Semiparametric Multivariate and Multiple Change-Point Modelling

Stefano Peluso* Siddhartha Chib† Antonietta Mira‡

June 18, 2018

Abstract

We develop a general Bayesian semiparametric change-point model in which separate groups of structural parameters (for example, location and dispersion parameters) can each follow a separate multiple change-point process, driven by time-dependent transition matrices among the latent regimes. The distribution of the observations within regimes is unknown and given by a Dirichlet process mixture prior. The properties of the proposed model are studied theoretically through the analysis of inter-arrival times and of the number of change-points in a given time interval. The prior-posterior analysis by Markov chain Monte Carlo techniques is developed on a forward-backward algorithm for sampling the various regime indicators. Analysis with simulated data under various scenarios and an application to short-term interest rates are used to show the generality and usefulness of the proposed model.

Keywords. Bayesian Semiparametric Inference, Dirichlet Process Mixture, Heterogeneous Transition Matrices, Interest Rates.

1 Introduction

Multiple change-point models allow for changes of model distributions at multiple, unknown, time points. These models have been intensively studied in statistics and econometrics over the last several years. The earliest Bayesian change-point models are explored by Chernoff and Zacks [1964], who assume a constant probability of regime change, and by Smith [1975], who investigate the single change-point model under the assumption of exchangeable intra-regime observations and inter-regime independence. The Bayesian approaches usually model the change-point as a stochastic process and conduct inference through Markov chain Monte Carlo (MCMC) algorithms (Carlin et al. 1992, Albert and Chib 1993 among many others) or through the reversible jump MCMC (Green 1995, Green and Mira 2001). Other recent papers on Bayesian change-point problems include

*Corresponding author. Università Cattolica del Sacro Cuore, Department of Statistical Sciences, Largo Gemelli 1, 20123 Milan (Italy). E-mail: stefano.peluso@unicatt.it. Tel.: +39 02 7234 2952

†Washington University in St. Louis, Olin Business School. Email: chib@wustl.edu

‡Università della Svizzera italiana, Institute of Computational Science and Università dell'Insubria, Department of Science and High Technology, E-mail: antonietta.mira@usi.ch

Giordani and Kohn [2008], Pesaran et al. [2006], Maheu and Gordon [2008], Geweke and Jiang [2011].

Our aim is to extend the literature by proposing a new model in which groups of structural parameters (for example location and dispersion parameters) are subject to change-points at different times. Both the number and locations of the change-points are unknown, and are estimated separately for each group of parameters. Our model takes advantage of the widely used formulation of Chib [1998], where the change-point modeling is in terms of a latent discrete state variable that follows a unidirectional Markov process and indicates the regime from which a particular observation has been drawn. In applying this formulation to the proposed model, our multivariate change-point process is described by p latent state variables, one for each group of structural parameter, where each state variable evolves in the manner of the Chib [1998] model. To achieve some further flexibility in the evolution of these state variables we assume that the transition matrix (one for each state variable) depends on time. Thus, the change-point probability depends not only on the regime between two adjacent change points, but also, realistically, on time. In addition, to robustify the distribution of the outcomes we suppose that the prior distribution of each group of structural parameters is unknown and model each prior distribution by a separate Dirichlet process (DP) prior. The Dirichlet concentration parameters are state-dependent and inferred from the data, so that the extent to which estimates deviate from the base parametric measures can differ among groups of structural parameters. We refer to this model as a Bayesian semiparametric multivariate multiple change-point model.

All the model features above have practical relevance in our motivating financial application where we provide a semiparametric Bayesian analysis of a model of short-term risk-less interest rates in which a) the conditional mean and variance are subject to separate change-points b) the probability of regime changes is time-dependent and c) the distribution of the data is heavy-tailed and skewed. We show that our Bayesian semiparametric multivariate multiple change-point model produces improved inferential results and better predictions.

Surveys of related frequentist semiparametric and nonparametric change-point modeling are given by Brodsky and Darkhovsky [1993] and Chen and Gupta [2011]. Within the Bayesian semiparametric literature, Muliere and Scarsini [1985] extend Smith [1975] to two independent DPs on the conditional distributions of the observations and for two different regimes, separated by an independent random change-point. Such independence assumptions were subsequently relaxed by Mira and Petrone [1996] through a mixture of products of DPs that introduces dependence among the two regimes. Barry and Hartigan [1992, 1993], Quintana and Iglesias [2003], Loschi et al. [2003], Loschi and Cruz [2005] and Martinez and Mena [2014] propose a different approach for change-point models, based on random partition distributions, whilst Park et al. [2012] present a Bayesian Poisson change-point regression model with a nonparametric step function as baseline rate.

Closer to our approach, Ko et al. [2015] and Maheu and Yang [2015] use the hierarchical DP to model the rows of the transition matrix of the regimes. DPs on both structural parameters and latent transition matrix are implemented in change-point models by Dufays [2016]. This last framework is extended to multivariate change-points in Bauwens et al. [2015], following the lines of the parametric model of Eo [2012]. We differ from the other approaches, since we model the prior distributions of the structural parameters as DPs, letting the state indicator of the regimes follow Chib [1998]. This allows for greater parsimony, while still permitting recurrence of regimes and an unknown number of regime changes. With no DPs on transition rows, we further detach from the framework of the infinite hidden Markov model of Teh et al. [2006]: the hierarchical DP structure on the transition rows potentially allows for an infinite number of change-points, a property that in finite sample sizes shows its usefulness in the possibility of not specifying a priori the number of regime changes. The price we pay is that we need to specify a *maximum* number of regimes. Furthermore, with the exception of Bauwens et al. [2015], in the mentioned works all structural parameters follow the same change-point process, whilst we distinguish different types of regimes for different structural parameters. Finally, we contribute to the extension of existing approaches through the introduction of more realistic heterogenous time-dependent latent transition matrices, a property that can be important as we show in simulation experiments. By being anchored to the Chib (1998) model, the proposed semiparametric multivariate multiple change-point model still keeps the attractive feature of being completely tractable MCMC techniques.

The rest of the paper is organized as follows. We introduce the model and study its properties in Section 2, with focus on interarrival times and number of change-points, respectively in Section 2.2 and 2.3. Posterior sampling of change-points and structural parameters are discussed in Sections 3.1 and 3.2. The algorithm is applied in Section 4 to simulated data: in Section 4.1, we omit, one at the time, the different components of the proposed model, to informally gauge the relevance of each component; in Section 4.2 we stress the robustness of our method in the face of heavy-tailed data and skewness; finally we implement and compare our method on autocorrelated data in Section 4.3 and in terms of prediction performance in Section 4.4. We present the empirical application to short-term riskless rates in Section 5 and finally provide summary comments in Section 6, highlighting avenues for further work. All the codes, available upon request, have been written in the `R` programming language and run on a PC with core i7-7500U CPU @ 2.70GHz.

2 Model

2.1 Introduction to the model

Let $Y_T := (y_1, y_2, \dots, y_T)$ denote the observed time series, and let $p(y_t|Y_{t-1}, \xi_t)$ denote the conditional distribution of y_t given the history Y_{t-1} and time-specific parameters $\xi_t = (\xi_{1t}, \dots, \xi_{pt})$. In a univariate multiple change-point model, ξ_t changes at latent time points $\tau_1, \dots, \tau_m$, so that $\xi_t = \theta_j$

for all $t \in [\tau_{j-1}, \tau_j)$. The modeling of the distinct values θ_j, for $j = 1, \ldots, m$, is given by a prior distribution F:

$$\begin{aligned} y_t|Y_{t-1}, \xi_t &\sim p(y_t|Y_{t-1}, \xi_t) \\ \xi_t &= \begin{cases} \theta_1 & \text{if } t < \tau_1 \\ \theta_2 & \text{if } \tau_1 \le t < \tau_2 \\ \vdots & \\ \theta_{m+1} & \text{if } \tau_m \le t \le T \end{cases} \\ \theta_j|F &\sim F, \quad j = 1, \ldots, m+1. \end{aligned} \tag{1}$$

The Chib [1998] model is a conceptually and computationally useful re-parameterization of this change-point model. The model is defined in terms of a state variable that follows a uni-directional Markov process. Specifically, let s_t denote a discrete-time discrete-state latent state variable that takes values $\{1, 2, ..., m+1\}$ such that $s_t = j$ means that the structural parameter at time t belongs to the j-th regime. Next, suppose that this state variable is Markovian, and that it can either stay in the current state or move to the next higher state with transitions governed by the probabilities

$$p(s_{t+1} = j + 1|s_t = j) = 1 - p(s_{t+1} = j|s_t = j) = 1 - w_j.$$

Thus, transitions of this state variable from one state to the next higher state isolate the change-points $(\tau_1, \ldots, \tau_m)$.

Now suppose that the structural parameters are grouped as $\xi_t = (\xi_{1t}, \ldots, \xi_{pt})$, where each group of parameters ξ_{it} $(i \le p)$ changes at idiosyncratic latent time points $\tau_{i,1}, \ldots, \tau_{i,m_i}$. Then, $\xi_{it} = \theta_{ij}$ for all $t \in [\tau_{i,j-1}, \tau_{i,j})$. Following the preceding Chib [1998] model, we introduce p discrete-state, discrete time state variables $\{s_t^i\}$, one for each group of structural parameter, such that $s_t^i = j$ now indicates that the i-th structural parameter at time t belongs to the j-th regime, and let each state variable progress as above. To achieve further flexibility in the unidirectional evolution of these state variables we assume that the transition matrix (one for each state variable) depends on time, so that, for $i = 1, \ldots, p$,

$$p(s_{t+1}^i = j + 1|s_t^i = j) = 1 - p(s_{t+1}^i = j|s_t^i = j) = 1 - w_{j,t}^i.$$

Then $1 - w_{j,t}^i$ denotes the probability that the i-th structural parameter moves from the j-th regime at time t to the $(j+1)$-th regime at time $t+1$. For brevity, we denote the change-point process driven by the heterogeneous transition matrices formed by $w_{j,t}^i$ as

$$(\tau_{i1}, \ldots, \tau_{im_i}) \sim P\left(\left\{w_{j,t}^i\right\}\right).$$

We complete the model with a prior F_i on θ_{ij} and on $w_{j,t}^i$. Instead of assuming that each F_i is

parametric, we assume that F_i, $i = 1, \ldots, p$, are unknown. We model these unknown distributions by separate DP priors. More specifically, we suppose that $F_i \sim DP(M_i F_{i0})$, where M_i is the concentration parameter and F_{i0} is the base distribution function. Thus, under this formulation, the distribution of the observations within each regime is a DP mixture (Lo 1984).

