Ca' Foscari University of Venice

Department of Economics

Working Paper

Federico Bassetti, Roberto Casarin, and Fabrizio Leisen

Beta-Product Dependent Pitman-Yor Processes for Bayesian Inference

ISSN: 1827-3580
No. 13/WP/2013

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 13/WP/2013
ISSN 1827-3580

Beta-Product Dependent Pitman-Yor Processes for Bayesian Inference

Federico Bassetti
University of Pavia

Roberto Casarin
University Ca' Foscari of Venice

Fabrizio Leisen
Universidad Carlos III de Madrid

11 June 2013

Abstract
Multiple time series data may exhibit clustering over time and the clustering effect may change across different series. This paper is motivated by the Bayesian non–parametric modelling of the dependence between clustering effects in multiple time series analysis. We follow a Dirichlet process mixture approach and define a new class of multivariate dependent Pitman-Yor processes (DPY). The proposed DPY are represented in terms of a vector of stick-breaking processes which determines dependent clustering structures in the time series. We follow a hierarchical specification of the DPY base measure to accounts for various degrees of information pooling across the series. We discuss some theoretical properties of the DPY and use them to define Bayesian non–parametric repeated measurement and vector autoregressive models. We provide efficient Monte Carlo Markov Chain algorithms for posterior computation of the proposed models and illustrate the effectiveness of the method with a simulation study and an application to the United States and the European Union business cycles.

Keywords
Bayesian non–parametrics; Dirichlet process; Panel Time-series non–parametrics; Pitman-Yor process; Stick-breaking process; Vector autoregressive process; Repeated measurements non-parametrics.

JEL Codes
C11, C14, C32

Address for correspondence:
Roberto Casarin
Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta S.Giobbe
30121 Venezia - Italy
Phone: (++39) 041 2349149
Fax: (++39) 041 2349210
e-mail: r.casarin@unive.it

BETA-PRODUCT DEPENDENT PITMAN-YOR PROCESSES FOR BAYESIAN INFERENCE

FEDERICO BASSETTI, ROBERTO CASARIN, AND FABRIZIO LEISEN

Abstract. Multiple time series data may exhibit clustering over time and the clustering effect may change across different series. This paper is motivated by the Bayesian non–parametric modelling of the dependence between clustering effects in multiple time series analysis. We follow a Dirichlet process mixture approach and define a new class of multivariate dependent Pitman-Yor processes (DPY). The proposed DPY are represented in terms of a vector of stick-breaking processes which determines dependent clustering structures in the time series. We follow a hierarchical specification of the DPY base measure to accounts for various degrees of information pooling across the series. We discuss some theoretical properties of the DPY and use them to define Bayesian non–parametric repeated measurement and vector autoregressive models. We provide efficient Monte Carlo Markov Chain algorithms for posterior computation of the proposed models and illustrate the effectiveness of the method with a simulation study and an application to the United States and the European Union business cycles.

JEL: C11,C14,C32

Keywords: Bayesian non–parametrics, Dirichlet process, Panel Time-series non–parametrics, Pitman-Yor process, Stick-breaking process, Vector autoregressive process, Repeated measurements non–parametrics.

1. Introduction

This paper focuses on vectors of dependent Pitman-Yor (PY) processes for Bayesian inference on repeated measurements and panel data models. The PY process of parameters $\alpha > 0, l \in [0, 1)$ and base measure H_0, in short $PY(\alpha, l, H_0)$, is defined by

$$G = \sum_{k \geq 1} W_k \, \delta_{\tilde{\vartheta}_k}, \tag{1}$$

where δ_x is a point mass at x, $(\tilde{\vartheta}_k)_k$ is a sequence of independent random variables (atoms) with common distribution H_0, and the weights W_k's are defined by the stick-breaking construction:

$$W_k = S_k \prod_{j<k} (1 - S_j), \tag{2}$$

S_k being independent random variables with beta distribution of parameters $(1 - l, \alpha + lk)$. Such a process has been introduced in Pitman and Yor [1997] and it can be seen as a generalization of the (one parameter) Dirichlet process (DP) defined in Ferguson [1973]. If $l = 0$ the Pitman-Yor process corresponds to a DP and (1) reduces to the well-known Sethuraman [1994] representation.

The Dirichlet process (and the PY process as well) is usually employed as a prior for a mixing distribution, resulting in the so-called DP mixture model (see for example Lo [1984]). More specifically, one defines a random density

$$f(y) = \int \mathcal{K}(y|\vartheta) G(d\vartheta) = \sum_{k \geq 1} W_k \mathcal{K}(y|\tilde{\vartheta}_k), \tag{3}$$

where $\mathcal{K}$ is a suitable density kernel. Due to the availability of simple and efficient methods for posterior computation, starting from Escobar [1994] and Escobar and West [1995], DP mixture models are now routinely implemented and used in many fields. A recent account of Bayesian non-parametric inference can be found in Hjort et al. [2010].

Date: November 29, 2012.

Although DP mixture models are extremely flexible, in many real applications data arise under different conditions and hence assuming a DP mixture model can be too restrictive. For example, using covariates, data may be divided into different units. In this case, one would like to consider different densities for different units instead of a single common density for all the units. For this reason, mixtures driven by vectors of random probability measures could be used as an alternative to the DP mixture model. After the seminal papers of MacEachern [1999, 2001], the problem of modelling a finite number of dependent densities, allowing information pooling across units, has become an active area of research. For the interested reader, a brief account of the state of the art is given at the beginning of Section 2.

The first contribution of this paper is to introduce a new class of dependent Pitman-Yor processes. We start from the general definition of dependent stick-breaking processes of MacEachern [1999] and use the multivariate beta distributions of Nadarajah and Kotz [2005] for the stick-breaking weights to define the beta-product dependent Pitman-Yor (DPY) process. Another contribution is the derivation of a simple and efficient method for posterior computation that constitutes a novel extension of the slice sampling algorithm for DP mixture models introduced in Walker [2007] and Kalli et al. [2011]. Recently, Hatjispyrosa et al. [2011] introduce a different vector of dependent DP mixture and propose a suitable multidimensional slice sampling algorithm, which does not apply to our model.

Another contribution of the paper is the definition of two Bayesian non–parametric models for groups of time series, based on our beta-product DPY processes. The first model is an infinite mixture Gaussian model for repeated measurements which accounts for different clustering structures in the series and for dependence between the series-specific clustering effects. The model is a novel extension of the finite mixture models for repeated measurements which are present in the literature (e.g., see Früwirth-Schnatter [2006] for a review).

Finally, we contribute to the literature on Bayesian vector autoregressive models (VAR) (e.g., see Sims [1980, 1992] and Sims and Zha [1998]) and introduce a Bayesian non–parametric VAR model. In time series analysis DP and dependent DP have been employed in different ways. Rodriguez and ter Horst [2008] used a dependent DP to define an infinite mixture of time series models. Taddy and Kottas [2009] propose a Markov-switching finite mixture of independent Dirichlet process mixtures. Jensen and Maheu [2010] consider Dirichlet process mixture of stochastic volatility models. Griffin [2011] proposed a continuous-time non–parametric model for volatility. Griffin and Steel [2011] propose a time-varying stick-breaking process to capture time-variations in the clustering structure of a set of time series. In this paper, we focus on Bayesian VAR models for time series which are collected from different countries, i.e. a multi-country panel VAR (e.g., see Chib and Greenberg [1995] and Canova and Ciccarelli [2004]) and proposed a non–parametric panel VAR models based on our DPY process prior. Our model accounts for shifts in the intercept, autoregressive coefficients and covariance matrix of the country-specific equations and for cross-country dependence through a hierarchical specification of the base measure. We apply the model to the analysis of two well studied business cycles of the international economic system: the United States (US) and the European Union (EU) cycles.

The structure of the paper is as follows. Section 2 introduces vectors of dependent stick-breaking processes for prior modelling in repeated measurements and vector autoregressive models. Section 3 introduces beta-product dependent Pitman-Yor processes and studies their properties. Section 4 discusses the hierarchical specification of the the base measure. Section 5 proposes a Monte Carlo Markov Chain (MCMC) algorithm for approximated inference for DPY mixture models. Section 6 provides some applications to both simulated and real data. Specifically, we consider a joint analysis of the United States and the European Union business cycles. Section 7 concludes the paper.

2. Dependent stick-breaking processes and infinite mixture models

Some of the first developments of vectors of dependent random distributions are in Cifarelli and Regazzini [1978]. More recently, MacEachern [1999, 2001] introduce the so-called dependent DP, incorporating dependence on covariates through both the atoms and the weights, although

these papers mainly consider dependent atoms. In particular, the random variables $\tilde{\vartheta}_k$'s in the Sethuraman's representation (1)-(3) are replaced with stochastic processes $\tilde{\vartheta}_{zk}$, z being a set of covariates. Following this line, De Iorio et al. [2004] propose an ANOVA-type dependence for the law of the atoms, while Gelfand et al. [2004] introduce a spatial dependence structure. Later, Griffin and Steel [2006] defined a class of DP with both dependent atoms and weights. Many alternative constructions that incorporate dependence in the weights have been proposed, see, for instance Duan et al. [2007], Chung and Dunson [2011], Dunson and Peddada [2008], Dunson et al. [2008], Rodriguez et al. [2010].

Other approaches to the definition of dependent vectors of random measures rely upon suitable convex combinations of independent DPs (e.g., Müller et al. [2004], Pennell and Dunson [2006], Hatjispyrosa et al. [2011], Kolossiatis et al. [2011]), hierarchical structures of stick-breakings (e.g., Teh et al. [2006]), normalization of dependent completely random measures (e.g., Ishwaran and Zarepour [2009], Epifani and Lijoi [2010], Leisen and Lijoi [2011]) or suitable multivariate extensions of the Polya tree prior (e.g., Trippa et al. [2011]).

2.1. **Vectors of stick-breaking processes.** Following the general definition of dependent stick-breaking processes, proposed in MacEachern [1999, 2001], we let

$$G_i(\cdot) = \sum_{k\geq 1} W_{ik}\ \delta_{\tilde{\vartheta}_{ik}}(\cdot) \qquad i = 1, \ldots, r, \tag{4}$$

where the weights $W_k = (W_{1k}, \ldots, W_{rk})$ and the atoms $\tilde{\vartheta}_k = (\tilde{\vartheta}_{1k}, \ldots, \tilde{\vartheta}_{rk})$ satisfy the following hypotheses:

- $(\tilde{\vartheta}_k)_k$ and $(W_k)_k$ are stochastically independent;
- $(\tilde{\vartheta}_k)_k$ is an i.i.d. sequence of random elements taking values in a product space Θ^r with common probability distribution G_0;
- the weigths W_{ik}s are determined via the stick-breaking construction

$$W_{ik} = S_{ik} \prod_{j<k} (1 - S_{ij}) \quad i = 1, \ldots, r,$$

with $\prod_{j<1}(1 - S_{ij}) = 1$, where $S_k = (S_{1k}, \ldots, S_{rk})$ are stochastically independent random vectors taking values in $[0, 1]^r$ such that $\sum_{k\geq 1} W_{ik} = 1$ a.s. for every i.

2.2. **Infinite mixture models.** Modelling based on mixture distributions is nowadays applied in many areas, especially in economics and finance (see Früwirth-Schnatter [2006] for a review). Stick-breaking processes can be used to extend finite mixture models to infinite mixture models, which can capture many specific properties of time series, such as multimodality, skewness, excess of kurtosis and presence of outliers (e.g., see Griffin [2011] and Griffin and Steel [2011]). Our interest in dependent stick-breaking processes is related to their use in building dependent infinite mixtures models which allows for modelling information pooling across different time series. More specifically, we develop non-parametric repeated measurements and panel (or longitudinal) dynamic models. Both models are widely used in economics for simultaneous inference on a set of parameters for similar units, e.g. firms or countries, when data are collected on several occasions or observed on a regular basis for several periods, e.g. years.

2.2.1. *Repeated measurements.* Consider a set of samples, taking values in an observations space $\mathbb{Y}$, divided in r sub–samples (groups of observations or units), that is:

$$Y_{it} \qquad i = 1, \ldots r, \quad t = 1, \ldots, T_i$$

where Y_{it} is the t-th observation within unit i. In repeated measurements models (see Crowder and Hand [1990], Davidian and Giltinan [1998]), i may correspond to a space label or predictor, and the observations of a unit i are independent with the same density f_i. In assessing a prior for $(f_1, f_2, \ldots, f_r)$, one aims at borrowing information across units and at assuming for each unit

an infinite mixture model. To do this, one can first introduce a density kernel $\mathcal{K} : \mathbb{Y} \times \Theta \to [0, 1]$, and then define

$$f_i(y) = \int \mathcal{K}(y|\vartheta) G_i(d\vartheta) \qquad i = 1, \dots, r, \tag{5}$$

where $(G_1, \dots, G_r)$ is a vector of dependent stick-breaking processes.

The dependence between the random probability measures affects the dependence structure underlying the densities $f_1, \dots, f_r$, which can be represented as infinite mixtures

$$f_i(y) = \sum_{k \geq 1} W_{ik} \; \mathcal{K}(y|\tilde{\vartheta}_{ik}) \qquad i = 1, \dots, r. \tag{6}$$

In this paper we focus on one of the most widely used kernel: the Gaussian kernel of parameter $\vartheta = (\mu, \sigma^2)$, i.e.

$$\mathcal{K}(y|\vartheta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big\{ \frac{1}{2\sigma^2}(y - \mu)^2 \Big\}. \tag{7}$$

Other kinds of kernels, such as binomial or multinomial, can also be applied within our framework for qualitative data modelling.

2.2.2. *Panel data models.* A Vector Autoregressive (VAR) structure allows for handling the dynamic properties of multiple time series and is now a standard tool for structural analysis and forecasting in macroeconomics (see Sims [1980, 1992]). Panel VAR models are used when different time series are collected from different units (e.g., countries). Such models can capture complex relationships between countries because they do not impose restrictions on the parameters, but this requires a large number of parameters and leads to over-parameterization. The resulting potential over-fitting problem calls for the use of a Bayesian approach to inference, which, through the use of the prior distributions, allows for prior constraints on the VAR parameters. See, e.g., Doan et al. [1984], Litterman [1986], Sims and Zha [1998] for Bayesian VAR, Chib and Greenberg [1995] for Bayesian Seemingly Unrelated Regression and Canova and Ciccarelli [2004] for panel Bayesian VAR.

In panel VAR model of the order p (VAR(p)), the subset of equations associated with the i-th unit of the panel is:

$$Y_{it} = \mu + \sum_{j=1}^{r} \sum_{l=1}^{p} \Phi_{jl} Y_{j\, t-l} + \varepsilon_{it}, \tag{8}$$

for $i = 1, \dots, r$ and $t = 1, \dots, T$, where $Y_{it} = (Y_{i1,t}, \dots, Y_{im,t})'$, $\mu = (\mu_1, \dots, \mu_m)'$ and Φ_{jl}, $j = 1, \dots, r$, $l = 1, \dots, p$, is a sequence of m-dimensional square matrices. Finally, $\varepsilon_{it} = (\varepsilon_{i1,t}, \dots, \varepsilon_{im,t})'$ follows a Gaussian distribution $\mathcal{N}_m(0, \Sigma)$ with mean 0 and covariance matrix Σ. We assume that ε_{it} and ε_{jt} are independent for every $i \neq j$ and $t \neq s$. As it is common in Bayesian VAR, the dependence between units is modelled through the dependence between the random parameters of the different unit-specific subsets equations. In this section, for the sake of simplicity we dropped the dependence of the parameters μ, Φ_{jl} and Σ on the unit index i. We will discuss later on in the paper possible assumptions on the parameters of the different units and focus now on the infinite mixture representation of the unit-specific equations. First we write (8) in a stacked regression form

$$Y_{it} = (I_m \otimes X_t')\phi + \varepsilon_{it}, \tag{9}$$

where $X_t = (1, Y_{1\, t-1}', \dots, Y_{r\, t-1}', \dots, Y_{1\, t-p}', \dots, Y_{r\, t-p}')'$ is the vector of predetermined variables, $\phi = \mathrm{vec}(\Phi)$, $\Phi = (\mu, \Phi_{11}, \dots, \Phi_{r1}, \; \dots, \Phi_{r1}, \dots, \Phi_{rp})'$, $\otimes$ denotes the Kronecker product and vec the column-wise vectorization operator that stacks the column of a matrix in a column vector.

Then we assume a set of dependent infinite mixture priors G_i, $i = 1, \dots, r$, for the unknown parameters, ϕ and Σ, and obtain the following conditional distribution of Y_{it} (given X_t)

$$f_{it}(y) = \sum_{k \geq 1} W_{ik} \mathcal{K}_t(y|\tilde{\vartheta}_{ik}, X_t) \qquad i = 1, \dots, r \tag{10}$$

where

$$\mathcal{K}_t(y|\vartheta, X_t) = (2\pi)^{-m/2}|\Sigma|^{-1/2}\exp\left\{(y-(I_m\otimes X_t')\phi))'\Sigma^{-1}(y-(I_m\otimes X_t')\phi)\right\} \tag{11}$$

with parameter $\vartheta = (\phi, \Sigma)$. In this infinite mixture representation the sources of conditional dependence between two units, Y_{it} and Y_{jt}, are the random weights, $(W_{ik})_k$ and $(W_{jk})_k$, and the random atoms $(\tilde{\vartheta}_{ik})_k$ and $(\tilde{\vartheta}_{jk})_k$.

3. Beta-Product dependent Pitman–Yor Processes

In this section we propose a new class of dependent stick-breaking processes in such a way that each marginal random measure G_i is a Pitman–Yor process. This result follows from the Sethurman's representation (1) if one considers a multivariate distribution for $(S_{1k}, \ldots, S_{rk})$ such that $S_{ik} \sim Beta(1-l, \alpha_i + lk)$ for every i and k.

It is worth noticing that there are many possible definitions of multivariate beta distribution (e.g. Olkin and Liu [2003], Nadarajah and Kotz [2005], Nieto-Barajas and Walker [2007], Taddy [2010] and Trippa et al. [2011]), but not all of them have a tractable stochastic representation and lead to simple Bayesian inference procedures. For this reason we follow Nadarajah and Kotz [2005] and consider a suitable product of independent beta random variables. More specifically we apply the following result.

