Bayesian Analysis (2013) **8**, Number 3, pp. 723–740

A Time Series Model for Responses on the Unit Interval

A. Jara *, L. E. Nieto-Barajas † and F. Quintana ‡

Abstract. We introduce an autoregressive model for responses that are restricted to lie on the unit interval, with beta-distributed marginals. The model includes strict stationarity as a special case, and is based on the introduction of a series of latent random variables with a simple hierarchical specification that achieves the desired dependence while being amenable to posterior simulation schemes. We discuss the construction, study some of the main properties, and compare it with alternative models using simulated data. We finally illustrate the usage of our proposal by modelling a yearly series of unemployment rates.

Keywords: Autoregressive models, beta processes, latent variables, unemployment rates

1 Introduction

In this article we introduce a novel autoregressive model for time series of observations which are restricted to lie on the unit interval. Our proposal is based on a hierarchical representation using a series of beta random variables, which are linked via a set of exchangeable latent indicators in an autoregressive fashion. The construction implies a beta marginal distribution and includes strict stationarity and non stationarity as special cases.

Classical time series analysis focuses on models for observations on unbounded support. This is the case for traditional autoregressive and moving average models with white noise errors (Box and Jenkins 1970). However, time series with observations on bounded support has received little attention. In particular, Wallis (1987) uses a logistic transformation to convert the observables in the unit interval to an unbounded support and then apply standard models. While other transformations could be equally considered, a limitation of this modeling approach is that it may not be possible to adequately represent data sets that concentrate near the boundaries.

In the Bayesian framework, dynamic models present an alternative for time series analysis. West et al. (1985) developed the general theory for dynamic generalized linear models, and da Silva et al. (2011) concentrated on the beta case with a suitable parametrization. The latter proposal includes an observation equation with beta-distributed marginal responses, and a linear system equation that is only partially specified. Indeed, following West et al. (1985), they proposed estimating the system equation using only moments of order up to two, but without a distributional assumption.

*Departamento de Estadística, Pontificia Universidad Católica de Chile ajara@mat.puc.cl
†Departamento de Estadística, ITAM, Mexico lnieto@itam.mx
‡Departamento de Estadística, Pontificia Universidad Católica de Chile quintana@mat.puc.cl

 DOI:10.1214/13-BA844

For multivariate proportions in the simplex, Grunwald et al. (1993) use a generalized Dirichlet dynamic model, and Bekele et al. (2012) induce a dynamic model via a Markov structure. As will be shown later, our model provides a flexible autocorrelation structure, the posterior simulation is straightforward to implement, and the user can easily specify the desired order of dependence. Furthermore, the model performs well and can outperform the existing models.

Describing the layout of the paper, we define our model and study its properties in Section 2. We carry out a simulation study to assess the performance of model with respect to competing models in Section 3. Section 4 illustrates the model with a real dataset of unemployment rates in Chile, and also compares the proposed model against other alternatives. We finally discuss extensions of our model in Section 5.

Before we proceed we introduce notation: $\mathrm{N}(\mu, \sigma^2)$ denotes a normal density with mean μ and variance σ^2; $\mathrm{Be}(a, b)$ denotes a beta density with mean $a/(a+b)$; $\mathrm{Ga}(a, b)$ denotes a gamma density with mean a/b; $\mathrm{Un}(a, b)$ denotes a continuous uniform density on the interval (a, b); $\mathrm{Bin}(n, p)$ denotes a binomial density with size n and probability of success p; $\mathrm{Po}(\lambda)$ denotes a Poisson density with mean λ.

2 The Model

Let $\{y_t\}$, $t = 1, 2, \ldots$ be a sequence of random variables with values on the unit interval, $P(0 \leq y_t \leq 1) = 1$ for all t. We want to define a sampling probability model for $\{y_t\}$ that includes serial dependence on q lagged terms. Instead of defining the dependence directly through the y_t's, we will specify the model in terms of a sequence of non-negative integer-valued latent variables $\{u_t\}$. These latent variables will share a common parameter w. To have a better idea of how the variables are linked, Figure 1 shows a graphical representation of the relationship among the observables $\{y_t\}$, the latent sets $\{u_t\}$ and w for an autoregressive dependence of order two.

In general, we assume that

$$y_t \mid u_t, u_{t-1}, \ldots, u_{t-q} \overset{\text{ind}}{\sim} \mathrm{Be}\left(a + \sum_{j=0}^{q} u_{t-j}, b + \sum_{j=0}^{q} (c_{t-j} - u_{t-j})\right), \tag{1}$$

for $t = 1, 2, \ldots$, where a and b are positive parameters, the latent variables u_t are defined to be zero with probability one for $t \leq 0$, $c_t = 0$ for $t \leq 0$, and $c_t \in \mathbb{N}$ otherwise. This implies that for $t \leq q$, $y_1|u_1 \sim \mathrm{Be}(a + u_1, b + c_1 + u_1)$, $y_2|u_2, u_1 \sim \mathrm{Be}(a + u_1 + u_2, b + c_1 - u_1 + c_2 - u_2)$ and so on. In turn, we specify the distribution of the latent variables u_t in terms of a common parameter w such that

$$u_t \mid w \overset{\text{ind}}{\sim} \mathrm{Bin}(c_t, w), \tag{2}$$

for $t = 1, 2, \ldots$, and

$$w \sim \mathrm{Be}(a, b), \tag{3}$$

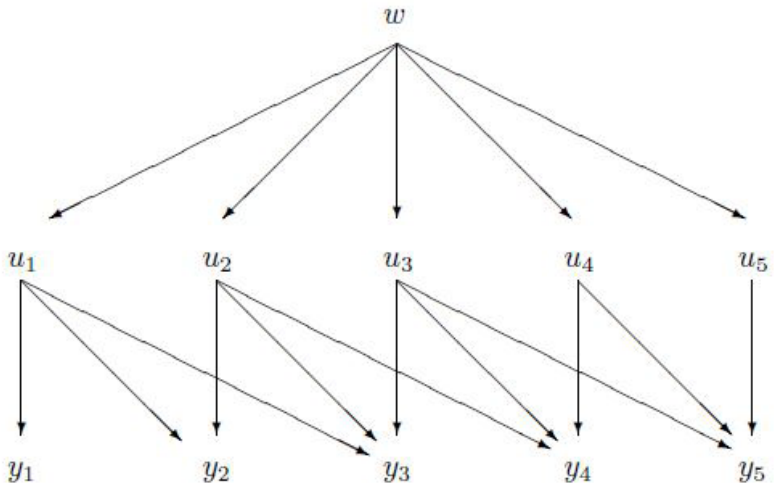

Figure 1: Graphical representation of autoregressive beta process of order $q = 2$, BeP_2.

with $a > 0$, $b > 0$ and $c_t \in \mathbb{N}$, and with the convention that $c_t = 0$ implies $u_t = 0$. Note that specifications (2) and (3) define an exchangeable sequence for $\{u_t\}$. The role of the latent u_t's is to establish a link between observations, whereas the latent w acts as anchor determining the overall level of the series. We refer to the process $\{y_t\}$ induced by (1)–(3), as an order-q dependent beta process with parameters a, b and $\mathbf{c} = \{c_1, c_2, \ldots\}$, and denote it by $\{y_t\} \sim \text{BeP}_q(a, b, \mathbf{c})$.

