Chilean Journal of Statistics
Vol. 1, No. 2, September 2010, 93–112

Special Issue "In Memory of Pilar Loreto Iglesias Zuazola"
Research Paper

Multiple change-point analysis for linear regression models

Rosangela H. Loschi*, Jeanne G. Pontel and Frederico R. B. Cruz

Departamento de Estatística, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

(Received: 21 April 2010 · Accepted in final form: 24 August 2010)

Abstract

The product partition model (PPM) is a powerful tool for clustering and change point analysis mainly because it considers the number of blocks or segments as a random variable. We apply the PPM to identify multiple change points in linear regression models extending some previous works. In addition, we provide a predictivistic justification for the within-block linear model. This way of modeling provides a non ad-hoc procedure for treating piecewise regression models. We also modify the original algorithm proposed by Barry and Hartigan (1993) in order to obtain samples from the product distributions –posteriors of the parameters in the regression model, say– in the contiguous-block case. Consequently, posterior summaries (including the posterior means or product estimates) can be obtained in the usual way. The product estimates are obtained considering both the proposed and Barry and Hartigan's algorithms, which are compared to least square estimates for the piecewise regression models. To illustrate the use of the proposed methodology, we analyze some financial data sets.

Keywords: de Finetti type-theorem · Least square estimates · Orthogonal invariance · Product partition model.

Mathematics Subject Classification: Primary 62F15 · Secondary 62E10.

1. Introduction

Hartigan (1990) proposed a powerful model, named product partition model (PPM), designed to analyze change point problems. Firstly, the PPM was so innovative because it unified many models previously developed, such as those introduced by Menzefricke (1981), Hsu (1984), Smith (1975) and Holbert (1982). Secondly, the PPM assumes the number of change points as well as their positions as random variables. Since then, the PPM has been considered for cluster and outlier detection by many authors. Barry and Hartigan (1993) and Crowley (1997), for instance, successfully applied the PPM for change point detection in normal means. Lately, Quintana et al. (2005) used the PPM for cluster analysis in the measurement error model and Quintana and Iglesias (2003) applied it to outlier detection in normal linear models. Later, Loschi and Cruz (2005) extended the PPM by providing a method to obtain the posterior distributions for the positions and number of change points as well as the posterior probability of each instant being a change point. Another approach to obtain such posteriors was provided by Fearnhead (2006) and Fearnhead and Liu (2007) with algorithms based on filtering for multiple change point problems.

* Corresponding author. Rosangela H. Loschi. Departamento de Estatística, Universidade Federal de Minas Gerais, 30.710-010, Belo Horizonte, MG, Brazil. Email: loschi@est.ufmg.br

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

An important extension of the PPM is provided by Quintana and Iglesias (2003) which presented a decision-theoretic formulation to specific change point problems, such as outlier detection, using the PPM. They also proved that PPM generalizes the Dirichlet process, which has been intensively used for clustering analysis (Müller and Quintana, 2010) and density estimation (Escobar and West, 1995). The PPM was also used by Tarantola et al. (2008) for row effects models. Quintana (2006) establishes connections between the PPM and other models that induce a partition structure, e.g, species sampling models. More recently, Hegarty and Barry (2008) used the PPM in the spatial context for Bayesian disease mapping, Demarqui et al. (2008) applied it in survival analysis to estimate the time grid in piecewise exponential models and Bormetti et al. (2010) present an approach to the value-at-risk (VaR) computation using the PPM to introduce the cluster structure in the mean, whenever the volatility is fixed, and in the volatility, whenever the mean return is a fixed and known value. Ruggeri and Sivaganesan (2005) and Booth et al. (2008) suggest other approaches on multiple change point identification. In the spatio-temporal setting, Majumdar et al. (2005) introduced a change point model attempting to capture changes in both the temporal and spatial associations.

Barry and Hartigan (1993) proposed a computationally intensive algorithm, based on a Gibbs sampling scheme (Barry and Hartigan's algorithm), for computing the product estimates (or posterior means) in the contiguous-block case. In spite of its efficiency, the product estimates is the only posterior information we could get from that method. Similar algorithms can be found in Yao (1984) and Loschi and Cruz (2005). However, posterior mean may not be the best posterior summary if, for instance, the posterior is asymmetric or else if it has two or more modes. In these cases other posterior summaries (modes and quantiles, for instance) can provide a better idea of the posterior behavior.

This paper aims at applying the PPM to identify multiple change points in normal linear regression models, which is an extension of some previous works (Diniz et al., 2003) by considering the possibility of changes in all regression parameters. We provide a predictivistic characterization to the model which follows as a consequence of judgment of orthogonal invariance on the observable response variables. We also modify the Barry and Hartigan's original algorithm in order to sample from the product distributions or the posteriors of the parameters in the regression model. Consequently, an improvement in the analysis is reached since other posterior summaries besides the posterior means or product estimates can be obtained in the usual way. A comparison among the least square estimates and the product estimates obtained by using both Barry and Hartigan's algorithm and the algorithm proposed here, is performed for two real data sets including the Brazilian Industrial Production Index (BIP) and the Brazilian Employment Index (BEI), as well as the Dow Jones Industrial Average (DJIA) and the BOVESPA Index (IBOVESPA), which is the São Paulo stock exchange broad-based stock index, the most representative indicator of the performance of the prices of the Brazilian stock market. The ultimate goal is to analyze the linear behavior of BIP and BEI throughout time and evaluate if it experiences changes in the period analyzed. A similar study is performed considering the indexes DJIA and BOVESPA.

The paper is organized as follows. In Section 2, we present a full Bayesian approach for analyzing piecewise linear regression models. In Section 3, we introduce the computational procedures to handle the problem of change point identification in regression models. Barry and Hartigan's algorithm to obtain the product estimates is briefly reviewed and we provide a modification of such algorithm in order to obtain samples from the posterior of the regression parameters. In Section 4, we discuss a comparison between Bayesian and least square estimates for the piecewise linear regression model. Finally, in Section 5, we sketch some concluding remarks.

2. Change Point Analysis for the Multiple Linear Regression Model

Some challenges related to the use of piecewise regression models are the identification of the number of clusters to be considered in the model and their positions. Ad-hoc choices for these parameters are the usual practice in classical procedures such as least square methods. GAM and splines are alternative procedures to accommodate the clustered structure in the data. Here, we consider the PPM to model the uncertainty about such parameters and to handle piecewise regression models in a full Bayesian context.

2.1 The PPM for linear regression model

Let us denote by $\mathbf{X}_k$ the $1 \times l$ vector $(1, X_{k,1}, \ldots, X_{k,l-1})$, for $k = 1, \ldots, n$. Let $(\mathbf{X}_1, Y_1)$, $\ldots$, $(\mathbf{X}_n, Y_n)$ be a sequence of vectors of observations. Assume that vector $(\mathbf{X}_k, Y_k)$ obeys the usual regression model specifications, that is,

$$Y_k = \mathbf{X}_k \boldsymbol{\beta}_k + \varepsilon_k, \quad k = 1, \ldots, n,$$

where $\boldsymbol{\beta}_k \in \mathbb{R}^l$ denotes the $l \times 1$ vector of parameters $(\beta_{k,0}, \ldots, \beta_{k,l-1})^\top$ at position (or instant) k and $\mathbf{A}^\top$ denotes the transpose of vector $\mathbf{A}$. Assume also that the errors are independent and such that $\varepsilon_k \sim \mathrm{N}(0, \sigma_k^2)$. Consequently, conditionally on $\mathbf{X}_k, \boldsymbol{\beta}_k$, and σ_k^2, for $k = 1, \ldots, n$, we have that the variables $Y_1, \ldots, Y_n$ are independent and such that

$$Y_k | \mathbf{X}_k, \boldsymbol{\beta}_k, \sigma_k^2 \overset{\text{ind.}}{\sim} \mathrm{N}(\mathbf{X}_k \boldsymbol{\beta}_k, \sigma_k^2).$$

Let us consider $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ where $\boldsymbol{\theta}_k^\top = (\boldsymbol{\beta}_k^\top, \sigma_k^2)$. Following Quintana et al. (2005), the cluster structure can be introduced into the model by rewriting the vector θ as follows

$$\boldsymbol{\theta} = \sum_{j=1}^{b} \Big(\boldsymbol{\theta}_{[i_{j-1} i_j]} \mathbf{1}\{i_{j-1} < 1 \leq i_j\}, \ldots, \boldsymbol{\theta}_{[i_{j-1} i_j]} \mathbf{1}\{i_{j-1} < n \leq i_j\} \Big),$$

where $\{i_0, \ldots, i_b\}$ is a particular value of the random partition ρ that denotes the instants when changes occur and satisfies the condition $0 = i_0 < \cdots < i_b = n$; $B = b$ denotes the number of blocks in ρ; $\mathbf{1}\{A\}$ is the indicator function of event A; and $\boldsymbol{\theta}_{[ij]}^\top = (\boldsymbol{\beta}_{[ij]}^\top, \sigma_{[ij]}^2)$ denotes the common value for $\boldsymbol{\theta}_k$, for $i < k \leq j$.

