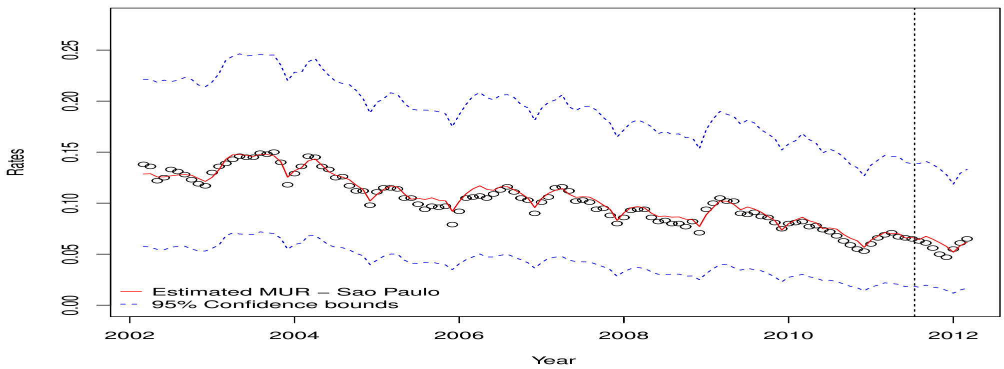
Model	**MSE**	**MAD**	d_i	**DIC**
Model 1	0.00158	0.03934	36.602	−1510.622
Model 2	0.00150	0.03559	229.819	−2445.007

In order to gain a better perspective of the real advantages of using Model 2 as opposed to Model 1, we present Figures 4 and 5 which display the estimated proportions or rates for each of the sub-populations and their corresponding confidence bands.

(a)



(b)



(c)

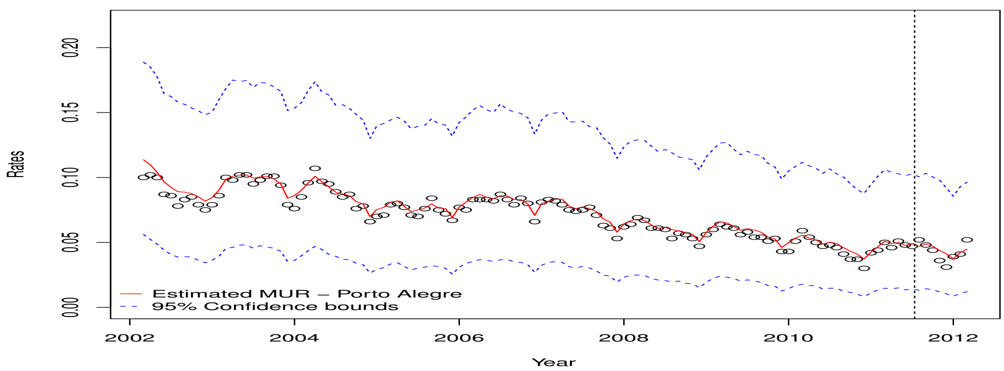
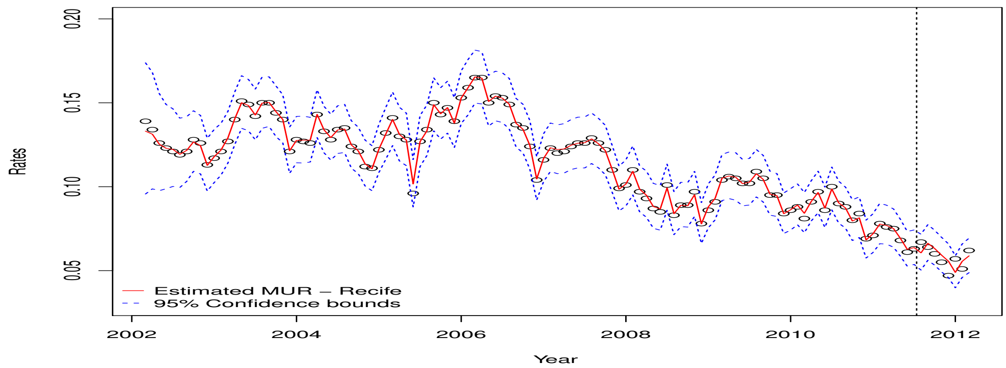


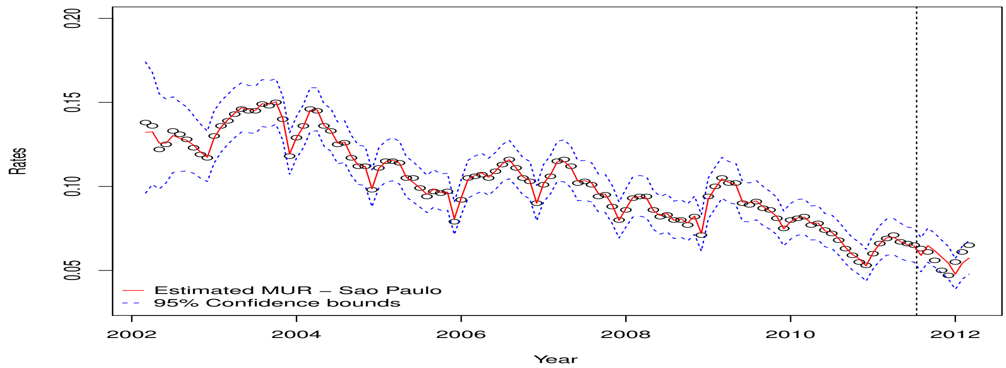
Figure 4: MUR data (Model 1 with static precision parameter) — estimated proportions and 95% credibility bounds for the three sub-populations: (a) Recife, (b) São Paulo and (c) Porto Alegre.
The forecast rates are presented after the dotted vertical lines.

