Bayesian Analysis (2024) **19**, Number 1, pp. 21–48

Joint Random Partition Models for Multivariate Change Point Analysis*

José J. Quinlan†, Garritt L. Page‡, and Luis M. Castro§,¶,‖

Abstract. Change point analyses are concerned with identifying positions of an ordered stochastic process that undergo abrupt local changes of some underlying distribution. When multiple processes are observed, it is often the case that information regarding the change point positions is shared across the different processes. This work describes a method that takes advantage of this type of information. Since the number and position of change points can be described through a partition with contiguous clusters, our approach develops a joint model for these types of partitions. We describe computational strategies associated with our approach and illustrate improved performance in detecting change points through a small simulation study. We then apply our method to a financial data set of emerging markets in Latin America and highlight interesting insights discovered due to the correlation between change point locations among these economies.

Keywords: correlated random partitions, multiple change point analysis, multivariate time series.

1 Introduction

Change point analyses identify times or positions of an ordered stochastic process that undergo abrupt local changes. These abrupt changes are typically seen as shifts in expectation, variability, or shape of an underlying distribution (or some combination of the three). Methods that detect change points have been employed in a variety of fields, including finance (Wood et al., 2021), climatology (Gupta et al., 2021), and ecology (Jones et al., 2021) to name a few. Due to this, many change point methods have been proposed in the statistical literature both in a univariate (see for example Arellano-Valle et al., 2013) and a multivariate setting (see Truong et al., 2020, for a comprehensive review).

The phenomenon that motivates our research is the so-called *financial contagion* or simply *contagion*. This phenomenon can be understood as the spread of financial crises from one country to another (see for example Lowell et al., 1998; Valdés, 2000; de P.

arXiv: 2201.07830

*José J. Quinlan and Luis M. Castro gratefully recognizes the financial support provided by the Agencia Nacional de Investigación y Desarrollo de Chile (ANID) through Grant FONDECYT 3190324. Luis M. Castro also acknowledges support from Grant FONDECYT 1220799.

†EY, Technology Consulting, Chile, jjquinla@uc.cl

‡Department of Statistics, Brigham Young University, Provo Utah, page@stat.byu.edu

§Department of Statistics, Pontificia Universidad Católica de Chile, Chile, Santiago de Chile, mcastro@mat.uc.cl

¶Millennium Nucleus Center for the Discovery of Structures in Complex Data, Chile.

‖Centro de Riesgos y Seguros UC, Pontificia Universidad Católica de Chile, Chile, Santiago Chile.

 https://doi.org/10.1214/22-BA1344

Filleti et al., 2008, among others). To illustrate *contagion* consider the price and returns of the five markets displayed in Figure 1. Note that the overall trend of the price in the Latin American markets (Argentina, Brazil, Chile, and Mexico) seem to coincide. However, the USA market (Dow Jones index) presents a different trend during the same observation period (1995 to 2001). It is important to note that, in the second half of 1998, Dow Jones suffered a slight crash (a change in volatility according to Figure 1, left column) due to the Russian financial crisis and the Long Term Capital Management episode. Consequently, we hypothesize that a change point in a mature market such as the US could produce change points in emerging markets such as those from Latin American, or simply, the *financial contagion* between the US market and Latin American markets could increase the chance of change points occurring in the later markets when they occur in the former. Consequently, the method we develop will incorporate dependence between change point probabilities across multiple processes which could potentially improve the ability of detecting a change point compared to an independent model.

One commonly used approach in the statistical literature for detecting time-series change points is based on product partition models (PPM). These models, which were introduced by Barry and Hartigan (1992), assume that (a) the number and positions of change points are random and, (b) observations within the same block are assumed to follow the same distribution. Thus, the inferential problem reduces to identifying a partition where each cluster is a collection of consecutive data points and then estimate parameters associated with each cluster's assumed data model. From a Bayesian viewpoint, a prior distribution on the space of partitions, which are restricted to be contiguous, is needed. Barry and Hartigan (1992) use a prior for which change point probabilities are based on the so-called cohesion function studied in Yao (1984). Since Barry and Hartigan (1992) many other PPM type approaches to change point analysis have been developed (see, for example Loschi and Cruz, 2002; Loschi et al., 2003; Loschi and Cruz, 2005; Loschi et al., 2005, 2010; Martínez and Mena, 2014; García and Gutiérrez-Peña, 2019; Pedroso et al., 2021).