Assuming that data are sequentially observed and that only contiguous blocks are possible, we construct the prior distributions for ρ and θ as follows. Assume that p denotes the probability of a change taking place at any instant. Define the Markov chain $\{Z_k \colon k \in \mathbb{N}\}$ generated by the instants when the changes occurred. Consequently, assuming that $\mathrm{P}(Z_0 = i_0 | p) = 1$, the prior distribution of ρ, given p, is

$$\mathrm{P}(\rho = \{i_0, \ldots, i_b\} | p) = p^{b-1} (1-p)^{n-b}, \tag{1}$$

for all $b \in I = \{1, \ldots, n\}$. As a consequence, the prior for B, given p is

$$\mathrm{P}(B = b | p) = \binom{n-1}{b-1} p^{b-1} (1-p)^{n-b}, \;\; b \in I.$$

The prior in Equation (1) and the objective one considered by Girón et al. (2007) belong both to the following general class of product prior distributions for ρ introduced by Hartigan (1990) expressed as

$$P(\rho = \{i_0, \ldots, i_b\}) = \frac{\Pi_{j=1}^{b} c_{[i_{j-1} i_j]}}{\sum_{\mathcal{C}} \Pi_{j=1}^{b} c_{[i_{j-1} i_j]}},$$

where $\mathcal{C}$ is the set of all possible partitions of the set I into b contiguous blocks, and $c_{[ij]}$ denotes the prior cohesion for the block $[ij]$. The prior cohesions are subjectively chosen and should disclose the similarity among the observations within the same block. Particularly, if the Yao's prior cohesions (Yao, 1984) are assumed, the prior for ρ is the one given in Equation (1). If $c_{[ij]} = 1$ for all blocks, we have the objective prior assumed by Girón et al. (2007), that is, $P(\rho = \{i_0, \ldots, i_b\}) = 2^{-(n-1)}$.

Remark 2.1 If $p \sim \text{Beta}(\alpha, \beta)$, the prior distributions for ρ and B are given, respectively, by

$$P(\rho = \{i_0, \ldots, i_b\}) = \frac{\Gamma(\alpha+\beta)\ \Gamma(\alpha+b-1)\ \Gamma(n+\beta-b)}{\Gamma(\alpha)\ \Gamma(\beta)\ \Gamma(n+\alpha+\beta-1)} \tag{2}$$

and

$$P(B = b) = \binom{n-1}{b-1} \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+b-1)\Gamma(n+\beta-b)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(n+\alpha+\beta-1)}. \tag{3}$$

It is noticeable from Equation (3) that $B \stackrel{\text{d}}{=} Y+1$, where " $\stackrel{\text{d}}{=}$ " means equal in distribution, Y is a random variable which has a binomial-beta distribution with parameters $n-1$, α and β, $\alpha > 0$, $\beta > 0$. Thus, the prior mean and variance of B are given, respectively, by

$$E(B) = \frac{(n-1)\ \alpha}{\alpha+\beta} + 1$$

and

$$\text{Var}(B) = \frac{(n-1)\ \alpha\ \beta\ (\alpha+\beta+n-1)}{(\alpha+\beta)^2\ (\alpha+\beta+1)}.$$

If $\alpha = \beta$, this prior stimulates a large number of clusters (around 50% of the observations) since its mean is $(n+1)/2$. If α is constant and $\beta \to 0$, then $E(B) \to n$ and $\text{Var}(B) \to 0$, that is, in the prior evaluation all the observations are in different clusters with probability one. If β is constant and $\alpha \to 0$, then $E(B) \to 1$, say, and we are expecting no changes in the data sequence.

Remark 2.2 Girón et al. (2007) proposed a joint objective prior for B and ρ assigning first a uniform prior for B, and then eliciting a uniform prior for $\rho = \{i_0, \ldots, i_b\}$, given $B = b$. Thus, it follows that

$$P(B = b, \rho = \{i_0, \ldots, i_b\}) = \left[n \binom{n-1}{b-1} \right]^{-1}.$$

Similarly to what is observed for the prior in Equation (2), assuming this joint prior for B and ρ, all partitions with $B = b$ blocks have the same probability which, for the prior considered by Girón et al. (2007), is given by

$$P(\rho = \{i_0, \ldots, i_b\} | B = b) = (b-1)!(n-b)![(n-1)!]^{-1}.$$

We should notice, however, that there is no choice of α and β in Equation (2) that leads to the same probability but we can obtain a close result assuming $\alpha = \beta = 1$.

In order to assign the joint prior distribution for $\boldsymbol{\theta}$, given ρ and p, we assume that:

(i) the common parameters $\boldsymbol{\theta}_{[i_0 i_1]}^{\top} = (\boldsymbol{\beta}_{[i_0 i_1]}^{\top}, \sigma_{[i_0 i_1]}^2), \ldots, \boldsymbol{\theta}_{[i_{b-1} i_b]}^{\top} = (\boldsymbol{\beta}_{[i_{b-1} i_b]}^{\top}, \sigma_{[i_{b-1} i_b]}^2)$ are independent and independent from p; and

(ii) for all i, with $j = 1, \ldots, b$ and $i < j$, we consider the following prior distributions for the common parameters:

$$\boldsymbol{\beta}_{[ij]} | \sigma_{[ij]}^2 \sim \mathrm{N}_l(\mathbf{m}_{[ij]}, \sigma_{[ij]}^2 \mathbf{V}),$$

$$\sigma_{[ij]}^2 \sim \mathrm{IGa}(\nu_{[ij]}/2, d_{[ij]}/2),$$

where $\mathrm{IGa}(a, b)$ denotes the inverse gamma distribution with parameters a and b, $\nu_{[ij]}$ and $d_{[ij]}$ are positive real values, $\mathbf{m}_{[ij]}$ is an $l \times 1$ real vector and $\mathbf{V}$ is an $l \times l$ definite positive constant matrix.

Let us consider $\mathbf{Y}_{[ij]} = (Y_{i+1}, \ldots, Y_j)^{\top}$ and denote by $\mathbf{X}_{[ij]}$ the $(j - i) \times l$ matrix given by

$$\mathbf{X}_{[ij]} = \begin{pmatrix} 1 & X_{i+1,1} & \ldots & X_{i+1,l-1} \\ 1 & X_{i+2,1} & \ldots & X_{i+2,l-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{j,1} & \ldots & X_{j,l-1} \end{pmatrix}.$$

Under these assumptions and considering results presented in Barry and Hartigan (1993) it follows that the posteriors for $\boldsymbol{\beta}_k$ and σ_k^2, for $k = 1, \ldots, n$, are mixtures of posterior-by-block distributions, which in our case are the l-variate Student-t distributions in Equation (4) and the inverse gamma distributions in Equation (5), respectively. The mixing measure is the posterior relevance $r_{[ij]}^*$ of block $[ij]$ that denotes the posterior probability of block $[ij]$ being into the partition ρ, that is, the posterior distributions for $\boldsymbol{\beta}_k$ and σ_k^2, for $k = 1, \ldots, n$, respectively, are

$$f(\boldsymbol{\beta}_k | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} r_{[ij]}^* f(\boldsymbol{\beta}_k | \mathbf{X}_{[ij]}, Y_{[ij]}),$$

$$f(\sigma_k^2 | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} r_{[ij]}^* f(\sigma_k^2 | \mathbf{X}_{[ij]}, Y_{[ij]}),$$

where

$$f(\boldsymbol{\beta}_k | \mathbf{X}_{[ij]}, \mathbf{Y}_{[ij]}) = \frac{(\nu_{[ij]}^*)^{d_{[ij]}^*/2} \Gamma\left(\frac{d_{[ij]}^* + l}{2}\right)}{\pi^{1/2} |\mathbf{V}^*|^{1/2} \Gamma\left(\frac{d_{[ij]}^*}{2}\right)} \times \left\{\nu_{[ij]}^* + (\boldsymbol{\beta}_k - \mathbf{m}_{[ij]})^{\top} (\mathbf{V})^{-1} (\boldsymbol{\beta}_k - \mathbf{m}_{[ij]})\right\}^{-(d_{[ij]}^* + l)/2}, \quad (4)$$

$$f(\sigma_k^2 | \mathbf{X}_{[ij]}, \mathbf{Y}_{[ij]}) = \frac{(\nu_{[ij]}^*)^{d_{[ij]}^*/2}}{\Gamma\left(\frac{d_{[ij]}^*}{2}\right)} (\sigma_k^2)^{-(d_{[ij]}^* + 2)/2} \exp\left\{-\frac{\nu_{[ij]}^*}{2\sigma_k^2}\right\}, \quad (5)$$

$$d^*_{[ij]} = d_{[ij]} + j - i,$$

$$\mathbf{V}^*_{[ij]} = \left(\mathbf{V}^{-1} + (\mathbf{X}_{[ij]})^\top \mathbf{X}_{[ij]}\right)^{-1},$$

$$\mathbf{m}^*_{[ij]} = \mathbf{V}^*_{[ij]} \left(\mathbf{V}^{-1}\mathbf{m}_{[ij]} + (\mathbf{X}_{[ij]})^\top \mathbf{X}_{[ij]}\right),$$

$$\nu^*_{[ij]} = \nu_{[ij]} + (\mathbf{m}_{[ij]})^\top \mathbf{V}^{-1}\mathbf{m}_{[ij]} + (\mathbf{Y}_{[ij]})^\top \mathbf{Y}_{[ij]} - (\mathbf{m}^*_{[ij]})^\top (\mathbf{V}^*)^{-1}\mathbf{m}^*_{[ij]}.$$

Consequently, the product estimates (or posterior means) for $\theta_k = (\boldsymbol{\beta}_k, \sigma^2_k)$, for all $k = 1, \ldots, n$, can be obtained by means of the expectations

$$\hat{\boldsymbol{\beta}}_k = \mathrm{E}(\boldsymbol{\beta}_k | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} r^*_{[ij]} \mathbf{m}^*_{[ij]},$$

$$\hat{\sigma}^2_k = \mathrm{E}(\sigma^2_k | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \sum_{i=0}^{k-1} \sum_{j=k}^{n} r^*_{[ij]} \frac{\nu^*_{[ij]}}{d^*_{[ij]} - 2}.$$

