DYNAMIC BAYESIAN BETA MODELS

C. Q. da-Silva [a,*], H. S. Migon [b], L. T., Correia [c],

[a] *Ph.D., Assoc. Prof. of Statistics, Department of Statistics, University of Brasília-UnB, Brasília 70910-900, Brazil*

[b] *Ph.D., Full prof. of Statistics, Department of Statistics, Federal University of Rio de Janeiro, Rio de Janeiro 21.945-970, Brazil*

[c] *M.Sc., Department of Statistics, University of Brasília-UnB, Brazil*

Abstract

We develop a dynamic Bayesian beta model for modeling and forecasting single time series of proportions. This work is related to the class of the so called dynamic generalized linear models (DGLM). We use non-conjugate priors and some forms of approximate Bayesian analysis, including Linear Bayesian estimation. Some applications to both real and simulated data are provided.

Key words: Dynamic models; Beta distribution; Logistic-normal distribution; Generalized linear models; Bayesian analysis.

1 Introduction

The beta distribution provides a useful tool for modeling data restricted to the interval (0,1), such as rates, percentages and proportions. In particular, one may be interested in modeling fluctuations in variables such as the proportion of a given fish species in a lake, the proportion of drug addicted adults in the population, the proportion of time the employees of a given company spend browsing the Internet, the monthly unemployment rates of a given country or the proportion of a given component in compositional data analysis.

The beta distribution is very flexible for modeling such data since its density can display quite different shapes depending on the parameter values.

* Corresponding author.
Email addresses: `cibeleqs@unb.br` (C. Q. da-Silva), `migon@im.ufrj.br` (H. S. Migon), `leandrotc@gmail.com` (L. T., Correia).

Preprint submitted to Computational Statistics and Data Analysis 12 April 2010

For a sample of size n of i.i.d. random variables following a beta distribution, some frequentist beta regressions (static beta regressions) have been proposed: Paolino (2001), Kieschnick and McCullough (2003) and Ferrari and Cribari-Neto (2004), among others. In the class of regression models introduced by Ferrari and Cribari-Neto (2004), the basic assumption is that the response follows a beta law whose expected value is related to a linear predictor through a link function. A Bayesian version of the static beta regression was proposed by Branscum *et al.* (2007) and also by Albi *et al.* (2009). The latter authors used a Laplace approximation formulation via INLA (see Rue *et al.*, 2009). For time series analysis, some models have been proposed: Azzalini (1984), McKenzie (1985), Wallis (1987), Grunwald, Raftery and Guttorp (1993), and Bruche and Gonzáles-Aguado (2010).

The present work is related to the class of so-called dynamic generalized linear models (DGLM). In West, Harrison and Migon (1985), Lindsey and Lambert (1995) and also in Godolphin and Triantafyllopoulos (2006), the dynamic linear models (DLM) (West and Harrison, 1997) are extended and generalized to various non-normal problems.

In the DGLM, time series $\{Y_t\}$ are modeled using the uniparametric exponential family of distributions and a specification similar to the well-known class of generalized linear models (McCullagh and Nelder, 1989). In order to simplify the calculations, the Bayesian formulation is completed with the use of conjugate priors. In this work we choose instead a more convenient form of prior. This prior has the advantage of simplifying the calculations more than a conjugate prior would, but allowing full exploration of the relationships with the moments of a linear function of the state vector, which is essential for DLM in general.

The article is organized as follows. In Section 2 we introduce some backgroud information about dynamic generalized linear models. In Section 3 we describe a new methodology to analyze data from dynamic beta processes, the dynamic Bayesian beta model (DBM), and we present the key steps necessary to estimate the parameters of the DBM for the case where ϕ is known. In Section 4 we discuss some more generalizations of the DBM (the case of unknown ϕ). In Section 5 we apply the methods to simulated and real data.

2 Dynamic generalized linear models

Dynamic linear models are parametric models where the parameter variation and the available data are described probabilistically. They are characterized by a pair of equations, named the *observational equation* and parameters evolution equation or *system equation*. The observational and the system equations are, respectively, given by

$$
\begin{aligned}
y_t &= \mathrm{F}_t'\theta_t + \epsilon_t,\ \ \epsilon_t \sim N(0, V_t) && (1)\\
\theta_t &= G_t\theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t) && (2)
\end{aligned}
$$

where y_t is a time sequence of observations, conditionally independent given the sequence of parameters θ_t, F_t is a vector of explanatory variables, θ_t is a $k \times 1$ vector of parameters, G_t is a $k \times k$ matrix describing the parameter evolution and, finally, V_t and W_t are the variances of the errors associated with the unidimensional observation and with the k-dimensional vector of parameters, respectively. A dynamic linear model is completely specified by the quadruple (F_t, G_t, V_t, W_t).

Based on the generalized linear models of Nelder and Wedderburn (1972), West, Harrison and Migon (1985) extended the DLMs to allow for observations in the exponential family. In such setting the observation equation (1) is described by

$$
p(y_t|\eta_t) \propto \exp[(y_t\eta_t - b(\eta_t))/\phi_t] \qquad (3)
$$

and, in addition, a suitable link function is introduced, relating the mean $\mu_t = E[y_t|\eta_t]$ to the natural parameter η via $\mu_t = b'(\eta_t)$, while η_t relates to the regressors F_t and to the linear function $\lambda_t = F_t'\theta_t$ of the state parameters through $g(\eta_t) = \lambda_t$. A conjugate prior for η_t is given as

$$
p(\eta_t|D_{t-1}) \propto \exp[(r_t\eta_t - b(\eta_t))/s_t]
$$

and the values of the pair of hyperparameters (r_t, s_t) have to be estimated.

The Bayesian inference in this class of models explores the sequential aspects of Bayesian inference combining the operations: *evolution* to build up the prior and *updating* to incorporate the new observation arrived at time t. Let $D_t = D_{t-1} \cup \{y_t\}$ denote the information until time t, including the values of F_t and G_t, $\forall t$, which are supposed to be known, with D_0 representing the prior information. Then for each time t the prior, predictive and posterior distribution are, respectivelly:

$$
\begin{aligned}
p(\theta_t|D_{t-1}) &= \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}\\
p(y_t|D_{t-1}) &= \int p(y_t|\theta_t)p(\theta_t|D_{t-1})d\theta_t\\
p(\theta_t|D_t) &\propto p(\theta_t|D_{t-1})p(y_t|D_{t-1})
\end{aligned}
$$

where the last distribution is obtained via Bayes' Theorem. These integrals cannot be obtained in closed form, and so the inference must be done using numerical approximations. Using linear Bayes estimation, a procedure allowing the sequential analysis of dynamic generalized linear models (DGLM) was implemented in West, Harrison and Migon (1985).

The evolution equation (2) is only partially specified. This means that the distributions of $(\theta_{t-1}|D_{t-1})$ and ω_t are only specified by the first- and second-order moments, that is: $(\theta_{t-1}|D_{t-1}) \sim [m_{t-1}, C_{t-1}]$ and $\omega_t \sim [0, W_t]$. Then the prior distribution of the state parameters is also partially specified as $(\theta_t|D_{t-1}) \sim [a_t, R_t]$, with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W_t$. Since

$$\lambda_t = g(\eta_t) = F_t'\theta_t,$$

this implies that the prior distribution $(\lambda_t|D_{t-1}) \sim [f_t, q_t]$, where $f_t = F_t' a_t$ and $q_t = F_t' R_t F_t$.

