UNIVERSIDAD CARLOS III DE MADRID

working papers

Beta-product Poisson-Dirichlet Processes

Federico Bassetti[1] Roberto Casarin[2] and Fabrizio Leisen[3]

Abstract

Time series data may exhibit clustering over time and, in a multiple time series context, the clustering behavior may differ across the series. This paper is motivated by the Bayesian non--parametric modeling of the dependence between the clustering structures and the distributions of different time series. We follow a Dirichlet process mixture approach and introduce a new class of multivariate dependent Dirichlet processes (DDP). The proposed DDP
are represented in terms of vector of stick-breaking processes with dependent weights. The weights are beta random vectors that determine different and dependent clustering effects along the dimension of the DDP vector. We discuss some theoretical properties and provide an efficient Monte Carlo Markov Chain algorithm for posterior computation. The effectiveness of the method is illustrated with a simulation study and an application to the United States and the European Union industrial production indexes.

Keywords: *Bayesian non--parametrics, Dirichlet process, Poisson-Dirichlet process,*

Multiple Time-series non--parametrics

[1] Depatments of Mathematics, Università di Pavia
via Ferrata, 1; 27100 Pavia, Italy
email: federico.bassetti@unipv.it
[2] Department of economics, Università Cà Foscari di Venezia
San Giobbe 873/b; 30121 Venezia, Italy
email: r.casarin@unive.it
[3] Departamento de Estadistica, Universidad Carlos III de Madrid
, Calle Madrid 126; 28903 - Getafe (Madrid), Spain
. e-mail: fabrizio.leisen@gmail.com

1. INTRODUCTION

This paper is concerned with some multivariate extensions of the Poisson-Dirichlet process. In this paper the class of models considered originates from the Ferguson Dirichlet process (DP) (see Ferguson [1973, 1974]) that is now widely used in non-parametric Bayesian statistics. Our extensions rely upon the so-called Sethuraman's representation of the DP. Sethuraman [1994] shows that, given a Polish space $(\mathbb{X}, \mathcal{X})$ and a probability measure H_0 on $\mathcal{X}$, a random probability measure G on $\mathcal{X}$ is a Dirichlet Process of precision parameter $\alpha > 0$ and base measure H_0, in short $DP(\alpha, H_0)$, if and only if it has the stick-breaking representation:

$$G(\cdot) = \sum_{k \geq 1} W_k \, \delta_{\tilde{\varphi}_k}(\cdot), \tag{1}$$

where $(\tilde{\varphi}_k)_k$ and $(W_k)_k$ are stochastically independent, $(\tilde{\varphi}_k)_k$ is a sequence of independent random variables (atoms) with common distribution H_0 (base measure), and the weights W_ks are defined by the stick-breaking construction:

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

S_j being independent random variables with beta distribution $Beta(1, \alpha)$.

The DP has been extended in many directions. In this paper, we will build on a generalization of the DP that is called the Poisson–Dirichlet process. A Poisson–DP, $PD(\alpha, l, H_0)$, with parameters $0 \leq l < 1$ and $\alpha > -l$ and base measure H_0, is a random probability measure that can be represented with a Sethuraman-like construction by taking in (1)-(2) a sequence of independent random variables S_k with $S_k \sim Beta(1 - l, \alpha + lk)$, see Pitman [2006] and Pitman and Yor [1997]. Further generalizations based on the stick-breaking construction of the DP can be found in Ishwaran and James [2001].

The DP process and its univariate extensions are now widely used in Bayesian non-parametric statistics. A recent account of Bayesian non-parametric inference can be found in Hjort et al. [2010]. The univariate DP is usually employed as a prior for a mixing distribution, resulting in a DP mixture (DPM) model (see for example Lo [1984]). The DPM models incorporate DP priors for parameters in Bayesian hierarchical models, providing an extremely flexible class of models. More specifically, the DPM models define a random density by setting:

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

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], DPM models are now routinely implemented and used in many fields.

The first aim of this paper is to introduce new class of vectors of Poisson-Dirichlet processes and of DPM models. Vectors of random probability measures arise naturally in generalizations of the

DP and DPM models that accommodate dependence of observations on covariates or time. Using covariates, data may be divided into different groups and this leads to consider group-specific random probability measures and, as we shall see, to assume that the observations are partially exchangeable.

Probably the first paper in this direction, that introduced vectors of priors for partially exchangeable data, is Cifarelli and Regazzini [1978]. More recently, MacEachern [1999, 2001] introduced the so-called dependent Dirichlet process (DDP) and the DDP mixture models. His specification of the DDP incorporates dependence on covariates through the atoms while assuming fixed weights. More specifically, the atoms in the Sethuraman's representation (1)-(3) are replaced with stochastic processes $\tilde{\varphi}_k(z)$, z being a set of covariates. There exist many applications of this specification of the DDP. For instance, De Iorio et al. [2004] proposed an ANOVA-type dependence structure for the atoms, while Gelfand et al. [2004] considered a spatial dependence structure for the atoms. Later, DDP with both dependent atoms and weights was introduced in Griffin and Steel [2006]. Other constructions that incorporate a dependence structure in the weights have been proposed, for instance, in Duan et al. [2007], Chung and Dunson [2011], Dunson and Peddada [2008], Dunson et al. [2008] and Rodriguez et al. [2010].

Other approaches to the definition of dependent vectors of random measures rely upon either suitable convex combinations of independent DPs (e.g., Müller et al. [2004], Pennell and Dunson [2006], Hatjispyrosa et al. [2011], Kolossiatis et al. [2011]) or hierarchical structures of stick-breakings (e.g., Teh et al. [2006], Sudderth and Jordan [2009]).

Finally, we should note that it is possible to follow alternative routes other than the Sethuraman's representation to the definition of vectors of dependent random probabilities. For example, Leisen and Lijoi [2011] used normalized vectors of completely random measures, while Ishwaran and Zarepour [2009] employed bivariate gamma processes.

In this paper, we introduce a new class of multivariate Poisson-DP and DP by using a vector of stick-breaking processes with multivariate dependent weights. In the construction of the dependent weights, we consider the class of multivariate beta distributions introduced by Nadarajah and Kotz [2005] that have a tractable stochastic representation and makes the Bayesian inference procedures easier. We discuss some properties of the resulting multivariate DP and Poisson-DP and show that our process has the appealing property that the marginal distributions are DP or Poisson-DP.

The second aim of the paper is to apply the new DP to Bayesian non–parametric inference and to provide a simple and efficient method for posterior computation of DDP mixture models. We follow a data-augmentation framework and extend to the multivariate context the slice sampling algorithm described in Walker [2007] and Kalli et al. [2011]. The sampling methods for the full conditional distributions of the resulting Gibbs sampling procedure are detailed and the effectiveness of the proposed algorithm is studied with a set of simulation experiments.

Another contribution of this paper is to present an application of the proposed multivariate DDP mixture models to multivariate time series modeling. In the recent years, the interest in Bayesian non-parametric models for time series has increased. In this context, DP have been recently employed in different ways. Rodriguez and ter Horst [2008] used a Dirichlet process to define an infinite mixture of time series models. Taddy and Kottas [2009] proposed a Markov-switching finite mixture of independent Dirichlet process mixtures. Jensen [2010] considered Dirichlet process mixture of stochastic volatility models and Griffin [2011] proposed a continuous-time non–parametric model for volatility. A flexible non–parametric model with a time-varying stick-breaking process has been recently proposed by Griffin and Steel [2011]. In their model, a sequence of dependent Dirichlet processes is used for capturing time-variations in the clustering structure of a set of time series. In our paper, we extend the existing Bayesian non–parametric models for multiple time series by allowing each series to have a possible different clustering structure and by accounting for dependence between the series-specific clusterings. Since we obtain a dynamic infinite-mixture model and since the number of components with negligible weights can be different in each series, our model represents a non–parametric alternative to multivariate dynamic finite-mixture models (e.g., Markov-switching models) that are usually employed in time series analysis.

The structure of the paper is as follows. Section 2 introduces vectors of dependent stick-breaking processes and defines their properties. Section 3 introduces vectors of Poisson-Dirichlet processes for prior modelling. Section 4 proposes a Monte Carlo Markov Chain (MCMC) algorithm for approximated inference for vector of DP mixtures. Section 5 provides some applications to both simulated data and to the time series of the industrial production index for the United States and the European Union. Section 6 concludes the paper.

2. Dependent stick-breaking processes

Consider a set of observations, taking values in a space $\mathbb{Y}$, say a subset of $\mathbb{R}^d$, divided in r sub–samples (or group of observations), that is:

$$Y_{ij} \qquad i = 1, \ldots r, \quad j = 1, \ldots, n_i.$$

Above Y_{ij} is the j-th observation within sub–sample i. For instance, i may correspond to a space label or predictors. Typically, one assumes that the observations of the block i have the same (conditional) density f_i and that the observations are (conditionally) independent. Hence, to perform a non–parametric Bayesian analysis of the data, one needs to specify a prior distribution for the vector of densities $(f_1, f_2, \ldots, f_r)$. Moreover, in assessing a prior for $(f_1, f_2, \ldots, f_r)$, a possible interest is on borrowing information across blocks. To do this, first we introduce a sequence of density kernels $\mathcal{K}_i : \mathbb{Y} \times \mathbb{X} \to [0,1]$ $(i = 1, \ldots, r)$ (where $\mathcal{K}_i$ is jointly measurable and $C \mapsto \int_C \mathcal{K}_i(y|x)\nu(dy)$ defines a probability measure on $\mathbb{Y}$ for any x in $\mathbb{X}$, ν being a dominating

measure on $\mathbb{Y}$). Secondly we define:

$$(4) \qquad f_i(y) := \int \mathcal{K}_i(y|\theta) G_i(d\theta) \qquad i = 1, \ldots, r,$$

where $(G_1, \ldots, G_r)$ is a vector of dependent stick breaking processes that will be defined in the next section.

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

$$(5) \qquad G_i(\cdot) := \sum_{k \geq 1} W_{ik} \, \delta_{\tilde{\varphi}_{ik}}(\cdot) \qquad i = 1, \ldots, r.$$

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

- $(\tilde{\varphi}_k)_k$ and $(W_k)_k$ are stochastically independent;
- $(\tilde{\varphi}_k)_k$ is an i.i.d. sequence taking values in $\mathbb{X}^r$ with common probability distribution G_0;
- $(W_k)_k$ are determined via the stick breaking construction

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

where $\prod_i^j = 1$ for $i > j$ and $S_k = (S_{1k}, \ldots, S_{rk})$ 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.

Note that $(G_1, \ldots, G_r)$ is a vector of dependent random measures whenever $(S_{11}, \ldots, S_{r1})$ or $(\tilde{\varphi}_{11}, \ldots, \tilde{\varphi}_{r1})$ are vectors of dependent random variables. The dependence between the measures affects the dependence structure underlying the densities $f_1, \ldots, f_r$, which can be represented as infinite mixtures

$$(6) \qquad f_i(y) = \sum_{k \geq 1} W_{ik} \, \mathcal{K}_i(y|\tilde{\varphi}_{ik}) \qquad i = 1, \ldots, r$$

functions of the atoms $(\tilde{\varphi}_k)_k$ and the weights $(W_k)_k$.

The above definition of dependent random measures is quite general. For the sake of completeness, we shall notice that our specification of vectors of stick-breaking processes can be extended, even if not straightforwardly, up to include more rich structure such as the matrix of stick-breaking processes proposed in Dunson et al. [2008]. In the rest of this section we briefly discuss the choice of atoms $\{\tilde{\varphi}_k\}_k$ and analyze some general features of vectors of stick breaking processes. While in the next section we focus on the main contribution of this works which is a new specification of the stick-vectors $\{S_k\}_k$ based on multivariate beta distribution.

2.2. **Atoms.** The simplest assumption for the atoms is that they are common to all the measures G_i. Otherwise stated this means that the base measure of the atoms is

$$(7) \qquad G_0(A_1 \times A_2 \cdots \times A_r) = H_0(\bigcap_i A_i)$$

for every $A_1, \ldots, A_r$ measurable subsets of $\mathbb{X}$, H_0 being a probability measure on $\mathbb{X}$, which corresponds to the case

$$(8) \qquad (\tilde{\varphi}_{1k}, \ldots, \tilde{\varphi}_{rk}) = (\tilde{\varphi}_{0k}, \tilde{\varphi}_{0k}, \ldots, \tilde{\varphi}_{0k})$$

with $\tilde{\varphi}_{0k}$ distributed according to H_0.

Eventually one can choose a more complex structure for the law of the atoms, including covariates (or exogenous effects) related to the specific block i. For instance one can assume an ANOVA-like scheme of the form

$$(9) \qquad (\tilde{\varphi}_{1k}, \ldots, \tilde{\varphi}_{rk}) = (\tilde{\mu}_{0k} + \tilde{\mu}_{1k}, \tilde{\mu}_{0k} + \tilde{\mu}_{2k}, \ldots, \tilde{\mu}_{0k} + \tilde{\mu}_{rk})$$

where $\tilde{\mu}_{0k}$ represents the overall "mean" (of the k-th mixture component) and $\tilde{\mu}_{ik}$ the specific "mean" for factor i (of the k-th mixture component). A similar choice has been used in De Iorio et al. [2004].

In many situations it is reasonable to assume that the components of the mixture are essentially the same for all the blocks but that they have different weights. In addition, the choice (7)-(8) yields a simple form of the correlation between the related random measure. This feature, which may be useful in the parameter elicitation, is discussed in the next subsection.