Since the prior for ρ, given p, is a product distribution and the likelihood is based on conditional independence of the observations, the posteriors of ρ and B are, respectively,

$$\mathrm{P}(\rho = \{i_0, \ldots, i_b\} | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \prod_{j=1}^{b} f(\mathbf{Y}_{[i_{j-1} i_j]} | \mathbf{X}_{[i_{j-1} i_j]}) \int_0^1 p^b (1-p)^{(n-b)} \pi(p) dp,$$

$$\mathrm{P}(B = b | \mathbf{X}_{[0n]}, \mathbf{Y}_{[0n]}) = \binom{n-1}{b-1} \prod_{j=1}^{b} f(\mathbf{Y}_{[i_{j-1} i_j]} | \mathbf{X}_{[i_{j-1} i_j]}) \int_0^1 p^b (1-p)^{(n-b)} \pi(p) dp,$$

where π is the prior for p and $f(\mathbf{Y}_{[ij]} | \mathbf{X}_{[ij]})$ is the following $(j-i)$-variate Student-t distribution, denoted by $\mathbf{Y}_{[ij]} | \mathbf{X}_{[ij]} \sim t_{j-i}(\mathbf{X}_{[ij]} \mathbf{m}_{[ij]}, \mathbf{C}_{[ij]}; \nu_{[ij]}, d_{[ij]})$, which density is

$$f(\mathbf{Y}_{[ij]} | \mathbf{X}_{[ij]}) = \frac{\nu_{[ij]}^{d_{[ij]}/2} \Gamma\left(\frac{d^*_{[ij]}}{2}\right)}{\pi^{(j-i)/2} \Gamma\left(\frac{d_{[ij]}}{2}\right) |\mathbf{C}_{[ij]}|^{1/2}} \times \left\{\nu_{[ij]} + (\mathbf{y}_{[ij]} - \mathbf{X}_{[ij]}\mathbf{m}_{[ij]})^\top (\mathbf{C}_{[ij]})^{-1} (\mathbf{y}_{[ij]} - \mathbf{X}_{[ij]}\mathbf{m}_{[ij]})\right\}^{-d^*_{[ij]}/2}, \quad (6)$$

where $\mathbf{C}_{[ij]} = \mathbf{I}_{(j-i)} + \mathbf{X}_{[ij]} \mathbf{V} \mathbf{X}^\top_{[ij]}$ and $\mathbf{I}_{(j-i)}$ denotes the identity matrix of order $(j-i)$.

Although we can consider the posterior means as the estimates for ρ and B –as we have done for the regression parameters– for discrete parameters, the posterior modes are commonly preferred since they have a more intuitive interpretation in terms of most probable values. Besides they are the Bayes estimators for ρ and B whenever the 0-1 loss function is assumed as the penalty function.

2.2 Predictivistic justification for liner regression model

Quoting from Arellano-Valle et al. (1994, p. 221) "*De Finetti type theorems characterize models in terms of invariance. The idea is to take observables, postulate symmetry and then represent the model as a mixture of standard parametric models.*". Parameters involved in this representation have an operational interpretation which can make easy the construction of appropriate priors. This approach for inference is named operational Bayesian or predictivistic approach. From the predictivistic point of view, there is no distinction between prior and likelihood – both are equally important components of the model and are a consequence of judgments (invariance, uniformity, and so on) about quantities that could be observed. More details can be found in Iglesias et al. (2009), Iglesias et al. (1998), Iglesias (1993), Wechsler (1993), Diaconis et al. (1992), Daboni and Wedlin (1982), de Finetti (1975) and de Finetti (1937) among many others.

However, judgment of invariance only is usually insufficient to have the mixing measure involved in de Finetti type theorems completely specified. For a full predictivistic approach, additional conditions on the observable quantities - such as linearity of the expectation – are needed and some results by Diaconis and Ylvisaker (1979) are used to obtain the prior.

In this section, using results obtained by Loschi et al. (2003), we provide a predictivistic characterization for the linear model that was considered in the previous section to model data behavior within blocks. Sometimes it is easier to get prior information about observable quantities. In this case, the predictivistic approach provide an alternative way of modeling; see also Loschi et al. (2007). Let us start defining orthogonal invariance.

Let O_n be the group of all $n \times n$ real orthogonal matrices and let $\mathcal{M}$ be a p-dimensional subspace of $\mathbb{R}^n$. Denote by $O_n(\mathcal{M}) = \{\mathbf{\Gamma} \in O_n \colon \mathbf{\Gamma}\mathbf{x} = \mathbf{x}, \mathbf{x} \in \mathcal{M}\}$ the subgroup of O_n that preserves the elements of $\mathcal{M}$. Thus, a random $n \times p$ matrix $\mathbf{Y}$ is $O_n(\mathcal{M})$-invariant if $\mathbf{Y}$ and $\mathbf{\Gamma Y}$ are identically distributed for all $\mathbf{\Gamma} \in O_n(\mathcal{M})$. Moreover, an infinite sequence of $p \times 1$ random vector $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ is $O(\mathcal{M})$-invariant if, for each $n \geq 1$, the random matrix $\mathbf{Y}^{(n)} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n)^\top$ is $O_n(\mathcal{M})$-invariant (Diaconis et al., 1992).

For the linear model considered here, let $\mathcal{M}$ be the column space of $\mathbf{X}_{[0k]}$, where $\mathbf{X}_{[0k]}$ is a $k \times l$ matrix with rank l and such that $\lim_{k\to\infty}(\mathbf{X}_{[0k]}^\top \mathbf{X}_{[0k]})^{-1} = 0$. The first stage in model construction, which corresponds to the specification of the likelihood for the normal linear model, is replaced by the judgment of $O(\mathcal{M})$-invariance on the Y's, say, we assume that the infinite sequence of response variables $Y_1, Y_2, \ldots$ is $O(\mathcal{M})$-invariant, which means that for each $k > l$, the random vector $\mathbf{Y}_{[0k]}$ is $O_k(\mathcal{M})$-invariant. As established by Diaconis et al. (1992), under this assumption the distribution of any k-dimensional vector $\mathbf{Y}_{[0k]}$ is represented as a

$$f(\mathbf{Y}_{[0k]}|\mathbf{X}_{[0k]}) = \int_{\mathrm{R}^l \times (0,\infty)} \mathrm{N}(\mathbf{X}_{[0k]}\boldsymbol{\beta}, \sigma^2\mathbf{I}_k)\, \pi(\boldsymbol{\beta}, \sigma^2)\, d\boldsymbol{\beta}\, d\sigma^2, \ \forall k > l.$$

Under the assumption of $O(\mathcal{M})$-invariance, there is an operational interpretation for the parameter $(\boldsymbol{\beta}, \sigma^2)$. It can be interpreted as $\lim_{k\to\infty}\ (B_k, E_k/(k-l))$ (a.s.), where $B_k = (\mathbf{X}_{[0k]}^\top \mathbf{X}_{[0k]})^{-1}\mathbf{X}_{[0k]}^\top \mathbf{Y}_{[0k]}$ and $E_k = (\mathbf{Y}_{[0k]} - \mathbf{X}_{[0k]}B_k)^\top(\mathbf{Y}_{[0k]} - \mathbf{X}_{[0k]}B_k)$. Thus, the mixing measure $\pi(\boldsymbol{\beta}, \sigma^2)$ can be interpreted as the prior distribution for the operational parameter $(\boldsymbol{\beta}, \sigma^2)$. It is remarkable that the likelihood becomes known from the $O(\mathcal{M})$-invariance assumption. However, for a full predictivistic approach for the modeling, we need additional assumptions on the observable quantities. Let us assume that different vectors of response variables tend to be similar if the same matrix of covariates are observed. Such condition reveals that the second stage in model construction, which corresponds to the prior elicitation for the operational parameters, is replaced by additional conditions on the expectation of observable response quantities.

Let $\mathbf{Y}_{[0k]}^{(i)} = (Y_{(i-1)k+1}, \ldots Y_{ik})$, where $Y_1, Y_2, \ldots$ is an infinite sequence of $O(\mathcal{M})$-invariant random variables. Loschi et al. (2003) established that if $\mathrm{P}(\mathbf{Y}_{[0k]}^{(1)} \in \mathcal{M}) = 0$ and

$$\mathrm{E}\left(\mathbf{Y}_{[0k]}^{(2)} | \mathbf{X}_{[0k]}, \mathbf{Y}_{[0k]}^{(1)}\right) = a\mathbf{Y}_{[0k]}^{(1)} + \mathbf{b}_1,$$

$$\mathrm{E}\left(\left(\mathbf{Y}_{[0k]}^{(2)}\right)^{\top} \mathbf{Y}_{[0k]}^{(2)} | \mathbf{X}_{[0k]}, \mathbf{Y}_{[0k]}^{(1)}\right) = a\left(\mathbf{Y}_{[0k]}^{(1)}\right)^{\top} \mathbf{Y}_{[0k]}^{1} + b_2,$$

for some real constants a and b_2 and constant vector $\mathbf{b}_1 \in \mathbb{R}^k$, then it can be proved that $a \in (0,1)$, $\mathbf{b}_1 = \mathbf{X}_{[0k]}\mathbf{b}$ for any $\mathbf{b} \in \mathbb{R}^p$, $b_2 > \mathbf{b}^{\top}\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]}\mathbf{b}/(1-a)$ and, for each $n \geq 1$ and $k > l$, it follows that the predictive distribution of $\mathbf{Y}_{[0k]}$ is

$$\mathbf{Y}_{[0k]} | \mathbf{X}_{[0k]} \sim t_k \left(\frac{1}{1-a}\mathbf{X}_{[0k]}\mathbf{b},\ \mathbf{I}_k + \frac{a}{1-a}\mathbf{P}^{(k)}; \frac{b_2}{a} - \frac{\mathbf{b}^{\top}\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]}\mathbf{b}}{a(1-a)}, \frac{k(1-a)+(l+2)a}{a} \right),$$

where $\mathbf{P}^{(k)} = \mathbf{X}_{[0k]}(\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]})^{-1}\mathbf{X}_{[0k]}^{\top}$.