The parameters (r_t, s_t), in the prior distribution of η_t, must be related to f_t and q_t through the equations

$$E[g(\eta_t) \mid D_{t-1}] = f_t \quad \text{and} \quad \mathrm{var}[g(\eta_t) \mid D_{t-1}] = q_t.$$

Then the posterior for η_t is in the same form of the prior distribution, with parameters

$$(r_t/s_t + y_t/\phi_t, 1/s_t + 1/\phi_t).$$

The posterior distribution of the linear predictor is $(\lambda_t|D_t) \sim [f_t^*, q_t^*]$, where, again,

$$f_t^* = E[g(\eta_t)|D_t] \quad \text{and} \quad q_t^* = \mathrm{var}[g(\eta_t)|D_t].$$

Moreover, to complete the analysis, the posterior distribution of the state parameters must be obtained.

Linear Bayes estimation is used to approximate the first- and second-order moments of this distribution, leading to:

$$\hat{E}[\theta_t|\eta_t, D_{t-1}] = a_t + R_t F_t[\eta_t - f_t]/q_t \quad \text{and} \quad \hat{\mathrm{var}}[\theta_t|\eta_t, D_{t-1}] = R_t - R_t F_t F_t' R_t/q_t.$$

The moments of $(\theta_t|D_t)$ are calculated using the iterated expectation law given by $(\theta_t|D_t) \sim [m_t, C_t]$, where $m_t = a_t + R_t F_t[f_t^* - f_t]/q_t$ and $C_t = R_t - R_t F_t F_t' R_t(1 - q_t^*/q_t)/q_t$.

3 Dynamic beta model

In this section we present a methodology for modeling a time series of proportions y_t considering the class of the DGLM. Using a new parametrization of the Beta distribution (Ferrari and Cribari-Neto, 2004), the model is defined by the following components:

- **Observation Equation:**

$$p(y_t \mid \mu_t, \phi) = \frac{\Gamma(\phi)}{\Gamma(\phi\mu_t)\Gamma(\phi(1-\mu_t))} y_t^{\phi\mu_t - 1}(1-y_t)^{\phi(1-\mu_t)-1}. \tag{4}$$

- **Prior:** $(\mu_t \mid D_{t-1}) \sim \text{Beta}(r_t, s_t)$.

- **Link function:** Consider, without loss of generality, the logit link

$$\lambda_t = g(\mu_t) = F_t'\theta_t = \log\left(\frac{\mu_t}{1-\mu_t}\right), \quad \text{such that} \quad \mu_t = \frac{\exp(\lambda_t)}{1+\exp(\lambda_t)}.$$

- **System Equation:**

$$\theta_t = G_t\theta_{t-1} + w_t; \quad w_t \sim (0, W_t).$$

- **Initial Information:**

$$(\theta_0 | D_0) \sim (m_0, C_0).$$

The parameter μ_t is the expected value of $(y_t \mid \mu_t, \phi)$ while the parameter ϕ, a precision parameter, may also be interpreted as a "sample size". We would like to stress that in this work we are not imposing a conjugate prior for the parameter of interest. Instead, we use a convenient prior to simplify certain calculations. We will first deal with the dynamic beta model for the case of ϕ known.

3.1 Inference for the dynamic beta model with ϕ known

In this section we describe the inferential procedure we formulate for estimating the model parameters of the Beta Dynamic Model, which can be viewed as a variant of the DGLM steps described in Section 3. The main steps involved in our procedure: the *evolution*, the parameters *equating* and the *updating*, are summarized below, by cycling through the steps (i) to (iii) from $t = 1, \ldots, T$, thus keeping the sequential nature of the dynamic models. Almost all the distributions studied here are only partially specified in terms of their moments. For a given time t, steps (i) to (iii) are described by:

i) Evolution step: Based on the values of the moments (m_{t-1}, C_{t-1}) of the partially specified posterior distribution $(\theta_{t-1}|D_{t-1}, \phi)$, then, following the procedures detailed in Section 2, we obtain the values of the first moments (f_t, q_t) of the linear predictor λ_t.

(ii) Equating parameters step: Since $\mu_t = g^{-1}(\lambda_t)$, its prior distribution must have parameters uniquely related to those of λ_t. This is accomplished by solving a nonlinear system of equations.

(iii) Updating step: Based on the values of the first two moments of the posterior distribution of μ_t or any suitable approximation of them, generically denoted by $(\tilde{\mu}_t, \tilde{V}_t)$, we want to obtain (f_t^*, q_t^*), the parameters of the posterior moments of λ_t. That enable estimating to estimate the posterior moments of $(\theta_t|D_t, \phi)$, (m_t, C_t), using linear Bayes estimation.

To complete the sequential analysis, the marginal distribution of $(y_t|D_{t-1}, \phi)$ can be obtained, at least approximately, as a function of the moments of $(\mu_t|D_{t-1})$.

Next, we provide detailed calculation of some of the key components described in steps (i) to (iii) for the inferential process of the dynamic beta model.

3.1.1 Evolution step

a) *Priors for state parameters θ_t and the linear predictor λ_t.* As described in Section 2, using the evolution equation we get $(\theta_t \mid D_{t-1}) \sim (a_t, R_t)$ and $(\lambda_t \mid D_{t-1}) \sim (f_t, q_t)$. The prior covariance between θ_t and λ_t is easily obtained as: $R_t F_t = \text{cov}(\theta_t, \lambda_t|D_{t-1})$.

b) *Prior for μ_t.* Although a conjugate prior distribution is always available for models in the exponential family, in our case the evaluation of its moments is quite cumbersome. Since the parameter μ_t is restricted to the interval $(0, 1)$, a natural choice for its prior is the beta family: $(\mu_t \mid D_{t-1}) \sim Beta(r_t, s_t)$, where $r_t, s_t > 0$ are known quantities, occasionally functions of D_{t-1}. Its first two moments are known and will be used in the solution of a simple nonlinear system in order to obtain the parameter values (r_t, s_t) consistent with (f_t, q_t), the moments of $(\lambda_t \mid D_{t-1})$.

3.1.2 Equating parameters

Since the linear predictor is related to the mean of the observation distribution through a nonlinear link function, some approximation is needed to determine the hyperparameters r_t and s_t of the prior distribution of μ_t.

Since $(\mu_t \mid D_{t-1}) \sim Beta(r_t, s_t)$, the pair (r_t, s_t) can be found in terms of the exact expected value of the log-beta distribution (West and Harrison, 1997, pp. 529-30), given by

$$E[\log(\mu_t/(1-\mu_t))] = \psi(r_t) - \psi(s_t) \quad \text{and} \quad V[\log(\mu_t/(1-\mu_t))] = \psi'(r_t) + \psi'(s_t),$$

where $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ and $\psi'(z) = \frac{d\psi(z)}{dz}$ are, respectively, the digamma and the trigamma functions. For the dynamic beta model we get:

$$\begin{aligned}
E(\lambda_t|D_{t-1}) &= E\left[\log\left(\frac{\mu_t}{1-\mu_t}\right)\middle| D_{t-1}\right] = \psi(r_t) - \psi(s_t) = f_t \\
V(\lambda_t|D_{t-1}) &= \psi'(r_t) + \psi'(s_t) = q_t.
\end{aligned}$$

For large values of z, $\psi(z) \approx \log(z)$ while $\psi'(z) \approx z^{-1}$. Therefore, $f_t \approx \log\left(\frac{r_t}{s_t}\right)$ and $q_t \approx \frac{1}{r_t} + \frac{1}{s_t}$. Then,

$$r_t = [1 + \exp(f_t)]/q_t \quad \text{and} \quad s_t = [1 + \exp(-f_t)]/q_t.$$

West and Harrison (1997) argue that the approximation gives satisfactory results even for small values of s_t and r_t. An alternative method for calculating r_t and s_t, which is based on a second-order Taylor approximation, is described in the Appendix.