Proposition 1 (Rao [1949])**.** *If $U_1, U_2, \ldots, U_p$ are independent beta random variables with shape parameters (a_i, b_i), $i = 1, 2, \ldots, p$ and if $a_{i+1} = a_i + b_i$, $i = 1, 2, \ldots, p-1$, then the product $U_1U_2\cdots U_p$ is a beta random variable with parameters $(a_1, b_1 + \cdots + b_p)$.*

We propose two alternative specifications of the multidimensional beta variables. Specifically, if we set

$$(S_{1k}, S_{2k}, \ldots, S_{rk}) = (V_{0k}V_{1k}, V_{0k}V_{2k}, \ldots, V_{0k}V_{rk}) \tag{12}$$

with $V_{0k}, \ldots, V_{rk}$ independent, $V_{ik} \sim Beta(\alpha_{0k}+\alpha_{1k}, \alpha_{2k})$, $i = 1, 2, \ldots, r$, and $V_{0k} \sim Beta(\alpha_{0k}, \alpha_{1k})$, then $S_{ik} \sim Beta(\alpha_{0k}, \alpha_{1k} + \alpha_{2k})$.

As an alternative we consider

$$(S_{1k}, S_{2k}, \ldots, S_{rk}) = (V_{0k}V_{1k}\ldots V_{r-1k}, V_{0k}V_{1k}\ldots V_{r-2k}, \ldots, V_{0k}) \tag{13}$$

with $V_{0k}, \ldots, V_{r-1k}$ independent and $V_{ik} \sim Beta(\alpha_{0k} + \cdots + \alpha_{ik}, \alpha_{i+1k})$, $i = 0, \ldots, r-1$, that gives $S_{ik} \sim Beta(\alpha_{0k}, \alpha_{1k} + \cdots + \alpha_{r+1-i,k})$.

If (12) holds, thanks to Lemma 1 in Ishwaran and James [2001], the process $(G_1, \ldots, G_r)$ is well-defined, i.e. $\sum_{k\geq 1} W_{ik} = 1$ a.s. for every $i = 1, \ldots, r$, when $\sum_{k\geq 1}\log(1+\alpha_{0k}/(\alpha_{1k}+\alpha_{2k})) = +\infty$. Analogously, assuming (13), the process is well-defined when $\sum_{k\geq 1}\log(1+\alpha_{0k}/\alpha_{1k}) = +\infty$.

3.1. Pitman-Yor process marginal.

In this section we show that by a suitable choice of the parameters in (12)-(13) we obtain a vector of dependent random measures with Pitman-Yor marginals.

Consider (12) and set $\alpha_{0k} = 1-l$, $\alpha_{1k} = \alpha_1$ and $\alpha_{2k} = \alpha_2 + lk$, with $\alpha_1 > 0$, $\alpha_2 > 0$ and $0 \leq l < 1$. Since the random variables $V_{0k}, \ldots, V_{rk}$ are independent and

$$V_{0k} \sim Beta(1-l, \alpha_1), \quad V_{ik} \sim Beta(1-l+\alpha_1, \alpha_2 + lk) \quad i = 1, \ldots, r, \tag{14}$$

Proposition 1 yields that $S_{ik} \sim Beta(1-l, \alpha_1 + \alpha_2 + lk)$.

Alternatively, in (13) set $\alpha_{0k} = 1-l$, $\alpha_{1k} = \alpha_1 + lk$, and $\alpha_{ik} = \alpha_i$ for $i \geq 2$ with $\alpha_i > 0$ and $0 \leq l < 1$. Since, in this case,

$$\begin{aligned} &V_{0k} \sim Beta(1-l, \alpha_1 + lk), V_{1k} \sim Beta(1+\alpha_1 + l(k-1), \alpha_2), \ldots, \\ &\ldots V_{r-1k} \sim Beta(1+\alpha_1+\cdots+\alpha_{r-1} + l(k-1), \alpha_r), \end{aligned} \tag{15}$$

$S_{ik} \sim Beta(1-l, \alpha_1 + \cdots + \alpha_{r-i+1} + lk)$.

Summarizing, if G_{0i} denotes the distribution of $\tilde{\vartheta}_{ik}$, we have proved the following

Proposition 2. *If (12) and (14) hold true, then G_i is a $PY(\alpha_1+\alpha_2, l, G_{0i})$ for every $i = 1, \ldots, r$. If (13) and (15) hold true, then G_i is a $PY(\alpha_1 + \cdots + \alpha_{r-i+1}, l, G_{0i})$ for every $i = 1, \ldots, r$.*

It is worth noticing that the idea of using product of beta random variables for the weights of a dependent Dirichlet process is not new. Similar constructions are detailed and applied in Taddy [2010] and Griffin and Steel [2011]. For instance, the autoregressive beta stick-breaking process defined in Taddy [2010] consists in taking $S_{1k} \sim Beta(a, b)$ for all $k \geq 1$, while, for $j > 1$,

$$S_{jk} = 1 - U_{jk}(1 - R_{jk}S_{j-1,k}),$$

where U_{jk} are i.i.d. $Beta(b, a-\rho)$ and R_{jk} are i.i.d. $Beta(\rho, a-\rho)$. In this way S_{jk} turns out to be $Beta(a, b)$ for all j and k. So that, for $a = 1$ one gets Dirichlet process marginals. The advantage of this autoregressive structure for the weights is that it allows for the use of a simple particle filter algorithm for on-line inference (see Taddy [2010]). Here we choose the different specifications (14) and (15) essentially for two reasons. First, with (14) and (15) the resulting process has Pitman-Yor marginals, which can not be obtained with the autoregressive beta stick-breaking structure given above. Secondly the stochastic representations (14) and (15) allow the derivation of a slice sampling algorithm in the spirit of Walker [2007] and Kalli et al. [2011]. As a side remark note that our beta-product models are more parsimonious in terms of number of latent variables. For constructing an r-dimensional vector of weights Taddy [2010] uses $2r - 1$ latent variables while (14) and (15) require $r + 1$ and r respectively. This may be computationally useful for large r although the more complex structure of the autoregressive beta stick-breaking can probably model in a richer way the dependence between weights.

For the sake of simplicity in what follows we will mainly consider $r = 2$. According to the above construction schemes, the two alternative specifications of (S_{1k}, S_{2k}) are:

(H1) $(S_{1k}, S_{2k}) = (V_{0k}V_{1k}, V_{0k}V_{2k})$, *with* V_{0k}, V_{1k}, V_{2k} *independent,* $V_{0k} \sim Beta(1 - l, \alpha_1)$ *and* $V_{ik} \sim Beta(1 - l + \alpha_1, \alpha_2 + lk)$, $i = 1, 2$, *where* $\alpha_1 > 0$, $\alpha_2 > 0$ *and* $l \in [0, 1)$;

(H2) $(S_{1k}, S_{2k}) = (V_{0k}V_{1k}, V_{0k})$, *with* V_{0k}, V_{1k} *independent,* $V_{0k} \sim Beta(1 - l, \alpha_1 + lk)$ *and* $V_{1k} \sim Beta(1 + \alpha_1 + l(k - 1), \alpha_2)$ *with* $\alpha_1 > 0$, $\alpha_2 > 0$ *and* $l \in [0, 1)$.

Since with this construction the G_i's are Pitman-Yor processes, we call (G_1, G_2) Beta-Product Dependent Pitman-Yor Process, $\beta_i-\mathrm{DPY}(\psi, G_0)$ for short, of parameters $\psi = (\alpha_1, \alpha_2, l)$ and base measure G_0, where $i = 1$ for H1 and $i = 2$ for H2. If one takes $l = 0$, the resulting process has Dirichlet marginals, in this case we shall write $\beta_i-\mathrm{DD}(\psi, G_0)$.

It should be noted that the two processes have different marginal behaviors. The β_1–$\mathrm{DPY}(\psi, G_0)$ process has marginals with the same precision parameter and thus should be used as a prior when the clustering is expected to be similar along the different vector components. In the $\beta_2-\mathrm{DPY}(\psi, G_0)$ process, the precision parameter decreases along the vector dimension. This process should be used when a priori one suspects that the clustering features are different across units.

We conclude this section with few remarks on the limiting cases that can help in the interpretation of the parameters. Under H1, (S_{1k}, S_{2k}) converges (in distribution) to (V_{1k}, V_{2k}) as $\alpha_1 \to 0$, where V_{1k} and V_{2k} are independent random variables with distribution $Beta(1-l, \alpha_2+lk)$. While, under H2, (S_{11}, S_{21}) converges to (V_0, V_0) as $\alpha_2 \to 0$, where V_0 is a $Beta(1 - l, \alpha_1 + lk)$ random variable. In particular, if one assumes H2 and $\tilde{\vartheta}_{1k} = \tilde{\vartheta}_{2k}$ for all k, when $\alpha_2 \to 0$, one gets the limit situation in which all the observations are sampled from a common mixture of Pitman-Yor processes. In other words, in this limit case, one considers the observations (globally) exchangeable, so no distinction between the two blocks are allowed. The other limiting case is when one assumes H1 and takes $(\tilde{\vartheta}_{1k}, \tilde{\vartheta}_{2k})$ to be independent random elements with probability distribution G_{01} and G_{02}. In the limit for $\alpha_1 \to 0$, one obtains two independent Pitman-Yor processes G_1 and G_2 with base measures G_{01} and G_{02}. With this choice, the observations are grouped into two independent blocks which do not share information.

3.2. **Dirichlet Process marginal.** For $l = 0$ in (14) and (15) one gets Dirichlet Process marginals and, for this special case, it is possible to find an explicit formula for the correlation between G_i and G_j as a function of the parameters $(\alpha_1, \alpha_2, \dots)$. These results can be used for parameter elicitation purposes.

Let us assume time- and unit-independent kernels $\mathcal{K}(y|\vartheta)$, and define, for every $y \in \mathbb{Y}$ and every $i, j = 1, \ldots, r$ with $i \neq j$,

$$\kappa_{G_{0i}}(y) = \int \mathcal{K}(y|\vartheta) G_{0i}(d\vartheta), \quad \kappa_{G_{0j}}(y) = \int \mathcal{K}(y|\vartheta) G_{0j}(d\vartheta),$$
$$\kappa_{G_{0ij}}(y) = \int \mathcal{K}(y|\vartheta_i)\mathcal{K}(y|\vartheta_j) G_{0ij}(d\vartheta_i d\vartheta_j)$$

where G_{0i} denotes the distribution of $\tilde{\vartheta}_{i1}$ and G_{0ij} the distribution of $(\tilde{\vartheta}_{i1}, \tilde{\vartheta}_{j1})$.

Proposition 3. *Let* (12) *and* (14) *or* (13) *and* (15) *hold with* $l = 0$*, then for all measurable sets* A *and* B *and every* y *in* $\mathbb{Y}$

$$Cor(G_i(A), G_j(B)) = C_{ij} \times \frac{G_{0ij}(A \times B) - G_{0i}(A)G_{0j}(B)}{\sqrt{G_{0i}(A)(1 - G_{0i}(A))G_{0j}(B)(1 - G_{0j}(B))}} \tag{16}$$

$$Cor(f_i(y), f_j(y)) = C_{ij} \times \frac{\kappa_{G_{0ij}}(y) - \kappa_{G_{0i}}(y)\kappa_{G_{0j}}(y)}{\sqrt{\kappa_{G_{0i}}(y)(1 - \kappa_{G_{0i}}(y))\kappa_{G_{0j}}(y)(1 - \kappa_{G_{0j}}(y))}}. \tag{17}$$

with

$$C_{ij} = \frac{(1 + \alpha_1 + \alpha_2)(1 + \alpha_1)}{(1 + \alpha_1)(1 + \alpha_1 + \alpha_2) + \alpha_2}$$

under (12) *and* (14) *and*

$$C_{ij} = \frac{2\sqrt{(1 + \alpha_1 + \cdots + \alpha_{r-i+1})}(1 + \alpha_1 + \cdots + \alpha_{r-j+1})^{\frac{3}{2}}}{2(1 + \alpha_1 + \cdots + \alpha_{r-j+1})^2 + (2 + \alpha_1 + \cdots + \alpha_{r-j+1})(\alpha_{r-j+2} + \cdots + \alpha_{r-i+1})}.$$

under (13) *and* (15)*.*

Note that the correlation between random measures and densities, given in the above proposition, has two components. The first one is C_{ij} which is affected by the correlation level between the stick variables (see proof of the Proposition 3 in Appendix A), and the second one is given by the remaining term which reflects the dependence level between the atoms. In particular, if one assumes that the atoms are common to all G_i (see Eq. (18) in the next section) then $Cor(G_i(A), G_j(A)) = C_{ij}$, which suggests a possible interpretation for C_{ij} as a correlation and highlights the contribution of the stick variables to the correlation between random measures. We will discuss the assumptions on the atoms in the next section and focus, in the rest of this section, on the effects of the stick-variables correlation on C_{ij}, for different choices of the parameters. Assuming for simplicity $r = 2$, in the cases $l = 0$ one obtains the following correlation between S_{1h} and S_{2h}

$$Cor(S_{11}, S_{21}) = \begin{cases} \frac{\alpha_1(2+\alpha_1+\alpha_2)}{(2+\alpha_1)(\alpha_1+\alpha_2)} & \text{for} \quad H1 \\ \sqrt{\frac{\alpha_1(2+\alpha_1+\alpha_2)}{(2+\alpha_1)(\alpha_1+\alpha_2)}} & \text{for} \quad H2. \end{cases}$$

Fig. 1 shows the correlation level between the stick-breaking components (left column) and the random measures (right column) for different values of α_1 and α_2. In these graphs, the white color is used for correlation values equal to one and the black is used for a correlation values equal to zero. The gray areas represent correlation values in the unit interval. According to the left graphs of Fig. 1, one can conclude that the two parameterizations used in this paper allows for covering the whole range of possible correlation values in the unit interval. For instance a low correlation between the components of the stick-breaking corresponds to low values of α_1, say between 0 and 0.1, for any choice of α_2. The right graphs of Fig. 1 show the effect of α_1 and α_2 on C_{12}.

4. A hierarchical structure for the atoms

The simplest assumption for the atoms is that they are common to all the measures G_i. Otherwise stated this means that

$$(\tilde{\vartheta}_{1k}, \ldots, \tilde{\vartheta}_{rk}) = (\tilde{\vartheta}_{0k}, \tilde{\vartheta}_{0k}, \ldots, \tilde{\vartheta}_{0k}) \tag{18}$$

with $\tilde{\vartheta}_{0k}$ distributed according to a given common probability measure on Θ. Although in some situations it is reasonable to assume that the components of the mixture are the same for all the

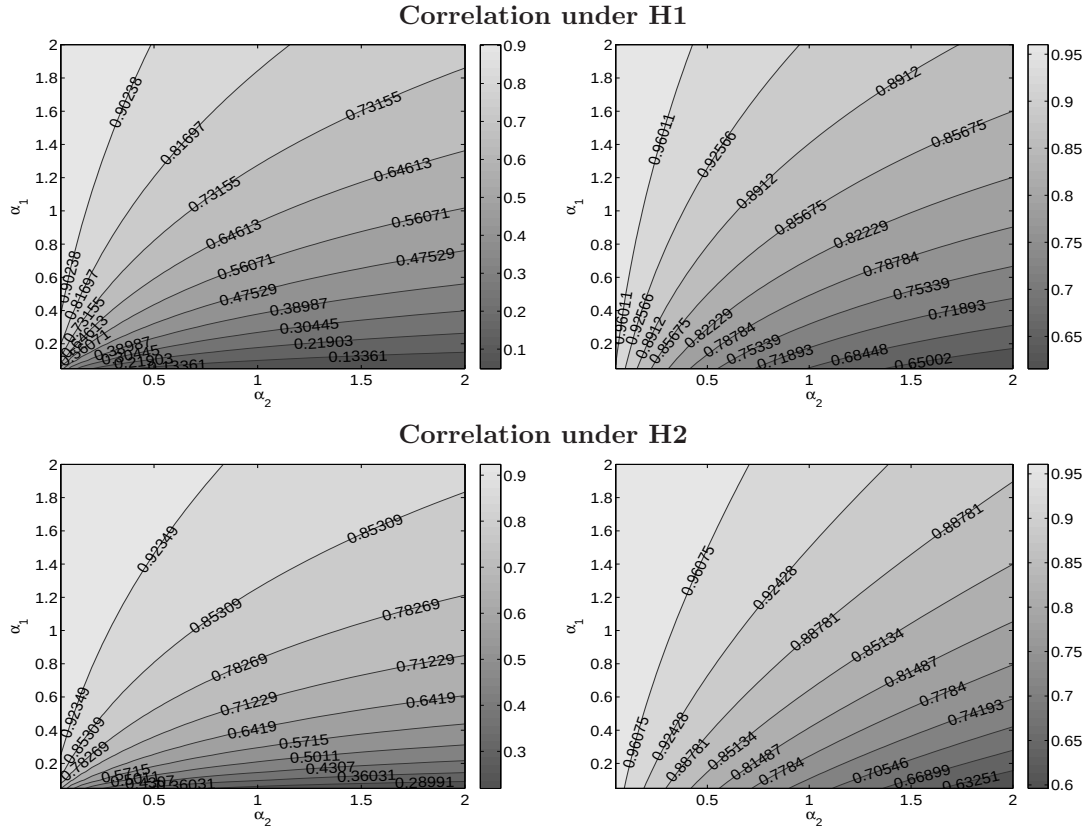

FIGURE 1. Left column: correlation between S_{11} and S_{21} under H1 (first row) and H2 (second row). Right column: C_{12}, that is correlation between G_1 and G_2, assuming common atoms, under H1 (first row) and H2 (second row)

units, the shared atom simplification (18) is a quite severe restriction when the units (or subsamples) exhibit a high degree of heterogeneity. For this reason, we consider a more elaborate hierarchical structure, that could include covariates (or exogenous effects) related to the specific block i in the law of $\tilde{\vartheta}_{ik}$.