It is not difficult to see that construction (1)–(3) defines a process whose marginal distributions are $y_t \sim \text{Be}(a, b)$ for all t. To see this we note that, from (2), the u_t are conditionally independent binomial variables with common success probability w. Therefore the sum of these variables is also binomial, conditional on w, that is

$$\sum_{j=0}^{q} u_j \mid w \sim \text{Bin}\left(\sum_{j=0}^{q} c_j, w\right),$$

so that, unconditionally, its distribution becomes a beta-binomial distribution $\text{BeBin}(a, b, \sum_{j=0}^{q} c_j)$. Finally, since the conditional distribution of y_t only depends on $u_t, u_{t-1}, \ldots, u_{t-q}$ through its sum $\sum_{j=0}^{q} u_{t-j}$, it follows (Bernardo and Smith 1994, pg. 436) that each y_t is marginally beta-distributed. Note that the role of the third latent variable w is to make the marginal distribution of any sum of u_t's to be beta-binomial. Note also that by adequately choosing a or b the model allows for data that accumulates near the boundaries.

The correlation structure induced by this construction can be computed in closed form. This is given in the following proposition.

Proposition 1. *The autocorrelation function of the order-q dependent beta process, defined by equations (1)–(3), depends on the three parameters* $(a, b, \mathbf{c})$ *and is given by*

$$Corr(y_t, y_{t+s}) = \frac{(a+b)\left(\sum_{j=0}^{q-s} c_{t-j}\right) + \left(\sum_{j=0}^{q} c_{t-j}\right)\left(\sum_{j=0}^{q} c_{t+s-j}\right)}{\left(a+b+\sum_{j=0}^{q} c_{t-j}\right)\left(a+b+\sum_{j=0}^{q} c_{t+s-j}\right)}, \tag{4}$$

for $t, s \geq 1$.

Proof. Using conditional expectation and the marginal beta distributions of the process we obtain the result. ⋄

The numerator of expression (4) can be split in two parts, namely, a function of a, b and the shared $\mathbf{c}$ parameters, $c_{t+s-q}, \ldots, c_t$, plus a function of the two whole sets of $\mathbf{c}$ parameters that define y_t and y_{t+s}. If these two variables do not share any of the c_j parameters, as is the case when $s > q$, then the first term in the numerator becomes zero. When we let both a and b approach 0, then (4) converges to 1 for all $s, q, \mathbf{c}$, and when either a or b go to infinity the autocorrelation function converges to zero for all $s, q, \mathbf{c}$.

A particular case of the process arises when $c_t = c$ for all t, which makes $\{y_t\}$ be strictly stationary with $\text{Be}(a, b)$ marginal distributions, and with autocorrelation function $\text{Corr}(y_t, y_{t+s})$ that depends only on the distance s. In fact, the autocorrelation function of Proposition 1 reduces to

$$\text{Corr}(y_t, y_{t+s}) = \frac{(a+b)\max\{q-s+1, 0\}c + (q+1)^2c^2}{\{a+b+(q+1)c\}^2}. \tag{5}$$

It is easy to see that (5) converges to 1 when $c \to \infty$ and to 0 when $c \to 0$ for any fixed values of a, b, q, s. On the other hand, (5) converges to 1 when $q \to \infty$ and to $\{c/(a+b+c)\}^2$ when $q \to 0$ for any fixed values of a, b, c, s.

Figure 2 shows various examples of the autocorrelation function (4), fixing $a = b = 0.5$, $q = 3$, $t = 1$ and varying $s = 1, \ldots, 17$. Focusing on the left panel, cases 1 (solid line) and 2 (dashed line) are non-stationary, while cases 3 (dotted line) and 4 (dotted-dashed line) are stationary versions of the BeP. Observe that the autocorrelations can drop all the way to zero, and go up and/or down along a given sequence.

By varying the values of the $a, b, \mathbf{c}$ parameters we can obtain quite different expressions for the autocorrelation function. A few very special cases of this are depicted in the right panel of Figure 2. Therefore, it is natural to consider a hierarchical approach that models these parameters as well. Concretely, in addition to (1)–(3) we assume

$$c_t \mid \lambda \overset{\text{iid}}{\sim} \text{Po}(\lambda),\ t = 1, 2, \ldots, \qquad \lambda \sim \text{Un}(0, L) \tag{6}$$

and

$$a \sim \text{Un}(0, A), \qquad b \sim \text{Un}(0, B), \tag{7}$$

for known values of hyper-parameters L, A and B. When genuine prior information is unavailable, these choices provide reasonable default vague prior distributions. Alternative choices in (6) and (7) are certainly possible, but we found these to also provide

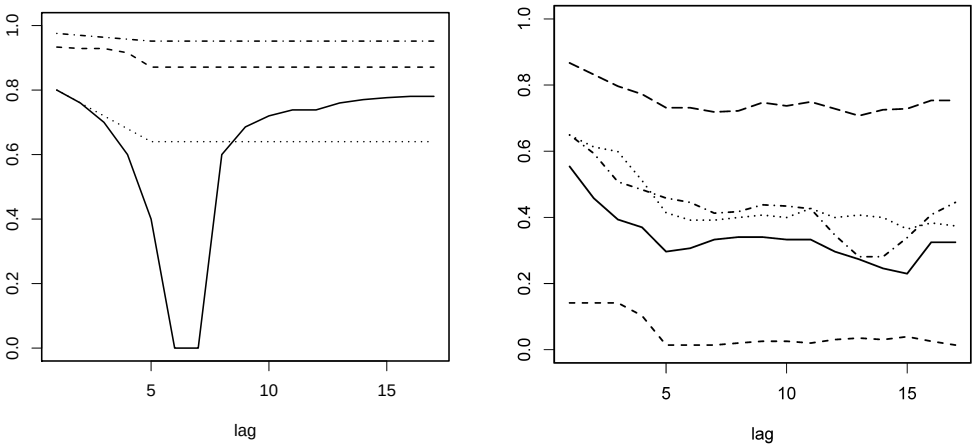

Figure 2: Autocorrelation function of $\text{BeP}_q(a, b, \mathbf{c})$ for $q = 3$. Left panel: Fixed values $a = b = 0.5$, $\mathbf{c} = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 3, 3, 3, 3, 3, 10, 10, 10, 10, 10, 1, 1, 1, 1, 1, \ldots)$ (solid line), $\mathbf{c} = (1, 0, 3, 10, 1, 0, 3, 10, \cdots)$ (dashed line), $c_t = 1$ for all t (dotted line), and $c_t = 10$ for all t (dotted-dashed line). Right panel: Five realizations using random values of $(a, b, \mathbf{c})$ according to (6) and (7) with $A = B = L = 10$.

more stable results in the posterior computations. Nevertheless, these prior distributions provide great flexibility in the model, and learning about them allows the $\text{BeP}_q(a, b, \mathbf{c})$ model to adapt to a greater range of datasets.