2.3. **Correlation and Dependence Structure.** In the general definition given in Subsection 2.1 the vectors S_k of stick variables are assumed to be independent. If in addition one assumes that they have the same distribution, i.e. that

$$(10) \qquad (S_k)_{k \geq 1} \text{ is a sequence of i.i.d. random vectors,}$$

then it is easy to compute the correlation of two elements of the vector $(G_1, \ldots, G_r)$ as well as the correlation between $f_i(y)$ and $f_j(y)$.

For the shake of simplicity we shall consider only the case $i = 1, j = 2$ and set

$$(11) \qquad C_{1,2} := \frac{\mathbb{E}[S_{11} S_{21}]}{1 - \mathbb{E}[(1 - S_{11})(1 - S_{21})]} \sqrt{\frac{(2\mathbb{E}[S_{11}] - \mathbb{E}[S_{11}^2])(2\mathbb{E}[S_{21}] - \mathbb{E}[S_{21}^2])}{\mathbb{E}[S_{11}^2]\mathbb{E}[S_{21}^2]}}.$$

and, for every y,

$$\kappa_{G_{01}}(y) := \int \mathcal{K}_2(y \mid x) G_{01}(dx), \quad \kappa_{G_{02}}(y) := \int \mathcal{K}_2(y \mid x) G_{02}(dx),$$

$$\kappa_{G_0}(y) := \int \mathcal{K}_1(y \mid x_1) \mathcal{K}_2(y \mid x_2) G_0(dx_1 \times dx_2 \times \mathbb{X}^{r-2})$$

where G_{0i} denotes the i-th marginal of G_0.

Proposition 1. *Assume that* (10) *holds true, then for all measurable set A and B*

$$(12) \qquad Cor(G_1(A), G_2(B)) = C_{1,2} \times \frac{G_0(A \times B \times \mathbb{X}^{r-2}) - G_{01}(A)G_{02}(B)}{\sqrt{G_{01}(A)(1 - G_{01}(A))G_{02}(B)(1 - G_{02}(B))}}$$

and for every y in $\mathbb{Y}$

$$(13) \qquad Cor(f_1(y), f_2(y)) = C_{1,2} \times \frac{\kappa_{G_0}(y) - \kappa_{G_{01}}(y)\kappa_{G_{02}}(y)}{\sqrt{\kappa_{G_{01}}(y)(1 - \kappa_{G_{01}}(y))\kappa_{G_{02}}(y)(1 - \kappa_{G_{02}}(y))}}.$$

The proof of Proposition 1 is in Appendix.

Corollary 2. *Assume that* (8) *and* (10) *hold true, then for every measurable set A*

$$(14) \qquad\qquad\qquad Cor(G_1(A), G_2(A)) = C_{1,2}.$$

where $C_{1,2}$ *is defined in* (11).

2.4. **Partial exchangeability.** We conclude this section by observing that $[Y_{ij} : i = 1, \ldots r, j \geq 1]$ is a partially exchangeable random array, indeed the joint law of the infinite process of observation is characterized by:

$$(15) \quad P\{Y_{ij} \in A_{ij}\ i = 1, \ldots r, j = 1, \ldots, n_i\} = \mathbb{E}\left[\prod_{i=1}^{r}\left[\int_{\mathbb{X}}\prod_{j=1}^{n_i}\int_{A_{ij}}\mathcal{K}_i(y_{ij}|x_i)\nu(dy_{ij})G_i(dx_i)\right]\right],$$

where the expectation is respect to the joint law of $(G_1, \ldots, G_r)$. Recall that an array $[Y_{ij} : i = 1, \ldots r, j \geq 1]$ is said to be row-wise partially exchangeable if, for every $n > 1$, every measurable sets A_{ij} and any permutations $\rho_1, \ldots, \rho_r$ of $\{1, \ldots, n\}$,

$$P\{Y_{ij} \in A_{ij}\ i = 1, \ldots r, j = 1, \ldots, n\} = P\{Y_{i\rho_i(j)} \in A_{ij}\ i = 1, \ldots r, j = 1, \ldots, n\}.$$

In other words, the joint law is not necessarily invariant to permutations of observations from different groups. From a practical point of view, the partial exchangeability represents a suitable model for sets of data that exhibit sub-samples with some possibly different features.

3. BETA-PRODUCT POISSON-DIRICHLET PROCESSES

We propose a new class of vector of dependent probability measures $(G_1, \ldots, G_r)$ in such a way that (marginally) G_i is a Dirichlet process for every i. This result follows from the Sethurman's representation (1) if one considers a multivariate distribution for $(S_{1k}, \ldots, S_{rk})$ such that $S_{ik} \sim Beta(1, \alpha_i), \forall i, k$.

It is worth noticing that there are many possible definitions of multivariate beta distribution (see for example Olkin and Liu [2003] and Nadarajah and Kotz [2005]), but not all of them has a tractable stochastic representation and leads 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 3 (Radhakrishna 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 $U_1 U_2 \cdots U_p$ is also a beta random variable with parameters $(a_1, b_1 + \cdots + b_p)$.*

Proof. It is easy to check that the Mellin transform of a beta random variable Z of parameters (a, b) is given by

$$\mathcal{M}_Z(s) := \mathbb{E}[Z^s] = \frac{\Gamma(a+b)\Gamma(a+s)}{\Gamma(a)\Gamma(a+b+s)}.$$

Hence, since $a_2 = a_1 + b_1$ and using the independence assumption

$$\mathcal{M}_{U_1 U_2}(s) = \mathbb{E}[U_1^s]\mathbb{E}[U_2^s] = \frac{\Gamma(a_1+b_1)\Gamma(a_1+s)}{\Gamma(a_1)\Gamma(a_1+b_1+s)}\frac{\Gamma(a_2+b_2)\Gamma(a_2+s)}{\Gamma(a_2)\Gamma(a_2+b_2+s)} = \frac{\Gamma(a_1+b_1+b_2)\Gamma(a_1+s)}{\Gamma(a_1)\Gamma(a_1+b_1+b_2+s)}.$$

Which gives the result for $p = 2$. The general case follows. $\qquad\square$

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

(16)
$$(S_{1k}, S_{2k}, \ldots, S_{rk}) := (V_{0k}V_{1k}, V_{0k}V_{2k}, \ldots, V_{0k}V_{rk})$$

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

As an alternative we consider

(17)
$$(S_{1k}, S_{2k}, \ldots, S_{rk}) := (V_{0k}V_{1k}\ldots V_{r-1k}, V_{0k}V_{1k}\ldots V_{r-2k}, \ldots, V_{0k})$$

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 $S_{ik} \sim Beta(\alpha_{0k}, \alpha_{1k} + \cdots + \alpha_{r+1-i,k})$.

It should be noted that (16) resembles the specification of the matrix stick-breaking process in Dunson et al. [2008]. In that paper all the components of the stick-breaking process are products of two independent beta variables with fixed parameters $(1, \alpha)$ and $(1, \beta)$, that precludes obtaining Poisson-Dirichlet marginals. For this reason we propose the specifications in (16) and (17) that, for a special choice of the parameters α_{ik}, allow for Poisson-Dirichlet marginals. In the first construction we obtain a random vector with identical marginal distributions, while in the second construction the vector has different marginals. Moreover, in the second case $S_{1k} \leq S_{2k} \leq \cdots \leq S_{rk}$, which induces an ordering on the concentration parameters of the Dirichlet marginals. These aspects will be further discussed in the following sections.

3.1. **Dirichlet Process marginal.** For the sake of clarity, we start with $r = 2$. We assume (10) and we discuss how to choose the parameters in (16)-(17) in order to get DP marginals. Note that since (10) holds true, then $S_k = (S_{1k}, S_{2k}) \sim (S_{11}, S_{21})$. According to the construction schemes given in (16) and (17), with $\alpha_{0k} = 1$, $\alpha_{1k} = \alpha_1$, $\alpha_{2k} = \alpha_2$, two possible and alternative specifications of (S_{11}, S_{21}) are:

(H1) $(S_{11}, S_{21}) := (V_0V_1, V_0V_2)$, *with* V_0, V_1, V_2 *independent,* $V_0 \sim Beta(1 + \alpha_1, \alpha_2)$ *and* $V_i \sim$ $Beta(1, \alpha_1)$, $i = 1, 2$, *where* $\alpha_1 > 0$ *and* $\alpha_2 > 0$;

(H2) $(S_{11}, S_{21}) := (V_0V_1, V_0)$, *with* V_0, V_1 *independent,* $V_0 \sim Beta(1, \alpha_1)$ *and* $V_1 \sim Beta(1 + \alpha_1, \alpha_2)$ *with* $\alpha_1 > 0$ *and* $\alpha_2 > 0$.

Thanks to Proposition 3, if (H1) holds, then $S_{11} \sim Beta(1, \alpha_1 + \alpha_2)$ and $S_{21} \sim Beta(1, \alpha_1 + \alpha_2)$, while if (H2) holds, then $S_{11} \sim Beta(1, \alpha_1 + \alpha_2)$ and $S_{21} = V_0 \sim Beta(1, \alpha_1)$. Hence, we have the following result.

Proposition 4. *Under* (10), *if* (H1) *holds true,* G_1 *and* G_2 *are (marginally) Dirichlet processes with the same precision parameter* $\alpha_1 + \alpha_2$ *and base measures* G_{01}, G_{02} *respectively. If* (H2) *holds true, then the first component of* (G_1, G_2) *is a Dirichlet process with precision parameter* $\alpha_1 + \alpha_2$ *and base measure* G_{01} *and the second component is a Dirichlet process with precision parameter* α_1 *and base measure* G_{02}.

Since with this construction the G_i's are Dirichlet processes, we call (G_1, G_2) Beta-Product Dependent Dirichlet Process of parameters $\alpha = (\alpha_1, \alpha_2)$ and base measure G_0, in short $\beta_i -$ DDP(α, G_0), where $i = 1$ for (H1) and $i = 2$ for (H2). In addition, when we assume (7), we denote the resulting process with $\beta_i -$DDP(α, H_0).

It should be noted that the two processes have different marginal behaviors. The β_1–DDP(α, G_0) process has marginals with the same precision parameter and should be used as a prior when the clustering along the different vector dimension is expected to be similar. In the $\beta_2 -$DDP(α, G_0) process, the precision parameter decreases along the vector dimension. This process should be used as a prior when a priori one suspects that the clustering features are different in the subgroups of observations.

For parameter elicitation purposes, it is useful to analyze how the choice of (α_1, α_2) affects the correlation between G_1 and G_2. Let us start by considering the correlation between the stick variables. From Theorem 4 and 6 in Nadarajah and Kotz [2005], one obtains the following correlation between the components S_{1h} and S_{2h} in the cases of (H1) and (H2)

$$Cor(S_{11}, S_{21}) = \begin{cases} \frac{\alpha_2}{(\alpha_1+1)(\alpha_1+\alpha_2)} & \text{for} \quad (H1) \\ \sqrt{\frac{\alpha_1(2+\alpha_1+\alpha_2)}{(\alpha_1+\alpha_2)(2+\alpha_1)}} & \text{for} \quad (H2). \end{cases}$$

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 value equal to zero. The gray areas represent correlation values in the $(0, 1)$ interval under both (H1) and (H2) beta models. According to the graph at the top-left of Fig. 1, one can conclude that the parametrization used in this paper allows for covering the whole range of possible correlation values in the $(0,1)$ interval. For instance, under (H1), a low correlation between the components
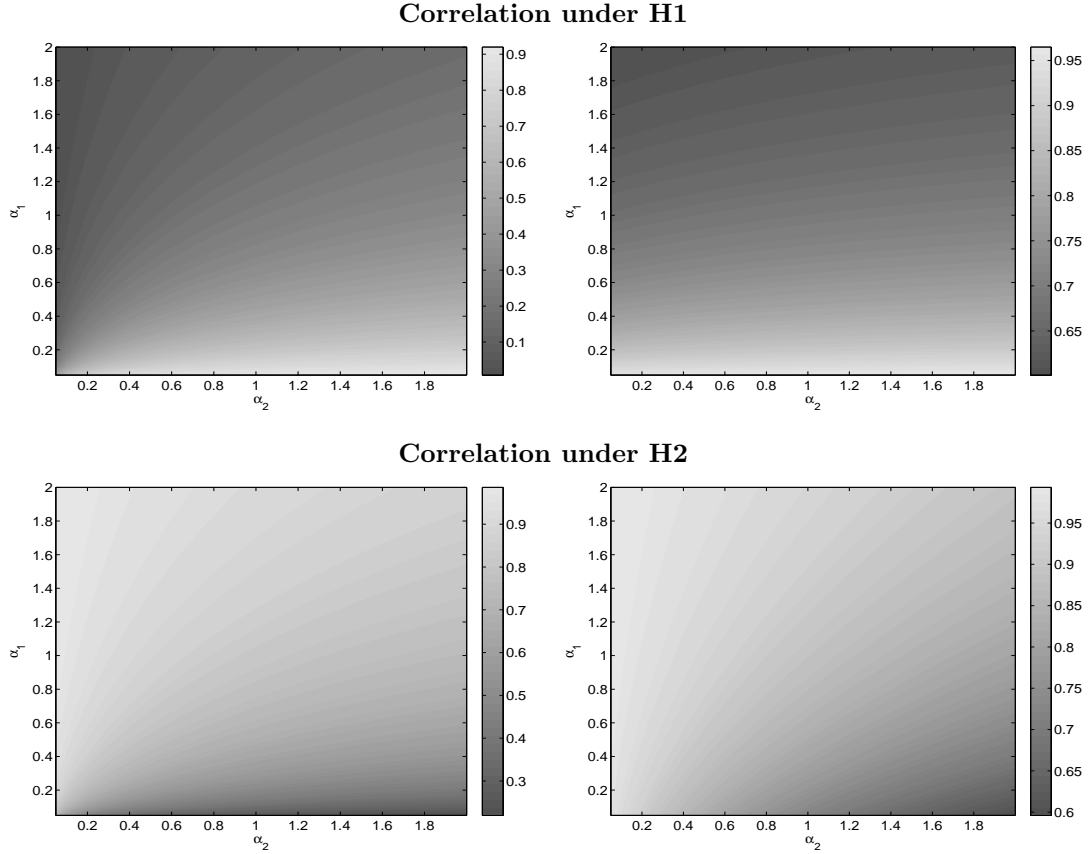
Correlation under H1

Correlation under H2

FIGURE 1. Left column: correlation between S_{11} and S_{21} under (H1) (first row) and (H2) (second row). Right column: correlation between G_1 and G_2, assuming (7), under (H1) (first row) and (H2) (second row)

of the stick-breaking corresponds to low values of the parameter α_1, say between 0 and 0.1, for any choice of the parameter α_2.