4.1. **ANOVA-like repeated measure models.** Let us cosider the Gaussian repeated measure model introduced in equations (6)-(7), Section 2.2.1. The components of the parameter $\vartheta = (\mu, \sigma^2)$ represent a mean and a variance respectively. To allow for various degrees of pooling of information across units, one can assume for the parameters of the i-th kernel the ANOVA-like scheme of De Iorio et al. [2004], i.e. $\vartheta = (\mu_0 + \mu_i, \sigma_0^2\sigma_i^2)$. With this choice, each atom is the vector

$$\tilde{\vartheta}_{ik} = (\tilde{\mu}_{0k} + \tilde{\mu}_{ik}, \tilde{\sigma}_{0k}^2\tilde{\sigma}_{ik}^2) \tag{19}$$

where $\tilde{\mu}_{0k}, \tilde{\mu}_{1k}, \ldots, \tilde{\mu}_{rk}$ and $\tilde{\sigma}_{0k}^2, \tilde{\sigma}_{1k}^2, \ldots, \tilde{\sigma}_{rk}^2$ are independent random variables. In this case $\tilde{\mu}_{0k}$ and $\tilde{\sigma}_{0k}^2$ represent the common mean and variance (of the k-th mixture component) and $\tilde{\mu}_{ik}$ and $\tilde{\sigma}_{ik}^2$ the i-th factor-specific mean and variance (of the k-th mixture component).

We complete the hierarchical specification of the base measure G_0 by setting

$$(\tilde{\mu}_{ik}, \tilde{\sigma}_{ik}^2) \sim \mathcal{N}(0, s_i^{-2})\mathcal{IG}(\lambda/2, \lambda/2) \qquad i = 1, 2$$
$$(\tilde{\mu}_{0k}, \tilde{\sigma}_{0k}^2) \sim \mathcal{N}(0, s_0^{-2})\mathcal{IG}(\varepsilon/2, \varepsilon/2)$$

where $\mathcal{N}(0, s^{-2})$ denotes a Gaussian distribution of mean 0 and precision s^2, $\mathcal{IG}(a, b)$ denotes an inverse gamma distribution of parameters (a, b).

4.2. **Hierarchical Bayesian VAR models.** In the panel VAR model given in Eq. (8) the number of parameters, $(N^2p+N+N(N+1)/2)$ with $N=mr$, increases rapidly with the number of units and variables possibly leading to an overfitting problem. Our Bayesian non–parametric approach deals with the overfitting problem through a suitable choice of the base measure of the Pitman-Yor process prior. We follow here a hierarchical specification of the base measure which can be used to incorporate cross-equation interdependences and various degrees of information pooling across units (e.g., see Chib and Greenberg [1995] and Min and Zellner [1993]).

For the i-th unit, the first stage equation of our hierarchical model is given by Eq. (9). As regard the second stage of hierarchy, our prior setting is

$$\tilde{\vartheta}_{ik} = (\tilde{\phi}_{0k} + \tilde{\phi}_{ik}, \tilde{\sigma}_{0k}\tilde{\Sigma}_{ik})$$

$i=1,\ldots,r$, where $\tilde{\phi}_{0k}$ and $\tilde{\sigma}_{0k}$ are common factors to all countries, while $\tilde{\phi}_{ik}$ and $\tilde{\Sigma}_{ik}$ are country-specific factors. Above $\tilde{\phi}_{0k}$,$\tilde{\sigma}_{0k}$,$\tilde{\phi}_{ik}$ and $\tilde{\Sigma}_{ik}$ are independent and

$$\begin{aligned} (\tilde{\phi}_{ik}, \tilde{\Sigma}_{ik}^{-1}) &\sim \mathcal{N}_m(0, \tilde{\Upsilon}_{ik})\mathcal{W}_2(\lambda, \Lambda) \qquad i=1,2 \\ (\tilde{\phi}_{0k}, \tilde{\sigma}_{0k}^{-1}) &\sim \mathcal{N}_m(0, \Upsilon_0)\mathcal{G}(\varepsilon/2, \varepsilon/2) \end{aligned} \tag{20}$$

where $m=2(2rp+1)$, $\mathcal{W}_2(\lambda,\Lambda)$ denotes a bivariate Wishart distribution of parameters λ and Λ and $\mathcal{G}(\varepsilon/2,\varepsilon/2)$ denotes the gamma distribution of parameters $\varepsilon/2$ and $\varepsilon/2$. A third stage of the hierarchy is imposed by specifying a prior also for all the $\tilde{\Upsilon}_{ik}$, $i=1,2$, $k\geq 1$. A specific choice of such a prior will be presented in Section 6.2.

5. Slice Sampling Algorithm for Posterior Simulation

For posterior computation, we propose an extension of the slice sampling algorithm introduced in Walker [2007] and Kalli et al. [2011]. For the sake of simplicity we shall describe the sampling strategy for a vector of Beta-Product DPY with $r=2$. The proposed algorithm can be easily extend to the case $r>2$.

Recall that the stick variables in the $\beta_i-\mathrm{DPY}(\psi, G_0)$ are defined by

$$(S_{1k}, S_{2k}) = (V_{0k}V_{1k}, V_{0k}V_{2k})$$

for a sequence of independent vectors $V_k=(V_{0k},V_{1k},V_{2k})$, with the convention $V_{2k}=1$ and $V_k=(V_{0,k},V_{1,k})$ under H2.

In order to deal in the same time with both the repeated measure model of Section 2.2.1 and the panel VAR model of Section 2.2.2, we assume that for any $t\geq 1$ the conditional distribution of Y_{it} given $Z_t=[Y_{is}: i=1,2, s=1,\ldots,t-1]$ is

$$f_{it}(y) = f_{it}(y|Z_t) = \sum_{k\geq 1} W_{ik}\mathcal{K}_t(y|\tilde{\vartheta}_{ik}, Z_t) \qquad i=1,2 \tag{21}$$

where $\mathcal{K}_t$ is a suitable kernel depending on some parameter $\tilde{\vartheta}$ and possibly on t and Z_t. As we have seen, both infinite mixtures of repeated measurement models (6)-(7) and infinte mixtures of panel VAR models (10)-(11) can be written in this way.

Starting from (21), the key idea of the slice sampling is to find a finite number of variables to be sampled. For each t, we introduce two latent variables (U_{it}, D_{it}) in such a way that the joint density of (Y_{it}, U_{it}, D_{it}) given $\left[Z_t, (U_{js}, D_{js}) : j=1,2, s=1,\ldots,t-1\right]$ is

$$f_{it}(y,u,d) = \mathbb{I}\{u\leq W_{i,d}\}\mathcal{K}_t(y|\tilde{\vartheta}_{id}, Z_t). \tag{22}$$

With this position, the marginal density of Y_{it} given Z_t is (21) and the likelihood function for the augmented variables is available as a simple product of terms. Crucially D_{it} is finite. Note also that the D_{it}'s $(i=1,2;t=1,\ldots,T_i)$ are the allocation variables for the observations Y_{it} and that the slice variables U_{it} $(i=1,2;t=1,\ldots,T_i)$ take values on $[0,1]$.

We shall use the notation

$$Y_i^{(T_i)} = (Y_{i1},\ldots,Y_{iT_i}), \quad D_i^{(T_i)} = (D_{i1},\ldots,D_{iT_i}), \quad U_i^{(T_i)} = (U_{i1},\ldots,U_{iT_i})$$

and we write: $\tilde{\vartheta}$ for $(\tilde{\vartheta}_k)_k$, V for $(V_k)_k$, U for $[U_1^{(T_1)}, U_2^{(T_2)}]$, D for $[D_1^{(T_1)}, D_2^{(T_2)}]$ and Y for $[Y_1^{(T_1)}, Y_2^{(T_2)}]$.

In the applications we further assume a prior for $\psi = (\alpha_1, \alpha_2, l)$. Following the standard practice for these Bayesian nonparametric models, we let $\tilde{\vartheta}$ be independent of $\tilde{\psi} = (\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{l})$, while the distribution of V_j depends on $\tilde{\psi}$ through H1 or H2, so we shall write $P\{V_j \in dv_j|\tilde{\psi}\}$.

For the posterior sampling of $[\tilde{\vartheta}, V, U, \tilde{\psi}, D]$ we propose block Gibbs sampler which iteratively simulates $\tilde{\vartheta}$ given $[V, U, D, \tilde{\psi}, Y]$, $[V, U, \tilde{\psi}]$ given $[D, \tilde{\vartheta}, Y]$ and D given $[V, U, \tilde{\vartheta}, \tilde{\psi}, Y]$.

In one dimension this blocking structure has been introduced by Papaspiliopoulos [2008] and Kalli et al. [2011] as an efficient alternative to the original algorithm of Walker [2007]. A multidimensional slice sampling has been proposed by Hatjispyrosa et al. [2011] for a different dependent DP mixtures model. In this paper, we follow an alternative route and elaborate further on the Kalli et al. [2011] blocking strategy by proposing a nested structure for sampling stick variables. It should be noted that we borrow from Kalli et al. [2011] only the main blocking structure of the algorithm and let for further research the use of auxiliary weights in the slicing scheme to increase the sampling efficiency. In order to describe the full-conditionals of our block Gibbs sampler we introduce some more notation. Define for $i = 1, 2$ and $k \geq 1$,

$$\mathcal{D}_{i,k} = \{t \in \{1, \ldots, T_i\} : D_{i,t} = k\},$$
$$A_{i,k} = \sum_{t=1}^{T_i} \mathbb{I}\{D_{i,t} = k\} = card(\mathcal{D}_{i,k}), \quad B_{i,k} = \sum_{t=1}^{T_i} \mathbb{I}\{D_{i,t} > k\}$$

and let

$$D^* = \max_{i=1,2} \max_{1 \leq t \leq T_i} D_{i,t}. \tag{23}$$

In our MCMC algorithm we shall treat V as three blocks of random length: $V = (V^*, V^{**}, V^{***})$, where

$$V^* = \{V_k : k \in \mathcal{D}^*\}, \qquad V^{**} = \{V_k : k \notin \mathcal{D}^*, k \leq D^*\}, \qquad V^{***} = \{V_k : k > D^*\}$$

and $\mathcal{D}^* = \{k : \mathcal{D}_{1,k} \cup \mathcal{D}_{2,k} \neq \emptyset\}$. In what follows, if the kernels $\mathcal{K}_t$ depend on $[Y_{js} : j = 1, 2; s = 1, \ldots, t-1]$, as in the case of the panel VAR model, we assume that $T_i = T$ for $i = 1, 2$. While, if the kernels $\mathcal{K}_t$ depend only on $[Y_{is} : s = 1, \ldots, t-1]$, as in the case of the repeated measurements, we allow T_1 and T_2 to be different. All the full conditionals can be deduced from the joint distribution given in (43). Further details are given in Appendix A and B.

5.1. **The full conditional of $\tilde{\vartheta}$.** The atoms $\tilde{\vartheta}$ given $[V, D, U, \tilde{\psi}, Y]$ are conditionally independent and the full conditionals are:

$$\begin{aligned} P\{\tilde{\vartheta}_k \in d\vartheta_k|D, U, V, \tilde{\psi}, Y\} &= P\{\tilde{\vartheta}_k \in d\vartheta_k|D, Y\} \\ &\propto G_0(d\vartheta_k) \prod_{t \in \mathcal{D}_{1,k}} \mathcal{K}_t(Y_{1,t}|\vartheta_{1k}, Z_t) \prod_{t \in \mathcal{D}_{2,k}} \mathcal{K}_t(Y_{2,t}|\vartheta_{2k}, Z_t); \end{aligned} \tag{24}$$

where $\vartheta_k = (\vartheta_{1k}, \vartheta_{2k})$. The strategy for sampling from this full conditional depends on the specific form of $\mathcal{K}_t$ and G_0. In Appendix B we will discuss a possible strategy for Gaussian kernels.

5.2. **The full conditional of $[V, U, \tilde{\psi}]$.** In order to sample from the conditional distribution of $[V, U, \tilde{\psi}]$ given $[D, \tilde{\vartheta}, Y]$ a further blocking is used:

- $[V^*, \tilde{\psi}]$ given $[D, \tilde{\vartheta}, Y]$. The joint conditional distribution of $[V^*, \tilde{\psi}]$ given $[D, \tilde{\vartheta}, Y]$ is

  $$P\{V^* \in dv^*, \tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|Y, \tilde{\vartheta}, D\} = P\{V^* \in dv^*, \tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|D\}$$

  moreover, under H1,

$$P\{V^* \in dv^*, \tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|D\} \propto \prod_{k \in \mathcal{D}^*} \frac{Q_k(v_k|D, \psi) dv_k \pi(d\alpha_1, d\alpha_2, dl)}{B^2(\alpha_1 + 1 - l, \alpha_2 + lk) B(1 - l, \alpha_1)} \tag{25}$$

  with $v_k = (v_{0k}, v_{1k}, v_{2k})$ and

$$Q_k(v_k|D, \psi) = v_{0k}^{-l + A_{1k} + A_{2k}} (1 - v_{0k})^{\alpha_1 - 1} \prod_{i=1,2} v_{ik}^{A_{ik} + \alpha_1 - l} (1 - v_{ik})^{\alpha_2 + lk - 1} (1 - v_{0k} v_{ik})^{B_{ik}},$$

while, under H2,

$$P\{V^* \in dv^*, \tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|D\} \propto \prod_{k\in\mathcal{D}^*} \frac{Q_k(v_k|D,\psi)dv_k\pi(d\alpha_1,d\alpha_2,dl)}{B(\alpha_1+1+l(k-1),\alpha_2)B(1-l,\alpha_1+lk)} \tag{26}$$

with $v_k = (v_{0k}, v_{1k})$ and

$$Q_k(v_k|D,\psi) = v_{0k}^{A_{1k}+A_{2k}-l}(1-v_{0k})^{\alpha_1+lk+B_{2k}-1}v_{1k}^{\alpha_1+l(k-1)+A_{1k}}(1-v_{1k})^{\alpha_2-1}(1-v_{0k}v_{1k})^{B_{1k}},$$

and $\pi(d\alpha_1, d\alpha_2, dl) = P\{\tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)\}$ is the prior on the stick-breaking parameters. To sample from (25)-(26), we iterate a two-step Metropolis-Hastings (M.-H.) within Gibbs with full conditionals

$$P\{V^* \in dv^*|\tilde{\psi}, D\} \propto \prod_{k\in\mathcal{D}^*} Q_k(v_k|D,\tilde{\psi})dv_k \tag{27}$$

and

$$\begin{aligned} P\{\tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|V^*, D\} \propto & \prod_{k\in\mathcal{D}^*} \frac{V_{0k}^{-l}(1-V_{0k})^{\alpha_1}}{B(1-l,\alpha_1)} \\ & \prod_{i=1,2} \frac{V_{ik}^{\alpha_1-l}(1-V_{ik})^{\alpha_2+lk}}{B(\alpha_1+1-l,\alpha_2+lk)}\pi(d\alpha_1,d\alpha_2,dl) \end{aligned} \tag{28}$$

under H1, and

$$\begin{aligned} P\{\tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|V^*, D\} \propto & \prod_{k\in\mathcal{D}^*} \frac{V_{0k}^{-l}(1-V_{0k})^{\alpha_1+lk}}{B(1-l,\alpha_1+lk)} \\ & \frac{V_{1k}^{\alpha_1+l(k-1)}(1-V_{1k})^{\alpha_2}}{B(\alpha_1+1+l(k-1),\alpha_2)}\pi(d\alpha_1,d\alpha_2,dl) \end{aligned} \tag{29}$$

under H2. For the elements of V^* we consider a multivariate Gaussian random walk proposal with diagonal scale matrix and with scale parameter in the interval $[0.05, 0.3]$, in order to have acceptance rates between 0.3 and 0.5 for the elements of V^*. We generate samples from the full conditional of $\tilde{\psi}$ by a M.-H. step. As the sampler depends on the prior specification we refer to Section 6 and Appendix B for further details.

- $[V^{**}, V^{***}]$ given $[D, V^*, \tilde{\vartheta}, \tilde{\psi}, Y]$. The V_k (with $k \notin \mathcal{D}^*$) are conditionally independent given $[D, V^*, \tilde{\vartheta}, \tilde{\psi}, Y]$ with $P\{V_k \in dv_k|\tilde{\psi}, D, V^*\} \propto Q_k(v_k|D,\tilde{\psi})dv_k$ if $k \leq D^*$ and $P\{V_k \in dv|V^*, \tilde{\vartheta}, D, \tilde{\psi}, Y\} = P\{V_k \in dv|\tilde{\psi}\}$ if $k > D^*$. Note that if $k \notin \mathcal{D}^*$ and $k \leq D^*$, then $A_{i,k} = 0$ in the definition of $Q_k(v_k|D,\tilde{\psi})$. In order to sample from $Q_k(v_k|D,\tilde{\psi})$ the same M.-H. step, used for the full conditional in (27), is employed.
- U given $[V, D, \tilde{\vartheta}, \tilde{\psi}, Y]$. The slice variables U are conditionally independent given $[V, D, \tilde{\vartheta}, \tilde{\psi}, Y]$ with

$$P\{U_{i,t} \in du|V, Y, \tilde{\vartheta}, D\} = P\{U_{i,t} \in du|V, D\} = \frac{\mathbb{I}\{u \leq W_{i,D_{i,t}}\}}{W_{i,D_{i,t}}}du. \tag{30}$$

5.3. **The full conditional of D.** The D's are conditionally independent given $[V, U, \tilde{\vartheta}, \tilde{\psi}, Y]$ with

$$P\{D_{i,t} = d|\tilde{\vartheta}, V, U, \tilde{\psi}, Y\} \propto \mathcal{K}_t(Y_{i,t}|\tilde{\vartheta}_{id}, Z_t)\ \mathbb{I}\{U_{i,t} \leq W_{i,d}\}. \tag{31}$$

Here an important remark is in order. As in the slice sampling proposed in Walker [2007] and Kalli et al. [2011], the full conditional (31) samples, almost surely, from a finite number of terms. More precisely, $d > N^*_{i,t}$ ensures that $W_{i,d} < U_{i,t}$ where $N^*_{i,t}$ $(i = 1, 2; t = 1, \ldots, T_i)$ is the smallest integer such that

$$\sum_{k=1}^{N^*_{i,t}} W_{i,k} > 1 - U_{i,t}. \tag{32}$$

6. Illustrations

6.1. **DPY(ψ, G_0) mixtures of Gaussian distributions.** We consider DPY(ψ, G_0) Gaussian mixture model for inference on synthetic data generated from finite Gaussian mixtures. For the atoms we assume the hierarchical structure described in Section 4.1.