Summarizing, our multivariate-multiple change model can be written, for $t = 1, \ldots, T$ and $i = 1, \ldots, p$, as

$$\begin{aligned} y_t|Y_{t-1}, \xi_t &\sim p(y_t|Y_{t-1}, \xi_t) \\ \xi_{it} &= \begin{cases} \theta_{i1} & \text{if } t < \tau_{i1} \\ \theta_{i2} & \text{if } \tau_{i1} \le t < \tau_{i2} \\ \vdots & \\ \theta_{i,m_i+1} & \text{if } \tau_{i,m_i} \le t \le T \end{cases} \\ \theta_{ij}|F_i &\sim F_i, \quad j = 1, \ldots, m_i + 1 \\ F_i &\sim DP(M_i F_{i0}). \end{aligned} \tag{2}$$

2.2 Interarrival times

We now derive some theoretical properties of our model. Here we focus on the implied prior probabilities of interarrival times and the distribution of the number of change-points. As a benchmark, we derive the equivalent results assuming that the transition matrix is time-homogeneous. Note that a priori

$$(\tau_{11}, \ldots, \tau_{1m_1}), \ldots, (\tau_{p1}, \ldots, \tau_{pm_p})$$

are independent, so that the analysis in the present subsection can ignore the multivariate nature of change-points: we can focus on the generic multiple change-points sequence $(\tau_{i1}, \ldots, \tau_{im_i})$, and suppress from all relevant quantities the notational dependence on i.

At time $t = 1$, no change-point can occur and $s_t = 1$. Define this first time point, immune to change by construction, as k_0, so that $\tau_1 > k_0$. In the time-homogeneous case, the transition matrix among the latent regimes is $P = (w_{ij})$, where w_{ij} is the probability (independent of time). Suppose that the single free element in each row of P is given a Beta prior distribution, $w_{ii} \sim Beta(\alpha, \beta)$ and $w_{i,i+1} = 1 - w_{ii}$ for $i = 1, \ldots, m$, $w_{m+1,m+1} = 1$ and all other elements of the matrix are null. It is important to realize that the bi-diagonal nature of the transition matrix is not restrictive in any way, because subsequent states may refer to distributions of previous regimes. Then, in the time-homogeneous setting, $p(\tau_1 = k_0 + k_1|P) = (1 - w_{11}) w_{11}^{k_1 - 1}$, so that $\tau_1 - k_0 \sim Geo(1 - w_{11})$, conditionally on the transition matrix. Unconditionally,

$$p(\tau_1 = k_0 + k_1) = \frac{B(\alpha + k_1 - 1, \beta + 1)}{B(\alpha, \beta)}.$$

Similarly, still in the same setting but with multiple change-points $\tau_1, \ldots, \tau_r$, for any $r \leq m$ and conditionally on P, the r-th interarrival time is $Geo(1 - w_{rr})$, whilst, once the transition matrix is marginalized out, the joint distribution of the interarrival times and the distribution of the r-th interarrival time given the previous ones are, respectively,

$$\begin{aligned} p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_r - \tau_{r-1} = k_r) &= \frac{\prod_{j=1}^{r} B(\alpha + k_j - 1, \beta + 1)}{B(\alpha, \beta)^r} \\ p(\tau_r - \tau_{r-1} = k_r | \tau_{r-1} - \tau_{r-2} = k_{r-1}, \ldots, \tau_1 - \tau_0 = k_1) &= \frac{B(\alpha + k_r - 1, \beta + 1)}{B(\alpha, \beta)}, \end{aligned}$$

where we have defined $\tau_0 = k_0$. We stress that, even if the number of change-points is bounded above by m, the interarrival times can assume any integer value lower or equal to m: a realized r-th interarrival time, $r \leq m$, inducing a change-point beyond the sample size, means that only $r - 1$ change-points actually occurred in the sample.

To ease the analysis of the proposed model, denote

$$k'_j := \sum_{i=1}^{j} k_i. \tag{3}$$

Proposition 2.1. *The probability mass function of the r-th interarrival time in Model* (2)*, conditionally to previous interarrival times and to transition probabilities, is*

$$p(\tau_r - \tau_{r-1} = k_r | \tau_{r-1} - \tau_{r-2} = k_{r-1}, \ldots, \tau_1 - \tau_0 = k_1) = (1 - w_{r,k'_r}) \prod_{t=k'_{r-1}+1}^{k'_r - 1} w_{r,t}.$$

In words, if the transition matrix is time-homogeneous, then the interarrival times are independent of each other, since the realization of the generic j-th interarrival time does not depend on the previous interarrivals. On the other hand, in the proposed model the time-dependent transition matrices create a dependence among interarrival times, more precisely a Markov dependence between an interarrival time and the previous one. Identical distributions of the transition matrices can be obtained as a special case if we choose $w_{j,t} \sim Beta(\alpha, \beta)$ for all j and t. If we further assume that $w_{j,t} = w_{j,t'}$ for all t and t', homogeneity is restored and it is possible to show that the j-th interarrival time $\tau_j - \tau_{j-1} \sim Geo(\alpha/(\alpha + \beta))$.

Corollary 2.2. *Fix $w_{j,t} \sim Beta(\alpha_t, \beta_t)$ for all $j = 1, \ldots, r$. Then, the probability mass function of the r-th interarrival time in Model* (2)*, conditionally to previous interarrival times, is*

$$p(\tau_r - \tau_{r-1} = k_r | \tau_{r-1} - \tau_{r-2} = k_{r-1}, \ldots, \tau_1 - \tau_0 = k_1) = \left(1 - \frac{\alpha_{k'_r}}{\alpha_{k'_r} + \beta_{k'_r}}\right) \prod_{t=k'_{r-1}+1}^{k'_r - 1} \frac{\alpha_t}{\alpha_t + \beta_t}.$$

2.3 Number of change-points

To study the number of change-points, define the stochastic process $\{N_t\}$ as

$$N_t := \{\#\text{ of change-points in } \{1,2,\dots,t\}\}.$$

To simplify the notation in the present subsection, we impose $m = T$, a maximum number of change-points fixed equal to the number of observations. This is a choice that avoids any restriction to the number of regimes. Similar results can be obtained for $m < T$, with truncated versions of the formulas below. If the transition probability from one latent state to the next one is not dependent on the current latent regime, i.e. $w_{ii} = w$ for all i, clearly $N_t|p \sim Bin(t-1, 1-w)$, and the beta prior on w implies a beta-binomial distribution for the number of change-points:

$$P(N_t = x) = \binom{t-1}{x} \frac{B(\alpha + x, \beta + t - 1 - x)}{B(\alpha, \beta)},$$

with an expected number of change-points in $\{1, \dots, t\}$ of $E(N_t) = (t-1)\beta/(\alpha+\beta)$. Chib [1998] generalizes from beta-binomial distributed number of change-points to regime-dependent transition probabilities, collected in the transition matrix P. Therefore, in the time-homogeneous case,

$$N_t|P = \begin{cases} 0, & w_{11}^{t-1} \\ 1, & w_{11}^{t-2}(1-w_{11}) + w_{11}^{t-3}(1-w_{11})w_{22} + \cdots + (1-w_{11})w_{22}^{t-2} \\ \vdots & \\ x, & \sum_{\Omega_{x,t}} \prod_{i=1}^{x+1} w_{ii}^{k_i \mathbb{1}_{x \geq i-1}} (1-w_{ii})^{\mathbb{1}_{x>i-1}} \\ \vdots & \\ t-1, & \prod_{i=1}^{t-1}(1-w_{ii}) \end{cases},$$

where $\Omega_{x,t} := \{k_1, \dots, k_{x+1} : \sum_j k_j = t - 1 - x\}$. Unconditionally, the previous expression can be written as

$$P(N_t = x) = \frac{1}{B(\alpha,\beta)^{x+1}} \sum_{\Omega_{x,t}} \prod_{i=1}^{x+1} B(\alpha + k_i \mathbb{1}_{x \geq i-1}, \beta + \mathbb{1}_{x>i-1}).$$

Proposition 2.3. *Fix $w_{it} \sim Beta(\alpha_t, \beta_t)$ for all $t \leq T$. In Model (2) the number of change-points in $[1, t]$ for a fixed t follows the distribution*

$$P(N_t = x) = \frac{1}{\prod_{i=1}^{t-1} B(\alpha_i \beta_i)} \sum_{k_1} \cdots \sum_{k_x} \left\{ \prod_{j=1}^{x} \left[B(\alpha_{k_j}, \beta_{k_j} + 1) \prod_{k \in (k_{j-1}, k_j)} B(\alpha_k + 1, \beta_j) \right] \cdot \prod_{j=i_x+1}^{t-1} B(\alpha_j + 1, \beta_j) \right\}$$

Remark 2.4. *Fixing $\alpha_t = \alpha$, $\beta_t = \beta$ for all t, the number of change-points N_t simplifies to a binomial random variable with parameters $t-1$ and $\beta/(\alpha+\beta)$. Prior information are reflected in the choice of α_t and β_t: $\beta_t/(\alpha_t+\beta_t)$ can be considered as the prior probability of observing a change-*

point at time t, conditionally to the absence of change-points before t. This implies an expected number of change-points in $[1,t]$ of $\sum_{i=1}^{t-1}\left\{\beta_i/(\alpha_i+\beta_i)\prod_{j=1}^{i-1}\alpha_i/(\alpha_i+\beta_i)\right\}$. Larger deviations from this average probabilty are accepted as reasonable for lower values of α_t and β_t.

3 Posterior Sampling

3.1 Posterior Sampling of Change-Points

We use the upper-case for defining sequence of random variables up to the subscript, for instance $Y_t := (y_1, \ldots, y_t)$. Recall that $w^i_{j,t}$ is the prior probability for the i-th structural parameter to move from the j-th regime at time t to the $(j+1)$-th regime at time $t+1$. Starting from $p = 1$, that is with all regimes parameters moving according to common change-points, $S_T = (s_1, s_2, \ldots, s_T)$ is sampled following Chib [1996], exploiting the decomposition of $p(S_T|Y_T, \theta)$ in

$$p(s_T|Y_T,\theta) \times p(s_{T-1}|Y_T, s_T, \theta) \times \cdots \times p(s_t|Y_T, S^{t+1}, \theta) \times \cdots \times p(s_1|Y_T, S^2, \theta), \tag{4}$$

where $\theta := (\theta_1, \ldots, \theta_{m+1})$ and $S^t := (s_t, s_{t+1}, \ldots, s_T)$. If $s_{t+1} = k+1$, s_t is restricted to be equal to k or $k+1$, then the term $p(s_t|Y_T, S^{t+1}, \theta)$ in (4) is

$$\begin{aligned} p(s_t = k|Y_T, S^{t+1}, \theta) &\propto p(s_t = k|Y_t, \theta)p(s_{t+1}|s_t = k, \theta) \\ &= p(s_t = k|Y_t, \theta)(1 - w_{k,t}). \end{aligned}$$

Note that we can write $p(s_t = k|Y_t, \theta) \propto p(s_t = k|Y_{t-1}, \theta)p(y_t|Y_{t-1}, \theta_k)$, where

$$\begin{aligned} p(s_t = k|Y_{t-1}, \theta) &= \sum_{i=1}^{m+1} p(s_t = k|s_{t-1} = i)p(s_{t-1} = i|Y_{t-1}, \theta) \\ &= w_{k,t-1}p(s_{t-1} = k|Y_{t-1}, \theta) + (1 - w_{k-1,t-1})p(s_{t-1} = k-1|Y_{t-1}, \theta). \end{aligned}$$

At time $t = 1$, $p(s_t = k|Y_{t-1}, \theta) = p(s_1 = k|\theta)$, with all the mass concentrated on $s_1 = 1$. Repeated updating and forecasting forward in time allow the computation of the generic $p(s_t = k|Y_T, S^{t+1}, \theta)$ in (4), starting from the last term $p(s_T = k|Y_T, \theta)$. Then, proceeding backwards, all the terms in (4) can be obtained. Finally, we recover τ from the sampled S_T.