3.1.3 Updating step

The posterior distribution of μ_t is often obtained by Bayes' Theorem,

$$\begin{aligned}
p(\mu_t|D_t, \phi) &\propto p(y_t|\mu_t, D_{t-1}, \phi)p(\mu_t|D_{t-1}) \\
&\propto \frac{\Gamma(\phi)}{\Gamma(\phi\mu_t)\Gamma(\phi(1-\mu_t))} y_t^{\phi\mu_t - 1}(1-y_t)^{\phi(1-\mu_t)-1} \\
&\times \frac{\Gamma(s_t + r_t)}{\Gamma(r_t)\Gamma(s_t)} \mu_t^{r_t - 1}(1-\mu_t)^{s_t - 1}. \qquad (5)
\end{aligned}$$

Observe that $p(\mu_t|D_t, \phi)$ has no known closed form. Therefore, it is necessary to obtain its first- and second-order moments approximately. A possibility is to use the Laplace approximation proposed by Tierney and Kadane (1986).

The first and second moments of $(\mu_t|D_t, \phi)$ allow us to evaluate (f_t^*, q_t^*), the posterior mean and variance of the linear predictor. These operations may involve the solution of a nonlinear system of equations.

a) Laplace approximation

The posterior moments of $(\mu_t \mid D_t, \phi)$ can be obtained by considering that

$$\begin{aligned} E(h(\mu_t)) &= \int h(\mu_t) p(\mu_t \mid D_t, \phi) d\mu_t \\ &= \frac{\int h(\mu_t) p(y_t \mid \mu_t, D_{t-1}, \phi) p(\mu_t \mid D_{t-1}) d\mu_t}{\int p(y_t \mid \mu_t, D_{t-1}, \phi) p(\mu_t \mid D_{t-1}) d\mu_t} \\ &= \frac{\int \exp(K^*(\mu_t)) d\mu_t}{\int \exp(K(\mu_t)) d\mu_t}, \end{aligned}$$

where $h(\mu_t)$ is a strictly positive function.

Since $p(\mu_t \mid D_t, \phi)$ is a smooth, bounded unimodal function with a maximum at $\hat{\mu}_t$ and μ_t is a scalar, the conditions for using Laplace approximation are all satisfied.

Applying a second-order Taylor expansion in K^*, we have

$$K^*(\mu_t) \approx K^*(m^*) - \frac{1}{2}(\mu_t - m^*)'(V^*)^{-1}(\mu_t - m^*),$$

where m^* is the point that maximizes $K^*(\mu_t)$ and V^* is given by minus the inverse of the Hessian matrix associated with $K^*(\mu_t)$ evaluated at m^*.

Thus, considering the Laplace aproximation proposed by Tierney and Kadane (1986),

$$E(h(\mu_t)) \approx \left(\frac{\mid V^* \mid}{\mid V \mid}\right)^{1/2} \exp(K^*(m^*) - K(m)).$$

Similar definitions apply to m and V.

Let $\ell(\mu_t) = \log[p(\mu_t | D_t, \phi)]$. Thus, the first and second derivatives of ℓ with respect to μ_t can be approximated, using $\psi(z) \approx \log(z)$ and $\psi'(z) \approx z^{-1}$, by:

$$\begin{aligned} \frac{\partial \ell}{\partial \mu_t} &\approx \phi \log\left(\frac{1-\mu_t}{\mu_t}\right) + \phi \log\left(\frac{y_t}{1-y_t}\right) + \frac{s_t - 1}{\mu_t} - \frac{r_t - 1}{1-\mu_t}, \\ \frac{\partial^2 \ell}{\partial \mu_t^2} &\approx -\frac{\phi}{\mu_t(1-\mu_t)} - \frac{s_t - 1}{\mu_t^2} - \frac{r_t - 1}{(1-\mu_t)^2}. \end{aligned}$$

Since it is not possible to find m and m^* analytically, we use the Newton-Raphson method.

In summary:

$$
\begin{aligned}
V &= -\left[\frac{\partial^2 K(\mu_t)}{\partial \mu_t^2}\bigg|_{\mu_t = m} \right]^{-1} \quad \text{and} \quad V^* = -\left[\frac{\partial^2 K^*(\mu_t)}{\partial \mu_t^2}\bigg|_{\mu_t = m^*} \right]^{-1}, \\
\tilde{\mu}_t &= E[\mu_t|D_t,\phi] \approx \left(\frac{|V^*|}{|V|}\right)^{1/2} \exp[K^*(m^*) - K(m)], \quad \text{with} \quad h(\mu_t) = \mu_t, \\
\tilde{V}_t &= V[\mu_t|D_t,\phi] \approx E[\mu_t^2|D_t,\phi] - (E[\mu_t|D_t,\phi])^2, \quad \text{with} \quad h(\mu_t) = \mu_t^2.
\end{aligned}
$$

b) Updating for λ_t. Using the the pair of values $\tilde{\mu}_t, \tilde{V}_t$ we evaluate the first moments f_t^* and q_t^* of the posterior distribution $(\lambda_t|D_t,\phi)$ by taking a first-order Taylor expansion of $E[\mu_t|D_t,\phi]$ and $V[\mu_t|D_t,\phi]$. The moments $\tilde{\mu}_t, \tilde{V}_t$ are expressed by

$$
\begin{aligned}
\tilde{\mu}_t &\approx E[\mu_t|D_t,\phi] = E\left[\frac{\exp(\lambda_t)}{1+\exp(\lambda_t)} \bigg| D_t,\phi \right] \approx \frac{e^{f_t^*}}{1+e^{f_t^*}}; \\
\tilde{V}_t &\approx V[\mu_t|D_t,\phi] \approx \left(\frac{e^{f_t^*}}{(1+e^{f_t^*})^2} \right)^2 q_t^*.
\end{aligned}
$$

Thus f_t^* and q_t^* are given by $f_t^* = \log(\tilde{\mu}_t/1-\tilde{\mu}_t)$ and $q_t^* = \tilde{V}_t(f_t^*/(1+f_t^*)^2)^{-2}$.

When using a second-order Taylor expansion, (f_t^*, q_t^*) can be obtained numerically by solving a nonlinear system. The function *"dfsane"* in the *BB* library of software *R* can be used in such case.

c) Updating for θ_t. The joint partially specified distribution of θ_t and λ_t is easily obtained from previous results in this section. The *linear Bayesian estimation* method (see West and Harrison (1997), Chapters 4 and 14) can be employed to get:

$$
\begin{cases}
\hat{E}[\theta_t|\lambda_t, D_{t-1}] = a_t + \frac{1}{q_t} R_t F_t (\lambda_t - f_t), \\
\hat{V}[\theta_t|\lambda_t, D_{t-1}] = R_t - \frac{1}{q_t} R_t F_t F_t' R_t.
\end{cases}
$$

Since the posterior distribution of $(\lambda_t|D_t,\phi)$ is available, we can obtain the posteriori moments of $(\theta_t|D_t,\phi)$ as:

$$
\begin{aligned}
m_t &= a_t + \frac{1}{q_t} R_t F_t (f_t^* - f_t) \\
C_t &= R_t - \frac{1}{q_t} \left[R_t F_t F_t' R_t \left(1 - \frac{q_t^*}{q_t}\right) \right].
\end{aligned}
$$

Intuitively this corresponds to channeling the information from y_t, via the posterior distribution of λ_t, to get the conditional moments

$$E[\theta_t \mid \lambda_t, D_{t-1}] \text{ and } V[\theta_t \mid \lambda_t, D_{t-1}].$$

Afterwoods, λ_t is eliminated by integration.