Posterior inference of our model is carried out by implementing a Gibbs sampler. The required full conditional distributions are postponed to the Appendix. Due to the prior choices, with distributions on bounded supports, sampling from the conditional distributions can be easily done by numerical integration of the density function and inverting. Alternatively, since the model is defined through standard distributions, posterior inference can also be obtained in `OpenBUGS` (available at `http://www.openbugs.info/`) or JAGS (available at `http://mcmc-jags.sourceforge.net/`). All the models considered here were implemented in JAGS. The codes are available upon request to the authors.

3 Simulation Study

We designed a simulation study to help us further understand the relative merits and performance of the proposed BeP model with respect to alternative models. Specifically, we compare how well these models fit batches of synthetic data, assessing at the same time their comparative predictive performances.

The simulated data were generated from the autoregressive model

$$y_i \mid y_{i-1}, \ldots, y_{i-q} \sim \text{Be}\left(a_0 + \sum_{j=1}^{q} y_{i-j},\, b_0 + \sum_{j=1}^{q}(1 - y_{i-j})\right), \qquad i = 1, \ldots, n \tag{8}$$

where $a_0 = b_0 = 1$, q is the true order of dependence and initially we choose $y_0 = y_{-1} = \cdots y_{-q+1} = 0$. We note that (8) is not a particular case of the proposed model (1), since the beta distribution parameters are computed in terms of actual lagged responses rather than latent variables.

We considered simulated time series of lengths $n = 30$ and $n = 50$ and values of q ranging in $\{1, \ldots, 5\}$ in (8). Each combination of n and q was repeated 100 times, leaving on every occasion the last 10 observations out of the analysis. These ten observations are used for assessing predictive performance. Therefore, the effective time series lengths for fitting the models are 20 and 40, respectively. For every combination of n, q and experiment repetition, we fitted three models.

We compare our model with two alternatives that we describe as follows. Let $z_t = \log\{y_y/(1 - y_t)\}$, for all t, the logit transformation of the original responses $y_t \in (0, 1)$. Then a simple autoregressive model on the $z_t's$ has the form

$$\begin{aligned} z_t \mid \mu_t, \sigma^2 &\overset{\text{ind}}{\sim} \text{N}(\mu_t, \sigma^2), \quad t = 1, \ldots, q \\ z_t \mid z_{t-1}, \ldots, z_{t-q}, \beta_0, \beta_1, \ldots, \beta_q &\overset{\text{ind}}{\sim} \text{N}\left(\beta_0 + \sum_{i=1}^{q} \beta_i z_{t-i}, \sigma^2\right), \quad t \geq q + 1 \\ \mu_1, \ldots, \mu_q \overset{\text{iid}}{\sim} \text{N}(m_0, \tau_m^2), \quad \sigma^{-2} &\sim \text{Ga}(s_0, s_1), \quad \beta_0, \ldots, \beta_q \overset{\text{iid}}{\sim} \text{N}(b_0, \tau_b^2). \end{aligned} \tag{9}$$

Model (9) is a normal $AR(q)$ model, with mean determined as a linear combination of the lagged responses for $t > q$, and a generic specification for the initial set of responses, $t = 1, \ldots, q$.

Generalizing the model proposed in da Silva et al. (2011) to the case of q lags, and modifying slightly the assumptions by taking instead a fully specified probability model, we get the following model

$$\begin{aligned} y_t \mid \mu_t, \phi &\overset{\text{ind}}{\sim} \text{Be}\left(\frac{\phi \exp(\mu_t)}{1 + \exp(\mu_t)}, \frac{\phi}{1 + \exp(\mu_t)}\right) \\ \mu_t = \beta_0 + \beta_1 \mu_{t-1} + \cdots + \beta_q \mu_{t-q} + \epsilon_t, &\quad \epsilon_t \overset{\text{iid}}{\sim} \text{N}(0, \tau_e^2), \quad t \geq q + 1 \\ \mu_1, \ldots, \mu_q \sim \text{N}(0, \tau_\mu^2), \quad \phi \sim \text{Ga}(a_\phi, b_\phi), &\quad \tau_e^{-2} \sim \text{Ga}(a_e, b_e), \quad \beta_0, \ldots, \beta_q \overset{\text{iid}}{\sim} \text{N}(b_0, \tau_b^2). \end{aligned} \tag{10}$$

Therefore, we denote the three competitors as: our beta autoregressive proposal (BeP), as in (1)-(3); the normal autoregressive with logit transformed responses (logitAR), as in (9); and the beta dynamic model (BDM), as in (10). In all cases, we used orders of dependence ranging from 1 to 6, so in total, there were 18 fitted models to each simulated dataset. In every case, we computed the log-pseudo marginal likelihood (LPML)

statistic, originally suggested by Geisser and Eddy (1979), and the deviance information criterion (DIC), proposed in Spiegelhalter et al. (2002). In addition, we consider the mean squared error (MSE) for observed responses, computed as the average over simulations of $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$, where y_i is the ith simulated datapoint and $\hat{y}_i$ is the corresponding estimated mean response. Similarly, we consider the MSE for predicted responses, which is computed as the average over simulations of $\sum_{i=n+1}^{n+10}(y_i - \hat{y}_i)^2$, with $\hat{y}_i = E(y_i \mid y_1, \ldots, y_n)$. Finally, we also consider credibility intervals. In particular, we report the average of credibility interval lengths (ACIL) for observed responses. This is defined as the average over simulations of the mean of all lengths $\ell_1, \ldots, \ell_n$ of 95% credibility intervals for observations. Similarly, we consider the ACIL measure for predictions as the average over simulations of mean lengths $\ell_{n+1}, \ldots, \ell_{n+10}$ of 95% credibility intervals for predictions. In all cases, we also estimated the order by considering the best value, for each of the six criteria considered.