(a)



[3mm]

(b)



[3mm]

(c)

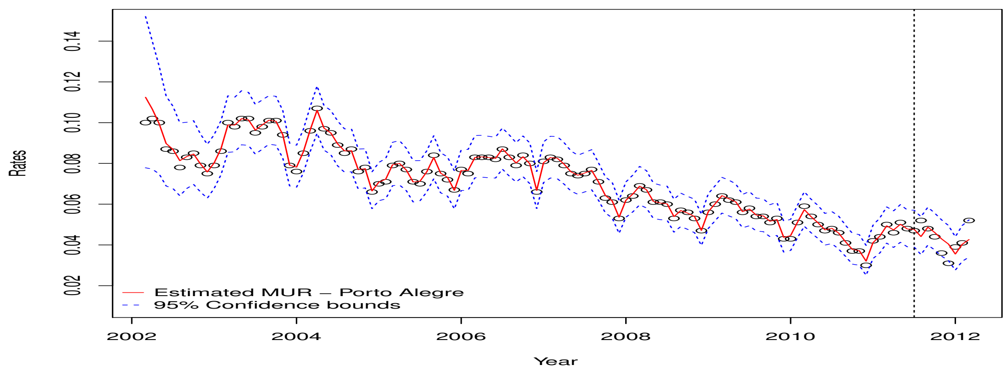


Figure 5: MUR data (Model 2 with dynamic precision parameter) — estimated proportions and 95% credibility bounds for the three sub-populations: (a) Recife, (b) São Paulo and (c) Porto Alegre.
The forecast rates are presented after the dotted vertical lines.

It is really reassuring the superiority of Model 2 compared to Model 1 in terms of both the precision of the credibility intervals, and how well Model 2 is able of describing the observed proportions for each of the sub-populations.

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 modelling 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 no 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*, makes 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.

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 March 2012. 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 by the inspection of 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 functions 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 type of regimes for both, the level of the mean process and the level of the dispersion process.

ACKNOWLEDGMENTS

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. We also want to thank the anonymous referees for their careful reading of this manuscript and also for their suggestions and criticisms which lead to very important improvements.

REFERENCES

[1] Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle.* In: "Proceedings of the Second International Symposium on Information Theory" (B.N. Petrov and F. Csáski, Eds.), Budapest, Akadémiai Kiadó, pp. 267–281.

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

[3] 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.

[4] Carter, C.K. and Kohn, R. (1994). On Gibbs sampling for state-space models, *Biometrika*, **81**, 541–553.

[5] Cuervo-Cepeda, E. and Gamerman, D. (2004). Bayesian modeling of joint regressions for the mean and covariance matrix, *Biometrical Journal*, **46**(4), 430–440.

[6] Da-Silva, C.Q.; Migon, H.S. and Correia, L.T. (2011). Dynamic Bayesian beta models, *Computational Statistics* & *Data Analysis*, **55**, 2074–2089.

[7] Espinheira, P.L. (2007). *Regressão Beta*, Doctoral Dissertation, Statistics Department, Institute of Mathematics and Statistics, University of São Paulo (USP).

[8] Ferrari, S.L.P. and Cribari-Neto, F. (2004). Beta regression for modeling rates and proportions, *Journal of Applied Statistics*, **31**, 799–815.

[9] Früwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *Journal of the Royal Statistical Society, Ser. B*, **55**, 629–642.

[10] Gamerman, D. and Migon, H. (1993). Dynamic hierarchical models, *Journal of the Royal Statistical Society, Series B*, **55**, 629–642.

[11] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.

[12] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemistry and Physics*, **21**, 1087–1091.

[13] 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.

[14] Petris, G. (2010). An R package for dynamic linear models, *Journal of Statistical Software*, **36**(12), 1–16. `http://www.jstatsoft.org/v36/i12/`

[15] Pole, A.; Mike West, M. and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*, CRC Press, USA.

[16] Pregibon, D. (1984). Review of generalized linear methods by McCullagh and Nelder, *Ann. Statist.*, **12**, 1589–1596.

[17] Rocha, A.V. and Cribari-Neto, F. (2009). Beta autoregressive moving average models, *Test*, **18**(3), 529–545.

[18] 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*, **54**, 348–366.

[19] Smithson, M. and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables, *Psychological Methods*, **11**(1), 54–71.

[20] Smyth, G.K. and Verbyla, A.P. (1999). Adjusted likelihood methods for modeling dispersion in generalized linear models, *Environmetrics*, **10**(6), 695–709.

[21] Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P. and Linde, A. Van der (2002). Bayesian measures of model complexity and fit, *Journal of the Royal Statistical Society, Series B*, **64**, 583–639.

[22] Schwarz, G. (1978). Estimating the dimension of a model, *Annals of Statistics*, **6**, 497–511.

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