3.1.4 One-step-ahead forecasting

The one-step-ahead forecasting distribution, given by

$$p(y_t|D_{t-1}, \phi) = \int_0^1 p(y_t|\mu_t, D_{t-1}, \phi)p(\mu_t|D_{t-1})d\mu_t \tag{6}$$

cannot be evaluated analytically, thus demanding the use of some approximation, like the use of Newton-Cotes type methods (quadrature approximations) (see Corbit, 1996). However, the first moments of $(y_t|D_{t-1}, \phi)$ can be found exactly, by first principles, and the fact that $(\mu_t \mid D_{t-1}) \sim Beta(r_t, s_t)$. Therefore,

$$\begin{aligned} E(y_t|D_{t-1}, \phi) &= \frac{r_t}{r_t + s_t}, \quad \text{and} \\ V(y_t|D_{t-1}, \phi) &= \frac{1}{1+\phi}\left[\frac{r_t}{r_t+s_t}\left(1-\frac{r_t}{r_t+s_t}\right) + \frac{\phi r_t s_t}{(r_t+s_t)^2(r_t+s_t+1)}\right]. \end{aligned} \tag{7}$$

These moments are very useful for model comparison purposes and to obtain credibility intervals. For example, based on $E(y_t|D_{t-1}, \phi)$, one can evaluate the mean absolute deviation, $MAD = \frac{1}{T}\sum_{t=1}^{T} |e_t|$, and the mean square error, $MSE = \frac{1}{T}\sum_{t=1}^{T} e_t^2$, with $e_t = y_t - E(y_t|D_{t-1}, \phi)$. Additionally, another summary statistics that is useful for model comparison is the *observed predictive density* or *observed likelihood*, which is described by

$$p(y_1, \dots, y_T \mid D_0, \phi) = \prod_{t=1}^{T} p(y_t \mid D_{t-1}, \phi). \tag{8}$$

4 The case of unknown ϕ

In this section we deal with the estimation of ϕ. Let $p(\phi \mid D_T)$ be a posterior distribution of ϕ considering the whole data, that is,

$$p(\phi \mid D_T) \propto \left[\prod_{t=1}^{T} p(y_t \mid D_{t-1}, \phi)\right] p(\phi). \tag{9}$$

The unconditional to ϕ one-step ahead forecasting distribution is given by

$$p(y_t \mid D_{t-1}) \approx \int p(y_t \mid D_{t-1}, \phi) p(\phi \mid D_T) d\phi. \tag{10}$$

As mentioned before, the parameter ϕ may be interpreted as a "sample size". Thus, we take a discrete uniform prior distribution for ϕ, i.e., $\phi \sim Unif(1, M)$, with M large. Using such prior, expression (10) is expressed as a mixture of distributions:

$$p(y_t \mid D_{t-1}) \approx \sum_{j=1}^{M} p(y_t \mid D_{t-1}, \phi_j) p(\phi_j \mid D_T). \tag{11}$$

Notice that to evaluate $p(y_t \mid D_{t-1}, \phi)$ it is also necessary to use a numerical integration with respect to μ_t (see equation 6).

5 Applications

5.1 The signal-to-noise ratio in the DBM

In this section we analyze the relationship between the variance of the observation error, V_t, and the variance of the evolution error, W_t, for the DBM. This is useful to provide some initial knowledge about the behavior of the time series we want to model. Next we present some analyzes for simulated data.

The errors, ω_t, of the system equation control its evolution through the variance W_t. That is, the larger (smaller) are the values of W_t the more erratic (smoother) will be the evolution over time. The fact that $E(\omega_t) = 0$ ensures a certain fixed level around which the values of θ_t will vary.

The behavior of the y_t and θ_t trajectories will be related to the magnitude of the ratio $r_t = W_t/V_t$ (a signal-to-noise ratio). When r_t is small, most of the series movements are due to the observations y_t, whereas when r_t is large the movements are due to both the variations in the y_t's and the θ_t's.

For normal dynamic models with constant observational and evolution variances, a value of $r_t = 0.05$, that is, $W = V/20$, indicates a time series that is typically smooth with locally constant level. On the other hand, a value of $r_t = 0.5$ indicates a series that behaves much more erractically.

Naturally, one cannot directly use the example above as a guide to describe the relationship between the observational and the evolution variances in the dynamic beta model, since the nature of the data analyzed in the Gaussian and in the beta case are too different. Besides this, the fact that V_t depends on μ_t (which is not the case in the normal model) adds additional complexities. So for the dynamic beta models, the r_t values mentioned above may not lead to the same signal-to-noise interpretation of the normal dynamic models.

For the dynamic beta model, the observational variance, V_t, is such that $V_t \leq \frac{1}{4(1+\phi)}$. The upper limit of V_t is reached for $\mu_t = 1/2$. As μ_t approaches either 0 or 1, V_t approaches 0. The magnitude of V_t is quite small and decreases with ϕ. The case of $V_t = 0$ leads to a static time series of observations. For the simulated data we used μ_t values as low as 0.01 and as high as 0.99. That leads to $\frac{0.0099}{1+\phi} \leq V_t \leq \frac{1}{4(1+\phi)}$.

In order to establish a correct comparision between V_t, which is defined on $(0,1)$, and W_t, which is defined on $(0,\infty)$, we need first to rescale W_t. Taking the transformation $w_t^* = \frac{e^{w_t}}{1+e^{w_t}}$, where w_t is the error term in the DBM system equation, we obtain $W_t^* = W_t/16$ as an approximated variance for w_t^*. Then $\frac{0.0099 r_t}{(1+\phi)} \leq W_t^* \leq \frac{r_t}{4(1+\phi)}$, where $r_t = W_t^*/V_t$, implying, for example, that $W_t^* \leq 0.0125/(1+\phi)$, for a signal-to-noise of $r_t = 0.05$.

5.2 Analysing simulated data

Taking into account the discussions in the previous section, we generated dynamic beta data for first-order models with $F_t = 1$ and $G_t = 1$ and considered three scenarios (or cases), which are summarized in Table 1.

Cases	I	II	III
ϕ	100	25	15
W_t^*	0.000625	0.009375	0.0125

Table 1
(ϕ, W_t^*) scenarios for dynamic beta simulated data.

Figure 1 illustrates the generated series and the estimated levels obtained using the DBM. For beta series generated under **Case I** (*well-behaved* ones), both W_t^* and V_t are small and the signal-to-noise ratio is around 0.2525 (considering the upper bound of V_t). Under **Case II**, the generated beta series is *reasonably stable* due to the intermediate values of V_t and W_t^*. Under **Case III**, the generated series present *sharp oscilations* due to high values of both V_t and W_t^*. For this case, the signal-to-noise ratio is around 0.80. For all the cases, we observe that the DBM provides a good description of the data. The values of W_t^* are, in general, unknown so that *discount factors*, δ, are used instead (see West and Harrison (1997) p. 51). Such values are defined in the interval $(0,1]$ and the degree of adaptation to new data increases as δ decreases, leading to more erractic forecast sequences. When $\delta = 1.0$ it corresponds to