It is relevant to stress that m_1 (m_2) does not represent a number of change-points in the location (variance) parameter fixed in advance, but it is only the maximum number of change-points that can occur: $s_{1,T} \leq m_1 + 1$ and $s_{2,T} \leq m_2 + 1$, and when the inequalities are strict, the number of change-points effectively sampled is less than m_1 and m_2. Therefore, the methodology proposed does not restrict to a fixed number of change-points. From the posterior samples of $S_{1,T}$ and $S_{2,T}$ we can recover the estimated number of change-points: in each Gibbs iteration the final values of the processes indicating the regimes, $s_{1,T}$ and $s_{2,T}$, are respectively the extracted number of

change-points for the first and second structural parameter, from which we can derive joint and marginal posterior distributions.

3.2 Posterior sampling of structural parameters

Assume for the moment that the change-points are univariate, that is all regime-specific parameters move according to common change-points ($p = 1$). For sampling θ, when $p(y_t|Y_{t-1}, \theta)$ and $F_0(\theta)$ are conjugate, Algorithm 2 in Neal [2000] can be slightly modified to be applied in our context. Consider θ^* as the distinct values in θ and $c = \{c_1, \ldots, c_{m+1}\}$ as the clustering configuration: $c_i = j$ tells that θ_i belongs to the j-th cluster, that is, $\theta_i = \theta^*_j$. Sampling of θ is conducted through sampling of θ^* and c. In particular, defining $c_{-i} := \{c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_{m+1}\}$, c is sampled one element at the time according to

$$\begin{aligned} p(c_i = c|Y_T, S_T, c_{-i}, \theta^*) &\propto \frac{m_{-i,c}}{m+M} \prod_{t\in(i)} p(y_t|Y_{t-1}, \theta^*_c) \text{ if some } c_j = c, j \neq i \\ p(c_i \neq c_j|Y_T, S_T, c_{-i}, \theta^*) &\propto \frac{M}{m+M} \int \prod_{t\in(i)} p(y_t|Y_{t-1}, \theta) dF_0(\theta) \text{ for all } j \neq i, \end{aligned} \tag{5}$$

where $m_{-i,c} = |\{j \in \{1, \ldots, m+1\} : \theta_j = \theta^*_c, j \neq i\}|$ is the number of elements in θ_{-i} equal to θ^*_c and $(i) = \{t \in \{1, \ldots, T\} : \xi_t = \theta_i\}$ is the set of times belonging to the i-th regime. Once c is given, θ^* is sampled element wise with probability proportional to the prior F_0, updated with all the data belonging to those regimes which share the same structural parameters:

$$p(\theta^*_i = u|Y_T, S_T, c) \propto \prod_{\{j:c_j=i\}} \prod_{t\in(j)} p(y_t|Y_{t-1}, u) dF_0(u). \tag{6}$$

Note that when $p(y_t|Y_{t-1}, \theta)$ and $F_0(\theta)$ are not conjugate, sampling of θ can be performed under similar considerations, with Algorithm 8 of Neal [2000] replacing the Algorithm 2 mentioned above. Furthermore, the extension to separate change-points for different parameters ($p > 1$) can be handled. For instance, if $\xi_t = (\xi_{1t}, \xi_{2t}) = (\mu, \sigma^2)$, then $\mu|Y_T, S_{1,T}, S_{2,T}, \sigma^2$ and $\sigma^2|Y_T, S_{1,T}, S_{2,T}, \mu$ each follows the scheme above.

Finally, we highlight that, when the sampled latent processes $S_{1,T}$ and $S_{2,T}$, indicator of the regimes, show a number of change-points that is less than the maximum allowed, all the remaining structural parameters (corresponding to the regimes not sampled) are drawn from their priors, following the jump method of Carlin and Chib [1995] .

4 Simulated Examples

4.1 Gaussian Independent Data

We first apply our methodology to simulated Gaussian data, with fixed structural parameters and change-points in the conditional mean and standard deviation. One of the generated samples is depicted in the top left plot of Figure 1: a sample of size $T = 400$, with 3 change-points in the mean and 3 in the standard deviation occurring at time points $(\tau_{11}, \tau_{12}, \tau_{13}) = (100, 200, 300)$ and $(\tau_{21}, \tau_{22}, \tau_{23}) = (122, 223, 325)$, and with structural parameters equal to $(\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}) = (0.5, 1, 0.25, 0.75)$ and $(\theta_{21}, \theta_{22}, \theta_{23}, \theta_{24}) = (0.3, 0.6, 0.15, 0.45)$. Change-points are highlighted by the vertical lines in the plot.

Our full model, denoted *FullMod*, is summarized, for $t = 1, \ldots, T$ and for $i = 1, 2$ as

$$
\begin{aligned}
y_t|\xi_t &\sim N(y_t|\xi_{1t}, \xi_{2t}) \\
\xi_{it} &= \begin{cases} \theta_{i1} & \text{if } t < \tau_{i1} \\ \theta_{i2} & \text{if } \tau_{i1} \leq t < \tau_{i2} \\ \vdots & \\ \theta_{i,m_i+1} & \text{if } \tau_{i,m_i} \leq t \leq T \end{cases} \\
\theta_{ij}|F_i &\sim F_i, \quad j = 1, \ldots, m_i + 1 \\
F_1 &\sim DP(M_1 \cdot N(\cdot|\mu_0, \lambda^2)), \quad F_2 \sim DP(M_2 \cdot IG(\cdot|\alpha_\sigma, \beta_\sigma)) \\
(\tau_{i1}, \ldots, \tau_{im_i}) &\sim P\left(\left\{w^i_{j,t}\right\}\right),
\end{aligned}
$$

where $N(\cdot)$ and $IG(\cdot)$ are, respectively, the Gaussian and Inverse Gamma distributions. The hyperparameters are fixed to $\mu_0 = 0$, $\lambda = 1$, $\alpha_\sigma = 1$, $\beta_\sigma = 1$, $w^i_{j,t} \sim Beta(1, 1)$ for all i, t and $j < m_i$ and $w^i_{m_i,t} = 1$ for all i, t. Note that all priors are centered on values distant from the true ones, since we actually want to test if the proposed model is able to correct for prior information not necessarily coherent with the data. The concentration parameters M_1 and M_2 have a $Gamma(0.05, 0.0001)$ prior, centered on a highly noninformative value, and are sampled a posteriori following Escobar and West [1995]. Finally, we always start the algorithm with no change-points for the structural parameters.

We run the algorithm for posterior sampling proposed in Section 3 for a total number of iterations $N = 1000$, of which the first $N/2$ are discarded as burn-in, over $K = 40$ simulated datasets. In the top right and in the bottom left plots of Figure 1 we report the posterior probabilities that s_{1t}, the latent regime for the conditional mean, is equal to $k = 1, \ldots, m_1 + 1$ and that s_{2t}, the latent regime for the conditional standard deviation, is equal to $k = 1, \ldots, m_2 + 1$, for $t = 1, \ldots, T$, averaged over the K simulated datasets and for fixed values of $m_1 = 3$, $m_2 = 3$. It is clear that the proposed multivariate algorithm perfectly identifies all the change-points, properly distinguishing between changes in mean from changes in variance. Furthermore, we correctly identify the unknown