It follows from Theorem 3 in Diaconis and Ylvisaker (1979) that the joint prior density of $(\boldsymbol{\beta}, \sigma^2)$ is

$$\pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}_{[0k]}) \propto \left\{\frac{1}{2\sigma^2}\right\}^{\frac{k(1-a)+(p+2)a}{2a}+1} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{b_2}{a} - \frac{\mathbf{b}^{\top}\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]}\mathbf{b}}{a(1-a)}\right)\right\}$$
$$\times \left(\frac{1}{\sigma^2}\right)^{\frac{p}{2}} \exp\left\{-\frac{1}{2a(1-a)^{-1}\sigma^2}\left(\boldsymbol{\beta} - \frac{1}{1-a}\mathbf{b}\right)^{\top} \mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]} \left(\boldsymbol{\beta} - \frac{1}{1-a}\mathbf{b}\right)\right\},$$

which is a normal-inverted-gamma distribution and can be hierarchically specified as follows

$$\boldsymbol{\beta} | \mathbf{X}_{[0k]}, \sigma^2 \sim \mathrm{N}\left(\frac{1}{1-a}\mathbf{b},\ \frac{a\sigma^2}{1-a}(\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]})^{-1}\right) \tag{7}$$

and

$$\sigma^2 | \mathbf{X}_{[0k]} \sim \mathrm{IGa}\left(\frac{b_2}{2a} - \frac{\mathbf{b}^{\top}\mathbf{X}_{[0k]}^{\top}\mathbf{X}_{[0k]}\mathbf{b}}{2a(1-a)},\ \frac{k(1-a)+(l+2)a}{2a}\right). \tag{8}$$

Notice that the priors in Equations (7) and (8) are particular members of the prior family elicited in Section 2.1. Moreover, the conditional prior for $\boldsymbol{\beta}$ is the same used in Zellner's prior for linear models (Marin and Robert, 2007), although Zellner's proposal assumes an improper prior for σ^2.

3. Computational Procedures for the PPM in the Regression Model

This section introduces the algorithm to obtain the posteriors (product distributions) of $\boldsymbol{\beta}_k$ and σ_k^2 at each instant $k = 1, \ldots, n$ and describes the algorithm proposed by Barry and Hartigan (1993) to obtain the product estimates for such parameters. The posterior distribution of p, B, and ρ can be estimated by using the algorithm in Loschi and Cruz (2005).

Assuming that, given ρ, $s \in [ij]$, for $s = 1, \ldots, n$ and $i, j \in I$, $i < j$, we notice that the full conditional distributions of ρ and $\boldsymbol{\theta}_s = (\boldsymbol{\beta}_s, \sigma_s^2)$, for $s = 1, \ldots, n$ are given, respectively, by

$$\pi(\rho|\boldsymbol{\theta}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]}) \propto \Pi_{j=1}^{b} f(\mathbf{Y}_{[i_{j-1}i_j]}|\mathbf{X}_{[i_{j-1}i_j]}) c_{[i_{j-1}i_j]},$$

$$f(\boldsymbol{\theta}_s|\rho, \boldsymbol{\theta}_{-s}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]}) \propto f(\boldsymbol{\theta}_s|\mathbf{Y}_{[0n]}, \mathbf{X}_{[ij]}),$$

where $\mathbf{X}_{[0n]} = (X_1, \ldots, X_n)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$, $\boldsymbol{\theta}_{-s}$ denotes the vector $(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{s-1}, \boldsymbol{\theta}_{s+1}, \ldots, \boldsymbol{\theta}_n)$, $f(\boldsymbol{\theta}_s|\mathbf{Y}_{[0n]}, \mathbf{X}_{[ij]})$ is obtained from Equations (4) and (5) and $f(\mathbf{Y}_{[ij]}|\mathbf{X}_{[ij]})$ is the block predictive distribution of $\mathbf{Y}_{[ij]}$. Thus, it is noticeable that using MCMC techniques we can sample from the posteriors of $\boldsymbol{\theta}_s$ improving the original algorithm proposed by Barry and Hartigan (1993).

3.1 Generating a partition

Let us consider the auxiliary random quantity U_i, suggested by Barry and Hartigan (1993), which reflects whether or not a change point occurs at time i, that is, $U_i = 0$ if $\theta_i \neq \theta_{i+1}$, and $U_i = 1$, otherwise, for $i = 1, \ldots, n-1$. Notice that knowing the random vector $\mathbf{U} = (U_1, \ldots, U_{n-1})$ is equivalent to knowing the random partition ρ. The s-th partition ρ^s, $s \geq 1$, is thus generated by using Gibbs sampling considering the full conditional distributions $U_r|U_1^s, \ldots, U_{r-1}^s, U_{r+1}^{s-1}, \boldsymbol{\theta}_{[0n]}^{s-1}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]})$. Since the U_i's are binary random variables, each component U_r of $\mathbf{U}$ is generated by considering the following ratio

$$\begin{aligned} R_r &= \frac{\mathrm{P}(U_r = 1|U_1^s, \ldots, U_{r-1}^s, U_{r+1}^{s-1}, \boldsymbol{\theta}_{[0n]}^{s-1}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]})}{\mathrm{P}(U_r = 0|U_1^s, \ldots, U_{r-1}^s, U_{r+1}^{s-1}, \boldsymbol{\theta}_{[0n]}^{s-1}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]})} \\ &= \frac{f(\mathbf{Y}_{[xy]}|\mathbf{X}_{[xy]}) \int_0^1 p^{b-2}(1-p)^{n-b+1} d\pi(p)}{f(\mathbf{Y}_{[xr]}|\mathbf{X}_{[xr]}) f(\mathbf{Y}_{[ry]}|\mathbf{X}_{[ry]}) \int_0^1 p^{b-1}(1-p)^{n-b} d\pi(p)}, \end{aligned}$$

for $r = 1, \ldots, n-1$, in which x denotes the last change point before r, y denotes the next change point following r, and, for the linear regression model, the joint density $f(\mathbf{Y}_{[ij]}|\mathbf{X}_{[ij]})$ is the $(j-i)$-variate Student-t distribution given in Equation (6). For more details, see Loschi and Cruz (2005).

3.2 Sampling from the product distributions: the proposed algorithm

To sample from the product distributions of the parameter in the regression model consider the following algorithm. For $s = 1, \ldots, T$,

- Step 1: generate a partition ρ^s from $\pi(\rho|\boldsymbol{\beta}_k^{s-1}, (\sigma_k^2)^{s-1}, \mathbf{Y}_{[0n]}, \mathbf{X}_{[0n]})$; see Section 3.1.
- Step 2: find the block $[ij]$ in ρ^s such that $k \in [ij]$, for $i, j = 1, \ldots, n$ and $i < j$;
- Step 3: generate a sample $(\boldsymbol{\beta}_k^s, (\sigma_k^2)^s)$ from the block posterior distributions of $(\boldsymbol{\beta}_k, \sigma_k^2)$ as follows:

$$\begin{aligned} (\sigma_k^2)^s &\sim \mathrm{IGa}(\nu_{[ij]}^*/2, d_{[ij]}^*/2), \\ \boldsymbol{\beta}_k^s &\sim \mathrm{N}_l(\mathbf{m}_{[ij]}^*, (\sigma_k^2)^s \mathbf{V}_{[ij]}^*). \end{aligned}$$

Notice that after T steps, we have a sample from the posterior distributions of $\boldsymbol{\beta}_k$ and σ_k^2. Then, the product estimates can be approximated by the averages

$$\hat{\boldsymbol{\beta}}_k = \frac{\sum_{s=1}^T \boldsymbol{\beta}_k^s}{T},$$
$$\hat{\sigma^2}_k = \frac{\sum_{s=1}^T (\sigma_k^2)^s}{T}.$$

3.3 Barry and Hartigan's algorithm

Consider Steps 1 and 2 of the proposed algorithm. Step 3 of Barry and Hartigan's algorithm consists of calculating

$$\bar{\boldsymbol{\beta}}_k^s = \mathbf{m}_{[ij]}{}^* \quad \text{and} \quad \bar{\sigma^2}_k^s = \frac{\nu_{[ij]}^*}{d_{[ij]}^* - 2},$$

for each generated sample ρ^s, $s = 1, \ldots, T$. Thus, the product estimates of $\boldsymbol{\beta}_k$ and $\sigma^2{}_k$, for $k = 1, \ldots, n$, are approximated, respectively, by

$$\hat{\boldsymbol{\beta}}_k = \frac{\sum_{s=1}^T \bar{\boldsymbol{\beta}}_k^s}{T},$$
$$\hat{\sigma^2}_k = \frac{\sum_{s=1}^T (\bar{\sigma}_k^2)^s}{T}.$$

Notice that, in this case, the product estimates are the averages of the posterior means by blocks.

In spite of introducing more variability in the computation of the product estimates as compared with that of Barry and Hartigan, the modified algorithm introduced here should be preferred since complete posterior information about regression parameters is available from it and, consequently, some other posterior summaries (besides the posterior means) can be obtained. For similar algorithms based on particle filtering, see Fearnhead and Liu (2007) and Fearnhead (2006).