Most of the existing approaches based on PPMs for detecting change points in time series treat the series independently. For example, the PPM-based proposal of Wang and Emerson (2015) examines a class of change point problems on connected graphs, assuming a regression model within each block of the partition of the graph. This proposal is implemented in the `bcp` package (Erdman and Emerson, 2007) available in software `R` (R Core Team, 2022). Another univariate approach based on the work of Barry and Hartigan (1992) is proposed by Loschi and Cruz (2002). Their method allows detecting multiple change points in a particular series using an easy-to-implement Gibbs sampling scheme. This method is available in the `R` package `ppmSuite` (Page and Quinlan, 2022).

However, in the presence of *contagion*, the information available from several series could improve the accuracy of the change point detection mechanism compared to when series are treated independently. In general, the strategies for detecting change points in the multivariate context focus on detecting changes in the joint distribution of the coordinates of a multivariate process across time. For example, Cheon and Kim (2010)

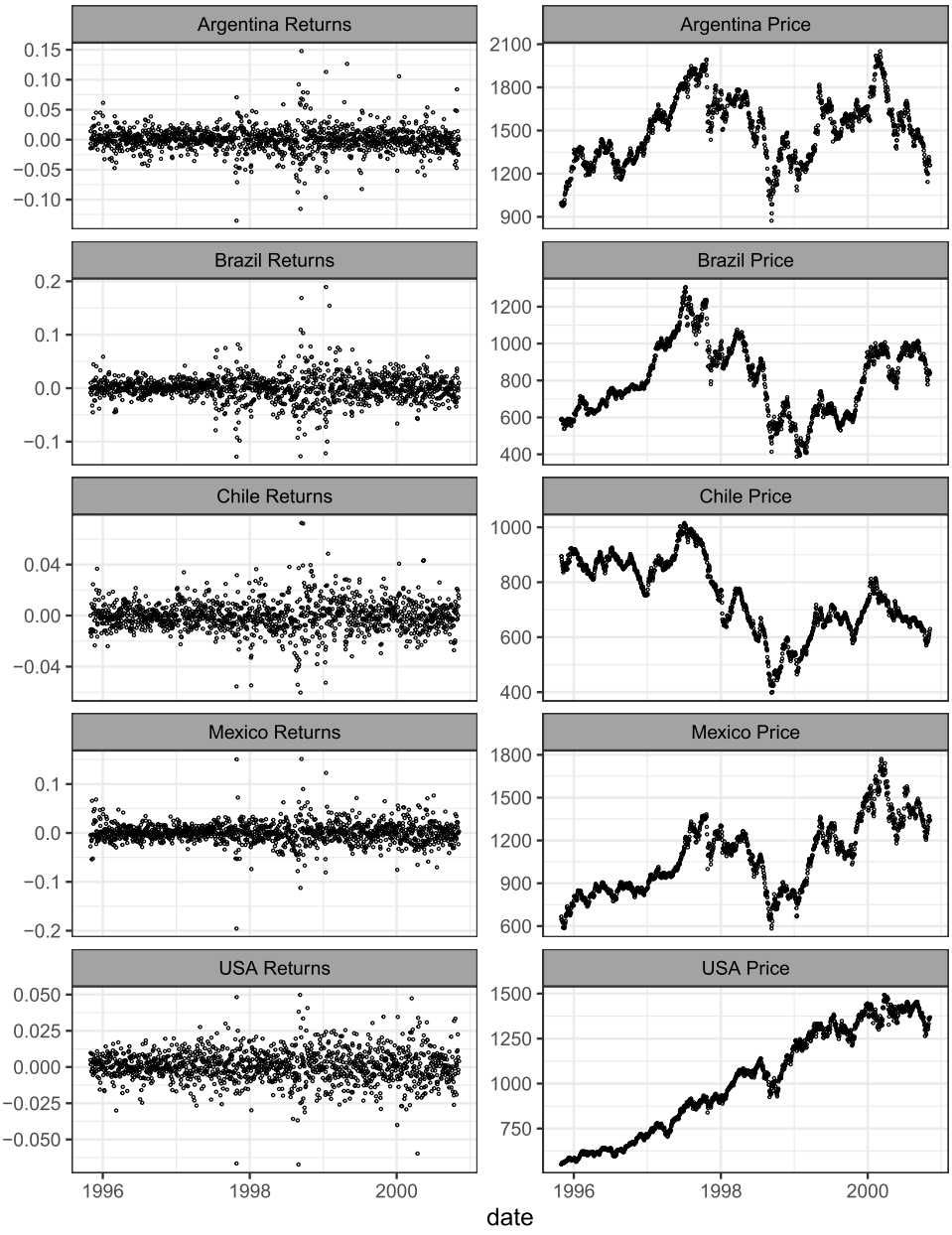

Figure 1: Daily returns and price of stock market indexes of Argentina, Brazil, Chile, Mexico and USA. Returns are calculated using $R_t = (P_t - P_{t-1})/P_{t-1}$ for $t \in \{2, \ldots, n\}$, where P_t is the price of day t.

developed a Bayesian model for detecting changes in the mean and variance when data follow a multivariate normal distribution. Their approach considers a latent vector that identifies change point positions for partitioning the observations. Nyamundanda et al. (2015) proposed an approach that combines a PPM with a probabilistic principal component analysis. This approach identifies change points by imposing a partition structure on parameters of interest, reducing the data dimension via the principal component analysis. The cluster structure is incorporated at the level of the observations and the latent variables used in the principal component setup. Therefore, the approach is called a product partition latent variable model (PPLVM). The key feature of the PPLVM is that it can be used to detect distributional changes in the mean and covariance of the series, even in high-dimensional settings. Recently, Jin et al. (2022) considered a Bayesian hierarchical model to detect mean shifts in