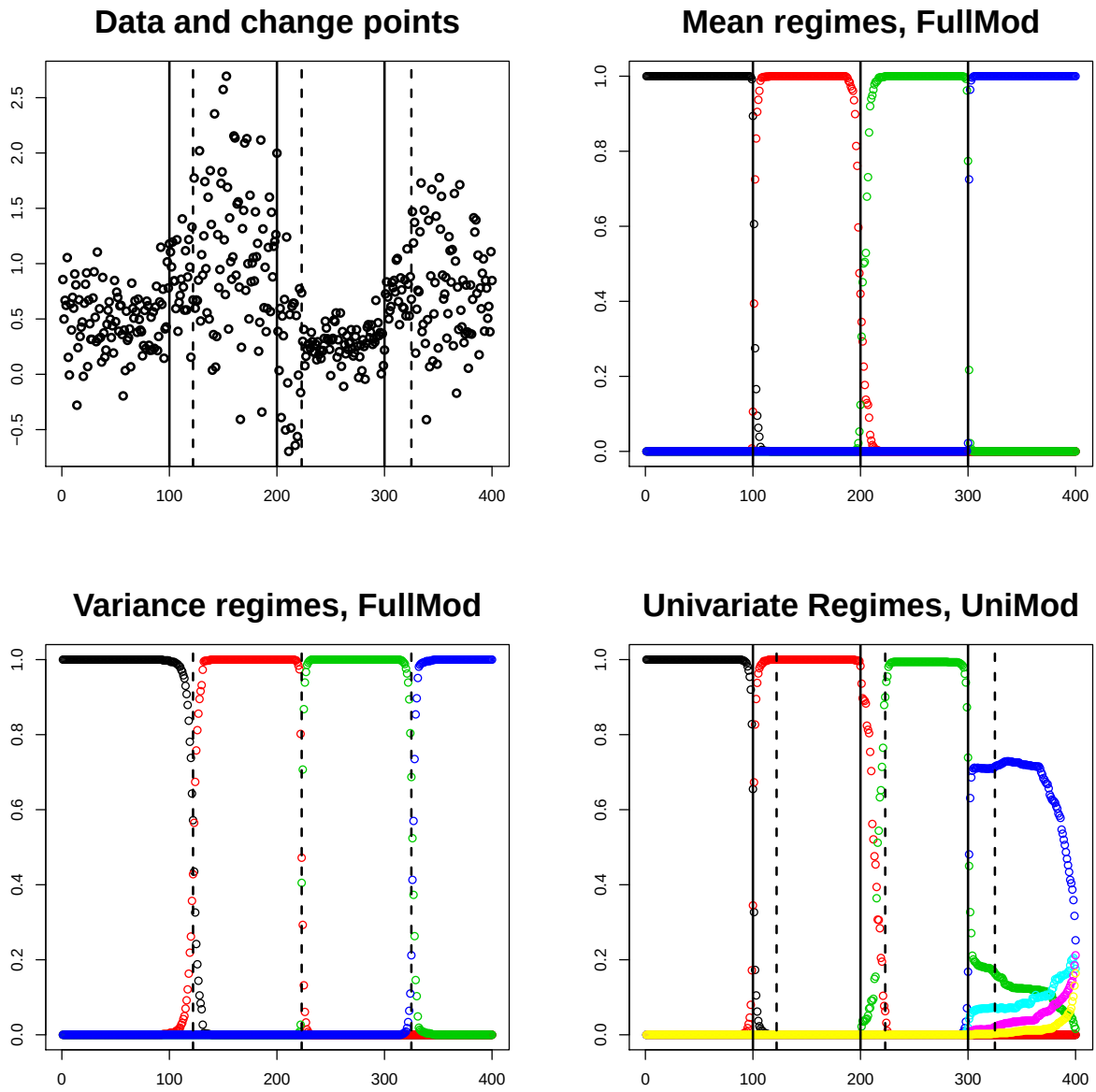

Figure 1: Independent Gaussian simulation study, 40 simulated datasets. Top left: simulated sample with three change-points in the mean (vertical continuous lines) and three change-points in the standard deviation (vertical dashed lines). Top right: posterior probabilities of $s_{1t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$. Bottom left: posterior probabilities of $s_{2t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$. Bottom right: posterior probabilities of $s_t = k$, for $k = 1, \ldots, 7$ and $t = 1, \ldots, 400$, for multiple univariate change-points simultaneously in mean and standard deviation.

number of regimes: the posterior marginal distributions for the number of mean regimes and for the number of variance regimes are correctly centered on the true values.

To highlight the importance of the single features of the proposed multivariate change-point model (multivariate change-point process, random structural parameters distributions, heterogeneous transition matrices), we compare, for equal hyperprior values and starting points, *FullMod* with alternative models having specific features turned off: a) model *UniMod*, with a single change-point process, so that conditional variances and means have common univariate change-points, b) model *NoDpMod*, with structural parameters θ drawn from a fixed, and not from a random DP, distribution, c) model *NoHetMod* with transition matrices among latent states homogeneous over time, d) model *CauchyMod*, where structural parameters θ are drawn from a fixed distribution, with, in particular, the variance following a priori the half-Cauchy distribution (see Polson and Scott 2012), characterized by tails fatter than those in the Inverse Gamma distribution.

In more details, the univariate model *UniMod* can be written in hierarchical form as

$$\begin{aligned}
y_t|\xi_t &\sim N(y_t|\xi_{1t},\xi_{2t}) \\
(\xi_{1t},\xi_{2t}) &= \begin{cases} (\theta_{11},\theta_{21}) & \text{if } t<\tau_1 \\ (\theta_{12},\theta_{22}) & \text{if } \tau_1 \le t < \tau_2 \\ \vdots & \\ (\theta_{1,m+1},\theta_{2,m+1}) & \text{if } \tau_m \le t \le T \end{cases} \\
\theta_{ij}|F_i &\sim F_i, \quad j=1,\dots,m_i+1 \\
F_1 &\sim DP(M_1 \cdot N(\cdot|\mu_0,\lambda^2)), \quad F_2 \sim DP(M_2 \cdot IG(\cdot|\alpha_\sigma,\beta_\sigma)) \\
(\tau_1,\dots,\tau_m) &\sim P\left(\{w_{j,t}\}\right),
\end{aligned}$$

whilst model *NoDpMod* can be represented as *FullMod*, with the exception that $\theta_{ij}|F_i$ and F_i in *FullMod* are replaced by

$$\theta_{1j} \sim N(\cdot|\mu_0,\lambda^2), \quad \theta_{2j} \sim IG(\cdot|\alpha_\sigma,\beta_\sigma).$$

Model *NoHetMod* replaces the last row in *FullMod* with

$$(\tau_{i1},\dots,\tau_{im_i}) \sim \prod_{j=1}^{m} B(\alpha+\tau_j-\tau_{j-1},\beta+1)/B(\alpha,\beta)^m,$$

where the change-points are implied by latent processes and transition matrices sampled as in Chib [1998]. Finally, model *CauchyMod* retraces model *NoDpMod*, but with a fat-tails half-Cauchy prior for the standard deviation structural parameters, for $j=1,\dots,m_2$, proportional to $1/(1+\theta_{2j})$.

For all models under comparison, we run the Gibbs samplers for N iterations (whose $N/2$ discarded as burn-in), over the same K simulated datasets. For the algorithm with univariate change-points (model *UniMod*), the results on the regime posterior probabilities are reported in

the bottom right plot of Figure 1. The univariate algorithm is not able to identify all the change-points, missing one change-point in the mean and two change-points in the variance. Every mean change-point is closely followed by a variance change-point, therefore a unique sampling step for both types of change-points makes the inferential problem more complicated, since the structural parameters of the regime between every two close change-points can rely on few data. For the Gibbs sampler of model *NoDpMod*, top left and right plots in Figure 2 show that the removal of the DP random distribution for the structural parameters causes some estimation problem in the variance change-points, whilst change-points in the mean are correctly identified. On average the model identifies the correct variance change-points, but with much lower posterior probabilities. The weakness of model *NoDpMod* is solved by the variance prior with fatter tails in the parametric model *CauchyMod* in Figure 3. Finally, the sampling algorithm for *NoHetMod* in bottom left and right plots of Figure 2 does not correctly identify the conditional mean change-points, whilst it is able to find the variance change-points, but with low posterior probabilities.

The performance of all models on the estimation of the structural parameters and of mean and variance change-point identification is summarized in Table 1, for various scenarios on the allowed maximum number of change-points. We let (m_1, m_2), ranging in $\{(3,3),(10,3),(3,10),(10,10)\}$, to test the robustness of the results to different specifications of m_1 and m_2. Root Mean Square Errors are reported, separated for mean and variance structural and change-point parameters. For instance, at the row denoted with θ_1, is reported the quantity $\frac{1}{K}\sum_{k=1}^{K}\sqrt{\sum_{j=1}^{m_1}(\hat{\theta}_{1j}^{(k)}-\theta_{1j})^2}$, in which θ_{1j} is the true mean structural parameter of the jth regime, and $\hat{\theta}_{1j}^{(k)}$ is the corresponding estimate in the kth dataset, and similarly for the other parameters. There is a clear distinction among the competing models as far as estimation of the regime means is concerned: models *FulMod* and *CauchyMod* perform well, with no clear ranking between the two: *FulMod* seems to better infer the mean structural parameters and the mean change-points, whilst *CauchyMod* has a better performance on the variance parameters, and there is no majority of scenarios in which one model prevails on the other. Models *NoDpMod*, *NoHetMod* and *UniMod* are more distant from the true parameter values. In the sequel, we focus on the comparison between *FullMod* and *CauchyMod*, omitting for brevity the results related to the other models, since those are the two best performing models.

4.2 Mispecified Heavy Tails and Skewness

We test the robustness of the simulation results to model mispecifications that do not account for heavy tails and skewness in the data generating mechanism. We first generate the data from a t-Student with 3 degrees of freedom, with location and scale given by the structural parameters in the previous sections, and with the same change-point locations. The models we compare are *FullMod* and *CauchyMod*, as presented in the previous subsection. The values of m_1 and m_2 are fixed to their true values, namely 3 and 3, but also alternative scenarios with $(m_1, m_2) \in \{(10,3),(3,10),(10,10)\}$

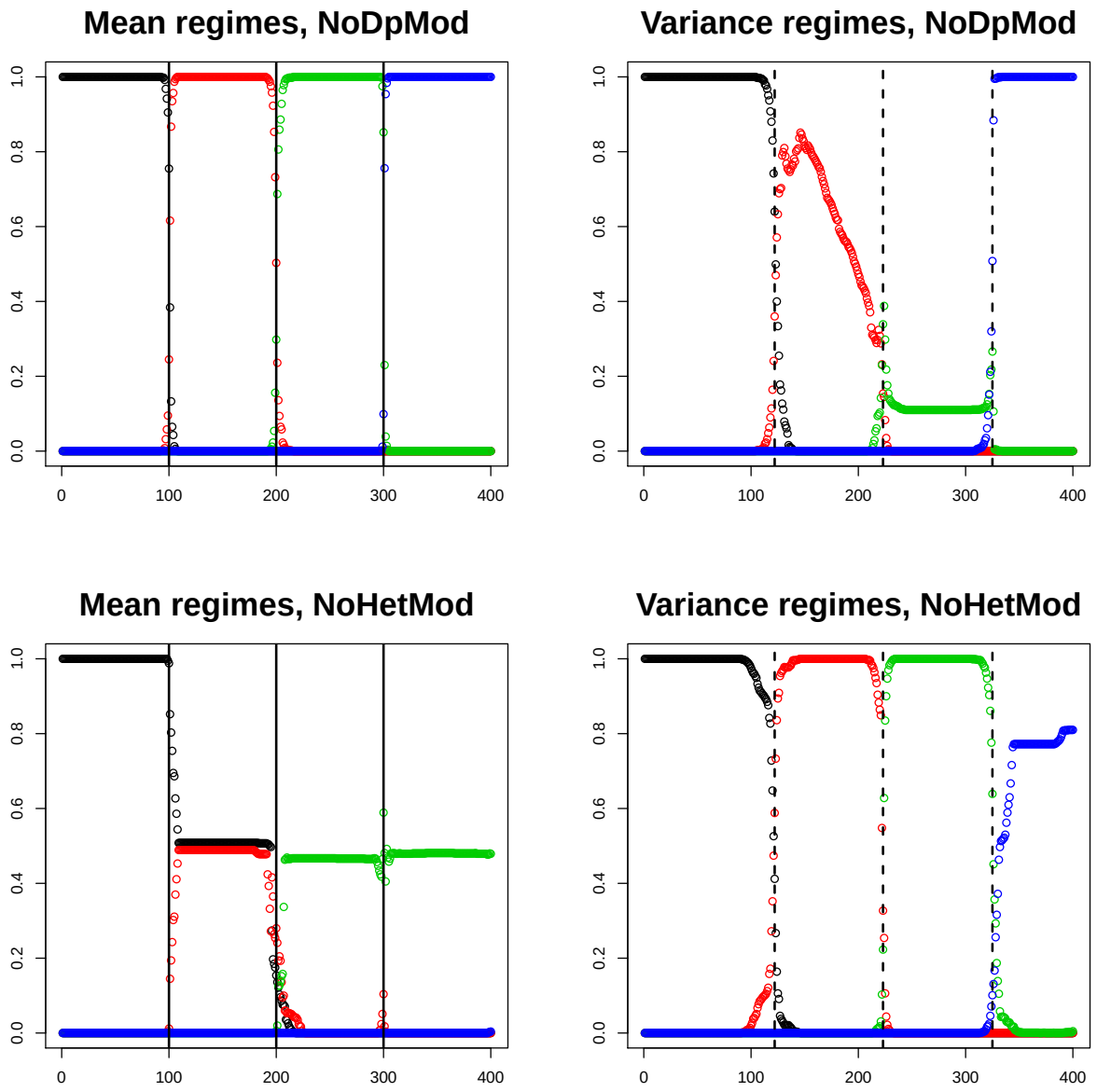

Figure 2: Independent Gaussian simulation study, 40 simulated datasets. Top left: posterior probabilities of $s_{1t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *NoDpMod* without structural parameters DP random distribution. Top right: posterior probabilities of $s_{2t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *NoDpMod* without structural parameters DP random distribution. Bottom left: posterior probabilities of $s_{1t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *NoHetMod* without heterogeneous change-points. Bottom right: posterior probabilities of $s_{2t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *NoHetMod* without heterogeneous change-points.