Remark 3.1 Notice that both algorithms can be adapted for the general PPM introduced by Hartigan (1990). The only thing that we have to take into consideration is the generation from the posterior distribution of ρ. For the contiguous-block case, Barry and Hartigan (1993) propose a Gibbs sampling scheme to generate from such posterior distribution, which is also considered here. A generalization of such a Gibbs sampling scheme for the general PPM (non contiguous-block case) can be found in Crowley (1997). For the general PPM, algorithms to sample from the Dirichlet process (DP) can also be used. Quintana and Iglesias (2003) proved that the DP reduces to the particular case of the general parametric PPM in which the prior cohesion of the set S is given by $c(S) = c(|S| - 1)!$ and c is the weight parameter of the DP. For such particular case, algorithms like those in Bush and MacEachern (1996), MacEachern and Müller (1998) or Neal (2000) can be used to obtain the posterior distributions. However, these algorithms are not applicable for the PPM considered in this paper since, as a consequence of Quintana and Iglesias (2003) statements, we can conclude that there is no association between the DP and the PPM in the contiguous-block case.

4. Case Studies

In this section, the goal is to analyze if the linear relationship between two variables changes throughout time. That is, we initially consider a linear model such that

$$Y = \alpha_k + \psi_k X + \varepsilon_k, \quad \varepsilon_k \overset{\text{ind.}}{\sim} \mathrm{N}(0, \sigma_k^2),$$

where α_k, ψ_k and σ_k^2 denote, respectively, the intercept, slope and variance at time k, $k = 1, \ldots, n$. We consider three data sets to illustrate the previous methodology. Firstly, we analyze the Brazilian industrial production index (BIP) and the Brazilian employment index (BEI). Also analyzed is the Dow Jones industrial average index (DJIA) and the BOVESPA index (IBOVESPA), an indicator that shows the behavior of the principal shares trade on the São Paulo stock exchange.

We compare the least square estimates with the product estimates obtained by applying both Barry and Hartigan's and the proposed algorithms. The least square estimates were computed assuming the blocks pointed out by the posterior most probable partition. The block information is included in the model through dummy variables. We consider R^2 as the goodness of fit measure. The results for the least square method may be obtained by using any basic statistical software.

All algorithms were coded in C++ and for all cases, 11,000 samples of 0-1 values were generated by the Gibbs sampling scheme, with the dimension of the time series, starting from a sequence of zeros. Since the convergence was reached before the 3,000th step, the initial 3,000 iterations were discarded as burn-in. In order to avoid correlation among the vectors a lag of 10 was selected.

According to some expert's opinion, changes in the economy behavior of emerging markets countries are mainly consequence of crises or events that occur in other countries. These important events are country specific. However, they can spread out across countries with a similar economy eventually producing changes in their behavior. Three great financial crises involving emerging markets occurred in the period we are considering for analysis: Mexico's crisis in January, 1995, Asia's crisis in August, 1997, and Russia's crisis in July, 1998. Also, policies adopted by the local governments can also produced changes in the behavior of the economy. From 1964 through 1994, Brazil used to experience huge inflation rates and to combat it, the Federal Government felt the necessity of adopting more extreme measures against it. One effective measure was the Real Plan launched in June 1993. The key elements of the Real Plan was a fiscal strategy creating the Social Emergency Fund, a monetary reform and the opening of the economy with trade liberalization and a new foreign exchange policy. It is well-known that Real Plan brought some stability for the Brazilian economy. Thus, we are expecting few changes in the data and in their relationship. Because we do not have much information about the relationship between BEI and BIP, as well as between the Down Jones and BOVESPA indexes, we also consider flat conjugate priors –that is, we assume priors with large variance– to describe the uncertainty about the intercept, slope and variance. For these reasons, in both case studies we assume the following prior specifications:

$$\begin{aligned}
\alpha_{[ij]}|\sigma^2_{[ij]} &\sim \mathrm{N}(0.0, \sigma^2_{[ij]}),\\
\psi_{[ij]}|\sigma^2_{[ij]} &\sim \mathrm{N}(0.0, \sigma^2_{[ij]}),\\
\sigma^2_{[ij]} &\sim \mathrm{IGa}(0.001/2, 0.001/2),\\
p &\sim \mathrm{Beta}(5.0, 50.0).
\end{aligned}$$

4.1 Case study 1: Brazilian Industrial Production Index (BIP) versus Brazilian Employment Index (BEI)

Here, we consider the Brazilian industrial production index (BIP) and the Brazilian employment index (BEI), recorded monthly from January, 1985, to April, 2001; see Figure 1. The data sets are available at `http://www.ipeadata.gov.br`.

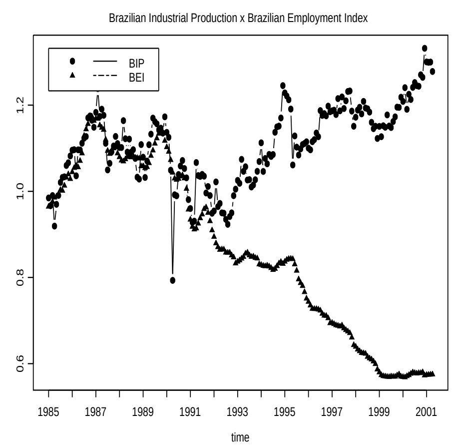

Figure 1. BIP × BEI

In this case, we analyze if the linear relationship given by

$$\text{BEI} = \alpha_k + \psi_k \text{BIP} + \varepsilon_k, \quad \varepsilon_k \overset{\text{ind.}}{\sim} \text{N}(0, \sigma_k^2)$$

changes throughout time. As a consequence of the above it follows that, *a priori*, the expected probability of a change taking place at any instant is close to 9.0% and the expected number of change points, $B-1$, is 17.7.

The posterior most probable partition indicates changes at instants 102 (June, 1993) and 149 (May, 1997) with posterior probability equal to one. We also observe in the posterior analysis that p and $B-1$ are smaller than in the prior evaluation (the posterior means are 2.8% and 2, respectively). The posterior mass function indicates that the instants 102 and 149 are change points with probability one. The other instants have probability equal to zero of being a change point.

From Figure 2, it is noticeable that the estimates by the Bayesian method and least squares are different. In contrast, our proposal and Barry and Hartigan's method provide similar estimates for all parameters. We can also observe that the three methods indicate two important changes in the three parameters at the same time. For the intercept all three methods indicate changes at the same direction, identifying an increase in June, 1993, and a decrease in May, 1997. However, for the slope and the variance the posterior and least square estimates present different behavior. The posterior estimates indicate that the slope is decreasing throughout and that the variance has the same behavior we observed for the intercept, that is, increasing in June, 1993, and decreasing in May, 1997. On the other hand, the least square estimates point out that the variance is decreasing throughout whereas the slope decreases in June, 1993, but increases in May, 1997. The values of R^2, for the proposed method, Barry and Hartigan's, and least squares, are 89.0%, 86.9%, and 87.4%, respectively.

Figure 3 presents the blocks of observations, corresponding to the posterior most probable partition, along with the fitted models. From this figure, we notice that by using the proposed method we will have a large number of fitted models (actually, one slightly different fitted model for each BIP × BEI pair). We can also obtain some other additional information. From Figure 4, we can observe that the variance for the distribution of the three parameters tends to be smaller after May, 1997; see also Table 1. The posterior distributions for the intercept and slope are more symmetric and present less atypical observations than the posterior distribution for the variance. Table 1 provides some descriptive statistics for the posterior distributions at instants 101 to 103, and 148 to 150. It can be noticed that the posterior means and medians are very close.

Table 1. Descriptive statistics for the intercept, slope and variance at some particular instants for BIP × BEI.

	Intercept			Slope			Variance		
Instant	Mean	Median	Variance	Mean	Median	Variance	Mean	Median	Variance
101	0.3718	0.3724	4.0×10^{-3}	0.6102	0.6104	3.5×10^{-3}	0.0099	0.0098	1.0×10^{-6}
102	0.4124	0.4149	6.1×10^{-3}	0.3253	0.3259	4.6×10^{-3}	0.0110	0.0109	2.7×10^{-6}
103	0.4111	0.4087	5.3×10^{-3}	0.3291	0.3274	4.6×10^{-3}	0.0110	0.0108	2.7×10^{-6}
148	0.4125	0.4118	5.7×10^{-3}	0.3228	0.3222	4.6×10^{-3}	0.0110	0.0109	2.7×10^{-6}
149	0.2643	0.2637	2.7×10^{-3}	0.2770	0.2763	1.9×10^{-3}	0.0050	0.0049	0.5×10^{-6}
150	0.2634	0.2658	2.8×10^{-3}	0.2789	0.2794	1.9×10^{-3}	0.0049	0.0049	0.4×10^{-6}