multiple sequences by modeling mean differences. The authors considered an exchangeable random order distribution (Martínez and Mena, 2014) to construct the prior distribution for the set of change points. Notably, they use a Pitman-Yor process (for more details see De Blasi et al., 2015). This prior provides an exciting balance between being informative and noninformative about the number of clusters and, therefore, the number of change points (Lijoi et al., 2007). The proposal of Jin et al. (2022) considers the use of nonlocal priors for the mean difference which informs the detection error of change points.

In contrast to the approaches described above, there are models that do not introduce explicitly a distribution for partitions of contiguous clusters. For example, Killick et al. (2012) considered a procedure for detecting change points by minimizing a cost function, which detects the optimal number and location of change points with a linear computational complexity under mild conditions. Tveten et al. (2022) also minimize a cost function when searching for change points in cross-correlated processes. Matteson and James (2014) proposed a robust nonparametric method using a divergence measure based on Euclidean distances. With this method, the authors showed that it is possible to detect any distributional change within an independent sequence of random variables without making any distributional assumption beyond the existence of the α-th absolute moment, for some $\alpha \in (0, 2)$. Padilla et al. (2021) proposed a novel change point detection algorithm based on the Kolmogorov–Smirnov statistic and showed that it is nearly minimax rate optimal under suitable conditions.

Other approaches for detecting changes in a multivariate process that is more inline with what we propose assume that each process has its own change point structure. Harlé et al. (2016) introduce a set of independent binary vectors whose entries indicate which coordinate of the multivariate process changes. They do this using a composite marginal likelihood based on Wilcoxon's rank-sum test and a suitable prior for the binary vectors. Their approach incorporates dependence among change points between different coordinates. Moreover, Fan and Mackey (2017) introduce a set of change point indicator variables for each coordinate and time such that the prior change point probabilities for all coordinates at a fixed time are the same but change through time.