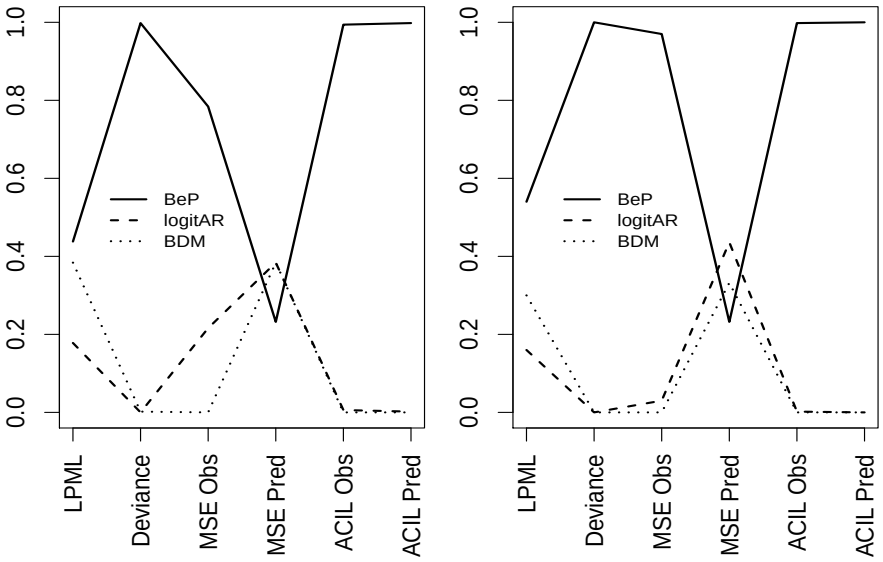

Figure 3: Simulation results. For each criterion considered (shown in the abscissas), we present the proportion of times (shown in the ordinates) each model obtained the best result compared to its competitors, averaging over number of simulations and true order of dependence q. Values are joined by lines to facilitate comparison. The left and right panels display the results for $n = 30$ and $n = 50$, respectively.

When running the simulation study, we used hyperparameter choices as described next. For model BeP we set $a = 1$, $b = 1$, $A = 1000$, $B = 100$ and $L = 1000$. For model logitAR, we chose $b_0 = 0$, $\tau_b^2 = 100$, $s_0 = 2.01$, $s_1 = 1.01$, $m_0 = 0$, and $\tau_0^2 = 100$. For model BDM we defined $a_\phi = 0.5$, $\mathrm{b}_\phi = 10$, $\tau_\mu^2 = 0.001$, $b_0 = 0$, $\tau_b^2 = 0.001$, $a_e = 50$, and $b_e = 1$. Models BeP and logitAR are thus fitted using vague prior

choices, but a more informative scenario was considered for model BDM. See further discussion on the definition of these hyperparameters in Section 4. The reason for these particular choices when dealing with model BDM in the simulation is the highly unstable Markov chain Monte Carlo (MCMC) behavior that frequently arises with other less informative alternatives. Otherwise, running an automated series of MCMC algorithms for simulated data would be unfeasible. The essential problem is produced when the imputed values of the auto-regression coefficients $\beta_0, \ldots, \beta_q$ lead, for some t, to large imputed $|\mu_t|$ values, thus creating illegal values for the Beta parameters in (10). In contrast, models BeP and logitAR do not require any special practical restrictions when defining the corresponding prior distributions.

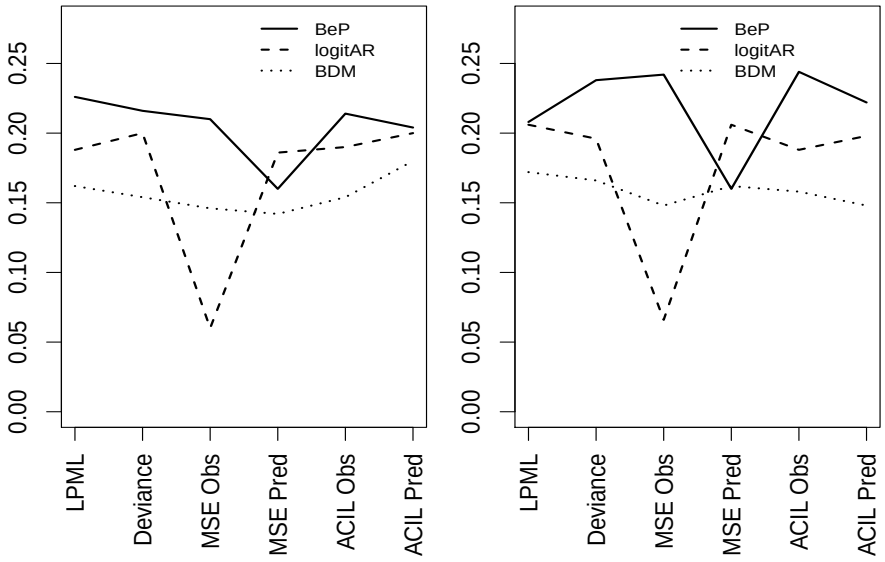

Figure 4: Simulation results. For each criterion considered (shown in the abscissas), we present the proportion of times (shown in the ordinates) each model selected the correct order of dependence q, averaging over number of simulations and q. Values are joined by lines to facilitate comparison. The left and right panels display the results for $n = 30$ and $n = 50$, respectively.

Summarizing, our simulation generated $1,000$ synthetic datasets (2 lengths and 5 true orders of dependence, 100 times each). For each dataset we fitted 18 models (the three competing models, each one applied with orders of dependence $1 \leq q \leq 6$), and using three parallel chains, which amounts to $54,000$ chains in total. The entire execution of the study took about 200 CPU hours in an Intel Xeon X3460 2.80GHz machine with 8Gb of RAM. Roughly speaking, about 40% of the time is needed to fit the proposed BeP model, about 10% is required for the logitAR model, and about 50% for the BDM. Figure 3 shows the proportion of times each of the three competing models obtained

the best performance in terms of the six criteria specified earlier. The results average over the 100 repetitions of the experiment and over the five true orders of dependence. Our proposed model BeP outperforms the other two competitors for both time series lengths and all criteria, except for the MSE for predictions. Here, models logitAR and BDM are essentially equivalent for short time series, with the former winning for long times series.

Turning now our attention to the estimation of the order of dependence, Figure 4 shows the proportion of times each of the three competing models selected the true simulation order q, averaging over all repeated simulations and the true value of q. The results are in terms of the fitted order that achieved the best value in each specific criterion. Again, model BeP outperforms the competitors, except in the case of MSE for predictions, where model logitAR is the overall winner.

Two facts are worth pointing out. First, when looking more carefully at the numerical values we find that the average over repeated simulations, true and fitted order of the MSE for predictions, were 0.0204 for BeP, 0.0232 for logitAR and 0.0187 for BDM, when the length was $n = 20$. When increasing the length to $n = 40$, the respective values were 0.0196, 0.0202 and 0.0192. Thus, the BeP model was, on average, very close to the winning model, especially for longer sequences. Interestingly, logitAR had the highest overall average, mostly explained by its rather inferior behavior for higher q values. And second, the proportion of times the true order is detected is never over 25% in all cases, though there is a slight improvement for longer time series.

In summary, the proposed BeP model performs generally better than the competitors logitAR and BDM, the exception being the case of MSE for predictions where small differences are observed, for both time series lengths that we studied.