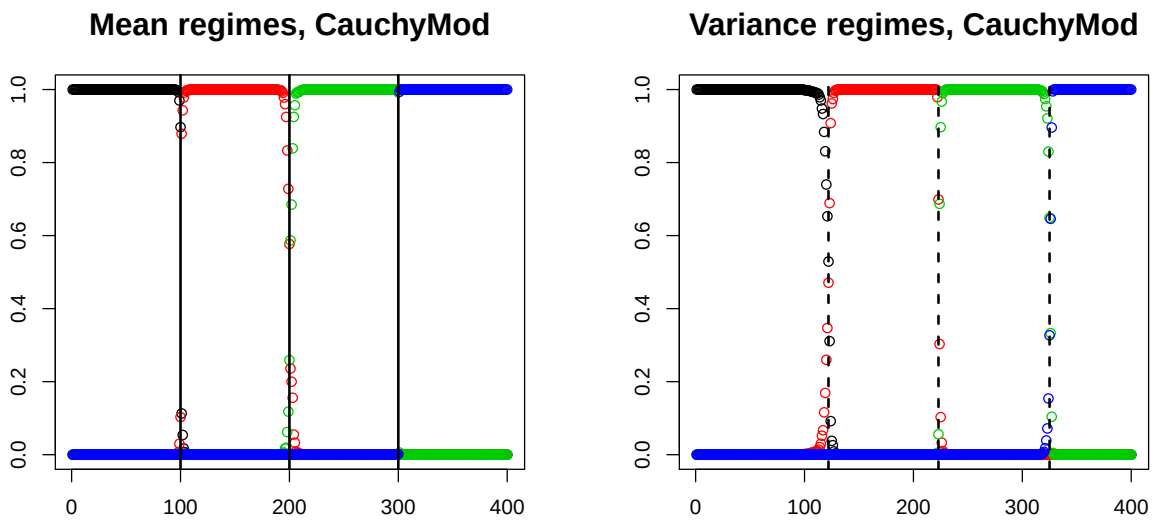

Figure 3: Independent Gaussian simulation study, 40 simulated datasets. Left: posterior probabilities of $s_{1t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *CauchyMod* without structural parameters DP random distribution and half-Cauchy variance prior. Right: posterior probabilities of $s_{2t} = k$, for $k = 1, \ldots, 4$ and $t = 1, \ldots, 400$, for model *CauchyMod* without structural parameters DP random distribution and half-Cauchy variance prior.

RMSE	*FullMod*	*CauchyMod*	*NoHetMod*	*NoDpMod*	*UniMod*
$(m_1, m_2) = (3, 3)$					
θ_1	0.9488	0.5037	1.3039	**0.3231**	1.1087
θ_2	0.4907	**0.1146**	1.8357	0.4418	2.2416
τ_1	68.8422	31.3566	81.9368	**22.0065**	218.3378
τ_2	27.1280	**15.5053**	42.6697	71.7657	218.3378
$(m_1, m_2) = (10, 3)$					
θ_1	**0.0658**	0.3462	0.9261	0.1128	-
θ_2	**0.3484**	0.4711	1.1628	0.7572	-
τ_1	**3.5803**	51.0688	55.5649	12.1835	-
τ_2	117.3375	245.5747	**82.1960**	176.4352	-
$(m_1, m_2) = (3, 10)$					
θ_1	**0.2838**	0.6861	0.6185	0.6576	-
θ_2	0.6503	**0.1011**	1.5694	0.3741	-
τ_1	**60.9658**	203.9357	101.5693	193.1653	-
τ_2	35.7168	**16.5521**	66.7635	71.3163	-
$(m_1, m_2) = (10, 10)$					
θ_1	**0.3711**	0.6467	0.5622	0.6663	0.5754
θ_2	0.4588	**0.4521**	1.8970	0.7483	0.7069
τ_1	87.2464	217.0935	**66.1152**	209.9879	99.5400
τ_2	107.1734	232.5288	156.7875	171.1133	**99.5400**

Table 1: Estimation performances in the independent Gaussian simulation study. Root Mean Square Errors are reported, for the structural parameters θ_1 and θ_2, and for the mean and variance change-point locations τ_1 and τ_2, with true values $m_1 = 3$, $m_2 = 3$, averaged over 40 simulated datasets.

are analyzed, to measure the impact of a change in the allowed maximum number of change-points. The estimation of structural parameters and change-point locations for mean and variance are reported in the second and third column of Table 2 (*Full-H* and *Cauchy-H*). The conclusions are similar for both the models under study: there is some deterioration in the estimation of the structural parameters, relative to the i.i.d. case, all change-point locations are correcly identified (results omitted for brevity) and there is no clear ranking between the two models.

To test the robustness of the proposed model to mispecified skewed data, we generate the sample from a skew-normal distribution [Azzalini, 2013], with location and scale fixed to the structural parameters of the previous section, and with asymmetry parameter equal to 4, which is an average case among the simulation settings presented in Chapter 2 of Azzalini [2013]. The performance of *FullMod* and *CauchyMod* are reported, for the structural parameters and for change-point locations, in Table 2 (columns *Full-S* and *Cauchy-S*), from which we see that *FullMod* performs better: the heavy-tailed half-Cauchy prior is able to react to the mispecification induced by data with heavy tails, but not to skewed data.

RMSE	*Full-H*	*Cauchy-H*	*Full-S*	*Cauchy-S*	*Full-C*	*Cauchy-C*
$(m_1, m_2) = (3, 3)$					$(m_1, m_2) = (1, 2)$	
θ_1	0.8214	**0.1518**	**0.7510**	1.1234	(**0.16**,**0.09**)	(0.19,0.10)
θ_2	**1.5745**	2.3615	0.7937	**0.3065**	**0.6716**	0.9168
τ_1	63.5750	**12.3639**	**32.5520**	69.6523	**32.4992**	32.7980
τ_2	**62.8552**	88.4939	**33.1680**	51.2301	**3.4710**	6.0634
$(m_1, m_2) = (10, 3)$					$(m_1, m_2) = (10, 2)$	
θ_1	0.4552	**0.2466**	**0.5030**	0.9784	(0.20,**0.10**)	(**0.17**,0.10)
θ_2	1.7625	**1.4472**	**0.3512**	0.4956	**2.1517**	3.1494
τ_1	48.1766	**23.4285**	**6.0977**	114.4956	43.1919	**33.9257**
τ_2	**175.7571**	258.6519	**102.7696**	225.9175	**73.4370**	103.9039
$(m_1, m_2) = (3, 10)$					$(m_1, m_2) = (1, 10)$	
θ_1	1.6263	**1.2585**	**0.9025**	1.1140	(**0.19**,**0.08**)	(0.20,0.09)
θ_2	**2.0725**	7.9320	0.6370	**0.3361**	**0.8052**	0.8677
τ_1	**144.2693**	201.1020	**81.2992**	182.7112	**32.9816**	190.8363
τ_2	71.1654	**41.1893**	36.5586	**31.5398**	**4.6904**	6.3036
$(m_1, m_2) = (10, 10)$					$(m_1, m_2) = (10, 10)$	
θ_1	**1.0033**	1.1392	**0.7567**	1.0752	(**0.20**,**0.11**)	(0.22,0.11)
θ_2	1.7180	**1.3801**	**0.4180**	0.4812	**2.4276**	3.1172
τ_1	**137.5777**	220.6925	**72.5450**	196.6619	**356.4523**	555.1267
τ_2	**157.4713**	227.5572	**94.2336**	232.8854	**81.6437**	107.9413

Table 2: Estimation performances in the simulation study with heavy tails (columns *Full-H* and *Cauchy-H*), skewness (columns *Full-S* and *Cauchy-S*) mispecifications and with autocorrelated times series (columns *Full-C* and *Cauchy-C*). Root Mean Square Errors are reported, for the structural parameters θ_1 and θ_2, and for the mean and variance change-point locations τ_1 and τ_2, with true values $m_1 = 3$, $m_2 = 3$ ($m_1 = 1$ and $m_2 = 2$ in the autocorrelated scenarios), averaged over 40 simulated datasets.

4.3 Autocorrelated Scenarios

We now introduce autocorrelation in the generated data and into the models *FullMod* and *CauchyMod*. Therefore the hierarchical structure of *FullMod* can be written as

$$
\begin{aligned}
y_t|\xi_t &\sim N(y_t|\xi_{1t} + \xi_{2t} y_{t-1}, \sqrt{\xi_{3t}}) \\
(\xi_{1t}, \xi_{2t}) &= \begin{cases} \theta_{11} & \text{if } t < \tau_{11} \\ \theta_{12} & \text{if } \tau_{11} \le t < \tau_{12} \\ \vdots & \\ \theta_{m_1+1} & \text{if } \tau_{m_1} \le t \le T \end{cases}, \quad \xi_{3t} = \begin{cases} \theta_{21} & \text{if } t < \tau_{21} \\ \theta_{22} & \text{if } \tau_{21} \le t < \tau_{22} \\ \vdots & \\ \theta_{m_2+1} & \text{if } \tau_{m_2} \le t \le T \end{cases} \\
\theta_{ij}|F_i &\sim F_i, \quad j = 1, \ldots, m_i + 1 \\
F_1 &\sim DP(M_1 \cdot N(\cdot|\mu_\gamma, V_\gamma)), \quad F_2 \sim DP(M_2 \cdot IG(\cdot|\alpha_\sigma, \beta_\sigma)) \\
(\tau_{i1}, \ldots, \tau_{im_i}) &\sim P\left(\left\{w^i_{j,t}\right\}\right),
\end{aligned}
$$

where, for a sample size $T = 240$, we have fixed one change-point for θ_1 at $t = 120$, and two change-points for θ_2 at times 80 and 160, in order to have a simulation setting where the two structural parameters have a different number of change-points. The location structural parameter changes from $(0, 0)'$ to $(2, 0.3)'$, whilst the variance structural parameter changes from 3 to 0.125 and then to 3.5. Hyperprior parameters are chosen to have diffuse prior distributions for θ_1 and θ_2: $\mu_\gamma = (0, 0)'$, $V_\gamma = 10 I_2$, where I_2 is the 2-by-2 identity matrix, $\alpha_\sigma = \beta_\sigma = 0.1$.