We assume independent gamma priors $\mathcal{G}(\zeta_{11}, \zeta_{21})\mathcal{G}(\zeta_{12}, \zeta_{22})$ for the vector $\tilde{\alpha} = (\tilde{\alpha}_1, \tilde{\alpha}_2)$ of precision parameters and uniform prior $\mathcal{U}_{[0,1)}$ for the parameter $\tilde{l}$. In summary the Bayesian non–parametric model is

$$\begin{aligned}
&Y_{it}|(\mu^*_{it}, {\sigma^*_{it}}^2) \overset{ind}{\sim} \mathcal{N}(\mu^*_{it}, {\sigma^*_{it}}^2) \qquad i = 1, 2, t \geq 1\\
&(\mu^*_{it}, {\sigma^*_{it}}^2)|G_1, G_2 \overset{iid}{\sim} G_i \quad i = 1, 2\\
&(G_1, G_2)|\tilde{\psi} \sim \beta{-}\mathrm{DPY}(\tilde{\psi}, G_0)\\
&\tilde{\psi} \sim \mathcal{G}(\zeta_{11}, \zeta_{21})\mathcal{G}(\zeta_{12}, \zeta_{22})\mathcal{U}_{[0,1)}.
\end{aligned}$$

Recall that, as described in Section 4.1, the base measure G_0 is such that

$$(\mu^*_{it}, {\sigma^*_{it}}^2) = (\tilde{\mu}_{0D_{it}} + \tilde{\mu}_{iD_{it}}, \tilde{\sigma}^2_{0D_{it}}\tilde{\sigma}^2_{iD_{it}}),$$

where D_{it} is the allocation variable of the observation Y_{it}, and

$$\begin{aligned}
&(\tilde{\mu}_{ik}, \tilde{\sigma}^2_{ik}) \sim \mathcal{N}(0, s_i^{-2})\mathcal{IG}(\lambda/2, \lambda/2) \qquad i = 1, 2\\
&(\tilde{\mu}_{0k}, \tilde{\sigma}^2_{0k}) \sim \mathcal{N}(0, s_0^{-2})\mathcal{IG}(\varepsilon/2, \varepsilon/2).
\end{aligned}$$

In this example we bring into action the sampling procedure for U and D given in the previous section. As regard to the sampling strategy for the other variables Appendix B.1 shall describe it in more details.

We simulate $T = 100$ independent vectors, $(Y_{1,t}, Y_{2,t})$ with $t = 1, \ldots, T$, of observations. The components of the vectors $(Y_{1,t}, Y_{2,t})$ are independent and alternatively follow one of the following models.

- The same three-component mixture of normals (model Mix1)
$$\begin{aligned}
Y_{1,t} &\sim \frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1)\\
Y_{2,t} &\sim \frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1)
\end{aligned}$$
- The same three-component mixture of normals with different component probabilities (model Mix2)
$$\begin{aligned}
Y_{1,t} &\sim \frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1)\\
Y_{2,t} &\sim \frac{1}{6}\mathcal{N}(-10, 1) + \frac{4}{6}\mathcal{N}(0, 1) + \frac{1}{6}\mathcal{N}(10, 1)
\end{aligned}$$
- Two different mixtures with two common components (model Mix3)
$$\begin{aligned}
Y_{1,t} &\sim \frac{1}{4}\mathcal{N}(0, 0.5) + \frac{1}{4}\mathcal{N}(3, 0.25) + \frac{1}{4}\mathcal{N}(2, 0.25) + \frac{1}{4}\mathcal{N}(5, 0.5)\\
Y_{2,t} &\sim \frac{1}{4}\mathcal{N}(0, 0.5) + \frac{1}{4}\mathcal{N}(3, 0.25) + \frac{1}{4}\mathcal{N}(-3, 0.25) + \frac{1}{4}\mathcal{N}(7, 0.5)
\end{aligned}$$
- Two different mixtures with two common modes (model Mix4) and small modes in one of the mixtures
$$\begin{aligned}
Y_{1,t} \sim{}& \frac{2}{9}\mathcal{N}(-10, 0.4) + \frac{1}{12}\mathcal{N}(-5, 0.4) + \frac{1}{12}\mathcal{N}(-3, 0.4) + \frac{2}{9}\mathcal{N}(0, 0.4)\\
&+ \frac{1}{12}\mathcal{N}(3, 0.4) + \frac{1}{12}\mathcal{N}(5, 0.4) + \frac{2}{9}\mathcal{N}(10, 0.4)\\
Y_{2,t} \sim{}& \frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1).
\end{aligned}$$

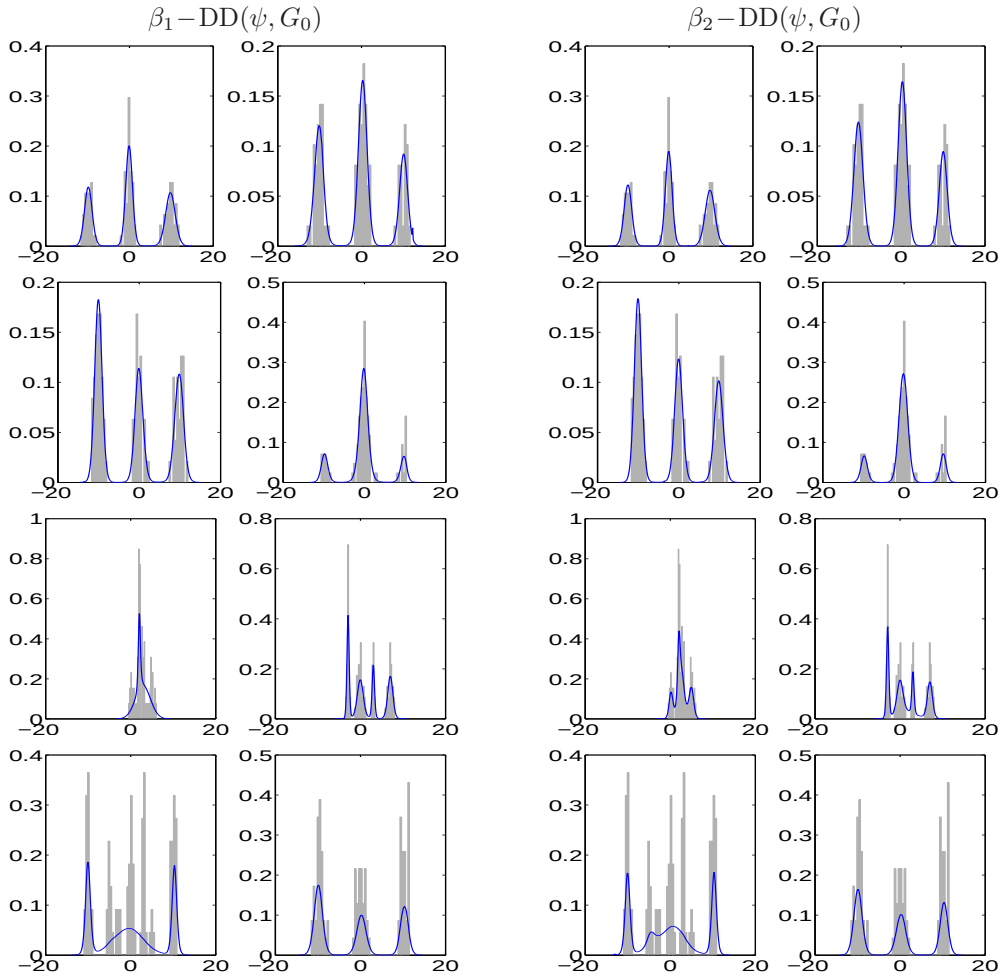

FIGURE 2. Predictive for different prior settings (panels $\beta_1-\mathrm{DD}(\psi, G_0)$ and $\beta_2-\mathrm{DD}(\psi, G_0)$). Data histograms for the first (left column) and second (right column) component and for the different models: Mix1 (first row), Mix2 (second row), Mix3 (third row) and Mix4 (fourth row).

The histograms of the simulated data considered in the experiments are given in Fig. 2-3.

We apply the $\beta_i-\mathrm{DD}(\psi, H_0)$ and the $\beta_i-\mathrm{DPY}(\psi, H_0)$, $i = 1, 2$, on the different set of data. In the inference exercise, we choose a fairly non–informative prior specification for the mean and precision parameters of the base measure and set $s_i^2 = 0.1$, $\lambda = 0.5$ (see for example Walker [2007]). Similarly we assume $s_0^2 = 0.01$ and $\varepsilon = 1$. For the concentration parameters (α_1, α_2) of the stick-breaking components, we follow Kalli et al. [2011] and consider weakly informative prior with hyperparameters $(\zeta_{1j} = 0.01, \zeta_{2j} = 0.01)$, for $j = 1, 2$, in all the experiments from Mix1 to Mix4. This setting corresponds to diffuse priors on the concentration parameters, with prior means $\mathbb{E}(\tilde{\alpha}_1) = \mathbb{E}(\tilde{\alpha}_2) = 1$ and variances $\mathbb{V}(\tilde{\alpha}_1) = \mathbb{V}(\tilde{\alpha}_2) = 100$. For each dataset and model, the Gibbs sampler, presented in the previous section, was run for 20,000 iterations. A burn-in period of 10,000 samples was discarded and at each Gibbs iteration from 10,000 onwards, a sample $(Y_{1,T+1}, Y_{2,T+1})$ from the predictive density was taken. Right and left panels in Fig. 2 show the histograms of the two components (right and left column of each panel) for each set from Mix1 to Mix4 of synthetic data (different rows). In the same figures the solid lines represent the estimated posterior predictive densities for $\beta_1-\mathrm{DD}(\psi, H_0)$ (left panel) and $\beta_2-\mathrm{DD}(\psi, H_0)$ (right panel). Fig. 3 shows data histograms and predictive densities estimated with $\beta_1-\mathrm{DPY}(\psi, H_0)$ (left panel) and $\beta_2-\mathrm{DPY}(\psi, H_0)$ (right panel). For all models and datasets the approximated posterior of

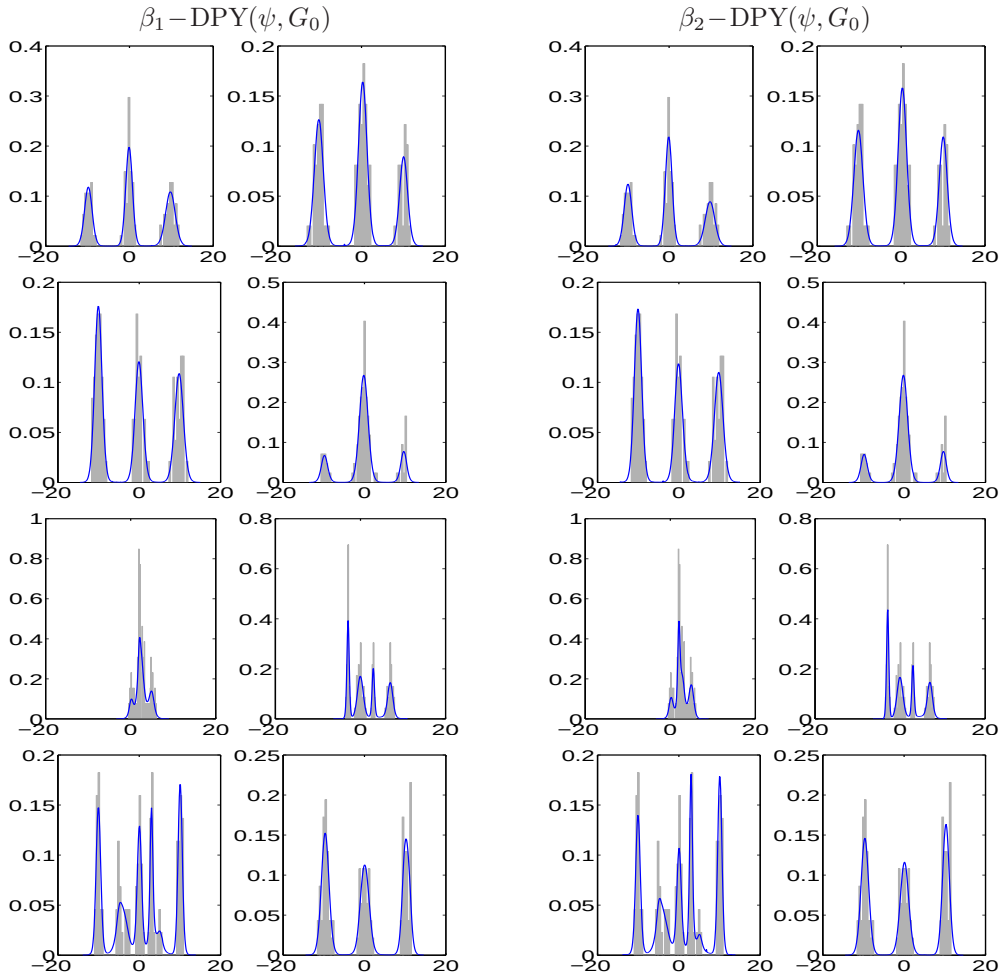

FIGURE 3. Predictive for different prior settings (panels $\beta_1-\mathrm{DPY}(\psi, G_0)$ and $\beta_2-\mathrm{DPY}(\psi, G_0)$). In each panel, data histograms for the first (left column) and second (right column) component and for the different models: Mix1 (first row), Mix2 (second row), Mix3 (third row) and Mix4 (fourth row).

the number of clusters is given in Fig. 4. The first and second row in Fig. 2 and 3 show that dependent Dirichlet and Pitman-Yor process priors have similar predictive densities for Mix1 and Mix2. The histograms in the first and second rows of each panels of Fig. 4 show that under H1 both the DD and the DPY overestimate the number of clusters, while under H2 they estimate the exact number of clusters. For Mix3 $\beta_i-\mathrm{DPY}(\psi, H_0)$, $i = 1, 2$ have similar predictive densities and $\beta_1-\mathrm{DPY}(\psi, H_0)$ is doing better than $\beta_1-\mathrm{DD}(\psi, H_0)$ in terms of number of clusters. DD and DPY under H2 give similar results. For dataset Mix4 DD and DPY give different results both in terms of predictive densities and posterior number of clusters. $\beta_i-\mathrm{DD}(\psi, H_0)$, $i = 1, 2$, are largely under-estimating the number of clusters while $\beta_i-\mathrm{DPY}(\psi, H_0)$, $i = 1, 2$, are doing better, especially $\beta_2-\mathrm{DPY}(\psi, H_0)$.

The results described above are confirmed in the statistical comparison between the different nonparametric models. In the comparison we consider a Bayesian model averaging approach (see Hoeting et al. [1999]). We focus on the predictive densities of the proposed models, due to the relevance of the density forecasts in many economics applications (e.g., see Granger [2006]). Although there are many alternative ways to assign to the models a numerical score based on the ability of the predictive density and to predict the variables of interest (e.g., see Gneiting and Raftery [2007]), we follow Jensen and Maheu [2010] and apply the model pooling approach

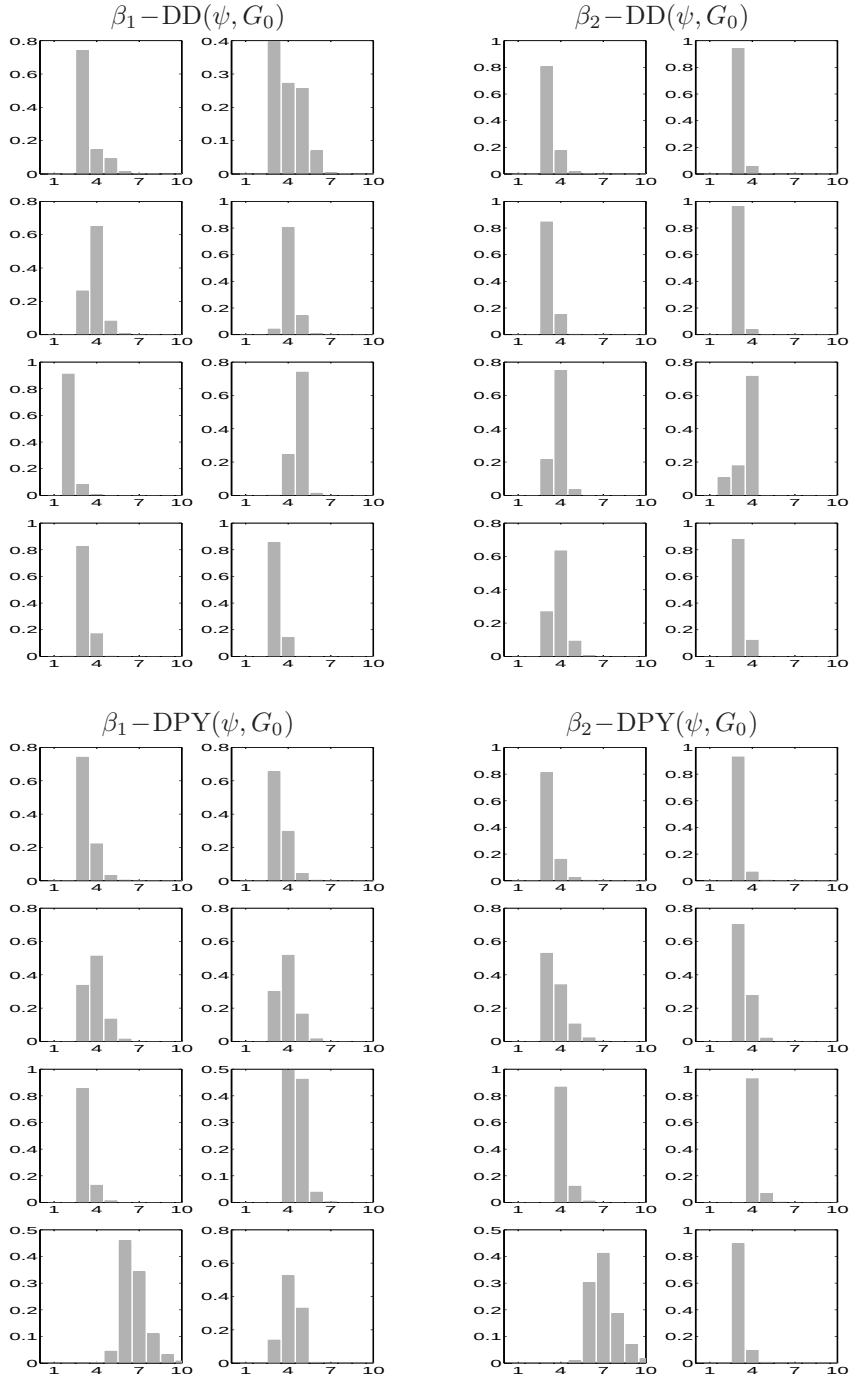

FIGURE 4. Number of clusters for different prior settings (panels $\beta_i-\mathrm{DD}(\psi, G_0)$, $i = 1, 2$ and $\beta_i-\mathrm{DPY}(\psi, G_0)$). In each panel, histograms for the first (left column) and second component (right column) and for the different models: Mix1 (first row), Mix2 (second row), Mix3 (third row) and Mix4 (fourth row).

of Geweke and Amisano [2010] based on log predictive score. Our set of predictive densities is $f(Y_t|Y_1, \ldots, Y_{t-1}, M_j)$, $j = 1, \ldots, J$, where $Y_t' = (Y_{1t}', Y_{2t}')$ and M_j indicates model $\beta_1-\mathrm{DD}(\psi, G_0)$ $(j = 1)$, $\beta_2-\mathrm{DD}(\psi, G_0)$ $(j = 2)$, $\beta_1-\mathrm{DPY}(\psi, G_0)$ $(j = 3)$ and $\beta_2-\mathrm{DPY}(\psi, G_0)$ $(j = 4)$. We define

TABLE 1. In each row: the simulation experiment (first column), the optimal pooled log-predictive score function (second column) and the associated optimal combination weights (remaining columns) for β_1-DD (w_1), β_2-DD (w_2), β_1-DPY (w_3) and β_2-DPY (w_4).