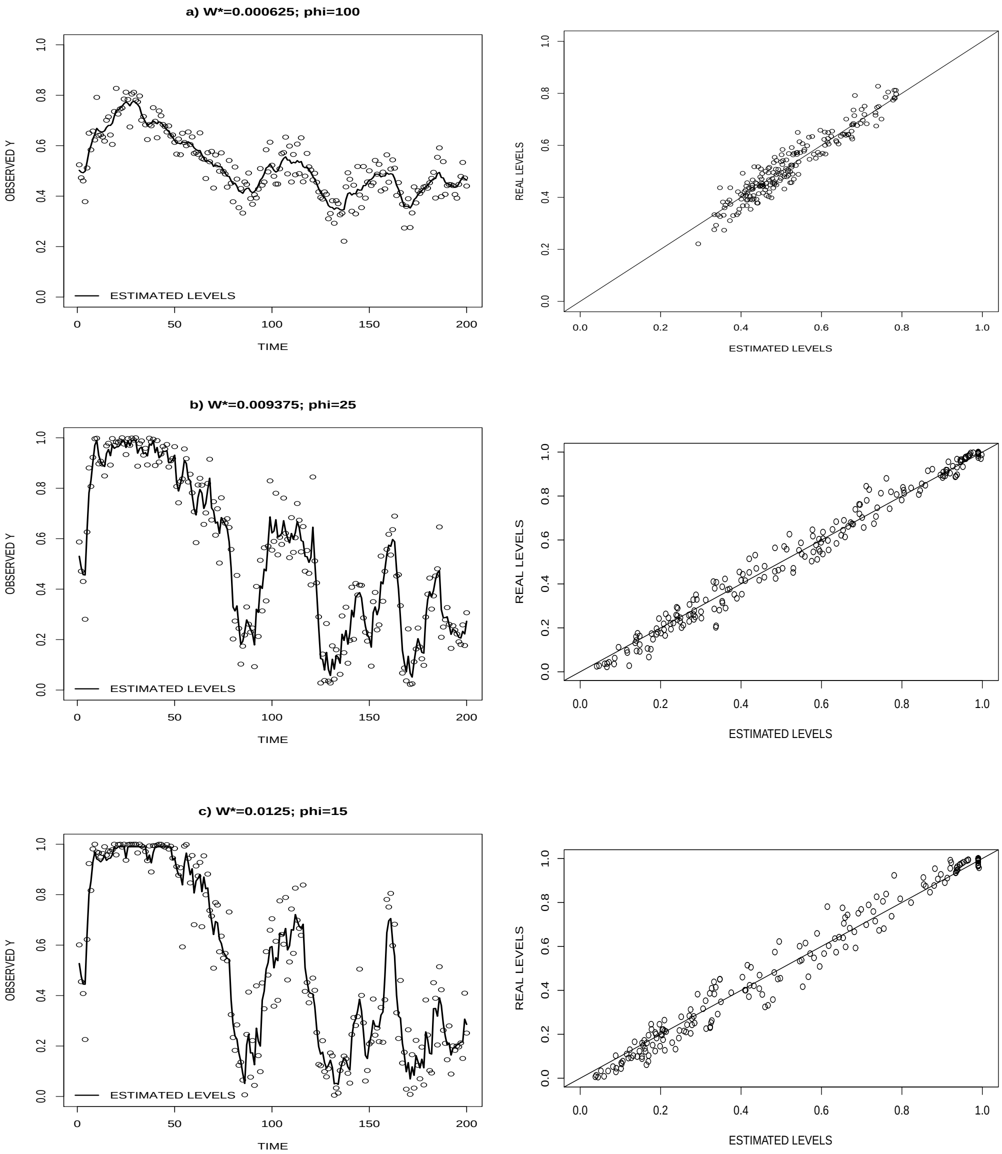

Fig. 1. Left pannels: observed data and estimated levels for generated dynamic beta series under the three cases. Right pannels: scatterplots of the true and estimated levels.

a degenerate static model with $W_t^* = 0$, characterizing the observations as a simple beta random sample. For generated data from Cases I to III, we worked with the real values of W_t^* as well as a choice of $\delta = 0.8$. This discount factor was found to provide the best predictive performance to the dynamic beta models we proposed in Section 3.

As a criterion to evaluate the goodness of fit, we used the MSE between the real μ_t and the estimated levels $\tilde{\mu}_t$ (the first k estimates are not used in the criterion since, quite certainly, the algorithm has not yet "learned" at observation $t = k$). From Table 2 we observe that for a fixed value of W_t^*, as ϕ increases, the magnitude of the MSE decreases. That is not surprising since as ϕ increases, the variance of the observations, V_t, decreases and the processes become more stable.

The estimated values of f_t^* and q_t^* (see Section 3.1.3 item (b)) obtained using either first- or second-order Taylor expansions were very close, justifying the use of the simpler approximation.

Cases	MSE $\times 10^2$
I: $\phi = 100$; $W^* = 0.000625$	0.2638
II: $\phi = 25$; $\mathrm{W}^* = 0.009375$	1.8628
III: $\phi = 15$; $W^* = 0.012500$	2.4142

Table 2
Mean squared errors (MSE) between the real, μ_t, and the estimated levels $\tilde{\mu}_t$. Discount factor: $\delta = 0.8$.

5.3 Retrospective Analysis

In time series analysis, in addition to learning and prediction, looking back at the end of a given period may provide a clearer understanding of what really happened during this period.

For the data set comprising observations from time 1 to T, such *retrospective or smoothed assessment* utilizes the filtered distributions $p(\theta_t \mid D_T)$, for $t = 1, \ldots, T$.

For a given linear dynamic model described by the quadruple (F_t, G_t, V_t, W_t), for every time t and $1 \leq h \leq t$, we have:

$$
\begin{aligned}
(\theta_{t-h} \mid D_t) &\sim [a_t(-h), R_t(-h)] \\
a_t(-h) &= m_{t-h} - B_{t-h}[a_{t-h+1} - a_t(-h+1)] \\
R_t(-h) &= C_{t-h} - B_{t-h}[R_{t-h+1} - R_t(-h+1)]B_{t-h}^{'} \\
B_t &= C_t G_{t+1}^{'} R_{t+1}^{-1} \\
a_t(0) &= m_t \quad \text{and} \quad R_t(0) = C_t.
\end{aligned}
$$

The above expressions can be found by induction arguments. More details can be found in Chapter 4 of West and Harrison (1997). In Section 14.3.4, West and Harrison (1977) indicate the details for the retrospective analysis for dynamic generalized linear models. The authors observe that for later models the expressions given above remain the same.

5.4 Applying the dynamic beta model to real data

In this section we apply the methodology described in Sections 3 and 4 to fit time series of proportions. Two datasets will be analyzed.

Application 1 - Brazilian monthly unemployment rates.

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/`.

We analyze monthly unemployment rates (MUR) based on PME data in the period from March 2002 to December 2009 (94 observations). The MUR are expressed in percentages, so we used the dynamic beta model. It is reasonable to presume that the MUR are affected by a systematic seasonal behavior that is dictated by yearly cycles caused, for instance, by climatic factors, Christmas festivities and school vacations. In the months of November and December the MUR drop substantially due to the growth in temporary employment. Besides this, due to some macroeconomic factors, the MUR might present either rising or declining trends in a given period. Taking into account these remarks, for the MUR data we used a second-order polynomial trend seasonal effects dynamic beta model (STSDBM). Thus, the parameter vector is given by

$$\theta_t = (\beta_{t1}, \beta_{t2}, \psi_{t1}, \ldots, \psi_{tp})',$$

where in the *trend vector* described by $(\beta_{t1}, \beta_{t2})'$, coordenate β_{t1} represents the current level while β_{t2} represents the rate of change in the level. For a seasonal cycle of size p, the *seasonal effects* are described by the remaining p parameters.

The following triple represents the STSDBM:

$$\left\{ F = \begin{pmatrix} E_2 \\ E_p \end{pmatrix}, G = \begin{pmatrix} J_2(1) & \mathbf{0} \\ \mathbf{0} & P \end{pmatrix}, \mathbf{W}_t = \begin{pmatrix} \mathbf{W}_{t,\beta} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{t,\psi} \end{pmatrix} \right\}.$$

where $\mathbf{W}_t$ is a block-diagonal covariance matrix and

$$E_p = (1, \mathbf{0}'_{p-1}), \quad J_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad P = \begin{pmatrix} \mathbf{0} & \mathbf{I}_{p-1} \\ 1 & \mathbf{0}' \end{pmatrix}.$$

The *permutation matrix* P is $p - cyclic$, so that $P^{np} = I_p$ and $P^{h+np} = P^h$, for $h = 1, \ldots, p$, and any integer $n \geq 0$.