The scenarios considered are averaged over $K = 40$ artificial datasets, and m_1 and m_2 span in $\{(1, 2), (10, 2), (1, 10), (10, 10)\}$. A Gibbs sampler is implemented for $N = 2000$ iterations, whose the first $N/2$ are discarded, for *FullMod* and *CauchyMod*. The RMSEs in the last two columns of Table 2 (*Full-C* and *Cauchy-C*) show that the estimation of the structural parameters and of the change-points is significantly better in *FullMod*.

4.4 Prediction Ability

We compare *FullMod* and *CauchyMod* in terms of their prediction capabilities, using the log predictive density: we estimate the log predictive density of each observation at time t^* conditionally to the whole past, for t^* in $[T/2 + 1, T]$, and use the average log predictive density over this range, as a measure of predictive ability. The log predictive density at time t^* is estimated as

$$
\sum_{i=1}^{N} \log p(y_{t^*}|y_{t^*-1}, \xi^*_{1t,i}, \xi^*_{2t,i}, \xi^*_{3t,i}) p(\xi^*_{1t,i}, \xi^*_{2t,i}, \xi^*_{3t,i}|y_1, y_2, \ldots, y_{t^*-1}),
$$

where N is the number of posterior samples, and $\xi^*_{1t,i}$, $\xi^*_{2t,i}$ and $\xi^*_{3t,i}$ are the extractions of the parameters at the i-th iteration of a Gibbs sampler that uses data only up to t^* excluded.

A model with better prediction capability presents a higher average log predictive density. The

results over the K simulated datasets with autocorrelated data, shown in Figure 4, strenghten the conclusions of the previous subsection since, in all scenarios $(m_1, m_2) \in \{(1,2),(10,2),(1,10),(10,10)\}$, *FullMod* performs better than *CauchyMod* in terms of predictive capability: its average log predictive density is higher than in *CauchyMod*, with an advantage which is more significant when m_2 is mispecified.

The fitting properties can be analysed through a comparison of the marginal likelihoods under the two models. From Chapter 9 of Müller et al. [2015], the only calculation of the marginal likelihood that is explicitly designed for DP mixture models is the one by Basu and Chib [2003]. Application of the Basu and Chib [2003] approach to our multiple DP set-up, however, requires a non-trivial extension of their approach for finding the likelihood ordinate, an extension which is beyond the scope of this paper.

5 Interest Rate Analysis

Short-term riskless rates p_t can be modelled according to the specification of Chan et al. [1992]:

$$dp_t = (\lambda + \beta p_{t-1})dt + \sigma p_{t-1}^x dW_t, \tag{7}$$

where dW_t is a Brownian motion. Different choices of x correspond to different well-known models, for instance the Vasicek [1977] model for $x = 0$ and the CIR model of Cox et al. [1985] for $x = 0.5$. Defining $y_t := p_t - p_{t-1}$ and $\gamma = (\lambda, \beta)'$, the discretized version of (7) can be incorporated in our proposed class of models, for $i = 1, 2$ and $t = 1, \ldots, T$, as follows:

$$\begin{aligned}
y_t|\xi_t &\sim N(y_t|\xi_{1t} + \xi_{2t}p_{t-1}, \sqrt{\xi_{3t}}p_{t-1}^x) \\
(\xi_{1t}, \xi_{2t}) &= \begin{cases} \theta_{11} & \text{if } t < \tau_{11} \\ \theta_{12} & \text{if } \tau_{11} \le t < \tau_{12} \\ \vdots & \\ \theta_{m_1+1} & \text{if } \tau_{m_1} \le t \le T \end{cases}, \quad \xi_{3t} = \begin{cases} \theta_{21} & \text{if } t < \tau_{21} \\ \theta_{22} & \text{if } \tau_{21} \le t < \tau_{22} \\ \vdots & \\ \theta_{m_2+1} & \text{if } \tau_{m_2} \le t \le T \end{cases} \\
\theta_{ij}|F_i &\sim F_i, \quad j = 1, \ldots, m_i + 1 \\
F_1 &\sim DP(M_1 \cdot N(\cdot|\mu_\gamma, V_\gamma)), \quad F_2 \sim DP(M_2 \cdot IG(\cdot|\alpha_\sigma, \beta_\sigma)) \\
(\tau_{i1}, \ldots, \tau_{im_i}) &\sim P\left(\left\{w_{j,t}^i\right\}\right).
\end{aligned}$$

The Gibbs sampler iterates between sampling from the full conditionals of S_{1T} and S_{2T} (latent regimes for the mean parameters θ_1 and for the variance parameter θ_2) and from the full conditionals of θ_1 and θ_2. Given θ_1 and θ_2, the processes S_{1T} and S_{2T} are sampled as in Section 3.1, fixing a maximum of $m_1 = 10$ and $m_2 = 10$ change-points. The clustering configuration and the distinct values are sampled following Section 3.2, for θ_1 and θ_2 separately and conditionally on each other. In sampling θ_1 and θ_2, the integrals involved in Equation (5) can be computed in closed form.

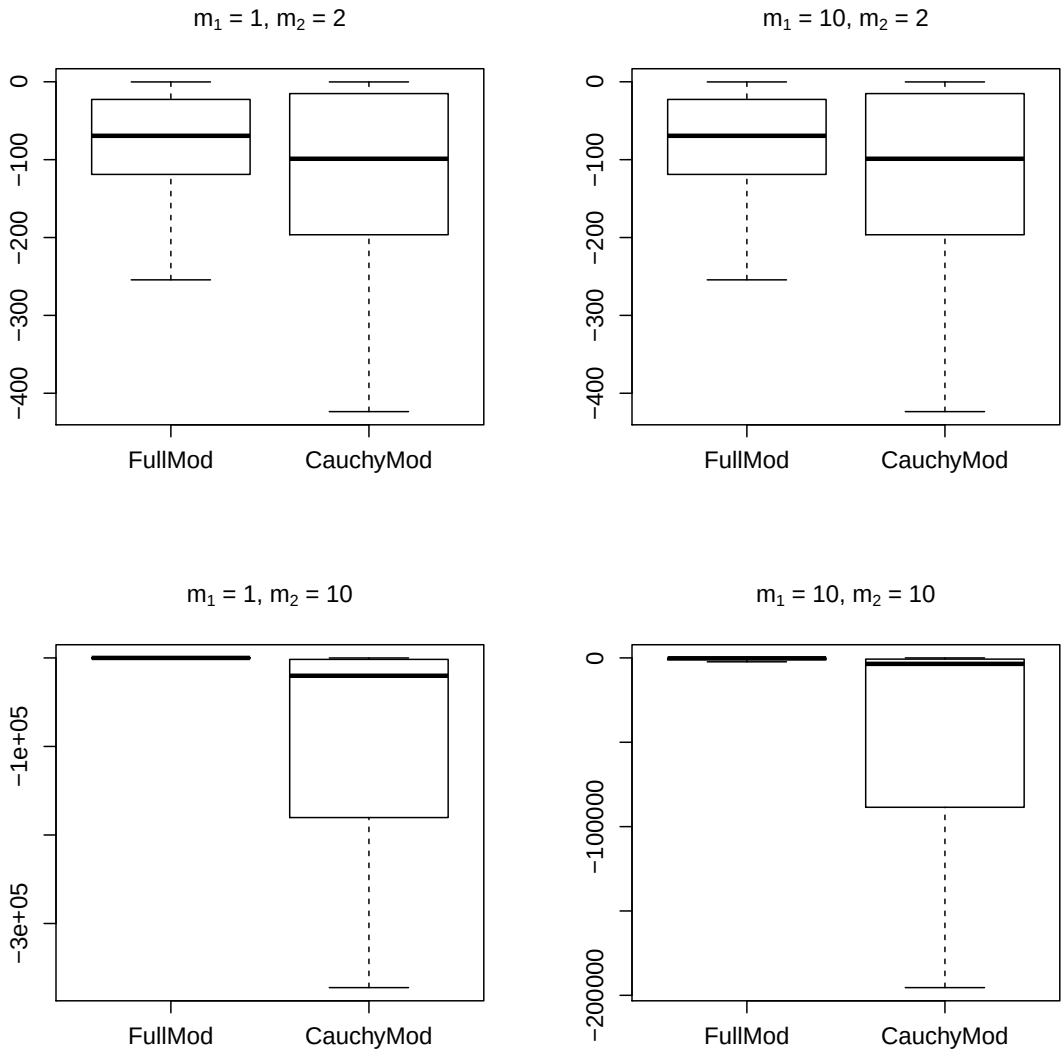

Figure 4: Autocorrelated simulation study. Log predictive densities for models *FullMod* and *CauchyMod*, in the autocorrelated simulation study, with $(m_1, m_2) \in (1, 2), (10, 2), (1, 10), (10, 10)$, over 40 simulated datasets, for true $m_1 = 1$ and $m_2 = 2$.

Regime j	$\hat{\tau}_{1j}$	$\hat{\tau}_{2j}$	$\hat{\lambda} \cdot 10^{-3}$	$\hat{\beta}$	$\hat{\sigma^2} \cdot 10^{-4}$
1	59	306	-2.18033	0.11823	0.14802
	(58, 61)	(304,306)	-(2.3175,2.0432)	(0.1162,0.1202)	(0.1259,0.1738)
2	165	385	-1.21484	0.11807	5.53763
	(163, 167)	(384,390)	-(1.2950,1.1352)	(0.1161,0.1200)	(4.0329,7.5746)
3	216	-	-3.00601	0.11796	0.39945
	(212, 226)	-	-(3.2695,2.7728)	(0.1160,0.1199)	(0.3297,0.4846)
4	338	-	-5.37077	0.11785	-
	(336, 338)	-	-(5.6364,5.1476)	(0.1159,0.1198)	-
5	485	-	-0.14755	0.11813	-
	(417, 509)	-	-(0.2675,0.0915)	(0.1161,0.1201)	-
6	-	-	-0.02664	0.11821	-
	-	-	-(0.0545,0.0004)	(0.1163,0.1202)	-

Table 3: Interest rate posterior estimates. Posterior 95% credible intervals in parenthesis.