Dataset	Log-score	w_1	w_2	w_3	w_4
Mix1	-250.25	0.0000	0.0000	0.9087	0.0913
Mix2	-179.61	0.2256	0.0000	0.7744	0.0000
Mix3	-265.00	0.4674	0.0000	0.5326	0.0000
Mix4	-273.25	0.0000	0.0790	0.3827	0.5383

the combined predictive density

$$f(Y_t|Y_1,\dots,Y_{t-1}) = \sum_{j=1}^{J} w_j f(Y_{t-1}|Y_1,\dots,Y_{t-1},M_j)$$

with $w_j \geq 0$ and $\sum_{j=1}^{J} w_j \leq 1$ and choose the combination weights $w = (w_1,\dots,w_J)$ to maximize the log pooled predictive score function

$$\max_{w} \sum_{t=\tau_1}^{\tau_2} \log \sum_{j=1}^{J} w_j f(Y_t|Y_1,\dots,Y_{t-1},M_j).$$

The model-specific predictive densities are not available in analytical form, but can be approximated by using the Gibbs sampling scheme given in the previous section. In this application, for a given dataset, we set $\tau_1 = 50$ and $\tau_2 = 100$ and for each model M_j, $j = 1,\dots,4$, we use 5,000 MCMC draws to approximate the predictive density $f(Y_t|Y_1,\dots,Y_{t-1},M_j)$ at the data point Y_t. We repeated the procedure for $t = \tau_1,\dots,\tau_2$ and obtain a sequence of densities for each model that can be used to find the optimal combination weights. The results in Tab. 1 show that for all datasets DPY models have the better predictive ability. More specifically for Mix1 β_1-DPY(ψ, H_0) has the highest combination weight, while for Mix2 and Mix3 both DD and DPY under assumption H1 are performing better than DD and DPY under assumption H2. Although, for both datasets β_1-DPY(ψ, H_0) has the highest combination weight. Finally on dataset Mix 4, which has a different number of modes for the two components, β_2-DPY is performing better than β_1-DPY.

In conclusion, the evidence is in favour of the use of both DD and DPY priors under the specification H1, with the exception of the cases where there is a strong heterogeneity across units, in terms of number of clusters. In this case β_2-DPY prior should be preferred.

6.2. β_2-**DPY**(ψ, H_0) **mixtures of vector autoregressive processes.** In business cycle modelling great advances have been made by allowing for separate parameter values in periods (called regimes) of recession and expansion. The seminal paper of Hamilton [1989] proposes to use a dynamic mixture model (regime-switching model) with two components for capturing clustering of observations during the recession and expansion phases in a business cycle. This simple model has been successfully extended in many directions and the issue of estimating the number of regimes has been considered in various papers (e.g., Kim and Murray [2002], Kim and Piger [2000] and Krolzig [2000]) following a parametric approach. Conversely, in this paper, we follow Otranto and Gallo [2002] and apply a non–parametric approach to the estimation of the number of regimes or structural breaks. We extend their approach to a multivariate set-up and propose a joint estimation procedure for the number of regimes or breaks in multiple time series. We assume a β_2-DPY(ψ, H_0) prior for the parameters of a two-country panel VAR and apply the resulting nonparametric model to two well studied business cycles: the United States (US) and the European Union (EU). Specifically it is interesting to verify whether the strong contraction in 2009 calls for the use of a higher number of regimes than three or four to achieve a better modelling of the cycles.

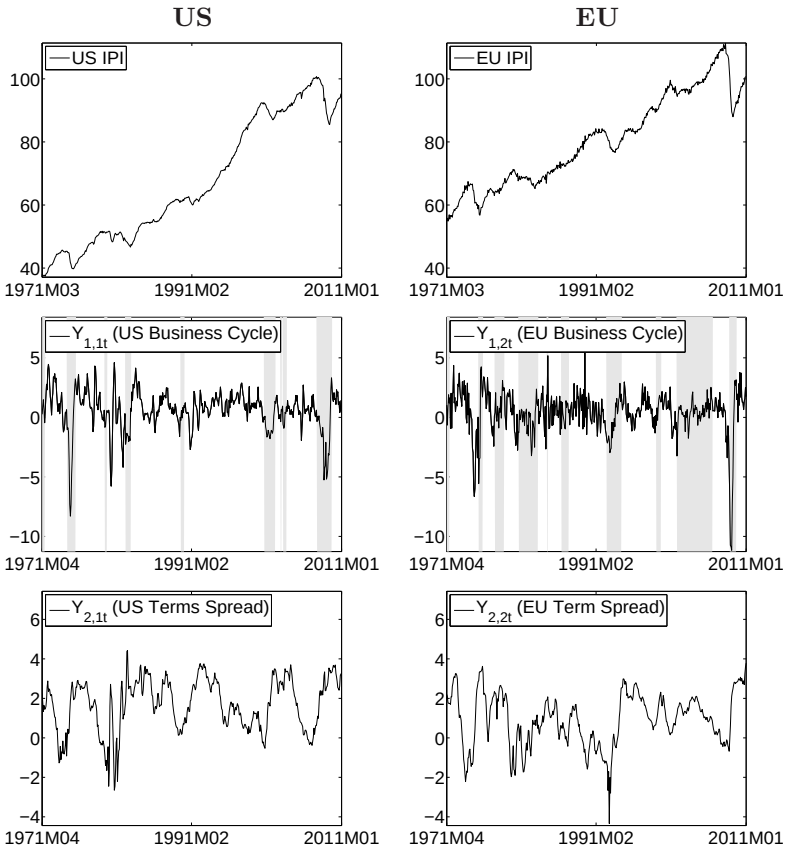

FIGURE 5. First row: Industrial Production Index (IPI) for the US and the EU at a monthly frequency for the period: March 1971 to January 2011. Second row: logarithmic quarterly changes in the US IPI ($Y_{1,1t}$) and the EU IPI ($Y_{1,2t}$) variables and NBER official recession phases (vertical gray bars). Third row: 10-year and 3-month interest rate spread (term spread) for the US ($Y_{2,1t}$) and the EU ($Y_{2,2t}$).

We consider seasonally and working day adjusted industrial production indexes (IPI), at a monthly frequency from the time of April 1971 to January 2011, for the US and the EU (see first row in Fig. 5). We take the quarterly growth rate of US and EU IPI, $\{Y_{1,1t}\}_{t=1}^{T}$ and $\{Y_{1,2t}\}_{t=1}^{T}$ respectively, (see second row in Fig. 5). To achieve a better modelling of the business cycle we consider the term spread (TS), that is the difference between the 3-month and the 10-year interest rates for the US and the EU, $\{Y_{2,1t}\}_{t=1}^{T}$ and $\{Y_{2,2t}\}_{t=1}^{T}$ respectively, (see third row in Fig. 5). As a preliminary analysis of the IPI data we apply the Bry and Boschan [1971] (BB) rule for the identification of the downward and upward turns of a time series. This nonparametric technique is now standard in the literature (see for example Billio et al. [2012]) and in the practice of business cycle analysis. It is currently used by the NBER institute for extracting the reference cycle for the US. Vertical bars in the second row of Fig. 5 show the time periods from an upturn to a downturn (recession phases) of the US and the EU economies. Following the BB rule, one can identify, in both series, many recession periods, and in particular the 2009 recession. This simple procedure lacks of a strong statistical foundation and cannot be used for forecasting purposes. This call for the use of suitable stochastic models, which are able to account for the relevant features of these series, such as skewness, heavy tails and multimodality in the unconditional distribution (see the histograms in Fig. 6). These departures from the normality (see solid lines in Fig. 6) may be due to the presence of regimes or breaks in the series, which are usually modelled with shifts in the parameters of a linear model (see Krolzig [2000] and Clements and Krolzig [1998]). Thus, in our Bayesian

nonparametric panel VAR model, we consider shifts in intercepts, autoregressive coefficients and covariance matrices. We assume cross-country conditional independence, as described in Section 2.2.2, and the hierarchical base measure given in Section 4.2 with $p = 4$. See Hamilton [1989] and Krolzig [2000] for a discussion on lag selection. In summary, the first stage equations of our hierarchical Bayesian model are

$$Y_{it} = (I_2 \otimes X_t')\phi_{it}^* + \varepsilon_{it} \tag{33}$$

for $i = 1, 2$ and $t = 1, \dots, T$, where $Y_{it} = (Y_{1,it}, Y_{2,it})'$, $X_t = (1, Y_{1\,t-1}', Y_{2\,t-1}', \dots, Y_{1\,t-p}', Y_{2\,t-p}')'$, $\varepsilon_{it} = (\varepsilon_{1,it}, \varepsilon_{2,it})'$ with $\varepsilon_{it} \sim \mathcal{N}_2(0, \Sigma_{it}^*)$. As regard to the second and third stage of hierarchy, the equations are

$$\begin{aligned} \phi_{it}^* &= \tilde{\phi}_{0D_{it}} + \tilde{\phi}_{iD_{it}} \\ \Sigma_{it}^* &= \tilde{\sigma}_{0D_{it}} \tilde{\Sigma}_{iD_{it}} \end{aligned} \tag{34}$$

$i = 1, 2$, where $\tilde{\phi}_{0k}$ is a common factor to all countries, $\tilde{\phi}_{ik}$ is a country-specific factor and $D_{it} \in \{1, \dots, D_i^*\}$ are the unit-specific allocation variables, which are generated by the Pitman-Yor model

$$\begin{aligned} &(\phi_{ik}^*, \Sigma_{ik}^{*\,-1}) | G_1, G_2 \overset{i.i.d.}{\sim} G_i \quad i = 1, 2 \\ &(G_1, G_2) \sim \beta_2 - \mathrm{DPY}(\tilde{\psi}, G_0) \\ &\tilde{\psi} \sim \mathcal{G}(\zeta_{11}, \zeta_{21}) \mathcal{G}(\zeta_{12}, \zeta_{22}) \end{aligned} \tag{35}$$

with base measure G_0 given by the following hierarchical structure

$$\begin{aligned} &(\tilde{\phi}_{ik}, \tilde{\Sigma}_{ik}^{-1}) \sim \mathcal{N}_m(0, \tilde{\Upsilon}_{ik}) \mathcal{W}_2(\lambda, \Lambda) \qquad i = 1, 2 \\ &(\tilde{\phi}_{0k}, \tilde{\sigma}_{0k}^{-1}) \sim \mathcal{N}_m(0, \Upsilon_0) \mathcal{G}(\varepsilon/2, \varepsilon/2), \end{aligned} \tag{36}$$

where $m = 2(4p + 1)$. In the same spirit of Chib and Greenberg [1995], we further assume that the hyperparameters λ, Λ, Υ_0 and ε are known and that $\tilde{\Upsilon}_{ik} = \tilde{\tau}_k^2 I_m$, $i = 1, 2$, are random with

$$\tilde{\tau}_k^2 \sim \mathcal{IG}(\nu_0/2, \nu_0/2). \tag{37}$$

The Gibbs sampling implementation for the model in Eq. (33)-(37) is given in Appendix B.2. The charts in the first row of Fig. 6 show the predictive distributions (solid lines) generated by the non–parametric approach conditioning on all values of $Y_t = (Y_{1t}', Y_{2t}')'$, $t = 1, \dots, T$, and the best normal fits (dashed lines) for the empirical distributions of the two series.

From a comparison with the empirical distribution, we note that the non–parametric approach, as opposed to the normal model, is able to capture skewness and excess of kurtosis in the data. The results from our model are in line with the practice of using time-varying parameter models (e.g., Markov-switching models) to capture asymmetry and non-linearity in both the US and the EU business cycles. The main results of our non–parametric approach can be summarized through the implied data clustering and the posterior predictive densities.

As regard to the data clustering, the posterior distribution of the number of clusters is given in the second row of Fig. 6. The location of the posterior mode of the histograms allows us to conclude that following our non–parametric approach there is evidence in favour of three clusters for the US cycle and at least of four clusters for the EU cycle. The result for the US data is coherent with the results available in the literature where three-regime Markov-switching models (see for example Krolzig [2000]) are usually considered. The result for the EU cycle is, in a certain way, coherent with the output of parametric studies which suggest to consider at least three regimes. Nevertheless, the effects of the 2009 recession on the past empirical findings is an open issue and a matter of research. The result from our non–parametric approach is interesting because it suggests a substantial evidence in favour of at least four components in the mixture (see Fig. 6) for the EU. The identification of the mixture components relies upon the implied data clustering. In order to estimate the implied data clustering of our DPY mixture model, we apply the least square clustering method proposed originally in Dahl [2006]. The method has been successfully used in many applications (see for example Kim et al. [2006] and Rodriguez et al. [2008]) and is based

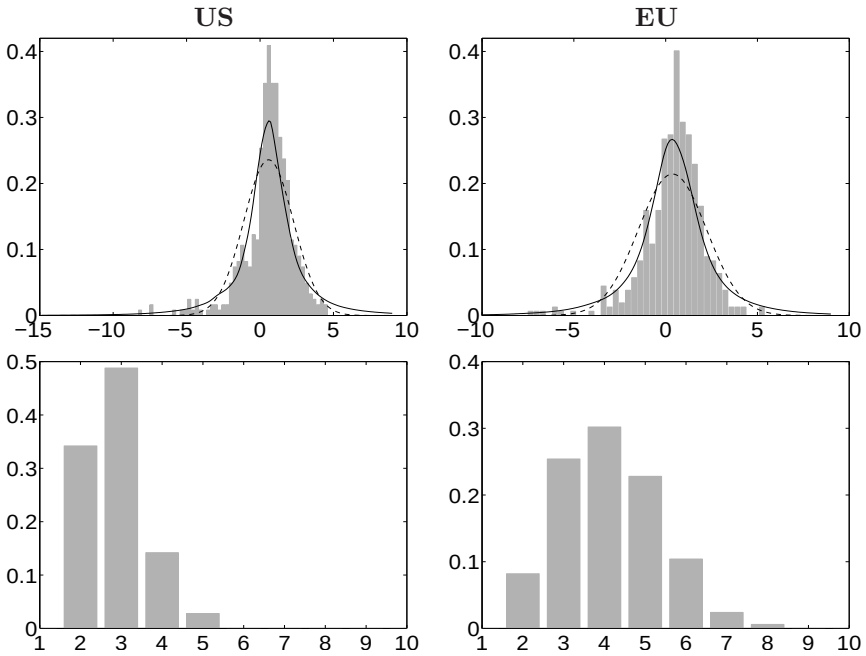

FIGURE 6. First row: IPI log-changes (histogram), predictive distribution (solid line) and best normal (dashed line) for the US (left column) and the EU (right column). Second row: posterior distribution of the number of clusters for the US (left column) and the EU (right column).

on the posterior pairwise probabilities of joint classification $P\{D_{is} = D_{it}|Y\}$. To estimate this matrix, one can use the following pairwise probability matrix:

$$P_{i,st} = \frac{1}{M}\sum_{l=1}^{M} \delta_{D_{is}^l}(D_{it}^l)$$

that is estimated by using every pair of allocation variables D_{is}^l D_{it}^l, with $s, t = 1, \ldots, T$ and over all the $l = 1, \ldots, M$ MCMC iterations.