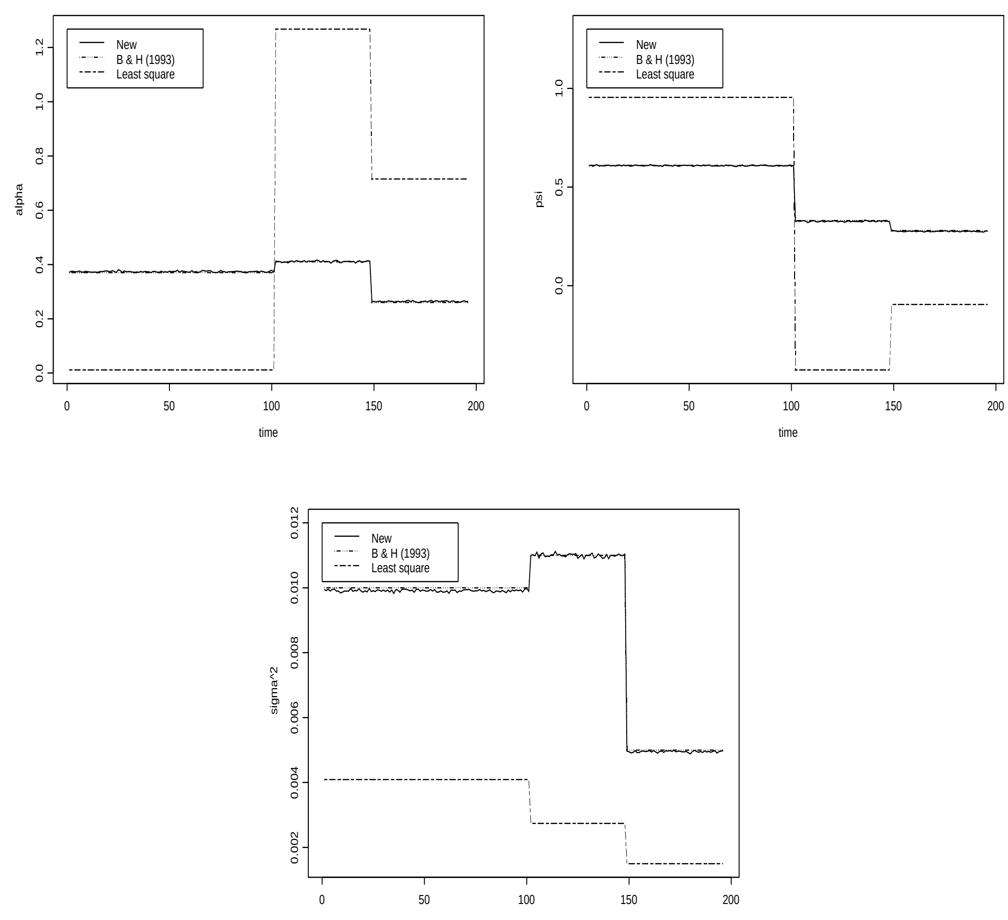

Figure 2. Least square and product estimates for BIP × BEI.

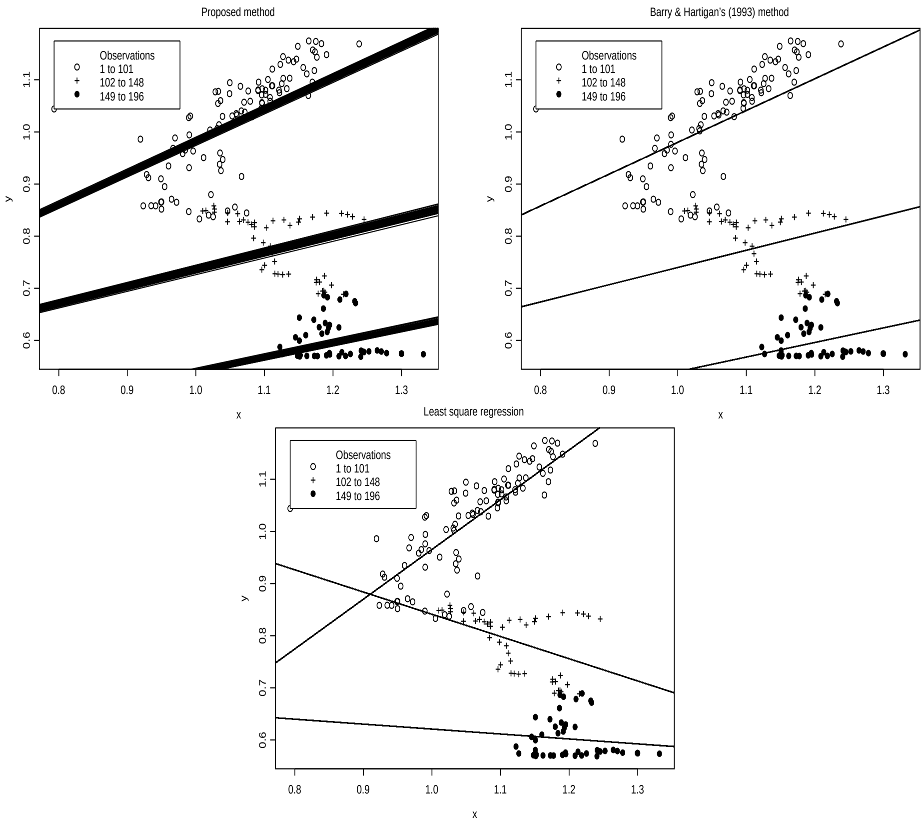

Figure 3. Fitted models for BIP × BEI.

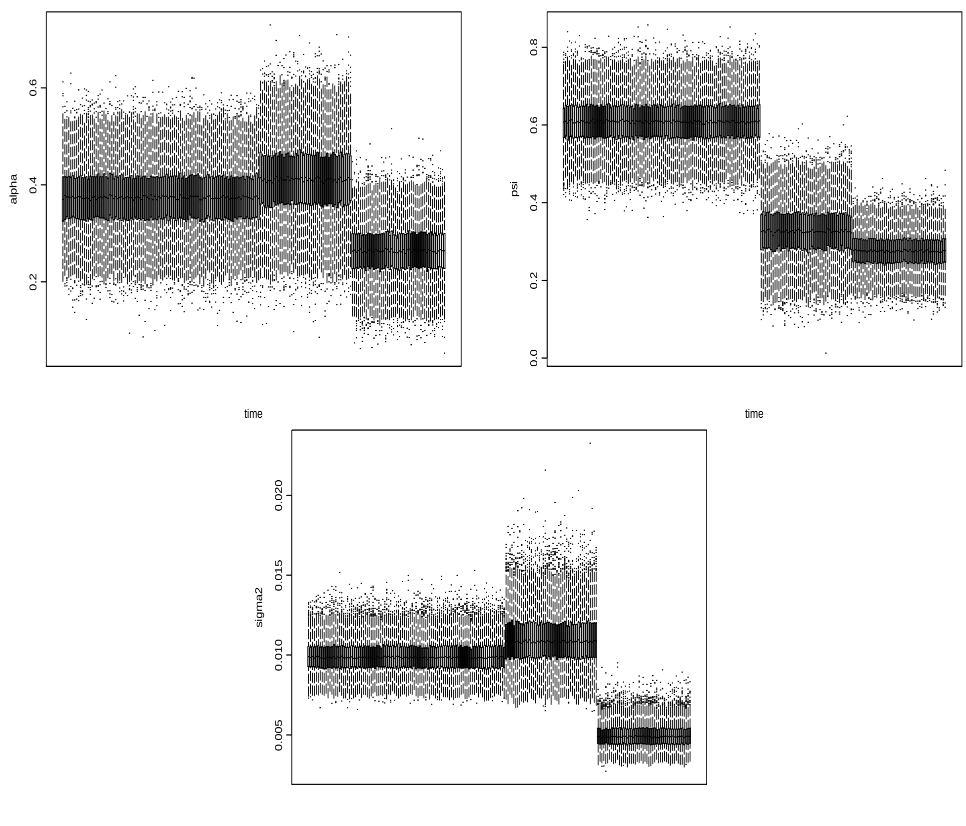

Figure 4. Boxplots of the posterior distributions at each instant for BIP × BEI.

4.2 Case study 2: Dow Jones versus BOVESPA indexes

Changes in the Dow Jones index usually leads to a change in the BOVESPA index. BOVESPA index tends to increase (decrease) whenever Dow Jones index experiences an increase (decrease). However, since the Brazilian economy is an emerging one, it can be affected by international events that do not influence the USA economy. Our interest is to evaluate how this other events influence the relationship between such indexes. In this section we analyze the linear relation through time between the returns of the Dow Jones industrial average (DJIA) and the BOVESPA index (IBOVESPA) recorded fortnightly from January, 1995, to October, 2000. That is, initially, we consider the model

$$\text{IBOVESPA} = \alpha_k + \psi_k \text{DJIA} + \varepsilon_k, \quad \varepsilon_k \overset{\text{ind.}}{\sim} \text{N}(0, \sigma_k^2).$$

As can be noticed from Figure 5, the relation between these two indexes could have experienced a change (at least in the variance) in the middle of 1997. Thus, there is no reason for not using the same prior specifications as in Case study 1.

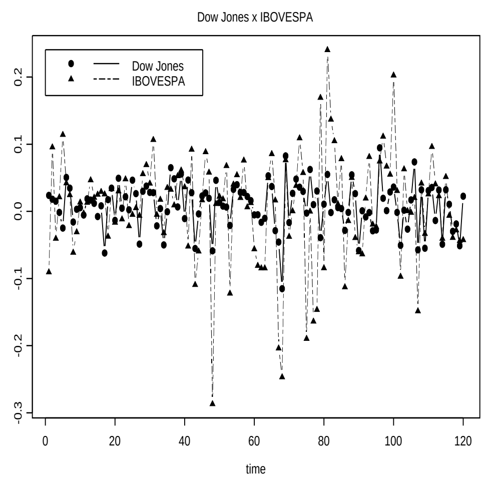

Figure 5. DJIA × IBOVESPA.

The posterior most probable partition indicates a change at the 2nd. fortnight of July, 1997 (instant 42), with low probability (25.6%). We also observe that the 2nd fortnight of July, 1997, has the highest posterior probability of being a change point (26.3%). Most probably, with posterior probability of 84.8%, we have a model with two blocks. Here, we also assume those blocks indicated by the most probable partition, to obtain the least square estimates shown in Figure 6, from which we notice that the most important change indicated by all three methods is close to the 2nd. fortnight of July, 1997. For each parameter, we also observe that all three methods indicate change in the same direction. The product estimates for the slope obtained by using the proposed and Barry and Hartigan's methods, are closer than those observed for the other parameters.