The methodology is applied to secondary market 3-month T-bills, publicly available from the Board of Governors of the Federal Reserve System. The data are weekly, Friday to Friday, from Sep-12-2003 to Jun-03-2015 ($T = 600$ observations). The Gibbs sampler is run for $N = 20000$ iterations, and the initial $N/2$ iterations are discarded as burn-in, with the chain initialized to the MLE estimates obtained by regressing the observed returns on their lagged values. The probabilities of changing the regime are as in the simulated examples, $x = 0.5$ (CIR model), and the remaining hyperparameters are fixed as in the autocorrelated simulation studies. A priori the concentration parameters of DPs are $M_i \sim Gamma(0.5, 0.0001)$ for $i = 1, 2$, centered on a high noninformative values, and are estimated following the method proposed in Escobar and West [1995].

Posterior estimates are reported in Table 3, with 95% credible intervals in parenthesis. The algorithm estimates 5 change-points for the location vector parameter, and 2 change-points for the variance. We report in Figure 5 the marginal posterior probability mass functions of τ_{11} and τ_{21}, the first change-point for both structural parameters.

The estimated conditional mean $\hat{\lambda} + p_{t-1}\hat{\beta}$ and conditional variance $p_{t-1}^{2x}\hat{\sigma^2}$ are shown in Figure 6 (top right), whilst the bottom graphs of the same figure evidence the regime changes through the posterior distribution of the latent processes S_{1T} and S_{2T}. From Table 3 the change-points show narrow credible intervals particularly for σ^2, for which the jumps are more visible. More volatile is the estimate of τ_{15}, coherently with the top right plot in Figure 6, which shows no visible change in the conditional mean around τ_{15}. Regime changes for (λ, β) are mainly driven by changes in λ, whilst credible intervals of βs are highly overlapped. This is also suggested from the behaviour of the estimated conditional mean, which evidences clear vertical jumps also in periods with no big price variations. The dynamics of the estimated variance in Figure 6 (top right) shows a feature that is not apparent from Table 3: whilst the change-points τ_{21} and τ_{22} could also be dirtily

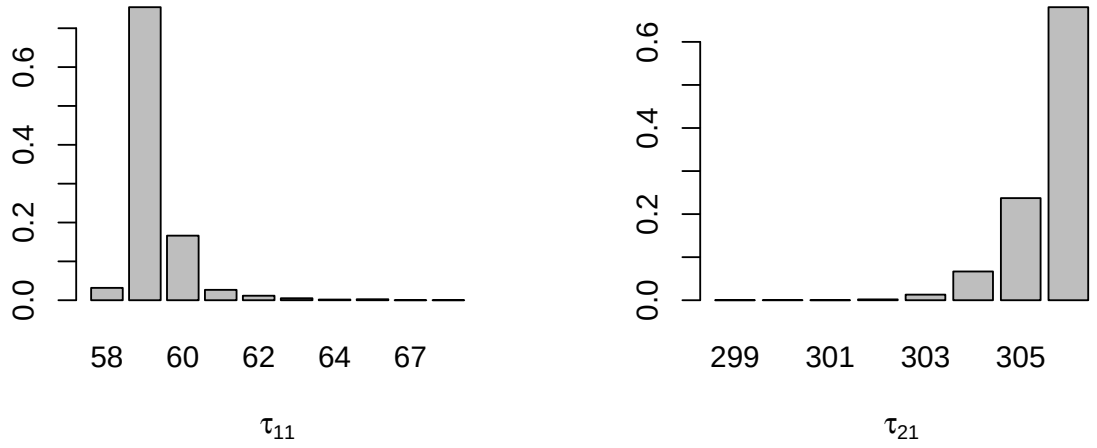

Figure 5: Marginal posterior probability mass function for the first change-point in the location structural parameter (left) and for the first change-point in the variance structural parameter (right), for the interest rate returns.

guessed by looking at the raw data, more latent is the asymmetric behaviour of the conditional variance around the two change-points, with the abrupt increase at τ_{21}, followed by the steep decrease towards τ_{22}. Finally, we compare the predictive performances of *FullMod* and *CauchyMod* in Figure 7, by comparing their predictive densities, estimated as in Section 4.3. It is clear that *FullMod* shows higher predictive density and therefore better prediction capability.

6 Conclusions and directions of investigation

We have introduced a new Bayesian semiparametric change-point model with various attractive features, to robustly analyse time series with unknown locations and number of multivariate change-points. A Dirichlet Process (DP) prior on each structural parameter (or group of structural parameters) provides robustness to mispecification in the conditional distributions of the structural parameters. Furthermore, the discrete random distribution assigns positive mass to previous realizations of structural parameters, in agreement with recurrent regimes. In simulation we evidence a relevant improvement in the inferential results, relative to the the model with no DP. Through a forward filtering backward sampling algorithm on the latent regimes, we introduce different change-point processes for different (or different groups of) structural parameters, and we are able to conduct the change-point analysis without specifying a priori the number of change-points. We

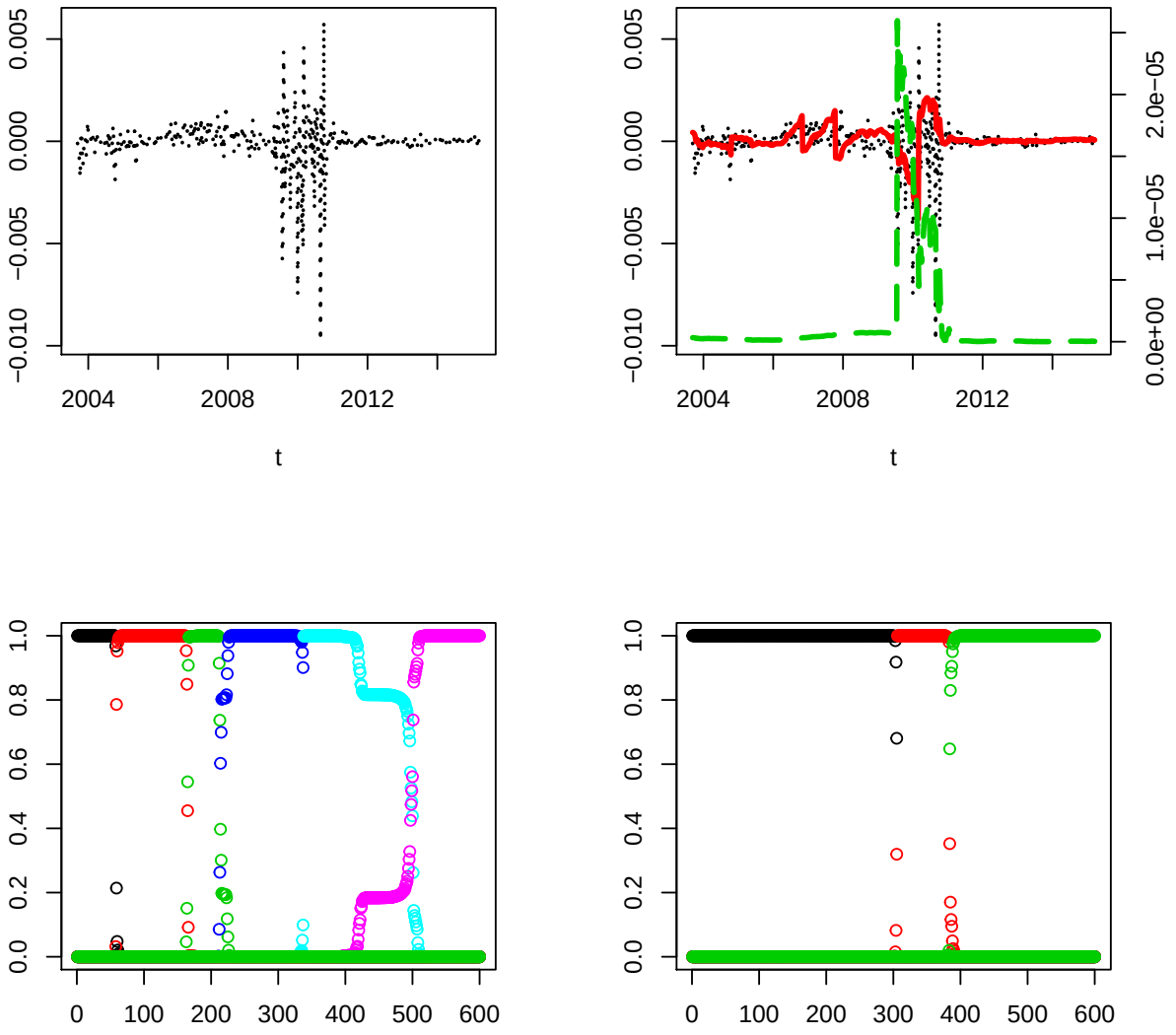

Figure 6: Upper-left: short-term riskless data. Upper-right: estimated conditional mean (red line) and conditional variance (green dashed) on interest rate returns (black dots). The left y-axis is for the conditional mean and for the returns, whilst the right y-axis is for the conditional variance. Bottom-left: posterior probabilities of $s_{1t} = k_1$, for $k_1 = 1, \ldots, 10$ and $t = 1, \ldots, 600$. Bottom-right: posterior probabilities of $s_{2t} = k_2$, for $k_2 = 1, \ldots, 10$ and $t = 1, \ldots, 600$.

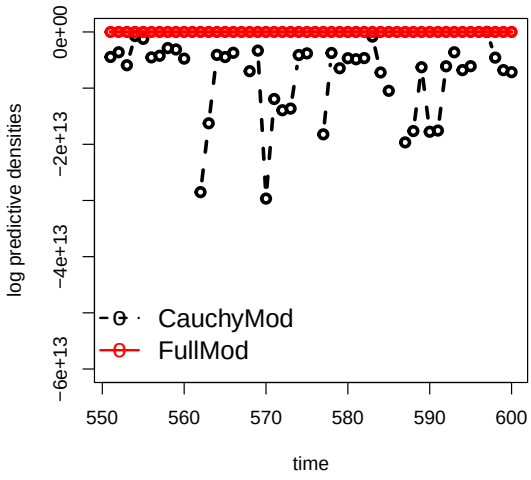

Figure 7: Short-term interest rate analysis. Log predictive densities for models *FullMod* (solid red) and *CauchyMod* (dashed black), with $(m_1, m_2) = (10, 10)$.

highlight the estimation advantage of a model that treats separately the change-points of different structural parameters, relative to a model with common change-points. Finally we introduce, to our knowledge for the first time in the literature, more realistic time-dependent transition matrices, and again demonstrate, in simulation, how turning off this feature has significant impact on change-points estimation. After deriving some distributional results on interarrival times between regime changes and on the number of change-points, we present an efficient MCMC algorithm for posterior inference, implemented on simulated data and on short-term interest rates. The results show that the proposed semiparametric model and an alternative parametric model with heavy-tailed prior on the variance perform better than alternative less realistic models in terms of estimation of the number and locations of the change-points and for the estimation of the structural parameters, when the data are iid Gaussian or heavy-tailed. When the generated data are skewed or autocorrelated, the proposed semiparametric model performs better than the benchmarks.