Corollary 5. *Under the same assumptions of* Proposition 4,

$$C_{1,2} = \begin{cases} \left(\frac{(\alpha_1+\alpha_2+1)(\alpha_1+2)}{2(\alpha_1+1)(\alpha_1+\alpha_2+1)-(\alpha_1+2)} \right) & \text{for} \quad (H1) \\ \frac{2(\alpha_1+1)}{(\alpha_1+2)(2\alpha_1+\alpha_2+1)-\alpha_1} \sqrt{(\alpha_1+1)(\alpha_1+\alpha_2+1)} & \text{for} \quad (H2). \end{cases}$$

Recall that if (7) holds true, then for any measurable set A:

$$Cor(G_1(A), G_2(A)) = C_{1,2}.$$

The graphs at the bottom of Fig. 1 show how the parameters α_1 and α_2 affect the correlation between the components S_{1h} and S_{2h} of the bivariate beta used in the stick breaking process and the correlation between the random measures G_1 and G_2 – when assuming (7).

It is worth noticing that, under (H1), (S_{11}, S_{21}) converges (in distribution) to (V_1, V_2) as $\alpha_2 \to 0$, where V_1 and V_2 independent random variables with distribution $Beta(1, \alpha_1)$. While, under (H2),

(S_{11}, S_{21}) converges to (V_0, V_0) as $\alpha_2 \to 0$, where V_0 is a $Beta(1, \alpha_1)$ random variable. In particular, if one assumes (7) and (H2), when $\alpha_2 \to 0$, one gets the limit situation in which all the observations are sampled from a common mixture of Dirichlet processes. In other words, in this limit case, as can be seen by (15) for $G_1 = G_2$, 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 $(\tilde{\varphi}_{1k}, \tilde{\varphi}_{2k})$ to be independent random elements with probability distribution G_{01} and G_{02}. In the limit for $\alpha_2 \to 0$, one obtains two independent Dirichlet processes G_1 and G_2 with base measures G_{01} and G_{02}. In other words, with this choice, one considers the blocks of observations as two independent blocks of exchangeable variables and no sharing of information is allowed.

The (H1) construction for the case $r > 2$ follows immediately from (16) with $\alpha_{0k} = 1$, $\alpha_{1k} = \alpha_1$, $\alpha_{2k} = \alpha_2$, that is by assuming $V_{0k}, V_{1k}, V_{2k}, \ldots, V_{rk}$ to be independent with $V_{0k} \sim Beta(1 + \alpha_1, \alpha_2)$ and $V_{ik} \sim Beta(1, \alpha_1)$, $i = 1, 2, \ldots, r$, where $\alpha_1 > 0$ and $\alpha_2 > 0$. In this case, S_{ik} has $Beta(1, \alpha_1 + \alpha_2)$ distribution for $i = 1, \ldots, r$ and hence G_i is marginally $DP(\alpha_1 + \alpha_2, G_{0i})$. Also the formula for the correlation between two measures easily extends to the case $r > 2$ under (7):

$$Corr(G_i(A), G_j(A)) = \frac{(\alpha_1 + \alpha_2 + 1)(\alpha_1 + 2)}{2(\alpha_1 + 1)(\alpha_1 + \alpha_2 + 1) - (\alpha_1 + 2)}.$$

The (H2) construction extends to $r > 2$ by setting in (17) $\alpha_{0k} = 1$ and $\alpha_{ik} = \alpha_i$ $(i \geq 1)$, that is by taking $V_{0k}, \ldots, V_{r-1\,k}$ to be independent random variables with $V_{0k} \sim Beta(1, \alpha_1)$, $V_{1k} \sim Beta(1 + \alpha_1, \alpha_2)$, $\ldots$, $V_{r-1\,k} \sim Beta(1 + \alpha_1 + \cdots + \alpha_{r-1}, \alpha_r)$, where $\alpha_i > 0$ for all $i = 1, \ldots, r$. In this last case $S_{ik} \sim Beta(1, \alpha_1 + \cdots + \alpha_{r-i+1})$ for every $i = 1, \ldots, r$. Hence G_i is marginally $DP(\alpha_1 + \cdots + \alpha_{r-i+1}, G_{0i})$. Under (7) the correlation between $G_i(A)$ and $G_j(A)$ with $1 \leq i < j \leq r$ is given by

(18)
$$Corr(G_i(A), G_j(A))$$
$$= \frac{2\sqrt{(1 + \alpha_1 + \cdots + \alpha_{r-i+1})}(1 + \alpha_1 + \cdots + \alpha_{r-j+1})^{\frac{3}{2}}}{2(1 + \alpha_1 + \cdots + \alpha_{r-j+1})^2 + (2 + \alpha_1 + \cdots + \alpha_{r-j+1})(\alpha_{r-j+2} + \cdots + \alpha_{r-i+1})}.$$

The proof of this last result is given in the Appendix.

3.2. **Poisson-Dirichlet process marginal.** Recall that a Poisson–Dirichlet process, $PD(\alpha, l, H_0)$, with parameters $0 \leq l < 1$ and $\alpha > -l$, and base measure H_0, is obtained by taking in (1)-(2) a sequence of independent random variables S_k with $S_k \sim Beta(1 - l, \alpha + lk)$. In this section we show that by a suitable choice of the parameters in (16)-(17) we obtain a vectors of dependent random measures with Poisson-Dirichlet marginals.

In the first case use (16) with $\alpha_{0k} = 1 - l$, $\alpha_{1k} = \alpha_1$ and $\alpha_{2k} = \alpha_2 + lk$, that is take $V_{0k}, \ldots, V_{rk}$ to be independent random variables such that

(19)
$$V_{0k} \sim Beta(1 - l + \alpha_1, \alpha_2 + lk), \quad V_{ik} \sim Beta(1 - l, \alpha_1) \quad i = 1, \ldots, r$$

where $\alpha_1 > 0$, $\alpha_2 > 0$, and $0 \leq l < 1$. Proposition 3 yields that $S_{ik} \sim Beta(1 - l, \alpha_1 + \alpha_2 + lk)$. In the second case use (17) with $\alpha_{0k} = 1 - l$, $\alpha_{1k} = \alpha_1 + lk$, and $\alpha_{ik} = \alpha_i$ for $i \geq 2$, that is take $V_{0k}, \ldots, V_{r-1k}$ to be independent random variables such that

(20)
$$V_{0k} \sim Beta(1 - l, \alpha_1 + lk), V_{1k} \sim Beta(1 + \alpha_1 + l(k - 1), \alpha_2), \ldots,$$

$$\ldots V_{r-1k} \sim Beta(1 + \alpha_1 + \cdots + \alpha_{r-1} + l(k - 1), \alpha_r)$$

with $\alpha_i > 0$ $i = 0, \ldots, r$ and $0 \leq l < 1$. In this last case S_{ik} has $Beta(1 - l, \alpha_1 + \cdots + \alpha_{r-i+1} + lk)$ for every $i = 1, \ldots, r$.

Summarizing we have proved the following

Proposition 6. *If* (16) *and* (19) *hold true* G_i *is a* $PD(\alpha_1 + \alpha_2, l, G_{0i})$ *for every* $i = 1, \ldots, r$, *while if* (17) *and* (20) *hold true* G_i *is a* $PD(\alpha_1 + \cdots + \alpha_{r-i+1}, l, G_{0i})$ *for every* $i = 1, \ldots, r$.

4. Slice Sampling Algorithm for Posterior Simulation

For posterior computation, we propose an extension of the slice sampling algorithm introduced in Walker [2007], Kalli et al. [2011]. For the sake of simplicity we shall describe the sampling strategy for a vector of Beta-Product DDP with $r = 2$ (β_i–DDP(α, G_0)), see Subsection 3.1. The proposed algorithm can be easily extend to the case $r > 2$ and to the Beta-Product dependent Poisson-DP.

Recall that in β_i–DDP(α, G_0) the stick variables 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 same distribution of (V_0, V_1, V_2) and the convention $V_{2k} = 1$ and $V_k := (V_{0,k}, V_{1,k})$ under (H2).

Starting from (6), the key idea of the slice sampling is to find a finite number of variables to be sampled. First we introduce a latent variable u in such a way that $f_i(y)$ is the marginal density of

$$f_i(y, u) = \sum_{k \geq 1} \mathbb{I}\{u \leq W_{ik}\} \, \mathcal{K}_i(y|\tilde{\varphi}_{ik}).$$

It is clear that given u, the number of components is finite. In addition we introduce a further latent variable d which indicates which of these finite number of components provides the observation, that is

(21)
$$f_i(y, u, d) := \mathbb{I}\{u < W_{i,d}\}\mathcal{K}_i(y|\tilde{\varphi}_d).$$

Hence, the likelihood function for the augmented variables (y, u, d) is available as a simple product of terms and crucially d is finite.

To be more precise we introduce the allocation variables D_{ij} ($i = 1, 2; j = 1, \ldots, n_i$) taking values in $\mathbb{N}$ and the slice variables U_{ij} ($i = 1, 2; j = 1, \ldots, n_i$) taking values on $[0, 1]$. We shall use

the notation

$$Y_i^{(n_i)} := (Y_{i1}, \ldots, Y_{in_i}), \quad D_i^{(n_i)} := (D_{i1}, \ldots, D_{in_i}), \quad U_i^{(n_i)} := (U_{i1}, \ldots, U_{in_i})$$

and we write: $\tilde{\varphi}$ for $(\tilde{\varphi}_k)_k$, V for $(V_k)_k$, U for $[U_1^{(n_1)}, U_2^{(n_2)}]$, $D^{(n)}$ for $[D_1^{(n_1)}, D_2^{(n_2)}]$ and Y for $[Y_1^{(n_1)}, Y_2^{(n_2)}]$.

The random elements $(\tilde{\varphi}, V)$ have the law already described. While conditionally on $(\tilde{\varphi}, V)$, the random vectors (Y_{ij}, U_{ij}, D_{ij}), $i = 1, 2; j = 1, \ldots, n_i$, are stochastically independent with the joint density (21).

We conclude by observing that it can be useful to put a prior distribution even on the hyperparameters (α_1, α_2). The law of $\tilde{\varphi}$ is assumed independent of the random vector of hyperparameters $\tilde{\alpha} := (\tilde{\alpha}_1, \tilde{\alpha}_2)$, while the distribution of V_j depends on $\tilde{\alpha}$ through (H1) or (H2), so we shall write $P\{V_j \in dv_j | \tilde{\alpha}\}$.

Our target is the exploration of the posterior distribution (given Y) of $[\tilde{\varphi}, V, U, \tilde{\alpha}, D]$ by Markov Chain Monte Carlo sampling.

Essentially we shall use a block Gibbs sampler which iteratively simulates $\tilde{\varphi}$ given $[V, U, D, \tilde{\alpha}, Y]$, $[V, U, \tilde{\alpha}]$ given $[D, \tilde{\varphi}, Y]$ and D given $[V, U, \tilde{\varphi}, \tilde{\alpha}, Y]$.

For the one dimensional, this blocking structure case has been introduced in Papaspiliopoulos [2008] and Kalli et al. [2011] as an alternative and more efficient version of the original algorithm of Walker [2007]. Our algorithm extends the one dimensional slice sampling of Kalli et al. [2011] to the multidimensional case. This extension is not trivial as it involves generation of random samples from vectors of allocation and slice variables of a multivariate stick-breaking process. We present an efficient Gibbs sampling algorithm by elaborating further on the blocking strategy of Kalli et al. [2011].

In order to describe in details the full-conditionals of the above sketched block Gibbs sampler, we need some more notation. Define for $i = 1, 2$ and $k \geq 1$,

$$\mathcal{D}_{i,k} := \{j \in \{1, \ldots, n_i\} : D_{i,j} = k\},$$

$$A_{i,k} := \sum_{j=1}^{n_i} \mathbb{I}\{D_{i,j} = k\} = card(\mathcal{D}_{i,k}), \quad B_{i,k} := \sum_{j=1}^{n_i} \mathbb{I}\{D_{i,j} > k\}.$$

Moreover, let

(22) $$D^* := \max_{i=1,2} \max_{1 \leq j \leq n_i} D_{i,j}.$$

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 $\mathcal{D}^* = \{k : \mathcal{D}_{1,k} \cup \mathcal{D}_{2,k} \neq \emptyset\}$. Note that $D^* = \max\{k \in \mathcal{D}^*\}$ and $|\mathcal{D}^*| \leq D^*$ almost surely. In the following subsections we give the details of the full conditionals of the blocking Gibbs sampler, further details on the algorithm are given in Appendix.