It can be noticed from Figure 7 that the distribution for the variance presents high variability and asymmetry for those instants close to the change point; see also Figure 8. The occurrence of a large number of atypical observations at instants close to the 2nd. fortnight of July, 1997 (instant 42) is also noticeable. From Figure 8 we can also observe that the posterior distribution for the variance at instants close to the change point have two modes.

Figure 9 presents the fitted models for all three methods. The value of R^2 is close to 1.5% for both the proposed and Barry and Hartigan's methods, and it is 27.2% for the least square method. Notice that, in this case, besides the Bayesian model being efficient for change point detection, a piecewise linear regression seems to explain the data poorly.

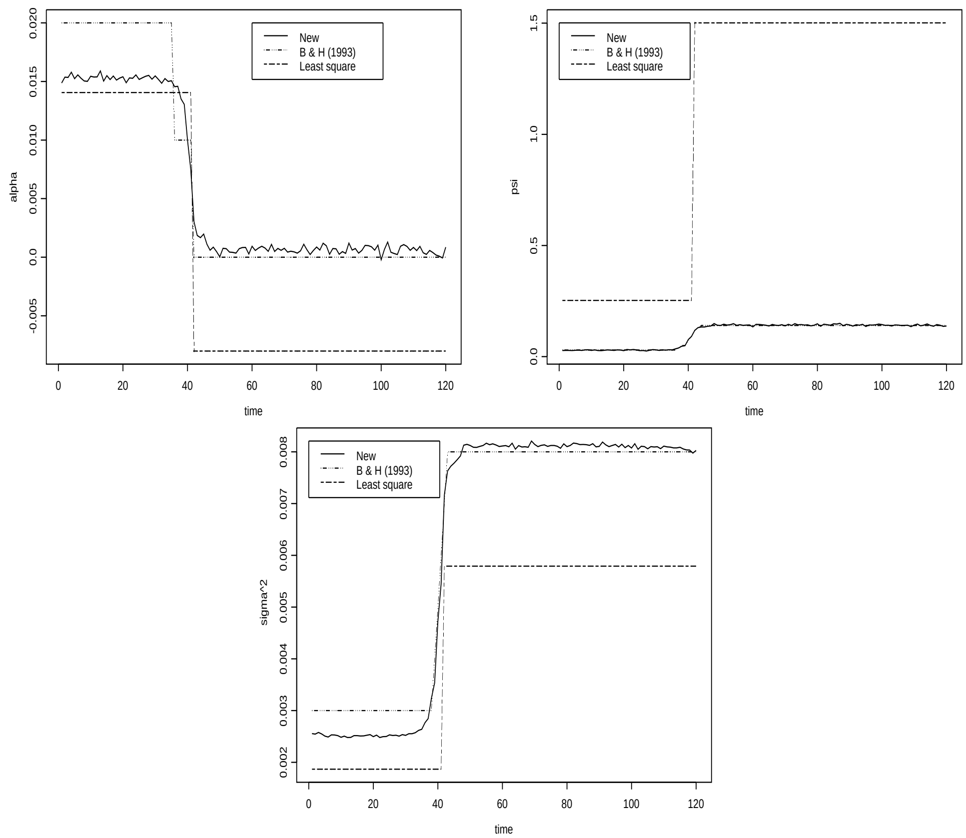

Figure 6. Least square and product estimates for DJIA × IBOVESPA.

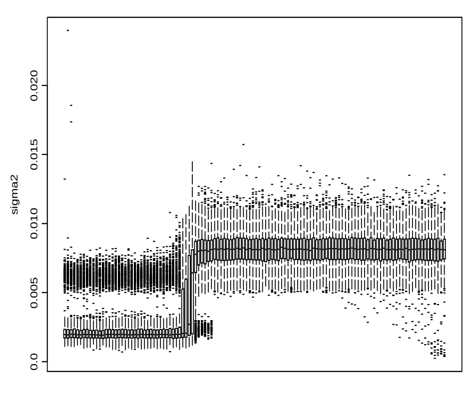

Figure 7. Boxplots of the posterior distributions for the variance at each instant for DJIA × IBOVESPA.

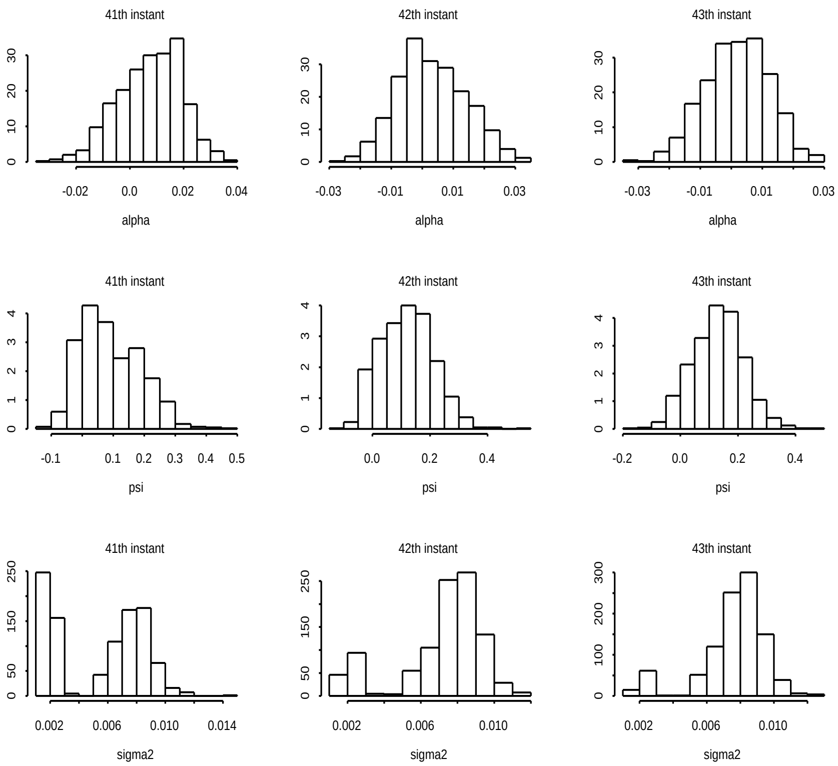

Figure 8. Posterior distributions at instants 41, 42, and 43 for DJIA × IBOVESPA.

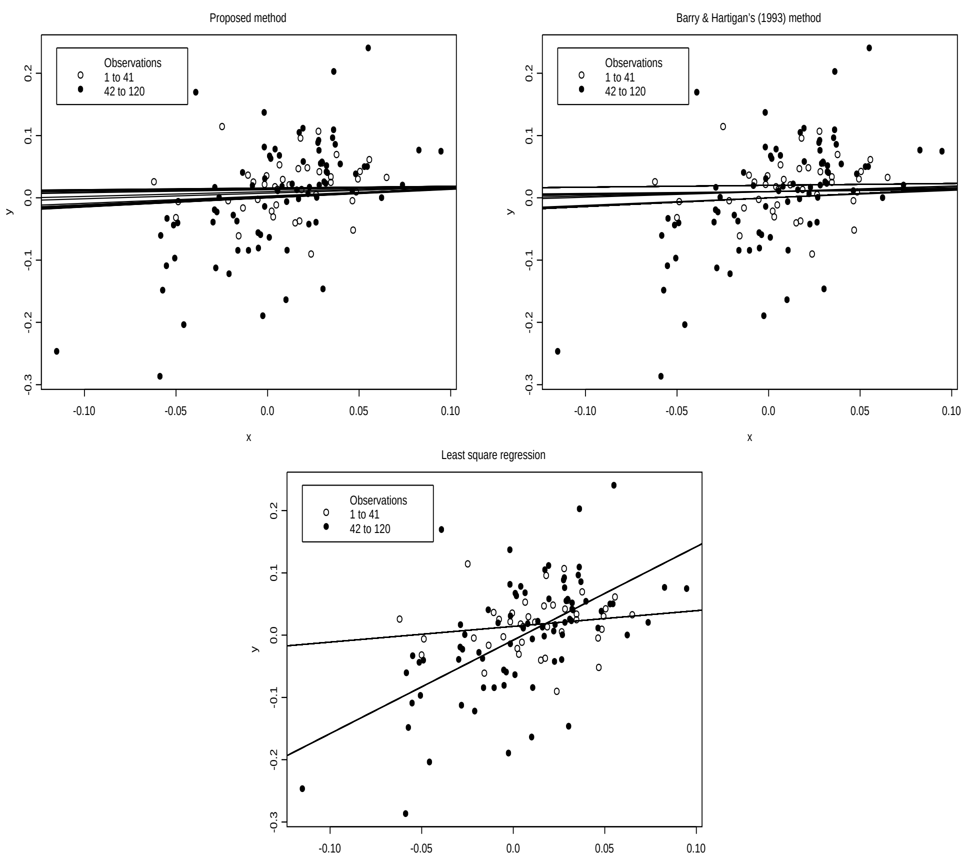

Figure 9. Fitted models for DJIA × IBOVESPA.