Our approach, which is motivated by the *contagion* phenomena, considers simultaneous changes in all parameters associated with a particular process dealing with multiple univariate time series (not all of which are necessarily the same type of response). Thus,

each series has its change point structure where the change points between them may or may not coincide in time. It is important to stress that several approaches for detecting change points in the context of contagion and financial data are proposed in the literature. For example, Zhu et al. (2013) propose a change point detection approach based on copula for detecting a financial contagion phenomena in Chinese banking. In this setup, the detection strategy is focused on the analysis of the dependence structure of the data, defining a change point where the dependence structure changes. Zhu et al. (2015) applied the same strategy for detecting change points for the subprime crisis in American banking. More recently, Song and Kang (2021a) propose the use of generalized autoregressive conditional heteroscedasticity of order (1,1), namely, GARCH (1,1) model for detecting change points combined with a hypothesis test proposed by Song and Kang (2021b). The approach is used for studying the price of Bitcoin from January 2013 to December 2020. Finally, we note that the works of R. Loschi (see for example Loschi and Cruz, 2002; Loschi et al., 2003; Loschi and Cruz, 2005, among others) are devoted to identifying change points sequences in Latin American emerging markets using PPM. They considered a contagion effect that occurs from the mature markets (markets belonging to developing countries) to the emerging ones.

We base our approach on elements from the method described in Fan and Mackey (2017) combined with a novel multivariate extension of the PPM approach of Barry and Hartigan (1992). The resulting method takes advantage of the existing correlation in change point locations between series. This strategy requires specifying a joint prior distribution for a collection of partitions. Constructing these types of dependent partition models over a series of partitions has only very recently been considered in the literature by, for instance, Zanini et al. (2019); Page et al. (2022). The method we present is the first work that we are aware of that considers jointly modeling contiguous partitions.

The remainder of the article is organized as follows. Section 2 provides notation and background to the change point PPM. In Section 3 we describe our approach of incorporating dependence between change point probabilities, provide some theoretical properties and details with regards to computation. Section 4 describes a numerical experiment designed to study our method's ability to detect change points, and in Section 5 we apply our approach to the finance data concerning stock market returns of five countries. We close the paper with some concluding remarks in Section 6.

2 Background and Preliminaries

To make the paper self-contained, we start this section with some background related to PPMs and introduce some notation we will use throughout the paper.

2.1 Partition Definition and Notation

Without loss of generality, consider $i = 1, \ldots, L > 1$ time series $\boldsymbol{y}_i = (y_{i,1}, \ldots, y_{i,n})^\top$, each of length $n > 2$. Change points occur when the behavior of $\boldsymbol{y}_i$ undergoes sudden changes at unknown times. These times of sudden changes partition $\{1, \ldots, n\}$ into k_i contiguous sets, say $\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\}$, for some $k_i \in \{1, \ldots, n\}$. Here, $S_{i,j}$ is the jth

block of the ith series and k_i is the number of blocks in ρ_i. The space of these types of partitions will be denoted by $\mathcal{C}_n$. If we let the set $\boldsymbol{\tau}_i = \{\tau_{i,1}+1, \ldots, \tau_{i,k_i-1}+1\}$ denotes the locations at which change points in $\boldsymbol{y}_i$ occur, with $\tau_{i,0} = 0$, then $S_{i,j} = \{\tau_{i,j-1}+1, \ldots, \tau_{i,j}\}$.

An alternative way to denote a partition of contiguous clusters that will facilitate model description is based on a set of change point indicators $\boldsymbol{c}_i = (c_{i,1}, \ldots, c_{i,n-1})^\top$, such that $c_{i,t} = 1$ if time $t+1$ is a change point in $\boldsymbol{y}_i$, and $c_{i,t} = 0$ otherwise. The number of change points can be identified using $\boldsymbol{c}_i$ by noticing that $k_i = 1 + \sum_{t=1}^{n-1} c_{i,t}$. In what follows, we will use ρ_i and $\boldsymbol{c}_i$ interchangeably. With the necessary notation introduced, we next describe the change point PPM.