4.1. The full conditional of $\tilde{\varphi}$. The atoms $\tilde{\varphi}$ given $[V, D, U, \tilde{\alpha}, Y]$ are conditionally independent and the full conditionals are:

$$
\begin{aligned}
P\{\tilde{\varphi}_k \in d\varphi_k | D, U, V, \tilde{\alpha}, Y\} &= P\{\tilde{\varphi}_k \in d\varphi_k | D, Y\} \\
&\propto G_0(d\varphi_k) \prod_{j \in \mathcal{D}_{1,k}} \mathcal{K}_1(Y_{1,j} | \varphi_{1k}) \prod_{j \in \mathcal{D}_{2,k}} \mathcal{K}_2(Y_{2,j} | \varphi_{2k});
\end{aligned}
\tag{23}
$$

where $\varphi_k = (\varphi_{1k}, \varphi_{2k})$. The strategy for sampling from this full conditional depends on the specific form of $\mathcal{K}_i$ and G_0. In the next section we will discuss a possible strategy for Gaussian kernels.

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

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

$$
\begin{aligned}
P\{V^* \in dv^*, \tilde{\alpha} \in (d\alpha_1, d\alpha_2) | Y, \tilde{\varphi}, D\} &= P\{V^* \in dv^*, \tilde{\alpha} \in (d\alpha_1, d\alpha_2) | D\} \\
&\propto \prod_{k \in \mathcal{D}^*} \frac{Q_k(v_k | D, \alpha_1, \alpha_2)}{B(\alpha_1 + 1, \alpha_2) B^2(1, \alpha_1)} dv_k \pi(d\alpha_1, d\alpha_2)
\end{aligned}
\tag{24}
$$

where $\pi(d\alpha_1, d\alpha_2) = P\{\tilde{\alpha} \in (d\alpha_1, d\alpha_2)\}$ is the prior on the concentration parameters and

$$
Q_k(v_k | D, \alpha) := \begin{cases}
v_{0k}^{\alpha_1 + A_{1k} + A_{2k}} (1 - v_{0k})^{\alpha_2 - 1} \prod_{i=1,2} v_{ik}^{A_{ik}} (1 - v_{ik})^{\alpha_1 - 1} (1 - v_{0k} v_{ik})^{B_{ik}} \\
\qquad \text{under (H1) with } v_k = (v_{0k}, v_{1k}, v_{2k}) \\
v_{0k}^{A_{1k} + A_{2k}} (1 - v_{0k})^{\alpha_1 + B_{2k} - 1} v_{1k}^{\alpha_1 + A_{1k}} (1 - v_{1k})^{\alpha_2 - 1} (1 - v_{0k} v_{1k})^{B_{1k}} \\
\qquad \text{under (H2) with } v_k = (v_{0k}, v_{1k})
\end{cases}
\tag{25}
$$

To sample from (24), we iterate a two-step Metropolis-Hastings (M.-H.) within Gibbs with full conditionals

$$
P\{V^* \in dv^* | \tilde{\alpha}, D\} \propto \prod_{k \in \mathcal{D}^*} Q_k(v_k | D, \tilde{\alpha}) dv_k
\tag{26}
$$

and

$$
\begin{aligned}
P\{\tilde{\alpha} \in (d\alpha_1, d\alpha_2) | V^*, D\} &\propto \prod_{k \in \mathcal{D}^*} \frac{1}{B(\alpha_1 + 1, \alpha_2)} v_{0k}^{\alpha_1} (1 - V_{0k})^{\alpha_2 - 1} \\
&\quad \cdot \prod_{i=1,2} \frac{1}{B(1, \alpha_1)^2} (1 - V_{ik})^{\alpha_1 - 1} \pi(d\alpha_1, d\alpha_2).
\end{aligned}
\tag{27}
$$

For the each element (V_{0k}, V_{1k}, V_{2k}) of V^* we consider a multivariate Gaussian random walk proposal with diagonal scale matrix $\tau^2 I_3$, with $\tau^2 = 0.05$ in order to have acceptance rates between 0.3 and 0.5 for the elements of V^*.

- $[V^{**}, V^{***}]$ given $[D, V^*, \tilde{\varphi}, \tilde{\alpha}, Y]$. The V_k (with $k \notin \mathcal{D}^*$) are conditionally independent given $[D, V^*, \tilde{\varphi}, \tilde{\alpha}, Y]$ with $P\{V_k \in dv_k | \tilde{\alpha}, D, V^*\} \propto Q_k(v_k | D, \tilde{\alpha}) dv_k$ if $k \leq D^*$ and $P\{V_k \in dv | V^*, \tilde{\varphi}, D, \tilde{\alpha}, Y\} = P\{V_k \in dv | \tilde{\alpha}\}$ if $k > D^*$. Note that if $k \notin \mathcal{D}^*$ and $k \leq D^*$, then $A_{i,k} = 0$ in the definition of $Q_k(v_k | D, \tilde{\alpha})$. In order to sample from $Q_k(v_k | D, \tilde{\alpha})$ the same M.-H. step, used for the full conditional in (26), is employed.

- U given $[V, D, \tilde{\varphi}, \tilde{\alpha}, Y]$. The slice variables U are conditionally independent given $[V, D, \tilde{\varphi}, \tilde{\alpha}, Y]$ with

$$(28) \qquad P\{U_{i,j} \in du | V, Y, \tilde{\varphi}, D\} = P\{U_{i,j} \in du | V, D\} = \frac{\mathbb{I}\{u \leq W_{i,D_{i,j}}\}}{W_{i,D_{i,j}}} du.$$

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

$$(29) \qquad P\{D_{i,j} = d | \tilde{\varphi}, V, U, \tilde{\alpha}, Y\} \propto \mathcal{K}_i(Y_{i,j} | \tilde{\varphi}_{id}) \, \mathbb{I}\{U_{i,j} \leq W_{i,d}\}.$$

Here an important remark is in order. As in the slice sampling proposed in Walker [2007], Kalli et al. [2011], the full conditional (29) samples, almost surely, from a finite number of terms. So again it is easy to sample from this full conditional. More precisely, following Walker [2007], we note that $d > N_{i,j}^*$ ensures that $W_{i,d} < U_{i,j}$ where $N_{i,j}^*$ $(i = 1, 2; j = 1, \ldots, n_i)$ is the smallest integer such that

$$(30) \qquad \sum_{k=1}^{N_{i,j}^*} W_{i,k} > 1 - U_{i,j}.$$

5. Illustrations

We apply our new Beta-product dependent Dirichlet process to make inference for mixture of normals and mixture of vector autoregressive processes. The resulting model and inference procedure have been applied on both simulated data and real data on the industrial production in the United States and the European Union.

5.1. $\beta_1 - \mathbf{DDP}(\alpha, H_0)$ **mixtures of Gaussian distributions.** In this section, we apply our $\beta_1 - \mathrm{DDP}(\alpha, H_0)$ Gaussian mixture model for inference on synthetic data generated from finite Gaussian mixtures. More precisely we assume (8)-(10), (H1) and Gaussian kernels $\mathcal{K}_i(y | \varphi)$ $(i = 1, 2)$ with means μ and variance σ^2 for $\varphi = (\mu, \sigma^2)$, i.e.

$$\mathcal{K}_i(y | \varphi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\}.$$

As base measure $H_0(d\varphi)$ we take the product of a normal $\mathcal{N}(0, s^{-2})$ and inverse gamma $\mathcal{IG}(\lambda, \lambda)$, which are conjugate distributions for the bivariate kernel at hand. For the vector $\alpha = (\alpha_1, \alpha_2)$ of the precision parameters of the bivariate Dirichlet process we consider independent gamma priors

$\mathcal{G}(\zeta_{11}, \zeta_{21})\mathcal{G}(\zeta_{12}, \zeta_{22})$. In summary the Bayesian non–parametric model is

$$Y_{ij}|(\tilde{\mu}_{ij}, \tilde{\sigma}_{ij}^2) \overset{ind}{\sim} \mathcal{N}(\tilde{\mu}_{ij}, \tilde{\sigma}_{ij}^2) \qquad i = 1, 2, j \geq 1$$

$$(\tilde{\mu}_{ij}, \tilde{\sigma}_{ij}^2)|G_1, G_2 \overset{iid}{\sim} G_i \quad i = 1, 2$$

$$(G_1, G_2)|\tilde{\alpha} \sim \beta_1 - \mathrm{DDP}(\tilde{\alpha}, H_0)$$

$$\tilde{\alpha} \sim \mathcal{G}(\zeta_{12}, \zeta_{22})\mathcal{G}(\zeta_{12}, \zeta_{22})$$

The sampling procedure for U and D given in the previous section applies straightforwardly to this example. We shall describe here in more details the sampling strategy for the other unknown quantities of the model. 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 $P\{\tilde{\varphi}_k \in d\varphi_k|V, D, Y^{(n)}, U\}$, for $k \geq 1$ we consider a two-step Gibbs sampler with full conditional distributions

$$P\{\tilde{\mu}_k \in d\mu_k|\tilde{\sigma}_k^{-2}, D, Y^{(n)}\} \propto$$

(31)
$$\propto \exp\left\{-\frac{1}{2s^{-2}}\mu_k^2\right\} \prod_{m=1,2} \prod_{i \in \mathcal{D}_{m,k}} \exp\left\{-\frac{1}{2\tilde{\sigma}_k^2}(Y_{mi} - \mu_k)^2\right\} d\mu_k$$

$$\propto \exp\left\{-\frac{1}{2}\mu_k^2 \left(s^2 + \tilde{\sigma}_k^{-2}(A_{1,k} + A_{2,k})\right) + \mu_k \frac{1}{\tilde{\sigma}_k^2}(\eta_{1,k}^{(1)} + \eta_{2,k}^{(1)})\right\} d\mu_k$$

and

$$P\{\tilde{\sigma}_k^{-2} \in d\sigma_k^{-2}|\tilde{\mu}_k, D, Y^{(n)}\} \propto$$

(32)
$$\propto \exp\left\{-\lambda\sigma_k^{-2}\right\} (\sigma_k^{-2})^{\lambda+1} \prod_{m=1,2} \prod_{i \in \mathcal{D}_{m,k}} (\sigma_k^{-2})^{1/2} \exp\left\{-\frac{1}{2\sigma_k^2}(Y_{mi} - \tilde{\mu}_k)^2\right\} d\sigma_k^{-2}$$

$$\propto \exp\left\{-(\lambda + \frac{1}{2}(\eta_{1,k}^{(2)} + \eta_{2,k}^{(2)}))\sigma_k^{-2}\right\} (\sigma_k^{-2})^{\lambda+\frac{1}{2}(A_{1,k} + A_{2,k})-1} d\sigma_k^{-2}$$

which are proportional to the density function of a normal

(33)
$$\mathcal{N}\left(\frac{\tilde{\sigma}_k^{-2}(\eta_{1,k}^{(1)} + \eta_{2,k}^{(1)})}{s^2 + \tilde{\sigma}_k^{-2}(A_{1,k} + A_{2,k})}, \frac{1}{s^2 + \tilde{\sigma}_k^{-2}(A_{1,k} + A_{2,k})}\right)$$

and an inverse gamma

(34)
$$\mathcal{IG}\left(\lambda + \frac{1}{2}(A_{1,k} + A_{2,k}), \lambda + \frac{1}{2}(\eta_{1,k}^{(2)} + \eta_{2,k}^{(2)})\right)$$

respectively, where $A_{m,k}$, $m = 1, 2$ have been defined in the previous section and

(35)
$$\eta_{m,k}^{(1)} = \sum_{i \in \mathcal{D}_{m,k}} Y_{im} \quad \text{and} \quad \eta_{m,k}^{(2)} = \sum_{i \in \mathcal{D}_{m,k}} (Y_{im} - \tilde{\mu}_k)^2.$$

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 4.2.