Remark 4.1 We also analyzed the volume of sales on Boston Stock Exchange (BSE) and the combined New York and American Stock Exchange (NYAMSE), shown in Holbert (1982), according to whom it was reported in the Business Week (issue of January 3, 1970) that regional stock exchanges were hurt by abolition of give-ups in December 5, 1968. Then Holbert (1982) identified a change point in November, 1968, with 26.1% of chance. Using our methodology, the posterior most probably partition indicates that there is no change points at all in the linear relation between NYAMSE and BSE, with probability of 97.8%. The chance of having two blocks is as low as only 2.3%. We also noticed that the probability of each month to be a change point is below 1.3%, being November, 1967, the month with the highest probability of being a change point.

5. Conclusions and Final Comments

In this paper, we applied the well-known PPM to analyze segmented linear regression models. A predictivistic characterization for the within block distributions was also provided. The advantage of using such approach is that the number of segments (or blocks) is a random variable and thus a non ad-hoc modeling for piecewise regression models can be obtained. We also modify the algorithm proposed by Barry and Hartigan (1993) to obtain the posterior distributions of the parameter of interest at each instant. By using the modified algorithm one can have additional information from the posterior distribution other than the posterior means. We applied the proposed algorithm to analyze two real data sets. In spite of introducing more variability in the computation of the product estimates than Barry and Hartigan's algorithm, we concluded that the new algorithm has good performance, providing posterior means –product estimates, say– very close to those obtainable by Barry and Hartigan's algorithm. Our algorithm should be preferred, since it provides samples from the posteriors; thus some other posterior summaries –besides the posterior means– can be obtained. In fact, this algorithm revealed that the posterior distributions at each instant can by asymmetric as well as multimodal. That is, the posterior means, which were the estimators suggested by Barry and Hartigan (1993), may not provide good estimates for the parameters. For the BIP and BEI data sets, we identified three blocks of observations and observed that the posterior means and least square estimates differ. For the DJIA and IBOVESPA data set, a change point was observed at the 2nd fortnight of July, 1997, and we noticed that the product and least square estimates were similar.

Acknowledgments

The topics discussed in this paper were brought to us by Pilar Iglesias. She did a relevant contribution in partition models but, for sure, in this paper she would be more excited with the predictivistic modeling. "Teoremas de de Finetti estão em meu sangue!": She would say. The authors would like to express their gratitude to the editors and referees for their careful refereeing of the paper. Rosangela H. Loschi would like to express her gratitude to Pilar for her friendship, and for her teachings. Her research has been partially funded by CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) of the Ministry for Science and Technology of Brazil, grants 306085/2009-7, 304505/2006-4. J. G. Pontel is funded by CNPq. The research of Frederico Cruz has been partially funded by CNPq (grants 201046/1994-6, 301809/1996-8, 307702/2004-9, 472066/2004-8, 304944/2007-6, 561259/2008-9, 553019/2009-0, 550207/2010-4), by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior; grant BEX-0522/07-4), and by FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais; grants CEX-289/98, CEX-855/98, TEC-875/07, CEX-PPM-00401/08).

References

Arellano-Valle, R.B., Bolfarine, H., Iglesias, P.L., 1994. A predictivistic interpretation of the multivariate t distribution. Test, 3, 221-236.

Barry, D., Hartigan, J.A., 1993. A Bayesian analysis for change point problem. Journal of the American Statistical Association, 88, 309-319.

Booth, J.G., Casella, G., Holbert, J.P., 2008. Clustering using objective functions and stochastic search. Journal of The Royal Statistical Society Series B – Statistical Methodology, 70, 119-139.

Bormetti, G., De Giuli, M.E., Delpini, D., Tarantola, C., 2010. Bayesian value-at-risk with product partition models. Quantitative Finance (in press). `http://dx.doi.org/10.1080/14697680903512786`

Bush, C.A., MacEachern, S.N., 1996. A semiparametic Bayesian model for randomised block designs. Biometrika, 83, 275-285.

Crowley, E.M., 1997. Product partition models for normal means. Journal of the American Statistical Association, 92, 192-198.

Daboni, L., Wedlin, A., 1982. Statistica-Un'Introduzione All'Impostazione Neo-Bayesiana. Unione Tipografico Editrice Torinese, Torino.

de Finetti, B., 1937. La prévision: ses lois logiques, ses sources subjectives, Annales de l'Institut Henry Poincaré, 7, 1-68. Translated in Kyburg, Jr., H.E., Smokler, H.E., (eds.), 1964. Studies in subjective probability. John Wiley and Sons, New York, pp. 93-158.

de Finetti, B., 1975. Theory of Probability. Volume 1, Wiley, New York.

Demarqui, F.N., Loschi, R.H., Colosimo, E.A., 2008. Estimating the grid of time-points for the piecewise exponential model. Lifetime Data Analysis, 14, 333-356.

Diaconis,P., Eaton, M.L., Lauritzen, S.L., 1992. Finite de Finetti theorems in linear models and multivariate analysis. Scandinavian Journal of Statistics, 19, 298-315.

Diaconis, P., Ylvisaker, D., 1979. Conjugate priors for exponential families. Annals of Statistics, 7, 269-281.

Diniz, C.A.R., Milan, L.A., Mazucheli, J., 2003. Bayesian inference of a linear segmented regression model. Brazilian Journal of Probability and Statistics, 17, 1-16.

Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association, 90, 577-588.

Fearnhead, P., 2006. Exact and efficient bayesian inference for multiple changepoint problems. Statistics and Computing, 16, 203-213.

Fearnhead, P., Liu, Z., 2007. Online inference for multiple changepoint problems. Journal of The Royal Statistical Society Series B – Statistical Methodology, 69, 589-605.

Girón, F.J., Moreno, E., Casella, G., 2007. Objective Bayesian analysis of multiple changepoints for linear models. In Bernardo, J.M., Bayarri, M.J., Berger, J.O., Dawid, A.P., Heckerman, D., Smith A.F.M., West, M., (eds.). Bayesian Statistics 8. University Press, Oxford, pp. 1-27.

Hartigan, J.A., 1990. Partition models. Communications in Statistics – Theory and Methods, 19, 2745-2756.

Hegarty, A., Barry, D., 2008. Bayesian disease mapping using product partition models. Statistics in Medicine, 27, 3868-3893.

Holbert, D., 1982. A Bayesian analysis of a switching linear model. Journal of Econometrics, 19, 77-87.

Hsu, D.A., 1984. A Bayesian robust detection of shift in the risk structure of stock market returns. Journal of the American Statistical Association, 77, 407-416.

Iglesias, P.L., 1993. Finite forms of de Finetti's theorem: a predictivistic approach to statistical inference in finite populations (in portuguese). Unpublished Ph.D. Dissertation. Dept. of Statistics, University of São Paulo, Brazil.

Iglesias, P.L., Pereira, C.A.B., Tanaka, N.I., 1998. Characterizations of multivariate spherical distributions in l_∞-norm. Test, 7, 307-324.

Iglesias, P.L., Loschi, R.H., Pereira, C.A.B., Wechsler, S., 2009. A note on extendibility and predictivistic inference in finite populations. Brazilian Journal of Probability and Statistics, 23, 216-226.

Loschi, R.H., Iglesias, P.L., Arellano-Valle R.B., Cruz, F.R.B., 2007. Full predictivistic modeling of stock market data: application to change point problems. European Journal of Operational Research, 180, 282-291.

Loschi, R.H., Cruz, F.R.B., 2005. Extension to the product partition model: computing the probability of a change. Computational Statistics and Data Analysis, 48, 255-268.

Loschi, R.H., Iglesias, P.L, Arellano-Valle, R.B, 2003. Predictivistic characterization of multivariate Student-t models. Journal of Multivariate Analysis, 85, 10-23.

MacEachern, S.N., Müller, P., 1998. Estimating mixture of Dirichilet process models. Journal of Computational and Graphical Statistics, 7, 223-238.

Majumdar, A., Gelfand, A.E., Banerjee, S., 2005. Spatio-temporal change-point modeling. Journal of Statistical Planning and Inference, 130, 149-166.

Marin, J.M., Robert, C.P., 2007. The Bayesian Core: A Practical Approach to Computational Bayesian Statistics. Springer, New York.

Menzefricke, U., 1981. A Bayesian analysis of a change in the precision of a sequence of independent normal random variables at an unknown time point. Journal of The Royal Statistical Society Series C – Applied Statistics, 30, 141-146.

Müller, P., Quintana, F., 2010. Random partition models with regression on covariates. Journal of Statistical Planning and Inference, 140, 2801-2808.

Neal, R.M., 2000. Markov chain sampling methods for Dirichlet process mixture models. Journal of Computational and Graphical Statistics, 9, 249-265.

Quintana, F.A., 2006. A predictive view of Bayesian clustering. Journal of Statistical Planning and Inference, 136, 2407-2429.

Quintana, F.A., Iglesias, P.L., 2003. Bayesian clustering and product partition models. Journal of The Royal Statistical Society Series B – Statistical Methodology, 65, 557-574.

Quintana, F.A., Iglesias, P.L., Bolfarine, H., 2005. Bayesian identification of outliers and change-points in measurement error models. Advances in Complex Systems, 8, 433-449.

Ruggeri, F., Sivaganesan, S., 2005. On modeling change points in non-homogeneous Poisson processes. Statistical Inference for Stochastic Processes, 8, 311-329.

Smith, A.F.M., 1975. A Bayesian approach to inference about a change-point in a sequence of random variables. Biometrika, 62, 407-416.

Tarantola, C., Consonni, G., Dallaportas, P., 2008. Bayesian clustering for row effects models. Journal of Statistical Planning and Inference, 138, 2223-2235.

Wechsler, S., 1993. Exchangeability and predictivism. Erkenntnis - International Journal of Analytic Philosophy, 38, 343-350.

Yao, Y., 1984. Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches. The Annals of Statistics, 12, 1434-1447.