The matrix $\mathbf{W}_t$ is described with the help of block discounting factors. Recalling that $(\theta_{t-1}|D_{t-1}) \sim (m_{t-1}, C_{t-1})$, let $C_{T,t-1}$ represent the posterior covariance submatrix for the trend components at time $t-1$ and $C_{S,t-1}$, the posterior covariance submatrix for the seasonal components. Thus, the blocks $\mathbf{W}_{t,\beta}$ and $\mathbf{W}_{t,\psi}$ in $\mathbf{W}_t$ are given by

$$\mathbf{W}_{t,\beta} = \left(\frac{1-\delta_T}{\delta_T}\right) J_2(1) C_{T,t-1} J_2(1)' \quad \text{and} \quad \mathbf{W}_{t,\psi} = \left(\frac{1-\delta_S}{\delta_S}\right) P C_{S,t-1} P',$$

where δ_T and δ_S are the discount factors associated with these components. For more details about seasonal models, see Chapter 8 of West and Harrison (1997).

For the monthly unemployment rates we used $\delta_T = 0.90$ and $\delta_S = 0.98$. These values were defined by comparing the MAD, MSE and the observed Log Likelihood for several choices of δ_T and δ_S, as summarized in Table 3.

Discounts	MSE	MAD	Log Likelihood
$\delta_T = 0.90$; $\delta_S = 0.98$	3.707653e-05	0.004955393	-427.4887
$\delta_T = 0.90$; $\delta_S = 0.99$	4.847885e-05	0.005747651	-427.8524
$\delta_T = 0.95$; $\delta_S = 0.98$	6.073741e-05	0.006260674	-427.7525
$\delta_T = 0.95$; $\delta_S = 0.99$	8.032208e-05	0.007105512	-427.8655

Table 3
Forecast performance summary

In Figure 2 (topmost figure) we present the back smoothed or retrospectively fitted MUR. The values shown represent the estimated component values for the best model for the MUR data. As can seen, the general aspect of the estimated values seems to describe the presence of seasonal effects and also a declining trend. According to a note posted on January 26, 2009, at the website `http://brazilportal.wordpress.com/2009/01/26/`, Brazil's MUR has been consistently dropping since 2002. "Brazil's unemployment rate dropped to its lowest point in seven years" Some economic analysts explain that this favorable scenario is due essentially to the continued consolidation of macroeconomic adjustment following the floating of the real (the Brazilian currency) in 1999, combined with some governmental policies that were effective in containing inflation and public debts, reducing external vulnerabilities. The good shape of the banking sector also helped to anchor the Brazilian economy. However, according to some economists, this favorable trend may change due to the global economic crisis and future government policies.

Figure 2 also presents the smoothed estimates for the level, growth and seasonal components. As can be observed, the estimated levels change

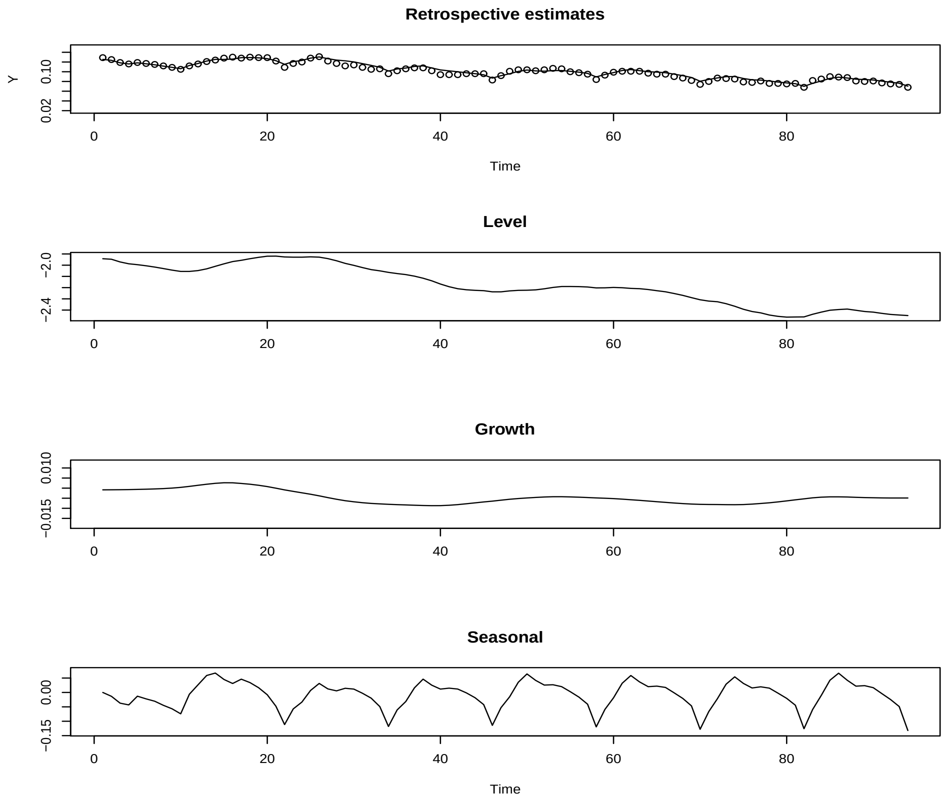

Fig. 2. Timeplots of smoothed seasonal, growth and level estimates for the Brazilian monthly unemployment rates (MUR) based on PME data in the period from March 2002 to December 2009.

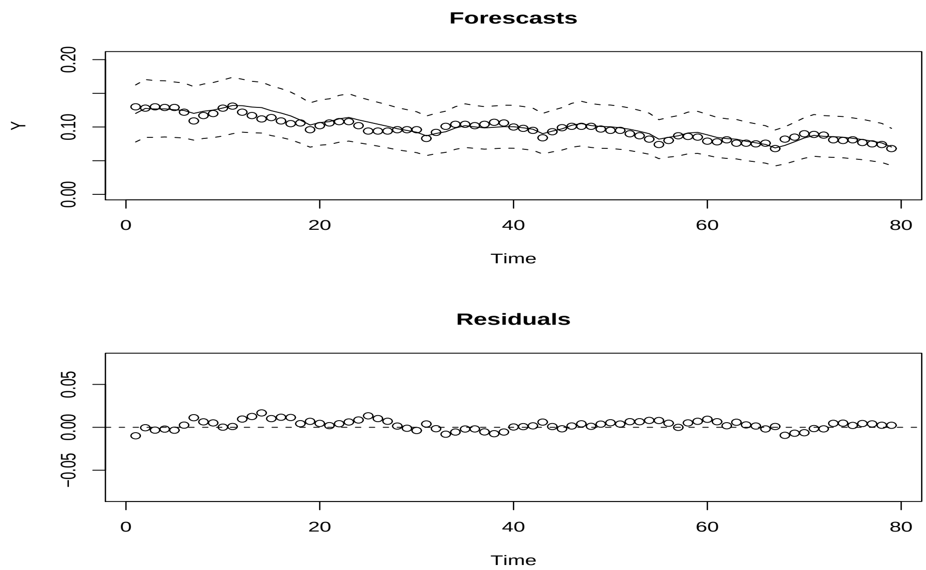

Fig. 3. Predictions for the Brazilian monthly unemployment rates (MUR) based on PME data in the period from March 2002 to December 2009.

very smoothly but with a downward sloping trend. The rates of changes in the levels (growth) of the MUR data reached their highest value between 2003 and 2004. Afterwards, the rates show a downward shift followed by a period in which the rates fluctuate around a given mean. For the seasonal component we observe that (i) there is a definitive seasonal variation in the MUR and (ii) the pattern is quite regular over the years.