4 Unemployment Rate Data Analysis

We consider the yearly sequence of unemployment rates in Chile from 1980 through 2010, available at http://www.indexmundi.com. These data have been plotted as big dots in Figure 5, or big dots joined by lines in Figure 10. The data feature a peak of 21% of unemployment in 1983, and then a steady decrease in the rates through the late 1980s. After that, the rates have oscillated between a minimum of 6.12% in 1997 and a maximum of 10.02% in 2004. We fitted our BeP model (1)–(3) with hierarchical prior (6)–(7) to this series of length $n = 31$, using $A = B = L = 1000$, which represents diffuse but proper priors. A conservative Gibbs sampler specification was considered by taking 220,000 iterations and a burn-in of 20,000. The posterior estimates were computed with a spacing of 200 to save disk space.

To assess the effect of choosing q in the posterior inference, we considered fits and predictions for the BeP_q model corresponding to $q = 1, \ldots, Q$. The results are shown in Figure 5 with $Q = 8$ (to simplify display, we omit the cases $q = 1, 2$). Notice the non-stationary behavior of the series of unemployment rates. The posterior predictive means and 95% credibility bands are indicated by thick and thin lines, respectively.

q	a	b	λ	w
1	6.44 (3.015)	57.39 (28.326)	67.73 (31.547)	0.099 (0.0148)
2	6.66 (3.264)	61.90 (30.745)	56.98 (23.476)	0.100 (0.0129)
3	6.73 (3.417)	63.42 (31.263)	50.49 (18.805)	0.099 (0.0116)
4	6.60 (3.536)	60.88 (31.881)	48.29 (17.410)	0.097 (0.0108)
5	6.33 (3.525)	55.61 (30.858)	48.13 (16.314)	0.095 (0.0102)
6	5.94 (3.410)	50.20 (28.911)	45.44 (14.979)	0.093 (0.0100)
7	6.17 (3.438)	49.41 (27.436)	38.60 (12.486)	0.091 (0.0104)
8	6.33 (3.554)	46.47 (26.751)	34.42 (11.275)	0.089 (0.0106)

Table 1: Posterior means (standard deviations) for some model parameters.

A stationary specification would seem inappropriate for these data, as there is a clear difference between the responses before and after the late 1980s. Therefore, the proposed model appears as a sensible alternative.

The posterior distributions for the hyperparameters a, b, λ and latent w can be seen in Figure 6 (only for $q = 3, \ldots, 8$), and posterior summaries are given in Table 1. Interestingly, the effect of increasing q seems to generally shift these distributions to the left, especially for w and most certainly for λ. This is explained by the fact that increasing q makes the dependence split in more lags, so the distribution of λ, which determines the common degree of dependence, is highly shifted to lower values. Additionally, the effect of larger q can be seen, to some extent, in Figure 5 with the tendency of the posterior predictive means to decrease with q. Numerically, this reduction is also appreciated in the overall level parameter of the series, w, in Table 1.

The choice of q is essential for the model definition, and we have seen that it affects the posterior inference as well (as expected). Rather than setting a prior for q and dealing with transdimensional MCMC algorithms, we compare the models for the different q values via the LPML statistic and the DIC criterion. High LPML values and low DIC values are indicative of the best models according to each criterion. The corresponding values are shown in Figure 7. In both cases the choice is the same, namely $q = 6$, which provides the best fit to the data. One interesting effect of q in the predicted values occurs when looking at the width of the 95% credibility bands for predictions. They also show a U-shape as a function of q, similar to the right panel of Figure 7. In fact, the average width over all observed times attains a minimum at $q = 6$ (values not shown), which is consistent with the best model obtained with the previous two criteria.

To better understand how our model adapts to non stationary settings, which is the case for the unemployment dataset at hand, Figure 8 shows a graph of u_t (left panel) and c_t (right panel) for the best fitting model obtained with $q = 6$. Point estimates of the c_t's do not change much along time, perhaps implying a (non significant) lower dependence for the first five years. This is due to the choice of the prior for the c_t's, which are assumed to be exchangeable (i.e., positively dependent). The data show a weaker dependence in the early years, which would correspond to lower c_t values, and a stronger dependence for the late years, which would correspond to higher c_t's, however,

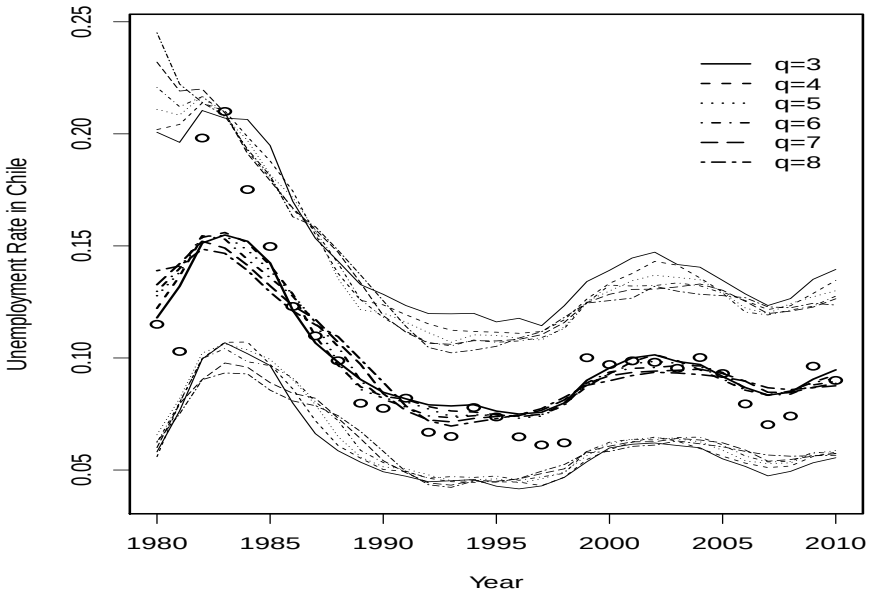

Figure 5: Fitted values for unemployment data set using BeP models with $q = 3, \ldots, 8$. The observed values are shown as big dots. Posterior predictive means are indicated with thicker lines, and 95% credibility bands by thinner lines.

since the c_t's are positively correlated a-priori, the different degrees of dependence shown in the data are translated into a huge uncertainty in the posterior distribution of the c_t's. On the other hand, the latents $\{u_t\}$ show the importance of each observation y_t in the model fit. The first four data points, for $t = 1980, \ldots, 1983$, are highly influential, with the observation at $t = 1982$ the most influential of all. Note that this data point does not correspond to the highest peak observed in the data, which was at $t = 1983$. Recall that the influence of each point vanishes after $q = 6$ times ahead, so y_{1982} is not influential from $t = 1989$ onwards, which is the precise time when the unemployment rate reduces its volatility and stabilizes. Additionally, observations at times $t = 1998, 1999, 2000$ are the most influential for the second part of the series.