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, (27) becomes

(36)
$$P\{\tilde{\alpha} \in (d\alpha_1, d\alpha_2)|V^*, D\} \propto$$
$$\frac{1}{B(\alpha_1 + 1, \alpha_2)^{d_1}} \frac{1}{B(1, \alpha_1)^{d_2}} \alpha_1^{\zeta_{11}-1} \exp\left\{-\alpha_1 \bar{\zeta}_{21}\right\} \alpha_2^{\zeta_{12}-1} \exp\left\{-\alpha_2 \bar{\zeta}_{22}\right\} d\alpha_1 d\alpha_2$$

where $\bar{\zeta}_{21} = \zeta_{21} - \sum_{k \in \mathcal{D}}(\log(V_{0k}) + \log(1-V_{1k}) + \log(1-V_{2k}))$ and $\bar{\zeta}_{22} = \zeta_{22} - \sum_{k \in \mathcal{D}} \log(1-V_{0k})$ and $d_1 = card(\mathcal{D}^*)$ and $d_2 = 2d_1$. We simulate from the full conditional by a M.-H. step. We considered two alternative proposals. First we assume independent proposals. At the j-th iteration, given $\alpha^{(j-1)}$, we simulate

(37)
$$\alpha_1^{(*)} \sim \mathcal{G}a(\zeta_{11}, \zeta_{21}), \quad \alpha_2^{(*)} \sim \mathcal{G}a(\zeta_{12}, \zeta_{22})$$

and accept with probability

(38)
$$\min\left\{1, \frac{B(\alpha_1^{(j-1)} + 1, \alpha_2^{(j-1)})^{d_1} B(1, \alpha_1^{(j-1)})^{d_2}}{B(\alpha_1^{(*)} + 1, \alpha_2^{(*)})^{d_1} B(1, \alpha_1^{(*)})^{d_2}}\right\}.$$

In our experiments this kind of proposal turns out to be highly inefficient and the M.-H. exhibits low acceptance rates, thus we consider a gamma random walk proposal. At the j-th iteration of the algorithm, given the previous value $(\alpha_1^{(j-1)}, \alpha_2^{(j-1)})$ of the chain, we simulate

(39)
$$\alpha_1^{(*)} \sim \mathcal{G}a(\kappa(\alpha_1^{(j-1)})^2, \kappa\alpha_1^{(j-1)}), \quad \alpha_2^{(*)} \sim \mathcal{G}a(\kappa(\alpha_2^{(j-1)})^2, \kappa\alpha_2^{(j-1)})$$

where κ represents the scale of random walk. The proposal is accepted with probability

(40)
$$\min\left\{1, \frac{P\{(\alpha_1^{(*)}, \alpha_2^{(*)})|V^*, D\}}{P\{\alpha_1^{(j-1)}, \alpha_2^{(j-1)}|V^*, D\}} \frac{g(\alpha_1^{(j-1)}|\alpha_1^{(*)})}{g(\alpha_1^{(*)}|\alpha_1^{(j-1)})} \frac{g(\alpha_2^{(j-1)}|\alpha_2^{(*)})}{g(\alpha_2^{(*)}|\alpha_2^{(j-1)})}\right\}$$

where

$$g(y|x) = \frac{1}{\Gamma(\kappa x^2)}(\kappa x)^{\kappa x^2} y^{\kappa x^2 - 1} e^{-\kappa xy}$$

is the conditional density of the gamma random walk proposals. We set the scale parameter κ in order to have acceptance rates close to 0.5.

We simulate $n = 50$ independent vectors, $(Y_{1,j}, Y_{2,j})$ with $j = 1, \ldots, n$, of observations. The component of the vectors $(Y_{1,j}, Y_{2,j})$ are independent and alternatively follow one of these models.

- The same three-component mixture of normals (model Mix1)

$$\frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1)$$
$$\frac{1}{3}\mathcal{N}(-10, 1) + \frac{1}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(10, 1)$$

- Two different mixtures with two common components (model Mix2)

$$\frac{1}{4}\mathcal{N}(0, 0.5) + \frac{1}{4}\mathcal{N}(3, 0.25) + \frac{1}{4}\mathcal{N}(2, 0.25) + \frac{1}{4}\mathcal{N}(5, 0.5)$$
$$\frac{1}{4}\mathcal{N}(0, 0.5) + \frac{1}{4}\mathcal{N}(3, 0.25) + \frac{1}{4}\mathcal{N}(-3, 0.25) + \frac{1}{4}\mathcal{N}(7, 0.5)$$
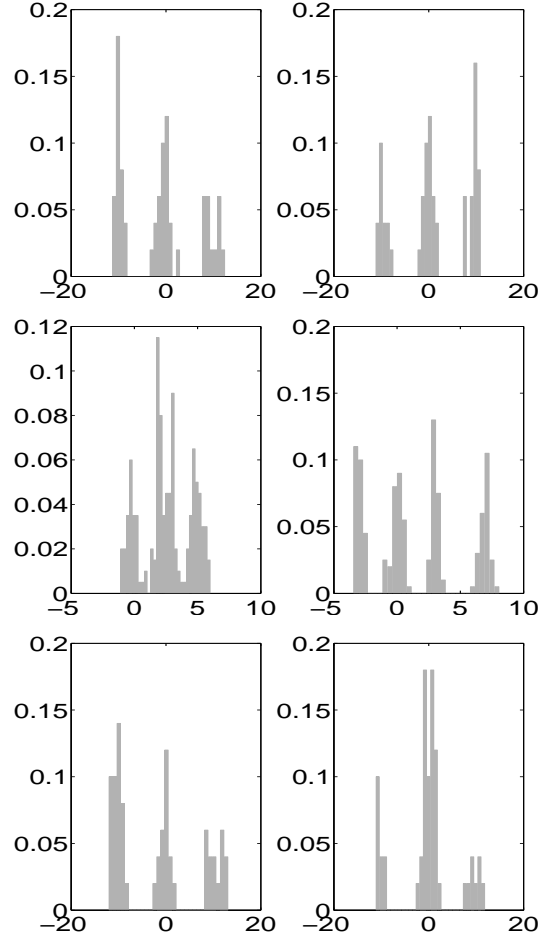
FIGURE 2. Sample of data for the first (left) and second (right) component generated under different model assumptions: Mix1 (first row), Mix2 (second row), and Mix3 (third row).

- The same three-component mixture of normals with different component probabilities (model Mix3)

$$\frac{1}{3}\mathcal{N}(-10,1) + \frac{1}{3}\mathcal{N}(0,1) + \frac{1}{3}\mathcal{N}(10,1)$$

$$\frac{1}{6}\mathcal{N}(-10,1) + \frac{4}{6}\mathcal{N}(0,1) + \frac{1}{6}\mathcal{N}(10,1)$$

The simulated set of data considered in the experiments are given in Fig. 2. In the left column is the histogram of the first component of the set of data and in the right column is the histogram of the second component. Then we estimate the $\beta_1-\text{DDP}(\alpha, H_0)$ mixture model on the different set of data. In the inference exercise, we choose a non–informative prior specification for the mean and precision parameters of the base measure and set $s^2 = 0.1$, $\lambda = 0.5$ (see for example Walker [2007]). For the concentration parameters (α_1, α_2) of the stick-breaking components, we follow Kalli et al. [2011] and consider two alternative specifications of the priors: weakly

informative (WI) prior and strongly informative (SI) prior. For the WI case, the hyperparameters setting is $(\zeta_{1j} = 0.01, \zeta_{2j} = 0.01)$, for $j = 1, 2$, in all the Mix1, Mix2 and Mix3 experiments. 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 $\mathbb{V}(\tilde{\alpha}_1) = \mathbb{V}(\tilde{\alpha}_2) = 100$, and to a low prior dependence level $(Cor(G_1, G_2) = 0.6902)$ between the random marginal densities. In our WI setup, a small amount of information exchange is allowed a priori, between the two marginal densities and the posterior level of information exchanged is heavily affected by the empirical evidence. For the SI case, a large amount of information exchange is desired instead between the two sets of data. In the SI, we set $(\zeta_{11} = 100, \zeta_{21} = 400)$ and $(\zeta_{12} = 100, \zeta_{22} = 200)$ in the Mix1 and Mix3 examples and $(\zeta_{11} = 10, \zeta_{21} = 100)$ and $(\zeta_{12} = 200, \zeta_{22} = 100)$ in the Mix2 example. These settings correspond to a very concentrated prior and a high prior dependence level between the two marginal densities $Cor(G_1, G_2) = 0.7609$ and $Cor(G_1, G_2) = 0.9343$ respectively.

For both the WI and SI settings, the Gibbs sampler, presented in the previous section, was run for 20,000 iterations. The raw output of the MCMC chain for the number of clusters is given in Fig. 3. For the estimation of the number of clusters, a burn-in period of 10,000 samples was discarded. At each Gibbs iteration from 10,000 onwards, a sample $(Y_{1,n+1}, Y_{2,n+1})$ from the predictive was taken. The solid lines in Figure 4 show the estimated predictive distributions using 10,000 samples from the Gibbs and the original sets of data.

Fig. 5 shows the raw output and the ergodic average of the MCMC chain for the parameters α_1 and α_2 in the WI and SI prior settings.

5.2. β_1–**DDP**(α, H_0) **mixtures of vector autoregressive processes.** In parametric models for the growth rate of the industrial production (business cycle), 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 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. In particular, the estimation of the number of regimes is an important issue studied in many papers (e.g., Kim and Murray [2002], Kim and Piger [2000] and Krolzig [2000]). The estimation of the number of regimes is still an open issue in the analysis of the business cycle. Moreover, specifically it is interesting to verify whether the strong contraction in 2009 calls for the use of a higher number of regime than three or four in business cycle models.

The above cited papers consider parametric models with a regime-switching mechanism and use some model selection criteria to estimate the number of regimes. Conversely, in this paper, we propose a non–parametric approach to the joint estimation of the number of regimes in multiple time series. We assume our Dirichlet mixture process β_1–DDP(α, H_0) as a prior for the parameters of a vector autoregressive model (VAR) for time series data. We consider two well studied cycles of

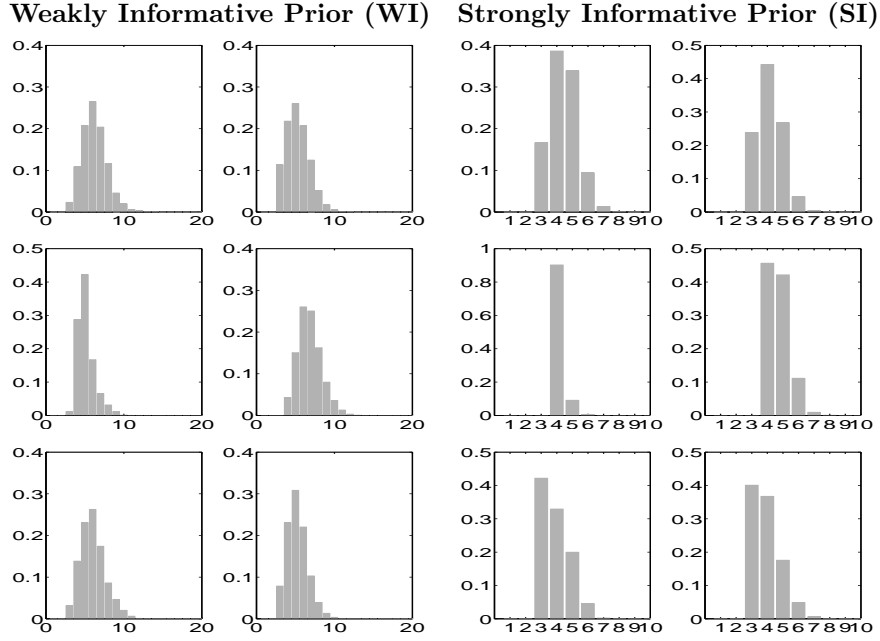**Weakly Informative Prior (WI)** **Strongly Informative Prior (SI)**

FIGURE 3. Number of clusters for different prior settings (panels WI and SI), for the first (left column) and second component (right column) and for the different models: Mix1 (first row), Mix2 (second row) and Mix3 (third row).
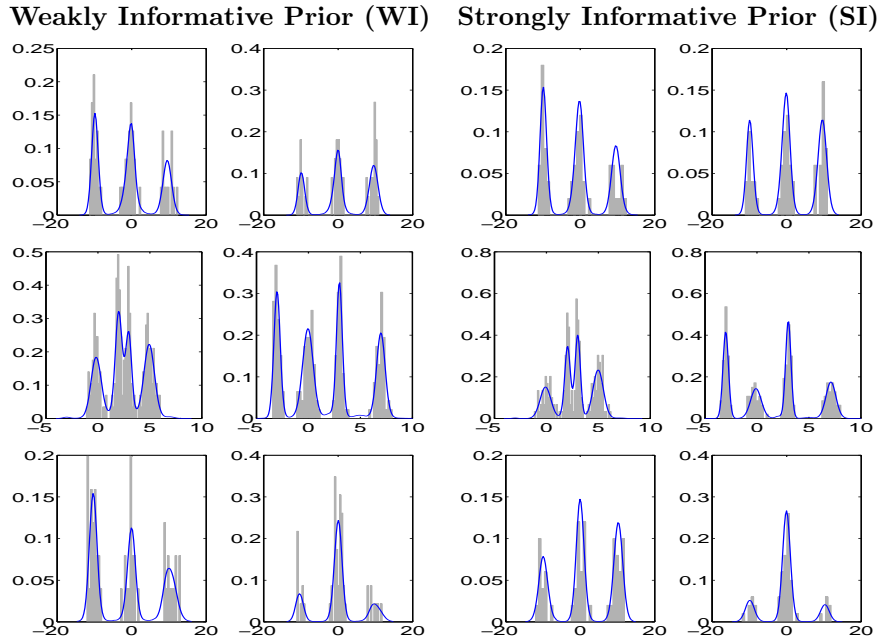
Weakly Informative Prior (WI) **Strongly Informative Prior (SI)**