Fig. 7 shows the pairwise posterior probabilities $P_{i,st}$, $i = 1, 2$, of the US and the EU data for $s, t \in \{1, \ldots, T\}$. The first row (second row) shows the posterior probabilities that two observations of the US cycle (EU cycle) belong to the same cluster. In the first column, one can detect the presence of vertical and horizontal dark gray bands. They correspond to observations that do not cluster frequently together with other observations. A similar remark is true for the light gray areas. In the second column of Fig. 7, one can see the different behavior of the clustering for the US and the EU during the 2009 crisis. As the observations about January 2009 group together in a cluster which excludes other observations of the sample, we can identify this component of the mixture as the one associated to the 2009 recession. The same interpretation of the implied data clustering and the identification of the components as the fourth regime for the 2009 recession can be also achieved through the least square marginal clustering $D_{i,LS}$, that is the clustering $D_i^{l_i} = (D_{i1}^{l_i}, \ldots, D_{iT}^{l_i})$ sampled at the l_i-th iteration which minimizes the sum of squared deviations from the pairwise posterior probability:

$$l_i = \underset{l\in\{1,\ldots,M\}}{\arg\min} \sum_{t=1}^{T}\sum_{s=1}^{T} \left(\delta_{D_{is}^l}(D_{it}^l) - P_{i,st}\right)^2.$$

As regard to the predictive densities generated by the DPY model, we observe that they have long left tails (solid lines in Fig. 6), fatter than the tails of the best normal (dashed lines in the

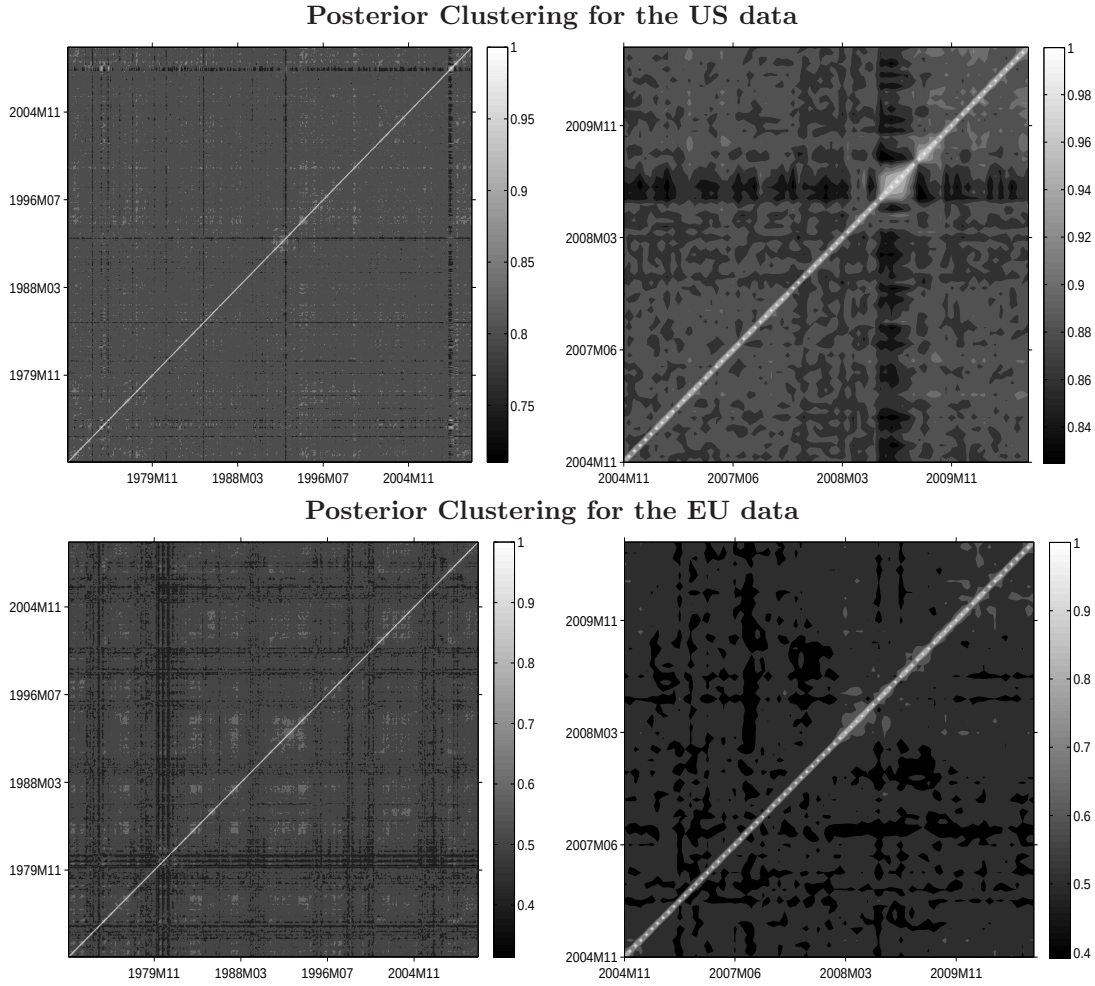

FIGURE 7. Pairwise posterior probabilities for the clustering of the US data $P_{1,st}$ and the EU data $P_{2,st}$ for $s, t \in \{1, \ldots, T\}$

same figure). Thanks to these features the DPY model is suitable for describing and predicting these data.

Fig. 8 shows the posterior predictive densities for $Y_{1,it}$, $i = 1, 2$, estimated on the whole set of data and evaluated sequentially over time at the current values of the explanatory variables $Y_{it-1}, \ldots, Y_{it-p}$ $(i = 1, 2)$. In the same plot, the gray area represents the heatmap sequence of the 95% high probability density region of the predictive densities (darker colours represent higher density values).

Fig. 9 shows a typical predictive density approximation of the US (left) and the EU (right) growth rates from our model during expansion (e.g., observation $t = 430$) and recession (e.g., observation $t = 450$) periods, where expansion and recession have been identified by applying the BB rule for business cycle classification. The posterior predictives during expansion phases exhibit skewness with right tails slightly fatter than left tails. During the recession period the predictives exhibit multimodality with at least two modes in the negative half. The main differences between the two cycles are that EU cycle exhibit more pronounced modes in the negative half, say around -11 and -5, and that the volatility of the mixture components is higher.

In order to asses the predictive ability of the proposed β_2-DPY(ψ, H_0) two-country panel VAR model, we compare it with other non–parametric specifications: two unit-specific and independent Dirichlet process priors, the dependent Dirichlet process priors, $\beta_i -$ DD(ψ, H_0), $i = 1, 2$

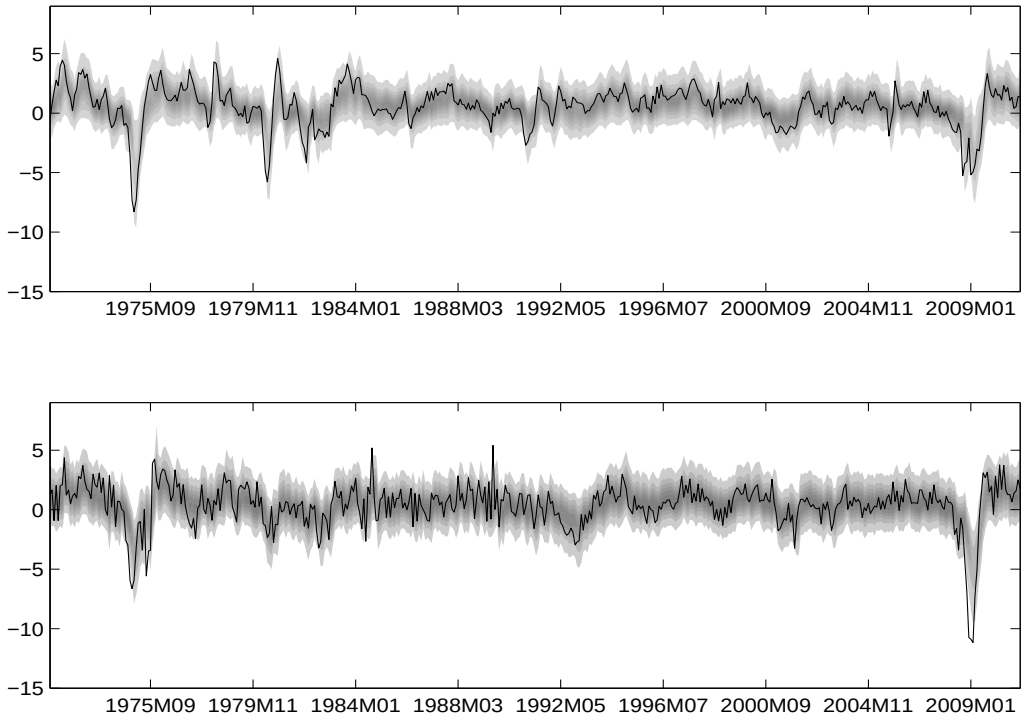

FIGURE 8. US (top) and EU (bottom) IPI growth rates (black lines) and heatmap (gray area) of the 99% high probability density region of the predictive density functions (darker colours represent higher density values) evaluated at each time point, for $t = 1, \ldots, T$, at the values of the predictors $Y_{t-1}, \ldots, Y_{t-p}$, for $i = 1, 2$.

and the dependent Pitman-Yor process prior, $\beta_1 - \text{DPY}(\psi, H_0)$. We apply, as in the previous section, the model pooling approach of Geweke and Amisano [2010] based on log predictive score. We consider two alternative set of predictive densities: the joint US-EU IPI density $f(Y_{1,1t}, Y_{1,2t}|Y_1, \ldots, Y_{t-1}, M_j)$, $j = 1, \ldots, J$, and the joint US-EU and IPI-TS predictive density $f(Y_t|Y_1, \ldots, Y_{t-1}, M_j)$, $j = 1, \ldots, J$, where $Y_t = (Y'_{1t}, Y'_{2t})'$. In the previous predictive densities, M_j indicates the following models: independent Dirichlet ($j = 1$), $\beta_1 - \text{DD}(\psi, G_0)$ ($j = 2$), $\beta_2 - \text{DD}(\psi, G_0)$ ($j = 3$), $\beta_1 - \text{DPY}(\psi, G_0)$ ($j = 4$) and $\beta_2 - \text{DPY}(\psi, G_0)$ ($j = 5$). We define the combined predictive and the log pooled predictive score function as in the previous section. We approximated the model-specific predictive densities by using Gibbs sampling scheme given in Appendix B. We evaluate the log pooled predictive score function on two different subsets of the data with the aim to check the effect of the 2009 crisis on model selection. More specifically, for the first sub-sample we consider observations from August 1975 to May 1996 and set $\tau_1 = 50$ and $\tau_2 = 200$ in the pooled score function, while for the second sub-sample we consider the time period January 1979-December 2010 (i.e., we set $\tau_1 = 50$ and $\tau_2 = 474$ in the pooled score function). We follow the same procedure of the previous section to obtain a sequence of predictive densities and an optimal combination weight for each model. The results in Tab. 2 show that for the two subsamples the DD and DPY models have the better predictive ability in terms of both US-EU IPI and US-EU IPI and TS joint predictive densities. As regard to the choice of the model within the class of vector of dependent Dirichlet, we found that in the first subsample $\beta_1 - \text{DD}(\psi, H_0)$ allows for a better modelling of the US-EU IPI joint predictive (see Panel (a) of Tab. 2) while on the whole sample $\beta_2 - \text{DD}(\psi, H_0)$ has better predictive ability. Finally, we note that the $\beta_1 - \text{DPY}(\psi, H_0)$ VAR for the joint prediction of IPI and TS, for US and EU, has the highest combination weight on the whole set of data.

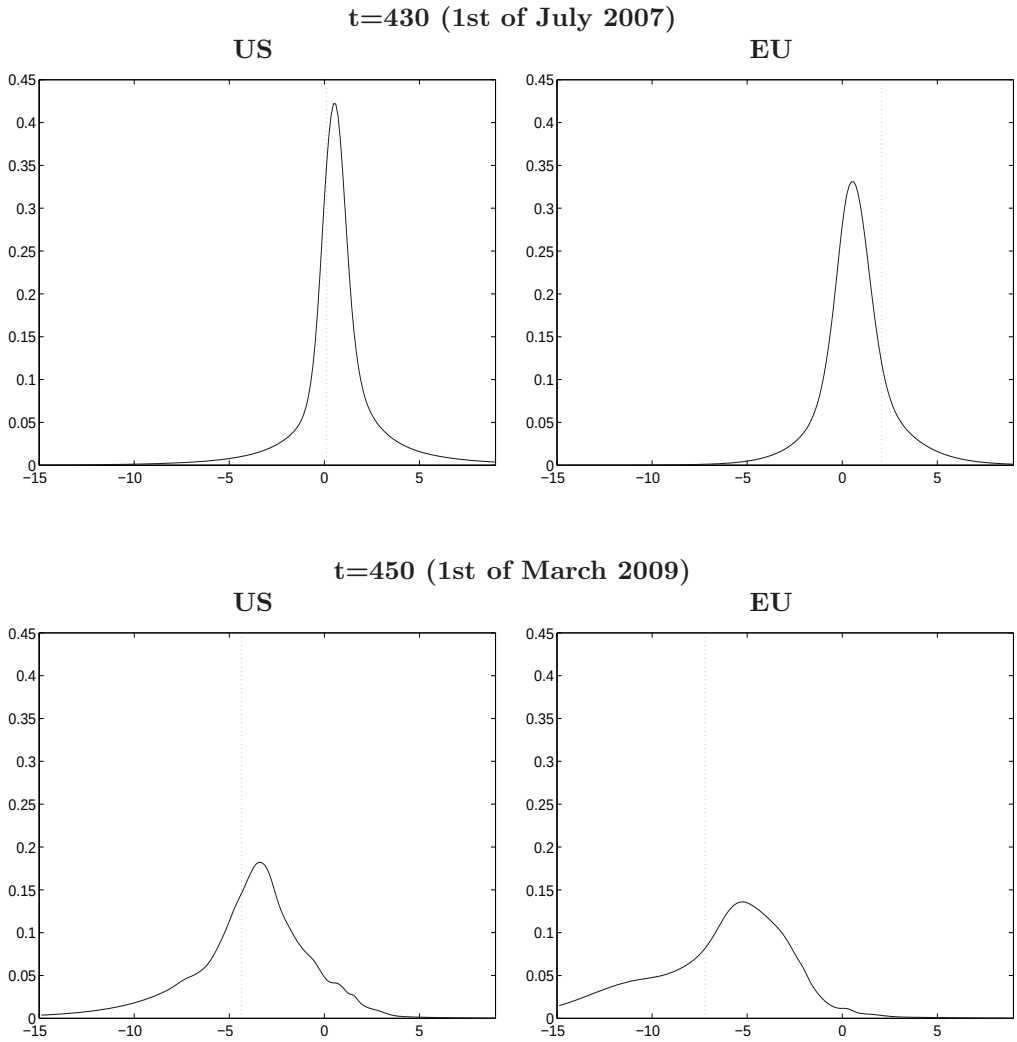

FIGURE 9. Predictive density approximations of the US (left column) and the EU (right column) growth rates $Y_{1,1t}$ and $Y_{1,2t}$ during expansion ($t = 430$) and recession ($t = 450$) periods.

In conclusion, the evidence is in favour of the use of both DD and DPY priors in VAR models to capture multiple breaks or regimes in the parameters of the model. From our experiments, the DPY is the prior specification for a multi-unit panel VAR that produces the best results in terms of prediction. The assumption H1 should be preferred to H2 when no prior information is available on the number of clusters for the different units of the panel. In our application both assumptions H1 and H2 lead to models with good predictive abilities. In the analysis of the EU and US IPI growth rates one can expect a priori, from the existing studies, that the number of clusters differs across the two countries and that the number of clusters for the EU is larger. Thus, the elicitation of a β_2-DPY prior can easily motivated in this example. On the whole sample, the β_2-DPY prior produces better prediction of the IPI growth rates than a β_1-DPY prior.

7. Conclusions

We define beta-product dependent Pitman-Yor processes with a hierarchical specification of the base measure, which allows for heterogeneous clustering effects and for information pooling

TABLE 2. In each row: the predictive considered (first column), the optimal pooled log-predictive score function(second column) and the optimal weights (remaining columns) for Independent Dirichlet (w_1), β_1-DD (w_2), β_2-DD (w_3), β_1-DPY (w_4) and β_2-DPY (w_5), for the periods 1975M08-1996M05 (panel (a)) and 1975M08-2010M12 (panel (b))

Panel (a)						
Joint Predictives	Log-score	w_1	w_2	w_3	w_4	w_5
US-EU IPI	-169.72	0.0000	0.5263	0.0000	0.3619	0.1118
US-EU IPI and TS	-350.91	0.0000	0.3203	0.0000	0.6797	0.0000

Panel (b)						
Joint Predictives	Log-score	w_1	w_2	w_3	w_4	w_5
US-EU IPI	-453.44	0.0004	0.1050	0.0000	0.3302	0.5644
US-EU IPI and TS	-912.52	0.0000	0.2512	0.0000	0.6275	0.1213

across different groups of time series. We use the proposed process to define new Bayesian non–parametric models for repeated measurement and multi-unit VAR models. We provide efficient Monte Carlo Markov Chain algorithms for posterior computation and show the effectiveness of the algorithm through a comparison, on simulated dataset, of various models in the proposed class of Pitman-Yor processes. We show the potentiality of our stick-breaking process prior by providing an original application to the joint analysis of the US and the EU business cycles. We compare independent Dirichlet processes and different specifications of the dependent Pitman-Yor process priors and found that the Pitman-Yor class leads to Bayesian non–parametric models with better prediction abilities. Moreover we found that they can capture some important features of the EU and US business cycles.