A limit of the proposed model is the assumed prior independence between the DPs. Still, a posteriori a dependence between the DPs is implied, since each DP is, a posteriori, centered on a base measure that depends on the structural parameters extracted from the other DPs. It would be interesting to extend the current setting to dependent DPs and test the gain in terms of inferential performance that would derive, for instance, through the mixture process of DPs [Cifarelli and Regazzini, 1978, Peluso et al., 2017], hierarchial DP [Beal et al., 2002, Teh et al., 2005] or the bivariate DP [Walker and Muliere, 2003]. Also, the DP induces a number of clusters that grows at an exponential rate, and having sizes with exponential tail behaviour. This clustering structure

can represent a limit in some applications, for instance in all those cases with cluster sizes decaying in power-law. An extension of the proposed model to more general Bayesian nonparametric priors [Lijoi and Prünster, 2010] can introduce a more flexible clustering of the structural parameters.

Acknowledgement

We thank Pietro Muliere, John Maheu and Qiao Yang for insightful discussions. Stefano Peluso acknowledges support from the Swiss National Science Foundation, grant P1TIP1_155031.

Appendix

Proof of Proposition 2.1
For the first interarrival time is immediate to see that

$$p(\tau_1 - \tau_0 = k_1) = (1 - w_{1,k_1}) \prod_{t=1}^{k_1} w_{1,t},$$

extended to the joint probability as

$$p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_r - \tau_{r-1} = k_r) = \prod_{j=1}^{r} \left\{ (1 - w_{j,k'_j}) \prod_{t=k'_{j-1}+1}^{k'_j - 1} w_{j,t} \right\}.$$

Now the result follows from the ratio

$$\frac{p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_r - \tau_{r-1} = k_r)}{p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_{r-1} - \tau_{r-2} = k_{r-1})}.$$

Proof of Corollary 2.2
Define the sets $\Omega_r := \{k'_1, \ldots, k'_r\}$ and $\Omega_r^c := \{1, 2, \ldots k'_r\} - \Omega_r$. For the first change-point we can write

$$\begin{aligned}
p(\tau_1 - \tau_0 = k_1) &= \int p(\tau_1 - \tau_0 = k_1 | \{w_{j,t} \text{ for all } j,t\}) d(w_{1,1}, \ldots, w_{1,k_1}) \\
&= \frac{1}{\prod_{t=1}^{k_1} B(\alpha_t, \beta_t)} \int_{[0,1]^{k_1}} (1 - w_{1,k_1}) \prod_{t=1}^{k_1 - 1} w_{1,t} \cdot \prod_{t=1}^{k_1} w_{1,t}^{\alpha - 1} (1 - w_{1,t})^{\beta - 1} d(w_{1,1}, \ldots, w_{1,k_1}) \\
&= \left(1 - \frac{\alpha_{k_1}}{\alpha_{k_1} + \beta_{k_1}}\right) \prod_{t=1}^{k_1} \frac{\alpha_{k_1}}{\alpha_{k_1} + \beta_{k_1}},
\end{aligned}$$

and, similarly, with multiple change-points,

$$p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_r - \tau_{r-1} = k_r) = \prod_{t \in \Omega_r} \left(1 - \frac{\alpha_t}{\alpha_t + \beta_t}\right) \cdot \prod_{t \in \Omega_r^c} \frac{\alpha_t}{\alpha_t + \beta_t}$$

As in the proof of Proposition 2.1, the result now follows from the ratio

$$\frac{p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_r - \tau_{r-1} = k_r)}{p(\tau_1 - \tau_0 = k_1, \tau_2 - \tau_1 = k_2, \ldots, \tau_{r-1} - \tau_{r-2} = k_{r-1})}.$$

Proof of Proposition 2.3

In the proposed model the number of change-points up to time t, conditionally on the probabilities of changing regime, has distribution

$$N_t|W = \begin{cases} 0, & \prod_{i=1}^{t-1} w_{1i} \\ 1, & \sum_{i=1}^{t-1} \left\{ (1 - w_{1i}) \prod_{j<i} w_{1j} \cdot \prod_{j>i} w_{2j} \right\} \\ \vdots & \\ x, & \sum_{k_1=1}^{t-1} \sum_{k_2=k_1+1}^{t-1} \cdots \sum_{k_x=k_{x-1}+1}^{t-1} \left\{ \prod_{j=1}^{x} \left[(1 - w_{j,k_j}) \prod_{k \in (k_{j-1}, k_j)} w_{jk} \right] \cdot \prod_{j=k_x+1}^{t-1} w_{x+1,j} \right\} \\ \vdots & \\ t-1, & \prod_{i=1}^{t-1} (1 - w_{ii}) \end{cases}.$$

With $w_{it} \sim Beta(\alpha_t, \beta_t)$ it is possible to write the result after the marginalization of w_{it}, as in the proof of Corollary 2.2.

References

J. Albert and S. Chib. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics*, 11:1–15, 1993.

A. Azzalini. *The skew-normal and related families*, volume 3. Cambridge University Press, 2013.

D. Barry and J. Hartigan. Product partition models for change point problems. *The Annals of Statistics*, 20:260–279, 1992.

D. Barry and J. Hartigan. A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88:309–319, 1993.

S. Basu and S. Chib. Marginal likelihood and bayes factors for dirichlet process mixture models. *Journal of the American Statistical Association*, 98(461):224–235, 2003.

L. Bauwens, J.-F. Carpantier, and A. Dufays. Autoregressive moving average infinite hidden markov-switching models. *Journal of Business and Economic Statistics*, 2015.

M.J. Beal, Z. Ghahramani, and C.E. Rasmussen. The infinite hidden markov model. In *Advances in neural information processing systems*, pages 577–584, 2002.

B. E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change-point problems*. Springer, 1993.

B. P. Carlin and S. Chib. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B*, 57:473–484, 1995.

P. B. Carlin, A. E. Gelfand, and A. F. M. Smith. Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society, Series C*, 41:389–405, 1992.

K. Chan, G. Karolyi, F. Longstaff, and A. Sanders. An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, 47:1209–1227, 1992.

J. Chen and A. K. Gupta. *Parametric statistical change point analysis: with applications to genetics, medicine, and finance*. Birkhauser, 2nd edition, 2011.

H. Chernoff and S. Zacks. Estimating the current mean of a normal distribution which is subjected to changes in time. *Annals of Mathematical Statistics*, 35:999–1018, 1964.

S. Chib. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75:79–97, 1996.

S. Chib. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86:221–241, 1998.

D. Cifarelli and E. Regazzini. Problemi statistici non parametrici in condizioni di scambiabilita parziale e impiego di medie associative. Technical report, Tech. rep., Quaderni Istituto Matematica Finanziaria dell?Universita di Torino, 1978.

J. Cox, J. Ingersoll, and R. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

A. Dufays. Infinite-State Markov-Switching for Dynamic Volatility. *Journal of Financial Econometrics*, 14: 418–460, 2016.

Y. Eo. Bayesian inference about the types of structural breaks when there are different breaks in many parameters. 2012.

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

J. Geweke and Y. Jiang. Inference and prediction in a multiple-structural-break model. *Journal of Econometrics*, 163:172–185, 2011.

P. Giordani and R. Kohn. Efficient Bayesian inference for multiple change-point and mixture innovation models. *Journal of Business and Economic Statistics*, 26:66–77, 2008.

Peter J Green and Antonietta Mira. Delayed rejection in reversible jump metropolis–hastings. *Biometrika*, 88(4):1035–1053, 2001.

P.J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

S.I.M. Ko, T.T.L. Chong, and P. Ghosh. Dirichlet process hidden Markov multiple change-point Model. *Bayesian Analysis*, 10:275–296, 2015.

A. Lijoi and I. Prünster. Models beyond the dirichlet process. *Bayesian nonparametrics*, 28:80, 2010.

A.Y. Lo. On a class of Bayesian nonparametric estimates. *Annals of Statistics*, 12:351–357, 1984.

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

R. H. Loschi, F. R. B. Cruz, P. L. Iglesias, and R. B. Arellano-Valle. A Gibbs sampling scheme to the product partition model: an application to change point problems. *Computers and Operations Research*, 30:463–482, 2003.

J. M. Maheu and S. Gordon. Learning, forecasting and structural breaks. *Journal of Applied Econometrics*, 23:553–583, 2008.

J.M. Maheu and Q. Yang. An infinite hidden Markov model for short-term interest rates. 2015.

A. F. Martinez and R. H. Mena. On a nonparametric change point detection model in Markovian regimes. *Bayesian Analysis*, 9:823–858, 2014.

A. Mira and S. Petrone. Bayesian hierarchical nonparametric inference for changepoint problems. In *Bernardo, J., Berger, J. O., Dawid, A. P., and Smith, A. F. M. (eds.) Bayesian Statistics 5*, pages 693–703, Oxford University Press, 1996.

P. Muliere and M. Scarsini. Changepoint problems: a Bayesian nonparametric approach. *Aplikace matematiky*, 30:397–402, 1985.

P. Müller, F.A. Quintana, A. Jara, and T. Hanson. *Bayesian nonparametric data analysis*. Springer, 2015.

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

T. Park, R. T. Krafty, and A.I. Sanchez. Bayesian semi-parametric analysis of Poisson change-point regression models: application to policy making in Cali, Colombia. *Journal of Applied Statistics*, 39:2285–2298, 2012.

S. Peluso, A. Mira, and P. Muliere. Robust identification of highly persistent interest rate regimes. *International Journal of Approximate Reasoning*, 83:102–117, 2017.

M. H. Pesaran, D. Pettenuzzo, and A. Timmermann. Forecasting time series subject to multiple structural breaks. *Review of Economic Studies*, 73:1057–1084, 2006.

N.G. Polson and J.G. Scott. On the half-cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4): 887–902, 2012.

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

A. F. M. Smith. A Bayesian approach to inference about a changepoint in a sequence of random variables. *Biometrika*, 62:407–416, 1975.

Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. In *Advances in neural information processing systems*, pages 1385–1392, 2005.

Y.W. Teh, M.I. Jordan, M.J. Beal, and D.M. Blei. Hierarchical dirichlet processes. *Journal of the American Statistical Association*, 101:1566–1581, 2006.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5: 177–188, 1977.

S. Walker and P. Muliere. A bivariate dirichlet process. *Statistics & probability letters*, 64(1):1–7, 2003.