FIGURE 4. Approximated predictive density (solid lines) for the two settings (panels WI and SI) and for the first (left) and second (right) component of the data from models Mix1 (first row), Mix2 (second row), and Mix3 (third row).
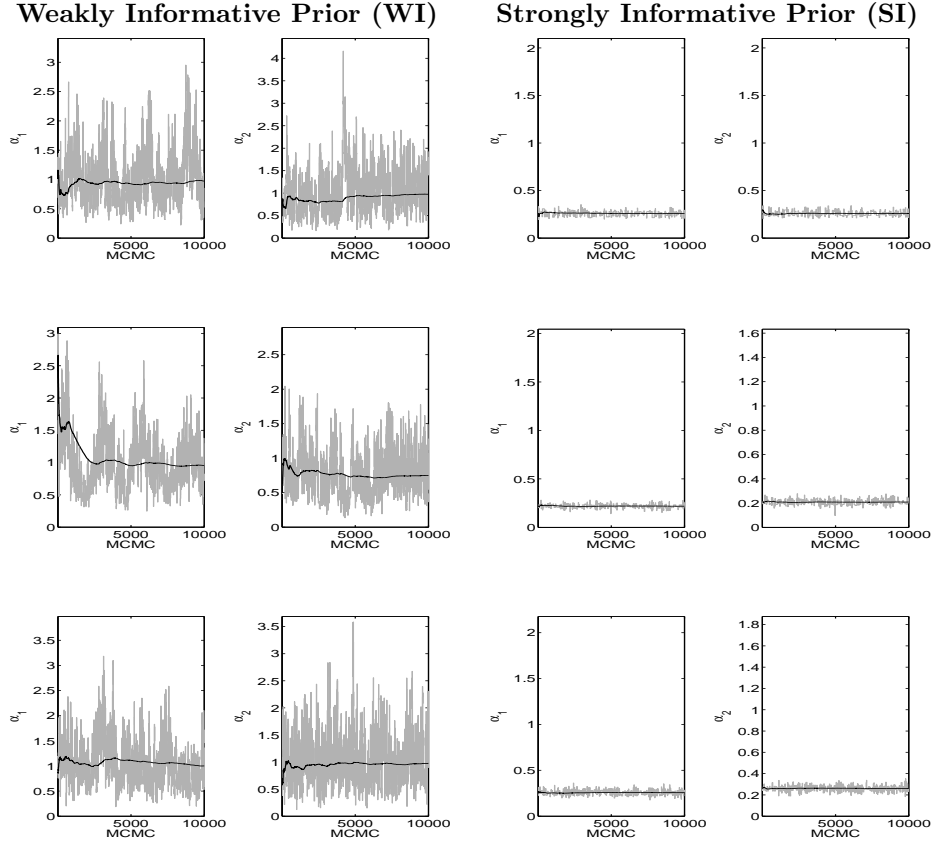
FIGURE 5. Output of the M.-H. within Gibbs for the two settings (panels WI and SI) for the parameters α_1 and α_2 of the models Mix1 (first row), Mix2 (second row), and Mix3 (third row).

the international economic system: the United States (US) and the European Union (EU) cycles. Even if the features of the regimes (or clusters) in the US and the EU growth rates are different, one could expect that the regimes in the two cycles also exhibit some dependence. For this reason, we apply a dependent multivariate Dirichlet process to account for the similarity between the clustering of the two series. In this sense, our model extends the existing literature on the use of univariate Dirichlet process prior for time series models. In this literature, the same clustering process is usually assumed for all the parameters of a multivariate model.

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, $\{X_{1t}\}_{t=1}^T$ and $\{X_{2t}\}_{t=1}^T$ respectively (see first row in Fig. 6). We take the quarterly growth rate: $Y_{it} = \log X_{it} - \log X_{it-4}$ (second row in Fig. 6). The histograms of these time series (see histograms Fig. 7) exhibit many modes that are the results of different regimes in the series. We consider the

following specification for the VAR model

$$\begin{pmatrix} Y_{1t} \\ Y_{2t} \end{pmatrix} = \begin{pmatrix} \tilde{\mu}_{1t} \\ \tilde{\mu}_{2t} \end{pmatrix} + \begin{pmatrix} Z'_t & O'_{2p} \\ O'_{2p} & Z'_t \end{pmatrix} \begin{pmatrix} \Upsilon_1 \\ \Upsilon_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

for $t = 1, \ldots, T$, where $O_{2p} = (0, \ldots, 0)' \in \mathbb{R}^{2p}$, $\Upsilon_i = (v_{1,1,i}, \ldots, v_{1,2p,i})'$ and $Z_t = (Y_{1t-1}, \ldots, Y_{1t-p}, Y_{2t-1}, \ldots, Y_{2t-p})'$ and $\varepsilon_{it} \sim \mathcal{N}(0, \tilde{\sigma}^2_{it})$ with ε_{1t} and ε_{2s} independent $\forall s, t$.

In this paper we consider four lags (i.e. $p = 4$) as for example in Hamilton [1989] and Krolzig [2000]. Moreover, as most of the forecast errors are due to shifts to the deterministic factors (see Krolzig [2000] and Clements and Krolzig [1998]), we propose a model with shifts in the intercept and in the volatility and assume a vector of Dirichlet processes as a prior for $(\mu_{it}, \sigma^2_{it})$, $i = 1, 2$

(41)

$$\begin{aligned} (\tilde{\mu}_{1t}, \tilde{\sigma}^2_{1t}) | G_1, G_2 &\overset{i.i.d.}{\sim} G_1 \\[4pt] (\tilde{\mu}_{2t}, \tilde{\sigma}^2_{2t}) | G_1, G_2 &\overset{i.i.d.}{\sim} G_2 \\[4pt] (G_1, G_2) &\sim \beta_1 - \mathrm{DDP}(\tilde{\alpha}, H_0) \\[4pt] \tilde{\alpha} &\sim \mathcal{G}(\zeta_{11}, \zeta_{21}) \mathcal{G}(\zeta_{12}, \zeta_{22}) \end{aligned}$$

where the base measure H_0 is a product of normal $\mathcal{N}(0, s^{-2} I_2)$ and inverse gamma $\mathcal{IG}(\lambda, \lambda)$.

Following the standard practice in Bayesian VAR modelling, for the parameters Υ_1 and Υ_2 we consider improper uniform prior on $\mathbb{R}^{4p}$ and obtain a multivariate normal as full conditional posterior distribution to be used in the Gibbs sampler. The charts in the first row of Fig. 7 show the predictive distributions (solid lines) generated by the non–parametric approach conditioning on all values of y_{it}, for $i = 1, 2$ and $t = 1, \ldots, T$ 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 asymmetry, excess of kurtosis, and multimodality in the data. The results from our non–parametric approach are in line with the practice of using of 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 posterior distribution of the number of clusters is given in the second row of Fig. 7. The location of the posterior mode of the histograms allows us to conclude that the non–parametric approach detects three clusters for the US cycle and 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. Moreover, we observe that the inclusion in the sample of the 2009 negative-growth (recession) period extends the validity of many past empirical findings that do not include the 2009 slowdown in the economic activity. An inspection of the posterior mean of the atoms and of the marginal clustering (see below in this section) allows us to conclude that the three clusters can have the economic interpretation of business cycle phases associated to substantially different levels of IPI growth-rates.
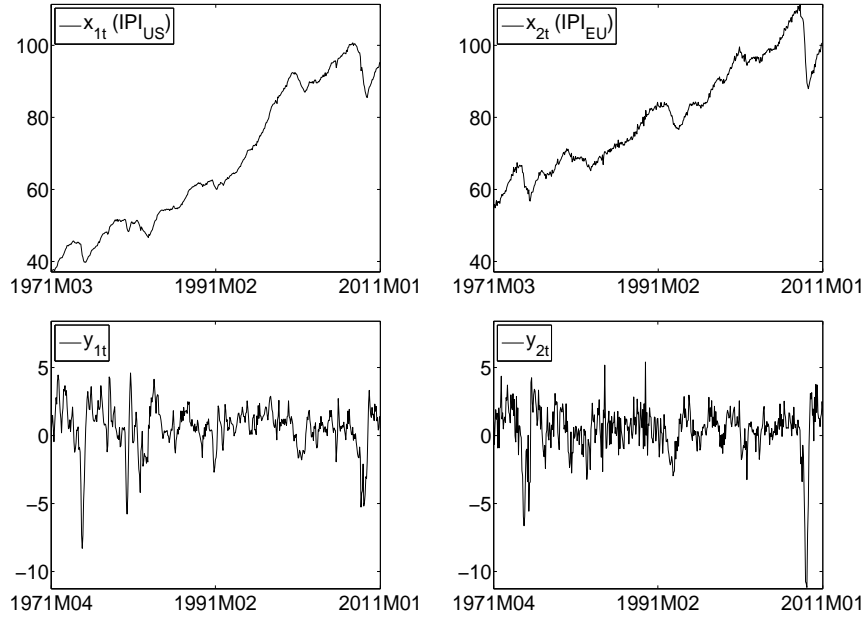
FIGURE 6. First row: Industrial Production Index (IPI) for the US (x_{1t}) and the EU (x_{2t}) at monthly frequency for the period: March 1971 to January 2011. Second row: logarithmic quarterly changes in the US IPI (y_{1t}) and the EU IPI (y_{2t}) variables.
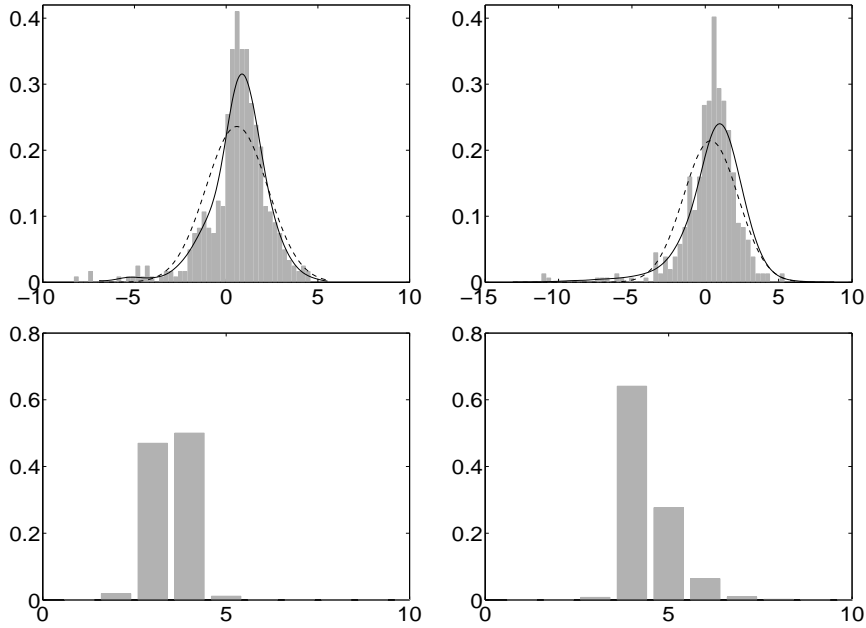

FIGURE 7. First row: IPI log-changes (histogram), predictive distribution (solid line), best normal (dashed line). Second row: number of clusters.
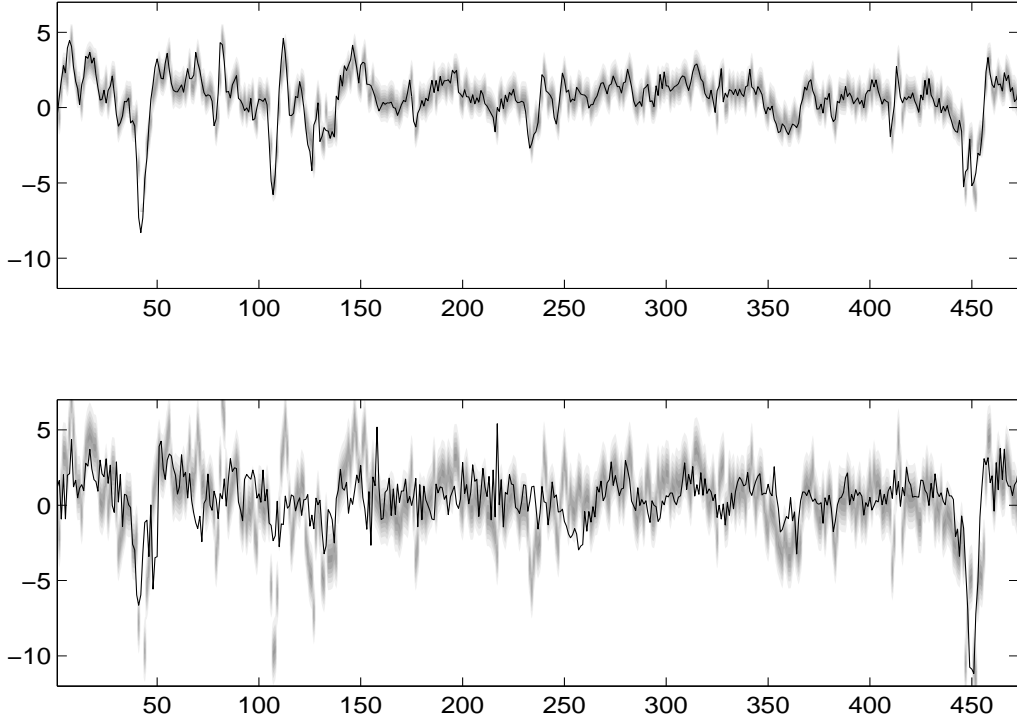
FIGURE 8. US and EU IPI growth rates (black lines) and predictive densities (gray areas) evaluated sequentially for $t = 1, \ldots, T$, at the values of the predictors $y_{it-1}, \ldots, y_{it-p}$, for $i = 1, 2$.

The results for the US and the EU cycles are, 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 an interesting one because it suggests that four components are needed in order to capture the effects of the 2009 recession phase (see Fig. 7). As a consequence of the 2009 recession, a long left tail present in the predictive (solid line in Fig. 7) is fatter than the tail of the best normal (dashed line in the same figure).