Figure 3 presents the one-step-ahead predictions together with one standard deviation limits using a first-order approximation for the variance of the predictive distribution at time t. For these calculations we used $V(y_t \mid D_{t-1}, \hat{\phi})$, with $\hat{\phi} = E(\phi \mid D_T) = 117$, which corresponds to a Bayesian estimator considering a quadratic loss. Due to the additional uncertainty induced by ϕ, the credibility intervals for the predictions calculated using $V(y_t \mid D_{t-1})$ were somewhat wider than the former, so we decided not to present them. In Figure 3 we observe that the residuals indicate a fairly reasonable fit.

Another possibility to obtain approximate credibility intervals is the use of highest posterior density (HPD) for $(y_t|D_{t-1})$. Since we are using a discrete approximation of $p(y_t|D_{t-1})$, a HPD can be constructed by adapting the definition, as well as the algorithm, given by Turkkan and Pham-Gia (1993):

Let $\{p(y_{ti} \mid D_{t-1}) = p_{ti},\ i = 1, \ldots, n \text{ with } \sum_{i=1}^{n} p_{ti} = 1\}$ and denote a 100(1-α)% HPD for y_t by $\mathbf{C}_{1-\alpha}$. This set will consist of points such that

$$\sum_{i\, \in\, \mathbf{C}_{1-\alpha}} p_{ti} \leq \alpha \text{ with } min\{p_{ti}\}_{i\, \in\, \mathbf{C}_{1-\alpha}} > max\{p_{ti}\}_{i\, \notin\, \mathbf{C}_{1-\alpha}}$$

and for any subset $\mathbf{C}' \supset \mathbf{C}_{1-\alpha}$, we have $\sum_{i\, \in\, \mathbf{C}'} p_{ti} > 1 - \alpha$.

The 100(1-α)% HPD region $\mathbf{C}_{1-\alpha}$ can be found by the following algorithm:

(a) Take a value k sufficiently large such that $p_{ti} < k$, $\forall\ i$.
(b) Compute $K = \sum_i p_{ti}$ after setting $p_{ti} = 0$ for any i such that $p_{ti} > k$.
(c) Check whether $K \geq \alpha$ for all i.
(d) The value $K = k_\alpha$ is determined as the smallest value K such that we have $K \geq \alpha$ and the resulting contour $\mathbf{C}$ reduces to a set of points i which is the $\mathbf{C}_{1-\alpha}$ itself.

For small forecasting horizons outside the original sample, the description of HPD is computationally viable since it is possible to use $\hat{\phi}$ (since ϕ is assumed constant over time) and to obtain updates using the evolution equation considering growth and seasonality.

Application 2 - Proportions of clay at different water depths in an Arctic lake.

Coakley and Rust (1968) present the composition of 39 sediment samples in terms of sand, silt and clay percentages. The samples have been taken at different water depths in an Arctic lake. These data were extensively studied, especially in compositional data analysis (Aitchison, 1982; Aitchison, 2003).

Some discussion about the dependency of the sediment compositions at water depths can be found in Aitchison (2003). In this work we analyze this same data set but we are interested in modeling the proportions of clay at different (increasing) water depths, unlike the usual applications in time series.

We have no reason to include other effects in the analysis besides trend. Therefore, we used a second-order polynomial trend effects dynamic beta model. The following triple represents the model for the proportions of clay, with $\mathbf{W}_{t,\beta}$ described in terms of discount factors as before:

$$\{F = E_2, G = J_2(1), \mathbf{W}_t = \mathbf{W}_{t,\beta}\},$$

For these data however, we used $\delta_{TL} = 0.80$ (as the discount factor for the level component), $\delta_{TG} = 0.90$ (as the discount factor for the seasonal component), and obtained $\hat{\phi} = 34$. This choice resulted in the lowest values of both MSE and MAD and highest value of the observed Log Likelihood (MSE=0.01045015, MAD=0.07770061 and Log Likelihood=-157.5418).

In Figure 4 we present the retrospectively fitted proportions of clay. As can observed, it is quite clear that the proportions of clay increase with the water depth, corroborating the findings in Aitchison (2003). Figure 4 also shows the smoothed estimates of the level and growth components. As can be seen, the estimated level changes are the main factors responsible for the upward sloping trend in the proportions of clay, while for the growth component, the rates of changes in the levels for proportions of clay present a downward behavior after a given depth.

The one-step-ahead predictions (see Figure 5) together with one standard deviation limits show that some points are not very well described by the fit. The residuals are also large, though no alarming patterns can be perceived. It is possible that by collapsing the compositional data into just one category (clay), some important data features could not be captured. This problem calls for the development of dynamic models for compositional data.

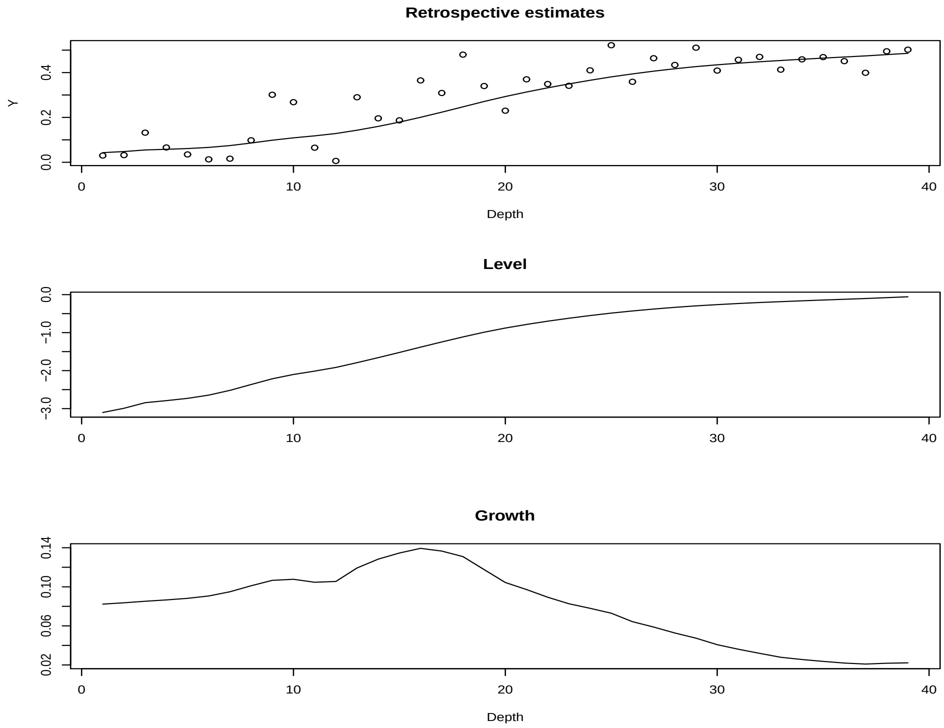

Fig. 4. Timeplots of smoothed growth and level estimates for the proportions of clay at different water depths in an Arctic lake.