Appendix A. Proofs

Proof of Proposition 3. First of all observe that

$$\begin{aligned}\mathbb{E}[G_i(A)G_j(B)] &= \sum_{h\geq 1,k\geq 1}\mathbb{E}[\mathbb{I}_A(\tilde{\vartheta}_{ik})\mathbb{I}_B(\tilde{\vartheta}_{jh})]\mathbb{E}[W_{ik}W_{jh}] \\ &= G_{0,i}(A)G_{0,j}(B)\sum_{h\geq 1,k\geq 1,h\neq k}\mathbb{E}[W_{ik}W_{jh}] + G_{0,ij}(A\times B)\sum_{h\geq 1}\mathbb{E}[W_{ih}W_{jh}].\end{aligned} \tag{38}$$

Now note that

$$\begin{aligned}\sum_{h\geq 1}\mathbb{E}[W_{ih}W_{jh}] &= \sum_{h\geq 1}\mathbb{E}\left[S_{1h}S_{2h}\prod_{m\leq h-1}(1-S_{1m})(1-S_{2m})\right] \\ &= \frac{\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}\end{aligned} \tag{39}$$

and

$$\begin{aligned}\sum_{h\geq 1,k\geq 1,h\neq k}\mathbb{E}[W_{ik}W_{jh}] &= \sum_{h\neq k}\mathbb{E}\left[S_{1h}S_{2k}\prod_{m\leq h-1}(1-S_{1m})\prod_{l\leq k-1}(1-S_{2l})\right] \\ &= \frac{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-2\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}.\end{aligned} \tag{40}$$

Combining (38) with (39)-(40) it follows that

$$\begin{aligned}\mathbb{E}[G_i(A)G_j(B)] = G_{0,ij}(A\times B\times)&\frac{\mathbb{E}[S_{i1}S_{j1}]}{\mathbb{E}[S_{i1}]+\mathbb{E}[S_{j1}]-\mathbb{E}[S_{i1}S_{j1}]} \\ +G_{0i}(A)G_{0j}(B)&\frac{\mathbb{E}[S_{i1}]+\mathbb{E}[S_{j1}]-2\mathbb{E}[S_{i1}S_{j1}]}{\mathbb{E}[S_{i1}]+\mathbb{E}[S_{j1}]-\mathbb{E}[S_{i1}S_{j1}]}.\end{aligned}$$

Since $\mathbb{E}[G_i(\cdot)] = G_{0i}(\cdot)$, $i = 1,\ldots,r$, one gets

$$Cov[G_i(A),G_j(B)] = \frac{\mathbb{E}[S_{i1}S_{j1}]}{\mathbb{E}[S_{i1}]+\mathbb{E}[S_{j1}]-\mathbb{E}[S_{i1}S_{j1}]}[G_{0,ij}(A\times B)-G_{0i}(A)G_{0j}(B)]. \tag{41}$$

In a similar way, for every $i = 1,\ldots,r$, one has

$$Var[G_i(A)] = G_{0i}(A)(1-G_{0i}(A))\frac{\mathbb{E}[S_{i1}^2]}{2\mathbb{E}[S_{i1}]-\mathbb{E}[S_{i1}^2]}. \tag{42}$$

Hence,

$$Cor(G_i(A),G_j(A)) = C_{ij}\frac{G_{0ij}(A\times B)-G_{0i}(A)G_{0j}(B)}{\sqrt{G_{0i}(A)(1-G_{0i}(A))G_{0j}(B)(1-G_{0j}(B))}}$$

with

$$C_{ij} = \frac{\mathbb{E}[S_{i1}S_{j1}]}{1-\mathbb{E}[(1-S_{i1})(1-S_{j1})]}\sqrt{\frac{(2\mathbb{E}[S_{i1}]-\mathbb{E}[S_{i1}^2])(2\mathbb{E}[S_{j1}]-\mathbb{E}[S_{j1}^2])}{\mathbb{E}[S_{j1}^2]\mathbb{E}[S_{j1}^2]}}.$$

Assuming (14), one gets

$$\begin{aligned}\mathbb{E}(S_{i1}) &= \frac{1}{1+\alpha_1+\alpha_2}, \quad \mathbb{E}(S_{i1}^2) = \frac{2}{(1+\alpha_1+\alpha_2)(2+\alpha_1+\alpha_2)} \\ \mathbb{E}(S_{i1}S_{j1}) &= \frac{2(1+\alpha_1)}{(2+\alpha_1)(1+\alpha_1+\alpha_2)^2}\end{aligned}$$

for every $i,j = 1,\ldots,r$ and then,

$$C_{ij} = \frac{(1+\alpha_1+\alpha_2)(1+\alpha_1)}{(1+\alpha_1)(1+\alpha_1+\alpha_2)+\alpha_2}.$$

Let us assume now (15). For the sake of simplicity write V_k in place of V_{k1}. Recall that $V_0 \sim Beta(1,\alpha_1)$ and, for $1 \le k \le r-1$, $V_k \sim Beta(1+\alpha_1+\cdots+\alpha_k, \alpha_{k+1})$. Let $1 \le i < j \le r$. Since $S_{i1} = V_0 V_1 \dots V_{r-i}$, one gets

$$S_{i1}S_{j1} = V_0^2 V_1^2 \dots V_{r-j}^2 V_{r-j+1} \dots V_{r-i}.$$

After some computations, using also the fact that the V_js are independent, one obtains

$$\mathbb{E}[S_{i1}S_{j1}] = \frac{2}{(2+\alpha_1+\cdots+\alpha_{r-j+1})(1+\alpha_1+\cdots+\alpha_{r-i+1})}$$

and

$$\mathbb{E}[S_{i1}] = \frac{1}{1+\alpha_1+\cdots+\alpha_{r-i+1}}, \quad \mathbb{E}[S_{i1}^2] = \frac{2}{(1+\alpha_1+\cdots+\alpha_{r-i+1})(2+\alpha_1+\cdots+\alpha_{r-i+1})}.$$

At this stage, simple algebra gives

$$\sqrt{\Big(\frac{2\mathbb{E}[S_{i1}]}{\mathbb{E}[S_{i1}^2]}-1\Big)\Big(\frac{2\mathbb{E}[S_{j1}]}{\mathbb{E}[S_{j1}^2]}-1\Big)} = \sqrt{(1+\alpha_1+\cdots+\alpha_{r-i+1})(1+\alpha_1+\cdots+\alpha_{r-j+1})}$$

and

$$\frac{\mathbb{E}[S_{i1}S_{j1}]}{\mathbb{E}[S_{i1}]+\mathbb{E}[S_{j1}]-\mathbb{E}[S_{i1}S_{j1}]} = \frac{2(1+\alpha_1+\cdots+\alpha_{r-j+1})}{2(1+\alpha_1+\cdots+\alpha_{r-j+1})^2+(2+\alpha_1+\cdots+\alpha_{r-j+1})(\alpha_{r-j+2}+\cdots+\alpha_{r-i+1})}.$$

That is

$$C_{i,j} = \frac{2\sqrt{(1+\alpha_1+\cdots+\alpha_{r-i+1})}(1+\alpha_1+\cdots+\alpha_{r-j+1})^{\frac{3}{2}}}{2(1+\alpha_1+\cdots+\alpha_{r-j+1})^2+(2+\alpha_1+\cdots+\alpha_{r-j+1})(\alpha_{r-j+2}+\cdots+\alpha_{r-i+1})}.$$

Formula (17) can be proved in an analogous way. □

Proof of (25)-(26). The joint distribution of $[V,\tilde\vartheta,U,D,Y,\tilde\psi]$ is

$$\begin{aligned} &P\{V\in dv,\tilde\vartheta\in d\vartheta, Y\in dy, U\in du, D=d,\tilde\psi\in(d\alpha_1,d\alpha_2,dl)\} \\ &= \Big[\prod_{i=1,2}\prod_{t=1}^{T_i}\mathbb{I}\{u_{it}<w_{i,d_{i,t}}\}\mathcal{K}_t(y_{i,t}|\vartheta_{d_{i,t}},z_t)\Big]dydu \\ &\qquad \times_{k\ge1}\Big[P\{V_k\in dv_k|\tilde\psi=(\alpha_1,\alpha_2,l)\}G_0(d\vartheta_k)\Big]P\{\tilde\psi\in(d\alpha_1,d\alpha_2,dl)\} \end{aligned} \tag{43}$$

where $w_{i,k} = v_{0k}v_{ik}\prod_{t<k}(1-v_{0t}v_{it})$, with the convenction that $v_{2k}=1$, for every k, under (H_2) and $z_t=[y_{js}: j=1,2; s=1,\dots,t-1]$. From (43) one gets

$$\begin{aligned} &P\{V\in dv,\tilde\psi\in(d\alpha_1,d\alpha_2 l)|Y,\tilde\vartheta,D\} \\ &\propto \Big[\prod_{i=1,2}\prod_{t=1}^{T_i} w_{i,D_{it}}\Big]\times_{t\ge1}P\{V_t\in dv_t|\tilde\psi=(\alpha_1,\alpha_2,l)\}P\{\tilde\psi\in(d\alpha_1,d\alpha_2,dl)\}. \end{aligned}$$

Now note that

$$\begin{aligned} \prod_{i=1,2}\prod_{t=1}^{T_i} w_{i,D_{it}} = &\prod_{t=1}^{D^*} v_{0t}^{A_{1t}+A_{2t}} v_{1t}^{A_{1t}} v_{2t}^{A_{2t}} \\ &(1-v_{0t}v_{1t})^{B_{1t}}(1-v_{0t}v_{2t})^{B_{2t}}. \end{aligned}$$

□

Appendix B. Computational details

In the Block Gibbs Sampler described in Section 5 in principle one needs to sample an infinite number of V_k and $\tilde\vartheta_k$. But in order to proceed with the chain it suffices to sample a finite number of V_ks to check condition (32) and the finite number of $\tilde\vartheta_k$ to be used in (31).

B.1. **Full conditionals for DPY(ψ, G_0) mixtures of Gaussian of Section 6.1.** For the sake of simplicity we will omit indicating the dependence of the full conditional on the hyperparameters. In order to sample from the full-conditional $P\{\tilde{\vartheta}_k \in d\vartheta_k | V, D, Y, U\}$, for $k \geq 1$ we consider a Gibbs sampler with normal and inverse gamma full conditional distributions

$$\begin{aligned} &P\{\tilde{\mu}_{ik} \in d\mu_{ik} | \tilde{\mu}_{0k}, \tilde{\sigma}_{ik}^2, \tilde{\sigma}_0^2, D, Y\} \propto \\ &\quad \propto \exp\left\{-\frac{1}{2s_i^{-2}}\mu_{ik}^2\right\} \prod_{t \in \mathcal{D}_{i,k}} \exp\left\{-\frac{1}{2\tilde{\sigma}_{ik}^2\tilde{\sigma}_{0k}^2}(Y_{it} - (\tilde{\mu}_{0k} + \mu_{ik}))^2\right\} d\mu_{ik} \\ &\quad \propto \exp\left\{-\frac{1}{2}\mu_{ik}^2\left(s_i^2 + \tilde{\sigma}_{ik}^{-2}\tilde{\sigma}_{0k}^{-2}A_{i,k}\right) + \mu_{ik}\tilde{\sigma}_{ik}^{-2}\tilde{\sigma}_{0k}^{-2}\eta_{ik}^{(1)}\right\} d\mu_{ik} \\ &\quad \propto \mathcal{N}\left(\frac{\tilde{\sigma}_{ik}^{-2}\tilde{\sigma}_{0k}^{-2}\eta_{ik}^{(1)}}{s_i^2 + \tilde{\sigma}_{ik}^{-2}\tilde{\sigma}_{0k}^{-2}A_{i,k}}, \frac{1}{s_i^2 + \tilde{\sigma}_{ik}^{-2}\tilde{\sigma}_{0k}^{-2}A_{i,k}}\right) \end{aligned} \tag{44}$$

and

$$\begin{aligned} &P\{\tilde{\sigma}_{ik}^2 \in d\sigma_{ik}^2 | \tilde{\mu}_{0k}, \tilde{\mu}_{ik}, \tilde{\sigma}_{0k}^2, D, Y\} \propto \\ &\quad \propto \exp\left\{-\frac{\lambda}{2}\sigma_{ik}^{-2}\right\} (\sigma_{ik}^{-2})^{\frac{1}{2}\lambda+1} \prod_{t \in \mathcal{D}_{i,k}} (\sigma_{ik}^{-2})^{1/2} \exp\left\{-\frac{1}{2\sigma_{ik}^2\tilde{\sigma}_{0k}^2}(Y_{it} - (\tilde{\mu}_{0k} + \tilde{\mu}_{ik}))^2\right\} d\sigma_{ik}^2 \\ &\quad \propto \exp\left\{-\left(\frac{\lambda}{2} + \frac{1}{2\tilde{\sigma}_{0k}^2}\eta_{ik}^{(2)}\right)\sigma_{ik}^{-2}\right\} (\sigma_{ik}^{-2})^{\frac{1}{2}\lambda + \frac{1}{2}A_{i,k}+1} d\sigma_{ik}^2 \\ &\quad \propto \mathcal{IG}\left(\frac{\lambda}{2} + \frac{1}{2}A_{i,k}, \frac{\lambda}{2} + \frac{1}{2\tilde{\sigma}_{0k}^2}\eta_{ik}^{(2)}\right) \end{aligned} \tag{45}$$

$i = 1, 2$, for the component-specific part of the atoms, where $A_{i,k}$, $i = 1, 2$ have been defined in Section 5 and

$$\eta_{ik}^{(1)} = \sum_{t \in \mathcal{D}_{i,k}} (Y_{it} - \tilde{\mu}_{0k}) \quad \text{and} \quad \eta_{ik}^{(2)} = \sum_{t \in \mathcal{D}_{i,k}} \left(Y_{it} - (\tilde{\mu}_{0k} + \tilde{\mu}_{ik})\right)^2. \tag{46}$$

The part of the atom that is common to all the components has full conditionals

$$\begin{aligned} &P\{\tilde{\mu}_{0k} \in d\mu_{0k} | \tilde{\mu}_{1k}, \tilde{\mu}_{2k}, \tilde{\sigma}_{0k}^2, \tilde{\sigma}_{1k}^2, \tilde{\sigma}_{2k}^2, D, Y\} \propto \\ &\quad \propto \exp\left\{-\frac{1}{2s_0^{-2}}\mu_{0k}^2\right\} \prod_{i=1,2} \prod_{t \in \mathcal{D}_{i,k}} \exp\left\{-\frac{1}{2\tilde{\sigma}_{ik}^2\tilde{\sigma}_{0k}^2}(Y_{it} - (\mu_{0k} + \tilde{\mu}_{ik}))^2\right\} d\mu_{0k} \\ &\quad \propto \mathcal{N}\left(\frac{\eta_k^{(0)}}{s_0^2 + \sum_{i=1,2} \tilde{\sigma}_{0k}^{-2}\tilde{\sigma}_{ik}^{-2}A_{ik}}, \frac{1}{s_0^2 + \sum_{i=1,2} \tilde{\sigma}_{0k}^{-2}\tilde{\sigma}_{ik}^{-2}A_{ik}}\right) \end{aligned} \tag{47}$$

with

$$\eta_k^{(0)} = \sum_{i=1,2} \tilde{\sigma}_{0k}^{-2}\tilde{\sigma}_{ik}^{-2} \sum_{t \in \mathcal{D}_{i,k}} (Y_{it} - \tilde{\mu}_{ik}) \tag{48}$$

and

$$\begin{aligned} &P\{\tilde{\sigma}_{0k}^2 \in d\sigma_{0k}^2 | \tilde{\sigma}_{1k}^2, \tilde{\sigma}_{2k}^2, \tilde{\mu}_0, \tilde{\mu}_1, \tilde{\mu}_2, D, Y\} \propto \\ &\quad \propto \exp\left\{-\frac{\varepsilon}{2}\sigma_{0k}^{-2}\right\} (\sigma_{0k}^{-2})^{\frac{\varepsilon}{2}+1} \prod_{i=1,2} \prod_{t \in \mathcal{D}_{i,k}} (\sigma_{0k}^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\tilde{\sigma}_{ik}^2\sigma_{0k}^2}(Y_{it} - (\tilde{\mu}_{0k} + \tilde{\mu}_{ik}))^2\right\} d\sigma_{0k}^2 \\ &\quad \propto \mathcal{IG}\left(\frac{\varepsilon}{2} + \frac{1}{2}\sum_{i=1,2} A_{ik}, \frac{\varepsilon}{2} + \frac{1}{2}\sum_{i=1,2} \frac{\eta_{ik}^{(2)}}{\tilde{\sigma}_{ik}^2}\right) \end{aligned} \tag{49}$$

respectively.

A sample from the conditional joint distribution of the precision parameters and the stick-breaking elements can be obtained following the blocking scheme described in Subsection 5.2.

Since we assume gamma priors, $\mathcal{G}(\zeta_{11}, \zeta_{21})$ and $\mathcal{G}(\zeta_{12}, \zeta_{22})$ for α_1 and α_2 respectively, and standard uniform prior for l, equations (28)-(29) become

$$
\begin{aligned}
&P\{\tilde{\psi} \in (d\alpha_1, d\alpha_2, dl)|V^*, D\} \propto \\
&\frac{1}{C(\psi)}\alpha_1^{\zeta_{11}-1}\alpha_2^{\zeta_{12}-1}\exp\left\{-\alpha_1\bar{\zeta}_{21} - \alpha_2\bar{\zeta}_{22} - l\bar{\zeta}_{32}\right\} d\alpha_1 d\alpha_2 dl
\end{aligned} \tag{50}
$$

where

$$
\begin{aligned}
\bar{\zeta}_{21} &= \zeta_{21} - \sum_{k\in\mathcal{D}^*}\Big(\log(1-V_{0k}) + \sum_{i=1,2}\log V_{ik}\Big) \\
\bar{\zeta}_{22} &= \zeta_{22} - \sum_{k\in\mathcal{D}^*}\sum_{i=1,2}\log(1-V_{ik}) \\
\bar{\zeta}_{32} &= \sum_{k\in\mathcal{D}^*}\Big(\log V_{0k} + \sum_{i=1,2}(\log V_{ik} - k\log(1-V_{ik}))\Big) \\
C(\psi) &= \prod_{k\in\mathcal{D}^*} B(1-l,\alpha_1)B^2(\alpha_1+1-l,\alpha_2+lk)
\end{aligned} \tag{51}
$$

under H1, and

$$
\begin{aligned}
\bar{\zeta}_{21} &= \zeta_{21} - \sum_{k\in\mathcal{D}^*}\Big(\log(1-V_{0k}) + \log(V_{1k})\Big) \\
\bar{\zeta}_{22} &= \zeta_{22} - \sum_{k\in\mathcal{D}^*}\log(1-V_{1k}) \\
\bar{\zeta}_{32} &= \sum_{k\in\mathcal{D}^*}\Big(\log(V_{0k}) - k\log(1-V_{0k}) - (k-1)\log(V_{1k})\Big) \\
C(\psi) &= \prod_{k\in\mathcal{D}^*} B(1-l,\alpha_1+lk)B(\alpha_1+1+l(k-1),\alpha_2)
\end{aligned} \tag{52}
$$

under H2.

We simulate from the full conditional by a M.-H. step. We considered a Gaussian random walk proposal for the transformed parameter vector $\xi = (\xi_1, \xi_2, \xi_3) = g(\psi)$ with $g(\psi) = (\log(\alpha_1), \log(\alpha_2), \log(l/(1-l)))$. At the j-th iteration, given $\psi^{(j-1)}$, we simulate

$$
\xi^{(*)} \sim \mathcal{N}_3(\xi^{(j-1)}, \kappa^2 I_3) \tag{53}
$$

where $\xi^{(j-1)} = g(\psi^{(j-1)})$ and κ^2 represents a scale parameter of random walk. The proposal $\psi^{(*)} = g^{-1}(\xi^{(*)})$, with g^{-1} inverse transform of g, is accepted with probability

$$
\min\left\{1, \frac{P\{\psi^{(*)}|V^*, D\}}{P\{\psi^{(j-1)}|V^*, D\}}\frac{|\nabla g^{-1}(\psi^{(*)})|}{|\nabla g^{-1}(\psi^{(j-1)})|}\right\} \tag{54}
$$

where ∇g^{-1} denotes the Jacobian of g^{-1}, that is a matrix with main diagonal

$$
\left(\exp(\xi_1), \exp(\xi_2), \frac{\exp(\xi_3)}{(1+\exp(\xi_3))^2}\right) \tag{55}
$$

and null elements out of the main diagonal and $|\nabla g^{-1}|$ its determinant. We set the scale parameter κ in order to have acceptance rates close to 0.5.