Fig. 8 shows the sequence of predictive densities (gray area) indexed by time t, for $t = 1, \ldots, T$. The predictive density for y_{it} has been estimated conditionally on the whole set of data and has been evaluated sequentially over time at the current values of the predictors $y_{it-1}, \ldots, y_{it-p}$, for $i = 1, 2$. In this figure, the effects of the recession are evident from the presence of non-negligible probability values in correspondence of extremely negative growth rates that were not realized before 2009. Similarly, we found that, in both the expansion and recession phases, posterior distribution of the atoms exhibit multimodality and asymmetry. As an example Fig. 9, shows the approximated posterior of the atoms μ_{it} and σ_{it} in periods of expansion ($t = 430$) and recession ($t = 450$). The posterior distribution of μ_{it} exhibits two modes in the positive half of the real

t=430 (1st of July 2007)
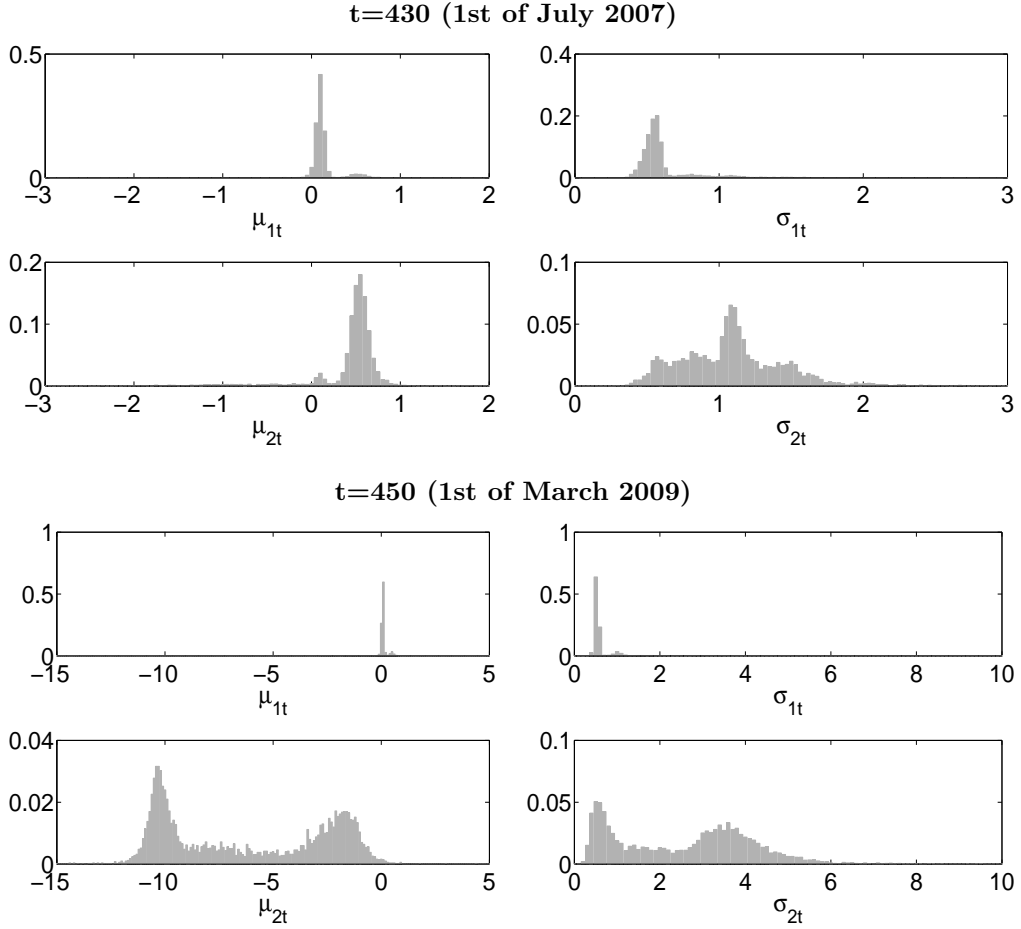

t=450 (1st of March 2009)

FIGURE 9. Posterior density approximations for the atoms of the US $((\mu_{1t}, \sigma_{1t}))$ and the EU $((\mu_{2t}, \sigma_{2t}))$ growth rates y_{1t} and y_{2t} during expansion $(t = 430)$ and recession period $(t = 450)$.

line during an expansion phase and two modes in the negative half during a recession phase (first column of Fig. 9). From the second column of the same figure, one can conclude that the volatility posterior distribution for both the US and the EU is more concentrated around lower values in expansion periods.

In order to identify the different components of our DP mixture model, we compute the posterior clustering of the data and the associated values of the atoms for each observations and country. 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 on the posterior pairwise probabilities of joint classification $P\{D_{is} = D_{jt}|Y\}$. To estimate this matrix, one can use the following pairwise probability matrix:

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

Posterior Clustering for the US data

Posterior Clustering for the EU data

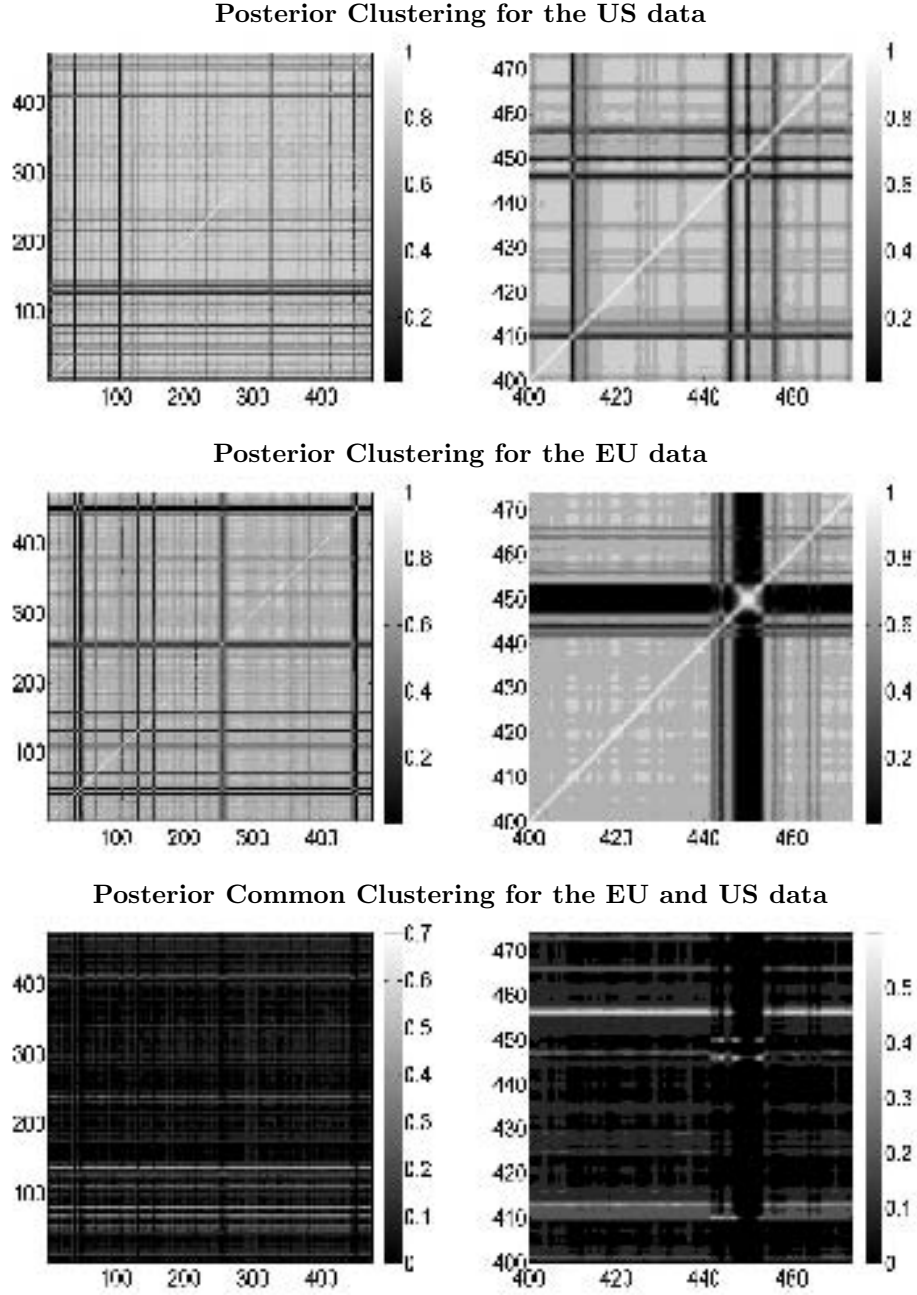Posterior Common Clustering for the EU and US data

FIGURE 10. Pairwise posterior probabilities for the clustering of the US data $P_{11,st}$ and the EU data $P_{22,st}$, and the common clustering between the US and EU data, $P_{12,st}$ for $s, t \in \{1, \ldots, T\}$

that is estimated by using every pair of allocation variable $D_{is}^l \, D_{jt}^l$, with $s, t = 1, \ldots, T$ and over all the $l = 1, \ldots, M$ MCMC iterations. In Dahl [2006]'s algorithm, one needs to evaluate $P_{ij,st}$ for $i = j$ and $i = 1, 2$. The least square marginal clustering $D_{i,LS}$ is the clustering $D_i^{l_i} = (D_{i1}^{l_i}, \ldots, D_{iT}^{l_i})$ (see Fig. 10) sampled at the l_i-th iteration which minimizes the sum of squared deviations from
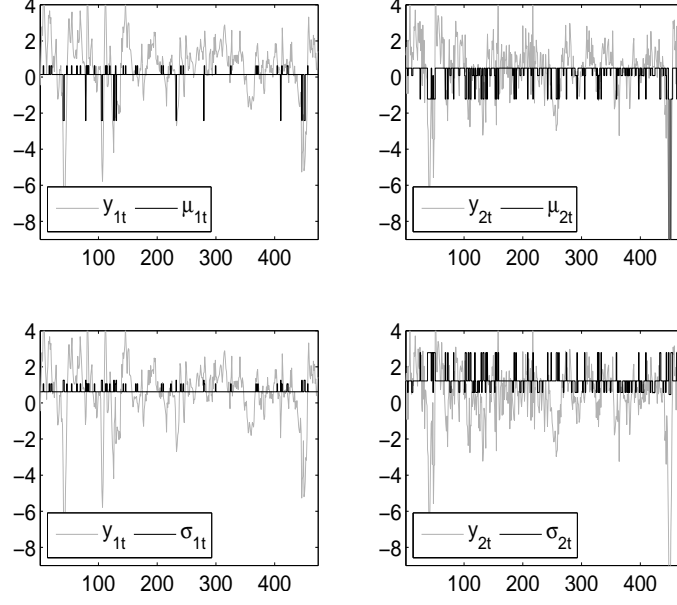
FIGURE 11. Atom $(\hat{\mu}_{it}^{(l_i)}, \sigma_{it}^{(l_i)})$ associated to the LS marginal clustering l_i for $i = 1, 2$.

the pairwise posterior probability:

$$l_i = \operatorname*{arg\,min}_{l \in \{1, \dots, M\}} \sum_{t=1}^{T} \sum_{s=1}^{T} \left(\delta_{D_{is}^l}(D_{it}^l) - P_{ii,st} \right)^2.$$

More specifically, 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 clearly detect the presence of vertical and horizontal dark gray bands. They correspond to observations that do not cluster frequently together with other observations and that are associated with negative growth rates. A similar remark is true for the light gray areas. In the second column of Fig. 10, one can see the different behavior of the clustering for the US and the EU during the 2009 crisis.

Finally, the assumption in Eq. (7) implies that the set of atoms sampled at every MCMC iteration is the same for the two series. This makes the allocation variables D_{1t}^l and D_{2s}^l comparable. For this reason, we apply Dahl [2006]'s algorithm to study the posterior probability $P_{ij,st}$ that two observations, each one from a different series (i.e. $i \neq j$), belong to the same cluster. That gives a measure of association between the two clustering, induced by the dependent DP, for the two series. The estimated pairwise probability is given in the third row of Fig. 10 that shows the probability that two observations, one of the US cycle and another one of the EU cycle, belong to the same cluster. The white and light gray lines show that the two marginal clustering share some atoms.

The least square clustering allows us to find the posterior clustering of the data and to identify the different clusters. For the US cycle, the observations cluster together in three groups (see Fig. 11) and the atoms associated with the three clusters are $(\mu_{1t}, \sigma_{1t}) \in \{(-2.4142, 1.3625), (0.136, 0.6166), (0.6216, 1.0618)\}$ and lead to the identification of the cluster as recession, normal expansion, and strong expansion phases. For the EU cycle, the observations are classified in four groups (see Fig. 11) and the atoms are $((-10.6170, 0.6600), (-1.2183, 2.7857), (0.0760, 0.5794), (0.4909, 1.2165))$ and are interpreted as strong recession, normal recession, normal expansion, and strong expansion phases. These results on the features of the cycle phases are coherent with the recent findings in the business cycle literature with an exception for the EU cycle, which presents a fourth cluster of observations with very high negative growth rate (-10.6170) corresponding to the 2009 recession period.