To place in proper context the performance of our BeP proposal for this particular dataset, we further compare with the two alternative models logitAR (9) and BDM (10), discussed in Section 3. For the case of the logitAR, we chose a rather vague prior, with $b_0 = m_0 = 0$, $\tau_m = \tau_b = 100$, and $s_0 = 2.01$, $s_1 = 1.01$ and applied these settings for $q = 1, \ldots, 8$. The best model was obtained with $q = 1$. For the

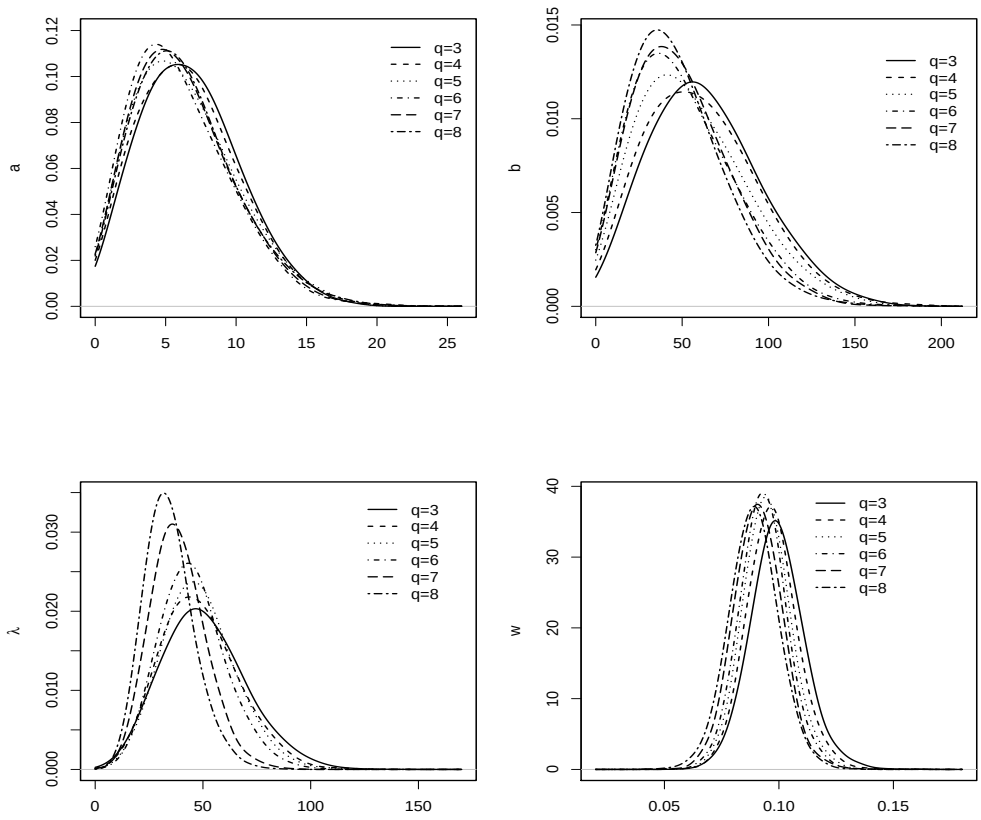

Figure 6: Posterior distributions of a (upper left), b (upper right), λ (lower left) and w (lower right) for each of the models corresponding to $q = 3, \ldots, 8$

BDM, we noticed that a vaguely defined prior leads to MCMC divergence problems. In particular, da Silva et al. (2011) point out that large ϕ values imply an increased log-likelihood. On the other hand, low ϕ values imply an unstable posterior simulation scheme. Therefore, we chose $a_\phi = 1$ and $b_\phi = 0.007$, just as da Silva et al. (2011) did for Brazilian monthly unemployment rates data. In addition, choosing a large τ_b value leads to MCMC divergence, so we chose $\tau_b = \tau_\mu = 0.1$. Finally, we picked $a_e = 2.0$ and $b_e = 0.5$. We also fitted this model for $q = 1, \ldots, 8$, obtaining the best fit with $q = 1$.

To compare among the best fitting cases for each of the three competing models, we computed the LPML and the DIC statistics, as well as the one-step-ahead log-predictive densities (Vivar and Ferreira 2009) for each observation. If we denote by $D_t = \{y_1, \ldots, y_t\}$ the set of observations up to time t, the one-step-ahead predictive density is defined as $f(y_t|D_{t-1})$. The first two statistics are presented in Table 2. According to the LPML

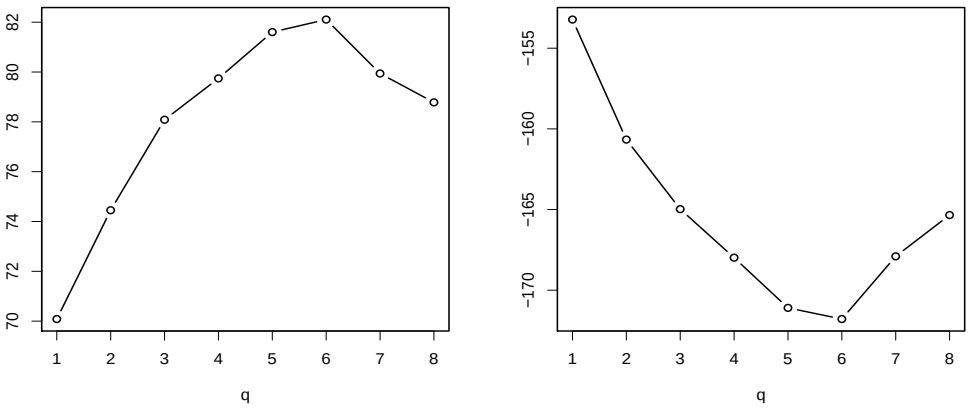

Figure 7: LPML (left panel) and DIC (right panel) values for model BeP_q, $q = 1, \ldots, 8$.

Model	LPML	DIC
BeP	82.11	−171.79
logitAR	73.75	−150.68
BDM	85.45	−191.26

Table 2: Goodness of fit measures for the unemployment data. The LPML and DIC values are reported for the following competing models: BeP ($q = 6$), logitAR ($q = 1$) and BDM ($q = 1$).

the best model is the BDM followed closely by the BeP. Now, looking at the DIC values, the best model is the BDM, and our BeP is again second place. For both criteria, the logitAR has been considered the worst fit. On the other hand, the one-step-ahead log-predictive densities for observations y_t, $t = 1985, \ldots, 2010$, are shown in Figure 9. From this figure we can see that apart from three observations around year 1997 and the observation in year 2007, the log-predictive densities favour the BeP model, leaving the BDM and the logitAR in second and third places, respectively. A possible explanation for this is that the BDM provides a better fit for the initial portion of the series (first five data points), which is precisely what Vivar and Ferreira (2009) recommend to discard, due to the instability of the predictions.