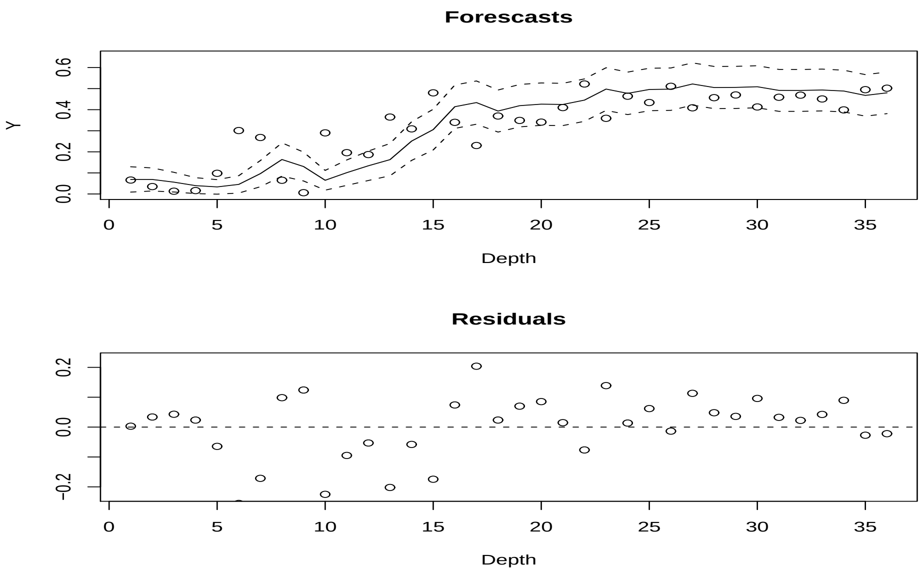

Fig. 5. Predictions for the proportions of clay at different water depths in an Arctic lake.

6 Discussion

We developed a Bayesian dynamic beta model (DBM) for modeling and forecasting single time series of proportions. The inferential process was based on some forms of approximate Bayesian analysis, including linear Bayesian estimation. The methodology was applied to both real and simulated data.

We envision some possible directions in order to improve/extend the current DBM:

(i) Some natural extensions of our model include the description of a hierarchical DBM using a common mean for a multivariate time series. The component means would be linked according to a random effects model.

(ii) Additionally, to enable the inclusion of different regimes for the level of the process, a possible approach to the problem is the use of Markov switching models. In the finance area there is an application proposed by Bruche and Gonzáles-Aguado (2010). In a financial setting, they developed an econometric model to describe the behavior of default probabilities to evaluate expected financial losses and recovery rates considering economic cycles.

(iii) Further and more general developments include the creation of dynamic models for compositional data and the development of the inferential process involved using linear Bayesian methods, as an extension of the work proposed by Grunwald, Raftery and Guttorp (1993).

Acknowledgements

The authors are deeply indebted to the Office to Improve University Research (CAPES-Brazil), for financial support via Project PROCAD-NF 2008. 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. Cibele da-Silva was supported by CNPq-Brazil, BPPesq.

References

Albi, M., Migon, H.S., da-Silva, C.Q. (2009). *Bayesian Inference for the Beta Distribution.* Technical Report - RTPE07 - Departamento de Estatística - Universidade de Brasília.

Aitchison, J., (1982). *The Statistical Analysis of Compositional Data. J.R. Statist. Soc. B*, **44**, 2, 139-177.

Aitchison, J., (2003). *A Concise Guide to Compositional Data Analysis.* CDA Workshop, Girona.

Azzalini, A. (1984). A Markov process with Beta marginal distribution. *Statistica*, **44**, 241-243.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory.* Chichester: Wiley.

Branscum A.J., Johnson W.O., Thurmond M.C. (2007). Bayesian beta regression: Applications to household expenditure data and genetic distance between foot- and-mouth disease viruses. *Australian & New Zealand Journal of Statistics*, **49**(3), 287-301.

Bruche, M. and Gonzáles-Aguado, C. (2010). Recovery Rates, Default Probabilities, and Credit Cycle. *Journal of Banking and Finance*, **34**, No. 4.

Coakley J.P. and Rust, B.R. (1968). Sedimentaion in an Arctic lake. *J. Sedimentary Petrology*, **38**, 1290-1300.

Corbit, D. (1996). Numerical Integration: From Trapezoids to RMS: Object-Oriented Numerical Integration. *Dr. Dobb's J.*, No. 252, 117-120.

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

Gamerman, D. and Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, 2nd. ed. Londres: Chapman Hall/CRC, v. 1. 336 p.

Godolphin, E.J. and Triantafyllopoulos, K. (2006). Decomposition of time series models in state-space form. *Computational Statistics and Data Analysis*, **50**, 2232-2246.

Grunwald, G.K. Raftery, A. E. and Guttorp, P. (1993). Time Series of Continuous proportions. *J.R. Statist. Soc.* **B**, **55**, No. 1, 103-116.

Kieschnick, R., McCullough, B.D., (2003). Regression analysis of variates observed on (0, 1): Percentages, proportions and fractions. *Statistical Modelling* **3**, 193-213.

Lindsey, J.K. and Lambert, P. (1995). Dynamic generalized linear models and repeated measurements. *Journal of Statistical Planning and Inference*, **47**, 129-139.

Migon, H.S. and Gamerman, D. (1999). *Statistical Inference: An Integrated Approach.* First edition, Hodder Arnold Publication, London.

McCullagh, P. and Nelder, J.A. (1994). *Generalized Linear Models*, The University Press, Cambridge: Monographs on Statistics and Applied Probability 37, 2nd. ed.

McKenzie, E. (1985). An autoregressive process for Beta random variables. *Mangmnt. Sci.*, **31**, 988-997.

Nelder, J. A.; Wedderburn, R. W. M. (1972). Generalized linear models. *Journal of the Royal Statistical Society*, A, **135**, 370-384.

Paolino, P., (2001). Maximum likelihood estimation of models with beta-distributed dependent variables. *Political Analysis* **9**, 325-346.

Rue, H., Martino, S., Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B, 71(2), 319-392.

Tierney, L. and Kadane, J.B. (1986). Accurate approximations for the posterior moments and marginal densities. *Journal of the American Statistical Association*, **81**, 82-6.

Turkkan, N. and Pham-Gia, T. (1993). Computation of the highest posterior density interval in Bayesian analysis. *journal of statistical computation and simulation*, **44**, 243-250.

Walis, K.F. (1987). Time series analysis of bounded economic variables. *J. Time Ser. Anal.*, **8**, 115-123.

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

West, M., Harrison, P.J. and Migon, H.S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). *Journal of the American Statistical Association*, **80**, 73-97.

Appendix

An alternative approach to the one presented in Section 3.1.2 for describing the hyperparameters r_t and s_t is based on a second-order Taylor approximation.

Remembering that $\mu_t = g^{-1}(\lambda_t)$, where $g(\mu_t) = \log(\mu_t/(1-\mu_t))$, we can get a second-order Taylor approximation of their mean and variance as:

$$E\left[\frac{\exp(\lambda_t)}{1+\exp(\lambda_t)}\bigg|D_{t-1},\phi\right] \approx \frac{e^{f_t}}{1+e^{f_t}} + \frac{q_t}{2}\frac{e^{f_t}(1-e^{f_t})}{(1+e^{f_t})^3}; \tag{12}$$

$$V\left[\frac{\exp(\lambda_t)}{1+\exp(\lambda_t)}\bigg|D_{t-1},\phi\right] \approx \left(\frac{e^{f_t}}{(1+e^{f_t})^2}\right)^2 q_t + \frac{(8q_t f_t^2 - q_t^2)}{4}\left(\frac{e^{f_t}(1-e^{f_t})}{(1+e^{f_t})^3}\right)^2. \tag{13}$$

Since $(\mu_t|D_{t-1},\phi) \sim Beta(r_t, s_t)$, then the following system of equations is obtained

$$\frac{r_t}{r_t+s_t} = h_t \quad \text{and} \quad \frac{r_t s_t}{(r_t+s_t)^2(r_t+s_t+1)} = v_t,$$

where h_t and v_t are, respectivelly, given by the expressions in the right-hand side of equations (12) and (13). The hyperparameters r_t and s_t of the beta prior for μ_t are obtained by solving the following system of equations:

$$r_t = \frac{(1-h_t)h_t^2}{v_t} - h_t \quad \text{and} \quad s_t = \frac{(1-h_t)^2 h_t}{v_t} - (1-h_t).$$

A first-order Taylor expansion is obtained when the last term in the right-hand side of (12) and (13) are omitted.