6. Conclusions

We introduce a new class of multivariate dependent Dirichlet processes for modeling vectors of random measures. We discuss some properties of the process. We apply the dependent Dirichlet process to the context of non–parametric Bayesian inference and provide an efficient MCMC algorithm for posterior computation. Since our process is particularly suitable for groups of data that exhibit a different clustering behavior, we apply it to multiple time series analysis. We provide an original application to the joint analysis of the US and the EU business cycles and show that our non–parametric Bayesian model is able able to highlight some important issues for this kind of data.

Appendix A. The algorithm

In the Block Gibbs Sampler described in Section 4 in principle one needs to sample an infinite number of V_k and $\tilde{\varphi}_k$. But in order to proceed with the chain it suffices to sample a finite number of V_ks to check condition (30) and the finite number of $\tilde{\varphi}_k$ to be used in (29).

For the sake of clarity we summarize here the blocked Gibbs sampling algorithm.

THE ALGORITHM

- INITIAL STEP. Initialize U, $D, \tilde{\alpha}$. With D compute D^* by using (22).

- UPDATING STEP. Suppose to have a sample of all the variables involved in the algorithm. This variables that comes from the previous step is labeled with "old". The variables that will be generated in the next step are labeled with "new".

  (1) The $(V^*, \tilde{\alpha})|new$ are sampled by using the $D|old$ and (24) with Metropolis within Gibbs step as described in Subsection 4.2.

  (2) The $V^{**}|new$ are sampled by using the $D|old$ and $\tilde{\alpha}|new$ by a Metropolis step as described in Subsection 4.2.

  (3) The $U|new$ are sampled by using the $(V^*, V^{**}, \tilde{\alpha})|new$ and (28);

  (4) $N_{i,j}^*$ are computed by using (30), with $U|new$ and $V_j|new$. If some $V_k|new$ with $k > D^*|old$ are needed they are sampled from the prior $P\{V_k \in dv_k|\tilde{\alpha}\}$.

  (5) The $\tilde{\varphi}_j|new$ for $j = 1, \ldots, N^*$, with $N^* := \max_{i=1,2} \max_{1 \leq j \leq n_i}(N_{i,j}^*)$, are sample by using (23) and $D|old$ as described in (23) and in Section 5.1.

  (6) The $D|new$ are sampled by using (29) with $U|new$, $V^{(n)}|new$ and $\tilde{\varphi}_1|new, \ldots, \tilde{\varphi}_{N^*}|new$.

Appendix B. Proofs

Proof of Proposition 12. First of all observe that

$$
\begin{aligned}
\mathbb{E}[G_1(A)G_2(B)] &= \sum_{h \geq 1, k \geq 1} \mathbb{E}[\mathbb{I}_A(\tilde{\varphi}_{1k})\mathbb{I}_B(\tilde{\varphi}_{2h})]\mathbb{E}[W_{1k}W_{2h}] \\
&= \sum_{h \geq 1, k \geq 1, h \neq k} \mathbb{E}[\mathbb{I}_A(\tilde{\varphi}_{1k})]\mathbb{E}[\mathbb{I}_B(\tilde{\varphi}_{2h})]\mathbb{E}[W_{1k}]\mathbb{E}[W_{2h}] \\
&\quad + \sum_{h \geq 1} \mathbb{E}[\mathbb{I}_{A \times B}(\tilde{\varphi}_{1h}, \tilde{\varphi}_{2h})]\mathbb{E}[W_{1h}W_{2h}].
\end{aligned}
$$
(42)

Now note that the following equalities hold

$$(43) \qquad \sum_{h \geq 1} \mathbb{E}\left[S_{1h}S_{2h} \prod_{m \leq h-1}(1-S_{1m})(1-S_{2m})\right] = \frac{\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}$$

$$(44) \qquad \sum_{h \neq k} \mathbb{E}\left[S_{1h}S_{2k} \prod_{m \leq h-1}(1-S_{1m}) \prod_{l \leq k-1}(1-S_{2l})\right] = \frac{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-2\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}$$

Combining (42) with (43)-(44) it follows that

$$\mathbb{E}[G_1(A)G_2(B)] = G_0(A \times B \times \mathbb{X}^{r-2})\frac{\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}$$

$$+G_{01}(A)G_{02}(B)\frac{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-2\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}.$$

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

$$(45) \quad Cov[G_1(A),G_2(B)] = \frac{\mathbb{E}[S_{11}S_{21}]}{\mathbb{E}[S_{11}]+\mathbb{E}[S_{21}]-\mathbb{E}[S_{11}S_{21}]}[G_0(A \times B \times \mathbb{X}^{r-2}) - G_{01}(A)G_{02}(B)].$$

In a similar fashion

$$(46) \qquad \mathbb{E}[G_1(A)^2] = \frac{G_{01}(A)\mathbb{E}[S_{11}^2]+2G_{01}^2(A)(\mathbb{E}[S_{11}]-\mathbb{E}[S_{11}^2])}{2\mathbb{E}[S_{11}]-\mathbb{E}[S_{11}^2]}$$

$$(47) \qquad \mathbb{E}[G_2(B)^2] = \frac{G_{02}(B)\mathbb{E}[S_{21}^2]+2\mathbb{G}_{02}^2(B)(\mathbb{E}[S_{21}]-\mathbb{E}[S_{21}^2])}{2\mathbb{E}[S_{21}]-\mathbb{E}[S_{21}^2]}$$

and then

$$(48) \qquad Var[G_1(A)] = G_{01}(A)(1-G_{01}(A))\frac{\mathbb{E}[S_{11}^2]}{2\mathbb{E}[S_{11}]-\mathbb{E}[S_{11}^2]}$$

$$(49) \qquad Var[G_2(B)] = G_{02}(B)(1-G_{02}(B))\frac{\mathbb{E}[S_{21}^2]}{2\mathbb{E}[S_{21}]-\mathbb{E}[S_{21}^2]}$$

$\square$

Proof of Corollary 5. By direct calculation or using the results in Nadarajah and Kotz [2005] one obtains

$$\mathbb{E}(S_{21}) = \mathbb{E}(S_{11}) = \frac{1}{1+\alpha_1+\alpha_2}, \quad \mathbb{E}(S_{21}^2) = \mathbb{E}(S_{11}^2) = \frac{2}{(1+\alpha_1+\alpha_2)(2+\alpha_1+\alpha_2)}$$

$$\mathbb{E}(S_{21}S_{11}) = \frac{B(2,\alpha_1)B(2,\alpha_1)B(\alpha_2,\alpha_1+3)}{B(1,\alpha_1)B(1,\alpha_1)B(\alpha_1+1,\alpha_2)} = \frac{\alpha_1+2}{(\alpha_1+1)(\alpha_1+\alpha_2+1)(\alpha_1+\alpha_2+2)}$$

for (H1) and

$$\mathbb{E}(S_{11}) = \frac{1}{1+\alpha_1+\alpha_2}, \ \mathbb{E}(S_{21}) = \frac{1}{1+\alpha_1},$$

$$\mathbb{E}(S_{11}^2) = \frac{2}{(1+\alpha_1+\alpha_2)(\alpha_1+\alpha_2+2)}, \ \mathbb{E}(S_{21}^2) = \frac{2}{(1+\alpha_1)(2+\alpha_1)}$$

$$\mathbb{E}(S_{21}S_{11}) = \frac{B(3,\alpha_1)B(2+\alpha_1,\alpha_2)}{B(1,\alpha_1)B(\alpha_1+1,\alpha_2)} = \frac{2}{(2+\alpha_1)(1+\alpha_1+\alpha_2)}$$

for (H2). Hence the correlation between the two random measures is

$$Cor(G_1(A), G_2(A)) = \frac{\mathbb{E}[S_{21}S_{11}]}{1 - \mathbb{E}[(1 - S_{21})(1 - S_{11})]} \sqrt{\frac{(2\mathbb{E}[S_{21}] - \mathbb{E}[S_{21}^2])(2\mathbb{E}[S_{11}] - \mathbb{E}[S_{11}^2])}{\mathbb{E}[S_{11}^2]\mathbb{E}[S_{21}^2]}}$$

$$= \left(\frac{(\alpha_1 + \alpha_2 + 1)(\alpha_1 + 2)}{2(\alpha_1 + 1)(\alpha_1 + \alpha_2 + 1) - (\alpha_1 + 2)} \right)$$

and

$$Cor(G_1(A), G_2(A)) = \frac{\mathbb{E}[S_{21}S_{11}]}{1 - \mathbb{E}[(1 - S_{21})(1 - S_{11})]} \sqrt{\frac{(2\mathbb{E}[S_{21}] - \mathbb{E}[S_{21}^2])(2\mathbb{E}[S_{11}] - \mathbb{E}[S_{11}^2])}{\mathbb{E}[S_{11}^2]\mathbb{E}[S_{21}^2]}}$$

$$= \frac{2(\alpha_1 + 1)}{(\alpha_1 + 2)(2\alpha_1 + \alpha_2 + 1) - \alpha_1} \sqrt{(\alpha_1 + 1)(\alpha_1 + \alpha_2 + 1)}$$

for (H1) and (H2), respectively.

$\square$

Proof of (18). 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 $S_{i1} = V_0 V_1 \ldots V_{r-i}$, one gets

$$S_{i,1} S_{j,1} = V_0^2 V_1^2 \ldots V_{r-j}^2 V_{r-j+1} \ldots V_{r-i}.$$

After some computations, using the fact that V_j are independent,

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

In addition, one has

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

Set

$$C_{i,j} := \frac{\mathbb{E}[S_{i,1} S_{j,1}]}{\mathbb{E}[S_{i,1}] + \mathbb{E}[S_{j,1}] - \mathbb{E}[S_{i,1} S_{j,1}]} \sqrt{\left(\frac{2\mathbb{E}[S_{i,1}]}{\mathbb{E}[S_{i,1}^2]} - 1 \right) \left(\frac{2\mathbb{E}[S_{j,1}]}{\mathbb{E}[S_{j,1}^2]} - 1 \right)}.$$

Simple algebra gives

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

and

$$\frac{\mathbb{E}[S_{i,1} S_{j,1}]}{\mathbb{E}[S_{i,1}] + \mathbb{E}[S_{j,1}] - \mathbb{E}[S_{i,1} S_{j,1}]}$$

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

That is

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

which gives (18) since $C_{i,j} = corr(G_i(A), G_j(A))$. $\square$

Full-conditionals. The joint distribution of $[V, \tilde{\varphi}, U, D, Y, \tilde{\alpha}]$ is

$$P\{V \in dv, \tilde{\varphi} \in d\varphi, Y \in dy, U \in du^{(n)}, D = d^{(n)}, \tilde{\alpha} \in (d\alpha_1, d\alpha_2)\}$$

(50)
$$= \Big[\prod_{i=1,2} \prod_{j=1}^{n_i} \mathbb{I}\{u_{ij} < w_{i,d_{i,j}}\} \mathcal{K}(y_{i,j}|\varphi_{d_{i,j}}) \Big] \otimes_{i=1,2} \otimes_{j=1}^{n_i} dy_{ij} du_{ij}$$

$$\otimes_{k \geq 1} \Big[P\{V_k \in dv_k | \tilde{\alpha} = (\alpha_1, \alpha_2)\} \otimes G_0(d\varphi_k) \Big] \otimes P\{\alpha \in (d\alpha_1, d\alpha_2)\}$$

where $w_{i,k} = v_{0k} v_{ik} \prod_{j<k}(1 - v_{0j} v_{ij})$, with the convenction that $v_{2k} = 1$, for every k, under (H_2).

Proof of (24)-(25). From (50) one gets

$$P\{V \in dv, \tilde{\alpha} \in (d\alpha_1, d\alpha_2) | Y^{(n)}, \tilde{\varphi}, D\}$$

$$\propto \Big[\prod_{i=1,2} \prod_{j=1}^{n_i} w_{i,D_{ij}} \Big] \otimes_{j \geq 1} P\{V_j \in dv_j | \tilde{\alpha} = (\alpha_1, \alpha_2)\} P\{\tilde{\alpha} \in (d\alpha_1, d\alpha_2)\}.$$

Now note that

$$\prod_{i=1,2} \prod_{j=1}^{n_i} w_{i,D_{ij}} = \prod_{j=1}^{D^*} v_{0j}^{A_{1j}+A_{2j}} v_{1j}^{A_{1j}} v_{2j}^{A_{2j}}$$

$$(1 - v_{0j} v_{1j})^{B_{1j}} (1 - v_{0j} v_{2j})^{B_{2j}}.$$

$\square$

REFERENCES

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.

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.

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. Cambridge University Press, 2006.

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.

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 Carin L. The Matrix Stick-Breaking Process. *Journal of the American Statistical Association*, 103:317–327, 2008.

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.

T.S. Ferguson. Prior distributions on spaces of probability measures. *The Annals of Statistics*, 1: 615–629, 1974.

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

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. 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. Nicolerisa, and Walker S.G. 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.

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, 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. Forthcoming in *Statistics and Computing*, 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.

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.

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.

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

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. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 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 Gelfand A.E. The nested Dirichlet process. *Journal of the American Statistical Association*, 103:1131–1144, 2008.

A. Rodriguez, D.B. Dunson, and Gelfand A.E. 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.

E. Sudderth and M.I. Jordan. Shared segmentation of natural scenes using dependent Pitman-Yor processes. In D. Schuurmans D. Koller, Y. Bengio and Editors L. Bottou, editors, *Advances in Neural information processing systems*, volume 21. 2009.

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.

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