For a final comparison between the BeP and the BDM models we show, in Figure 10, the observed time series together with predictions for the observed times and future predictions for ten years ahead. As can be seen, for the observed times, the BDM point predictions follow closely the path of the data, but provide enormous uncertainty for future predictions. This can be explained by the errors that quickly accumulate in the system equation, affecting even the short term predictions. In contrast, the BeP provides reasonable predictions for the observed times, with credible bands comparable

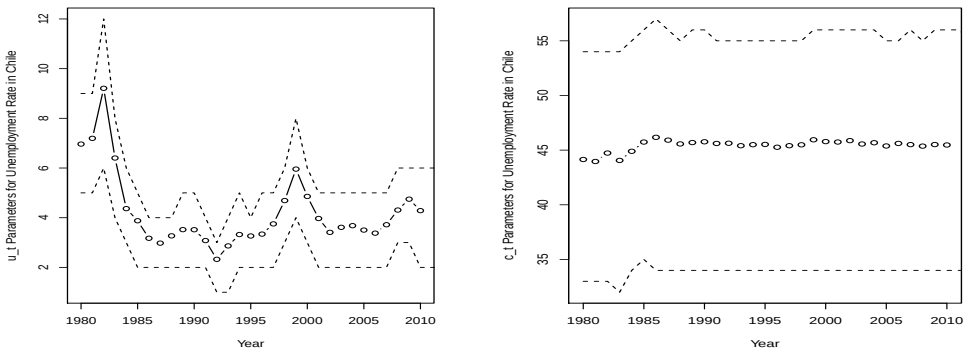

Figure 8: Posterior estimates for u_t (left panel) and c_t (right panel).

to those of the BDM, but produces much more stable uncertainty in future predictions. We can say that the BDM is too adaptive to the observations and provides poor future predictions, whereas our BeP proposal produces good predictions for both observed and future times.

Interpreting the results in the context of the data, the early years coincide with the generally bad (political and economic) situation in the Chile. After democracy was recovered, low rates in the early and mid 1990s were stimulated by the Construction Boom, quickly leading to higher values after the Asian Crisis of the late 1990s. The BeP model does provide good predictions over this period, with similar certainty to the BDM model. The quasi-cyclical behavior that follows in the 2000s is also captured by the BeP model and moreover projected in the predictions for the years to come. The autoregressive feature of BeP thus helps account for seasonal effects.

5 Discussion

We have proposed a model for a time series of responses constrained to the unit interval. The model is of the autoregressive type and has the feature that each response is marginally beta-distributed. The dependence on lagged terms is achieved by means of a sequence of latent random variables, conveniently chosen to retain the desired marginal distributions. Naturally, the extension to a series of outcomes lying on a bounded set is immediate. The model includes stationarity and non-stationarity as special cases.

Via a simulation study, we showed that our BeP model outperforms the logitAR and the BDM models in five out of six comparison criteria, for both fitting and determining the true level of dependence. Additionally, for the real data analysis we used a time series of unemployment rates in Chile. Here, the BeP model provides a reasonable fit

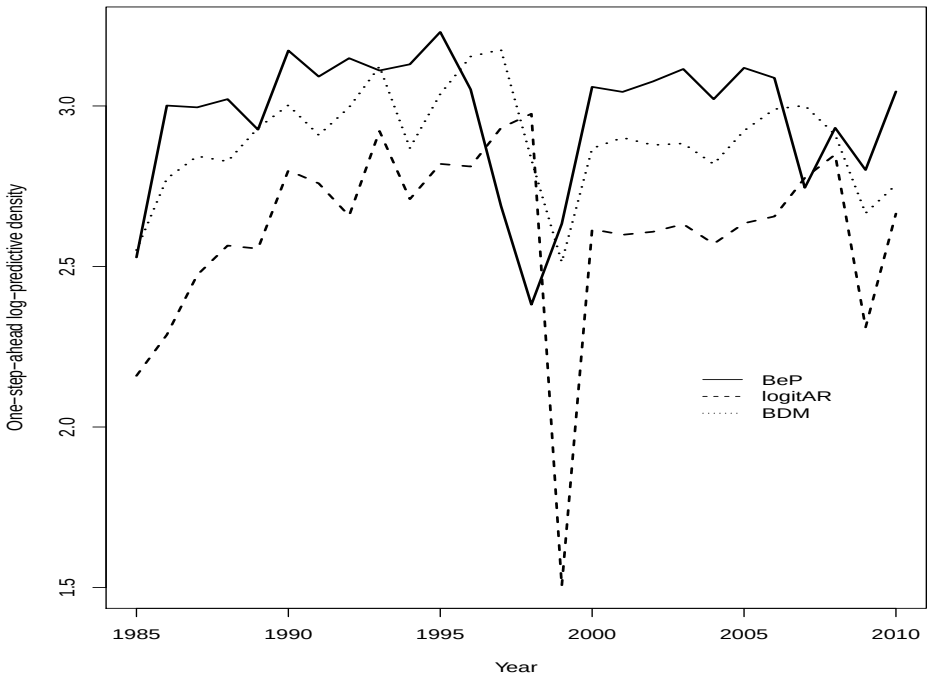

Figure 9: One-step ahead log-predictive densities using unemployment data, for $t = 1985, \ldots, 2010$. BeP with $q = 6$ (solid line), logitAR with $q = 1$ (dashed line), and BDM with $q = 1$ (dotted line).

to the data, as compared to the BDM, and produces predictions for future observations with similar uncertainty to that of the observed data.

More general autoregressive models, like seasonal (Nabeya 2001), or periodic (McLeod 1994), can also be defined. If the seasonality of the data is s, then a seasonal autoregressive model of order q would be

$$y_t \mid \mathbf{u} \sim \text{Be}\left(a + \sum_{j=0}^{q} u_{t-sj},\, b + \sum_{j=0}^{q}(c_{t-sj} - u_{t-sj})\right).$$

Moreover, re-writing the time index parameter as $t = t(r, m) = (r-1)s + m$, for $r = 1, 2 \ldots$ and $m = 1, \ldots, s$, allows us to define a periodic autoregressive model of order $(q_1, \ldots, q_s)$. For monthly data, $s = 12$ and r and m denote the year and month.

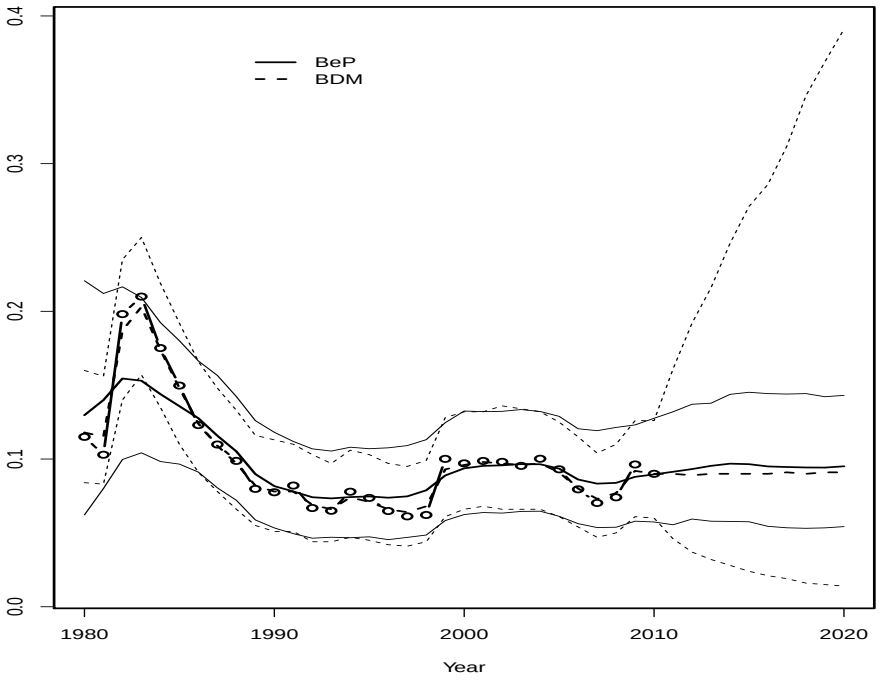

Figure 10: Fits and predictions for the BeP model (solid line), and the BDM model (dashed line). Posterior means (thick lines) and 95% credible bands (thin lines). Observed data are shown as big dots joined by lines.