B.2. **Full conditionals for DPY(ψ, G_0) mixtures of VAR of Section 6.2.** In order to sample from the full-conditional $P\{\tilde{\vartheta}_k \in d\vartheta_k|V, D, Y, U\}$, for $k \geq 1$, we use Gibbs sampling. The hierarchical structure of the base measure allows us to write the following Gibbs sampling updating scheme for the elements of $\tilde{\vartheta}_k$.

The full conditional of the country-specific VAR coefficients for the i-th unit is

$$\begin{aligned}
&P\{\tilde{\phi}_{ik} \in d\phi_{ik} | \tilde{\phi}_{0k}, \tilde{\Sigma}_{ik}, \tilde{\sigma}^2_{0k}, \tilde{\Upsilon}^2_{ik}, D, Y\} \propto \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big(\sum_{t\in\mathcal{D}_{i,k}} \Big(E_{0ikt} - (I_2 \otimes X_t')\phi_{ik}\Big)' \tilde{\Sigma}^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} \Big(E_{0ikt} - (I_2 \otimes X_t')\phi_{ik}\Big) \\
&\qquad + \phi_{ik}' \tilde{\Upsilon}^{-1}_{ik} \phi_{ik} \Big)\Big\} d\phi_{ik} \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big(\phi_{ik}' \Big(\sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} (I_2 \otimes X_t') + \tilde{\Upsilon}^{-1}_{ik} \Big) \phi_{ik} + \\
&\qquad - 2 \sum_{t\in\mathcal{D}_{i,k}} \phi_{ik}' \Big((I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} E_{0ikt} \Big) \Big) \Big\} d\phi_{ik} \\
&\quad \propto \mathcal{N}_m\left(m_{ik}, M_{ik}\right)
\end{aligned}$$

where $E_{0ikt} = Y_{it} - (I_2 \otimes X_t')\tilde{\phi}_{0k}$, $M^{-1}_{ik} = \tilde{\sigma}^{-2}_{0k} \sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} (I_2 \otimes X_t') + \tilde{\Upsilon}^{-1}_{ik}$ and $m_{ik} = M_{ik} \tilde{\sigma}^{-2}_{0k} \sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} E_{0ikt}$.

The country-specific variance matrix of the VAR equations for the i-th unit has full conditional

$$\begin{aligned}
&P\{\tilde{\Sigma}^{-1}_{ik} \in d\Sigma^{-1}_{ik} | \tilde{\phi}_{ik}, \tilde{\phi}_{0k}, \tilde{\sigma}^2_{0k}, \tilde{\Upsilon}^2_{ik}, D, Y\} \propto \\
&\quad \propto \exp\Big\{ -\frac{1}{2} \sum_{t\in\mathcal{D}_{i,k}} \mathrm{tr}\Big(E_{ikt}' \Sigma^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} E_{ikt} \Big) \Big\} |\Sigma_{ik}|^{-\frac{A_{ik}}{2}} |\Sigma^{-1}_{ik}|^{\frac{\lambda-3}{2}} \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big(\Lambda^{-1} \Sigma^{-1}_{ik} \Big) \Big\} d\Sigma_{ik} \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big(\Big(\Lambda^{-1} + \tilde{\sigma}^{-2}_{0k} \sum_{t\in\mathcal{D}_{i,k}} E_{ikt} E_{ikt}' \Big) \Sigma^{-1}_{ik} \Big) \Big\} |\Sigma^{-1}_{ik}|^{\frac{A_{ik}+\lambda-3}{2}} d\Sigma_{ik} \\
&\quad \propto \mathcal{W}_2\Big(\lambda + A_{ik}, \Lambda_{i,T} \Big)
\end{aligned}$$

where $A_{i,k} = card(\mathcal{D}_{i,k})$, $E_{ikt} = Y_{it} - (I_2 \otimes X_t')(\tilde{\phi}_{0k} + \tilde{\phi}_{ik})$ and $\Lambda^{-1}_{i,T} = \Lambda^{-1} + \tilde{\sigma}^{-2}_{0k} \sum_{t\in\mathcal{D}_{i,k}} E_{ikt} E_{ikt}'$.

The full conditionals of the common components $\tilde{\phi}_{0k}$ and σ^{-2}_{0k} are functions of the unit-specific components. More specifically $\tilde{\phi}_{0k}$ has full conditional

$$\begin{aligned}
&P\{\tilde{\phi}_{0k} \in d\phi_{0k} | \tilde{\phi}_{ik}, \tilde{\Sigma}_{ik}, \tilde{\sigma}^2_{0k}, \tilde{\Upsilon}^2_{ik}, D, Y\} \propto \\
&\quad \propto \exp\Big\{ -\frac{1}{2}\mathrm{tr}\Big(\phi_{0k}' \Big(\sum_{i=1,2} \sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} (I_2 \otimes X_t') + \tilde{\Upsilon}^{-1}_{0} \Big) \phi_{0k} + \\
&\qquad - 2 \sum_{i=1,2} \sum_{t\in\mathcal{D}_{i,k}} \phi_{0k}' \Big((I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} \tilde{\sigma}^{-2}_{0k} E_{1ikt} \Big) \Big) \Big\} d\phi_{0k} \\
&\quad \propto \mathcal{N}_m\Big(m_{0k}, M_{0k} \Big)
\end{aligned}$$

where $E_{1ikt} = Y_{it} - (I_2 \otimes X_t')\tilde{\phi}_{ik}$, $M^{-1}_{0k} = \tilde{\sigma}^{-2}_{0k} \sum_{i=1,2} \sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} (I_2 \otimes X_t') + \tilde{\Upsilon}^{-1}_{0}$ and $m_{0k} = M_{0k} \tilde{\sigma}^{-2}_{0k} \sum_{i=1,2} \sum_{t\in\mathcal{D}_{i,k}} (I_2 \otimes X_t) \tilde{\Sigma}^{-1}_{ik} E_{1ikt}$.

Finally the common factor of the variances, σ^{-2}_{0k}, has full conditional

$$\begin{aligned}
&P\{\tilde{\sigma}^{-2}_{0k} \in d\sigma^{-2}_{0k} | \tilde{\phi}_{ik}, \tilde{\Sigma}_{ik}, \tilde{\phi}_{0k}, \tilde{\Upsilon}^2_{ik}, D, Y\} \propto \\
&\quad \propto \exp\Big\{ -\frac{1}{2\sigma^2_{0k}} \sum_{i=1,2} \sum_{t\in\mathcal{D}_{i,k}} \Big(E_{ikt}' \tilde{\Sigma}^{-1}_{ik} E_{ikt} \Big) \Big\} (\sigma^{-2}_{0k})^{\frac{A_{1,k}+A_{2,k}}{2}} (\sigma^{-2}_{0k})^{\frac{\varepsilon-1}{2}} \exp\Big\{ -\frac{\varepsilon}{2}\sigma^{-2}_{0k} \Big\} d\sigma^{-2}_{0k} \\
&\quad \propto \mathcal{G}a\Big(\Big(\varepsilon + \sum_{i=1,2} A_{ik} \Big)/2, \varepsilon_T/2 \Big)
\end{aligned}$$

where $\varepsilon_T = \varepsilon + \sum_{i=1,2}\sum_{t\in\mathcal{D}_{i,k}} E'_{ikt}\tilde{\Sigma}_{ik}^{-1}E_{ikt}$ and

$$
\begin{aligned}
P\{\tilde{\tau}_k^2 \in d\tau_k^2 | \tilde{\phi}_{ik}, \tilde{\Sigma}_{ik}, \tilde{\phi}_{0k}, \tilde{\sigma}_{0k}^{-2}, D, Y\} &\propto \\
&\propto \exp\Big\{ -\frac{1}{2\tau_k^2}\mathrm{tr}\Big(\sum_{i=1,2}\tilde{\phi}'_{ik}\tilde{\phi}_{ik}\Big)\Big\}\tau_k^{-4}\tau_k^{-\nu-1/2}\exp\{-\nu/2\tau_k^{-2}\} \\
&\propto \exp\Big\{ -\frac{1}{2}\Big(\nu + \mathrm{tr}\Big(\sum_{i=1,2}\tilde{\phi}'_{ik}\tilde{\phi}_{ik}\Big)\Big)\tau_k^{-2}\Big\}\tau_k^{-2(2+\nu/2+1)} \\
&\propto \mathcal{IG}\Big(\nu/2+2, \nu/2 + \mathrm{tr}\Big(\sum_{i=1,2}\tilde{\phi}'_{ik}\tilde{\phi}_{ik}\Big)/2\Big)
\end{aligned}
$$

References

M. Billio, R. Casarin, F. Ravazzolo, and H. van Dijk. Combination schemes for turning point predictions. *Quarterly Reviews of Economics and Finance*, forthcoming:1–11, 2012.

G. Bry and C. Boschan. Cyclical Analysis of Time Series: Selected Procedures and Computer Programs. Technical report, NBER Technical Paper 20, 1971.

F. Canova and M. Ciccarelli. Forecasting and turning point prediction in a Bayesian panel VAR model. *Journal of Econometrics*, 120(2):327–359, 2004.

S. Chib and E. Greenberg. Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, 68:339–360, 1995.

Y. Chung and D. B. Dunson. The local Dirichlet process. *Annals of the Institute of Statistical Mathematics*, 63:59–80, 2011.

D. M. Cifarelli and E. Regazzini. Problemi statistici non parametrici in condizioni di scambiabilità parziale: impiego di medie associative. Quaderni Istituto di Matematica Finanziaria, Università di Torino, 12, 1978, 1978.

M. P. Clements and H. M. Krolzig. A comparison of the forecast performances of Markov-switching and threshold autoregressive models of US GNP. *Econometrics Journal*, 1:C47–C75, 1998.

M. J. Crowder and D. J. Hand. *Analysis of Repeated Measures*. Chapman & Hall, 1990.

D. B. Dahl. Model-based clustering for expression data via a Dirichlet process mixture model. In K.-A. Do, P. P. Muller, and M. Vannucci, editors, *Bayesian Inference for Gene Expression and Proteomics*, pages 201–218. Cambrdige University Press, 2006.

M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, 1998.

M. De Iorio, P. Müller, G. Rosner, and S. MacEachern. An ANOVA model for dependent random measures. *Journal of the American Statistical Association*, 99:205–215, 2004.

T. Doan, R. Litterman, and C. A. Sims. Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews*, 3:1–100, 1984.

J.A. Duan, M. Guidani, and A.E. Gelfand. Generalized spatial Dirichlet process models. *Biometrika*, 94:809–825, 2007.

D. B. Dunson and S. D. Peddada. Bayesian nonparametric inference on stochastic ordering. *Biometrika*, 95:859–874, 2008.

D.B. Dunson, Y. Xue, and L. Carin. The Matrix Stick-Breaking Process. *Journal of the American Statistical Association*, 103:317–327, 2008.

I. Epifani and A. Lijoi. Nonparametric priors for vectors of survival functions. *Statist. Sinica*, 20: 1455–1484, 2010.

M. D. Escobar. Estimating normal means with a Dirichlet process prior. *Journal of the American Statistical Association*, 89:268–277, 1994.

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

T. S. Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.

S. Früwirth-Schnatter. *Finite Mixture and Markov Switching Models*. Springer-Verlag, Berlin, 2006.

A. E. Gelfand, A. Kottas, and S. N. MacEachern. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association*, 100:1021–1035, 2004.

J. Geweke and G. Amisano. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, 26(2):216–230, 2010.

T. G. Gneiting and A. E. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102:477:359–378, 2007.

C. W. J. Granger. Invited review combining forecasts twenty years later. *Journal of Forecasting*, 8:167–173, 2006.

J. E. Griffin. Inference in Infinite Superpositions of Non-Gaussian Ornstein-Uhlenbeck Processes Using Bayesian Nonparametic Methods. *Journal of Financial Econometrics*, 1:1–31, 2011.

J. E. Griffin and M. F. J. Steel. Order-based dependent Dirichlet processes. *Journal of the American Statistical Association*, 101:179–194, 2006.

J. E. Griffin and M. F. J. Steel. Stick-breaking autoregressive processes. *Journal of Econometrics*, 162:383–396, 2011.

J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57:357–384, 1989.

S. J. Hatjispyrosa, T. N. Nicoleris, and S. G. Walker. Dependent mixtures of Dirichlet processes. *Computational Statistics & Data Analysis*, 55:2011–2025, 2011.

N. L. Hjort, C. Homes, P. Muller, and S. G. Walker. *Bayesian Nonparametrics*. Cambridge University Press, 2010.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian Model Averaging: A Tutorial. *Statistical Science*, 14:382–417, 1999.

H. Ishwaran and L. F. James. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96:161–173, 2001.

H. Ishwaran and M. Zarepour. Series representations for multivariate generalized gamma processes via a scale invariance principle. *Statistica Sinica*, 19:1665–1682, 2009.

J. M. Jensen and M. J. Maheu. Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics*, 157:306–316, 2010.

M. Kalli, J. E. Griffin, and S. G. Walker. Slice sampling mixture models. *Statistics and Computing*, 21:93–105, 2011.

C. J. Kim and C. J. Murray. Permanent and Transitory Components of Recessions. *Empirical Economics*, 27:163–183, 2002.

C. J. Kim and J. Piger. Common stochastic trends, common cycles, and asymmetry in economic fluctuations. Working paper, n. 681, Federal Reserve Board, September 2000.

S. Kim, M. G. Tadesse, and M. Vannucci. Variable selection in clustering via Dirichlet process mixture models. *Biometrika*, 93:877–893, 2006.

M. Kolossiatis, J. Griffin, and M. F. J. Steel. On Bayesian nonparametric modelling of two correlated distributions. *Statistics and Computing*, forthcoming, 2011.

H.-M. Krolzig. Predicting Markov-Switching Vector Autoregressive Processes. Working Papers, 2000-WP31, Nuffield College Economics, 2000.

F. Leisen and A. Lijoi. Vectors of Poisson Dirichlet processes. *The Journal of Multivariate Analysis*, 102:482–495, 2011.

R. Litterman. Forecasting With Bayesian Vector Autoregressions-Five Years of Experience. *Journal of Business and Economic Statistics*, 4:25–38, 1986.

A. Y. Lo. On a class of Bayesian Nonparametric Estimates: I. Density Estimates. *The Annals of Statistics*, 12:351–357, 1984.

S. N. MacEachern. Dependent Nonparametric Processes. In *In ASA Proceedings of the Section on Bayesian Statistical Science, Alexandria, VA*. American Statistical Association, 1999.

S. N. MacEachern. Decision theoretic aspects of dependent nonparametric processes. In E. George, editor, *Bayesian Methods with Applications to Science, Policy and Official Statistics*, pages 551–560. Creta: ISBA, 2001.

C. Min and A. Zellner. Bayesian and non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates. *Journal of Econometrics*, 56:89–118, 1993.

P. Müller, F. Quintana, and G. Rosner. A method for combining inference across related nonparametric Bayesian models. *Journal of the Royal Statistical Society B*, 66:735–749, 2004.

S. Nadarajah and S. Kotz. Some bivariate beta distributions. *Statistics*, 39:457–466, 2005.

L. E. Nieto-Barajas and S. G. Walker. Gibbs and autoregressive processes. *Statist. Prob. Lett.*, 77:1479–1485, 2007.

I. Olkin and R. Liu. A bivariate beta distribution. *Statistics & Probability Letters*, 62:407–412, 2003.

E. Otranto and G. M. Gallo. A nonparametric Bayesian approach to detect the number of regimes in Markov switching models. *Econometric Reviews*, 21:477–496, 2002.

O. Papaspiliopoulos. A note on posterior sampling from Dirichlet mixture models. Preprint N. 8-20, CRISM, University of Warwick, 2008.

M. L. Pennell and D. B. Dunson. Bayesian semiparametric dynamic frailty models for multiple event time data. *Biometrics*, 62:1044–1052, 2006.

J. Pitman and M. Yor. The two parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Annals of probability*, 25:855–900, 1997.

C. Radhakrishna Rao. On some problems arising out of discrimination with multiple characters. *Sankhya*, 4:343–366, 1949.

A. Rodriguez and E. ter Horst. Bayesian dynamics density estimation. *Bayesian Analysis*, 3: 339–366, 2008.

A. Rodriguez, D.B. Dunson, and A. E. Gelfand. The nested Dirichlet process. *Journal of the American Statistical Association*, 103:1131–1144, 2008.

A. Rodriguez, D. B. Dunson, and A. E. Gelfand. Latent Stick-Breaking Processes. *Journal of the American Statistical Association*, 105:647–659, 2010.

J. Sethuraman. A constructive definition of the Dirichlet process prior. *Statistica Sinica*, 2:639–650, 1994.

C. A. Sims. Macroeconomics and reality. *Econometrica*, 48:1–48, 1980.

C. A. Sims. Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 38:975–1000, 1992.

C. A. Sims and T. Zha. Bayesian Methods for Dynamic Multivariate Models. *International Economic Review*, 39(4):949–968, 1998.

M. A. Taddy. An auto-regressive mixture model for dynamic spatial poisson processes: Application to tracking the intensity of violent crime. *Journal of the American Statistical Association*, 105: 1403–1427, 2010.

M. A. Taddy and A. Kottas. Markov switching Dirichlet process mixture regression. *Bayesian Analysis*, 4:793–816, 2009.

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.

L. Trippa, P. Müller, and W. Johnson. The multivariate beta process and an extension of the Polya tree model. *Biometrika*, 98:17–34, 2011.

S. G. Walker. Sampling the Dirichlet Mixture Model with Slices. *Communications in Statistics - Simulation and Computation*, 36:45–54, 2007.

Università degli Studi di Pavia
E-mail address: federico.bassetti@unipv.it

University Ca' Foscari, Venice
E-mail address: r.casarin@unive.it

Universidad Carlos III de Madrid
E-mail address: leisen@gmail.com