Thus, the model would be

$$y_t \mid \mathbf{u} \sim \text{Be}\left(a + \sum_{j=0}^{q_m} u_{t(r,m)-j},\, b + \sum_{j=0}^{q_m} (c_{t(r,m)-j} - u_{t(r,m)-j})\right).$$

In both cases, the latents $\{u_t\}$ are defined by equations (2) and (3).

Several other extensions of the proposed model are of interest, including the incorporation of time-dependent covariates and a multivariate version of the models, retaining the marginal properties. These and other topics are the object of current research.

Appendix

Outline of conditional distributions

In all cases recall that for $t \leq 0$, $u_t = 0$ with probability one, and $c_t = 0$.

i) The conditional distribution of the latent variables u_t, for $t = 1, \ldots, T$, has the form

$$f(u_t|\,\text{rest}) \propto \frac{\binom{c_t}{u_t}\left\{\left(\frac{w}{1-w}\right)\prod_{i=0}^{q}\left(\frac{y_{t+i}}{1-y_{t+i}}\right)\right\}^{u_t} I(u_t \leq c_t)}{\prod_{i=0}^{q}\Gamma\left(a+\sum_{j=0}^{q}u_{t+i-j}\right)\Gamma\left(b+\sum_{j=0}^{q}(c_{t+i-j}-u_{t+i-j})\right)}.$$

ii) The conditional distribution of the latent w is

$$f(w|\,\text{rest}) = \text{Be}\left(w\,\middle|\,a+\sum_{t=1}^{T}u_t\,,\,b+\sum_{t=1}^{T}(c_t-u_t)\right).$$

iii) The conditional distribution of the parameter c_t, for $t = 1, \ldots, T$ has the form

$$f(c_t|\,\text{rest}) \propto \left\{\prod_{i=0}^{q}\frac{\Gamma\left(a+b+\sum_{j=0}^{q}c_{t+i-j}\right)}{\Gamma\left(b+\sum_{j=0}^{q}(c_{t+i-j}-u_{t+i-j})\right)}\right\} \times \frac{\{\lambda(1-w)\prod_{i=0}^{q}(1-y_{t+i})\}^{c_t}}{(c_t-u_t)!}I(c_t \geq u_t).$$

iv) The conditional distribution of the parameter a is

$$f(a|\,\text{rest}) \propto \left\{\prod_{t=1}^{T}\frac{\Gamma\left(a+b+\sum_{j=0}^{q}c_{t-j}\right)}{\Gamma\left(a+\sum_{j=0}^{q}u_{t-j}\right)}\right\}\frac{\Gamma(a+b)}{\Gamma(a)}\left(w\prod_{t=1}^{T}y_t\right)^a f(a).$$

with $f(a)$ the prior distribution, which in our case was $f(a) = \text{Un}(0, A)$.

v) The conditional distribution of the parameter b becomes

$$f(b|\,\text{rest}) \propto \left\{\prod_{t=1}^{T}\frac{\Gamma\left(a+b+\sum_{j=0}^{q}c_{t-j}\right)}{\Gamma\left(b+\sum_{j=0}^{q}(c_{t-j}-u_{t-j})\right)}\right\} \times\frac{\Gamma(a+b)}{\Gamma(b)}\left\{(1-w)\prod_{t=1}^{T}(1-y_t)\right\}^b f(b).$$

with $f(b)$ the prior distribution, which in our case was $f(b) = \text{Un}(0, B)$.

vi) The conditional distribution of the hyper-parameter λ is

$$f(\lambda|\,\text{rest}) \propto \text{Ga}\left(\lambda\,\middle|\,1+\sum_{t=1}^{T}c_t,\,T\right)I(\lambda < L).$$

which is a truncated gamma distribution.

Acknowledgments

Alejandro Jara was partially funded by grant FONDECYT 11100144. Luis Nieto-Barajas was partially funded by grant CONACYT 130991. Fernando Quintana was partially funded by grant FONDECYT 1100010.

References

Bekele, B. N., Nieto-Barajas, L. E., and Munsell, M. F. (2012). "Analysis of Partially Incomplete Tables of Breast Cancer Characteristics With an Ordinal Variable." *Journal of Statistical Theory and Practice*, 6: 1–20. 724

Bernardo, J.-M. and Smith, A. F. M. (1994). *Bayesian theory*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Chichester: John Wiley & Sons Ltd. 725

Box, G. E. P. and Jenkins, G. M. (1970). *Times series analysis. Forecasting and control*. San Francisco, Calif.: Holden-Day. 723

da Silva, C. Q., Migon, H. S., and Correia, L. T. (2011). "Dynamic Bayesian beta models." *Computational Statistics & Data Analysis*, 55(6): 2074–2089. 723, 728, 734

Geisser, S. and Eddy, W. F. (1979). "A predictive approach to model selection." *Journal of the American Statistical Association*, 74(365): 153–160. 729

Grunwald, G. K., Raftery, A. E., and Guttorp, P. (1993). "Time Series of Continuous Proportions." *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 55: 103–116. 724

McLeod, A. I. (1994). "Diagnostic Checking of Periodic Autoregression Models with Application." *Journal of Time Series Analysis*, 15: 221–233. 737

Nabeya, S. (2001). "Unit Root Seasonal Autoregressive Models with a Polynomial Trend of Higher Degree." *Econometric Theory*, 17(2): 357–385. 737

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). "Bayesian measures of model complexity and Fit." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 64(4): 583–639. 729

Vivar, J. C. and Ferreira, M. A. R. (2009). "Spatiotemporal models for Gaussian areal data." *Journal of Computational and Graphical Statistics*, 18(3): 658–674. 734, 735

Wallis, K. F. (1987). "Time Series Analysis of Bounded Economic Variables." *Journal of Time Series Analysis*, 8: 115–123. 723

West, M., Harrison, P. J., and Migon, H. S. (1985). "Dynamic generalized linear models and Bayesian forecasting." *Journal of the American Statistical Association*, 80(389): 73–97. With discussion. 723