2.2 Change Point Product Partition Models

For the i-th sequence, the PPM is a discrete distribution on space $\mathcal{C}_n$ such that

$$\mathbb{P}(\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\}) = \frac{\prod_{j=1}^{k_i} c(S_{i,j})}{\sum_{\rho_i \in \mathcal{C}_n} \prod_{j=1}^{k_i} c(S_{i,j})},$$

where $c(S_{i,j})$ is referred to as a cohesion function and measures the *a priori* belief that elements in $S_{i,j}$ co-cluster. The change point PPM as described in Barry and Hartigan (1992), Loschi et al. (2003), and others, uses Yao (1984)'s cohesion function to assign probabilities to each element in $\mathcal{C}_n$. Yao (1984)'s cohesion function applied to contiguous $S_{i,j}$ results in

$$c(S_{i,j}\,;\, p_i) = \begin{cases} p_i(1-p_i)^{\tau_{i,j}-\tau_{i,j-1}-1}, & \text{if} \quad \tau_{i,j} < n \\ (1-p_i)^{\tau_{i,j}-\tau_{i,j-1}-1}, & \text{if} \quad \tau_{i,j} = n \end{cases}, \tag{1}$$

for some $p_i \in [0,1]$ such that $p_i = \mathbb{P}(c_{i,t} = 1 \mid p_i)$. Based on this cohesion we have that

$$\sum_{\rho_i \in \mathcal{C}_n} \prod_{j=1}^{k_i} c(S_{i,j}\,;\, p_i) = 1.$$

Thus, the change point PPM takes on the following form

$$\mathbb{P}(\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\} \mid p_i) = \prod_{j=1}^{k_i} c(S_{i,j}\,;\, p_i) = p_i^{k_i-1}(1-p_i)^{n-k_i}. \tag{2}$$

Once the partition model has been specified, the key idea behind change point modeling from a partition perspective is that observations within the same block are assumed to follow a common distribution, whereas different distributions are assumed between blocks. Following Barry and Hartigan (1992)'s approach, given ρ_i, the joint density of $\boldsymbol{y}_i$ is written as a product of k_i *data factors*, also known as marginal likelihoods, which

measure the similarity of observations within each block. More precisely,

$$\begin{aligned} f(\boldsymbol{y}_i \mid \rho_i, \boldsymbol{\xi}_i) &= \prod_{j=1}^{k_i} \mathcal{F}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\xi}_i), \\ \mathcal{F}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\xi}_i) &= \int_{\boldsymbol{\Theta}_i} \mathcal{L}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\theta}_i, \boldsymbol{\xi}_i) \mathrm{d}G_i(\boldsymbol{\theta}_i \mid \boldsymbol{\xi}_i), \end{aligned} \tag{3}$$

where $\boldsymbol{y}_{i,j} = (y_{i,t} : t \in S_{i,j})^\top$ and $\mathcal{L}_i(\,\cdot \mid \boldsymbol{\theta}_i, \boldsymbol{\xi}_i)$ is a likelihood function indexed by the set of parameters $\boldsymbol{\theta}_i \in \boldsymbol{\Theta}_i$ which are block-specific, and $\boldsymbol{\xi}_i \in \boldsymbol{\Xi}_i$ a collection of parameters that are common to all blocks. Further, $G_i(\,\cdot \mid \boldsymbol{\xi}_i)$ is a suitable prior for $\boldsymbol{\theta}_i$. The data generating mechanism (3) along with prior distributions for ρ_i and $\boldsymbol{\xi}_i$ (if applicable) completely specify the Bayesian change point PPM.

It is common to select $\mathcal{L}_i(\,\cdot \mid \boldsymbol{\theta}_i, \boldsymbol{\xi}_i)$ and $G_i(\,\cdot \mid \boldsymbol{\xi}_i)$ such that they form a conjugate pair which results in $\mathcal{F}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\xi}_i)$ being available in closed form. That said, the choices for $\mathcal{L}_i(\,\cdot \mid \boldsymbol{\theta}_i, \boldsymbol{\xi}_i)$ and $G_i(\,\cdot \mid \boldsymbol{\xi}_i)$ in (3) can be quite general, depending on the nature of $\boldsymbol{y}_i$ and $\boldsymbol{\theta}_i$. Examples of this are data that follow an Ornstein-Uhlenbeck process with a Normal-Gamma prior for mean-precision parameters (Martínez and Mena, 2014) (here, $\boldsymbol{\xi}_i$ is a case dependency parameter with a Uniform$(0, 1)$ prior) and independent data belonging to the exponential family with a conjugate prior for the natural parameters (Loschi and Cruz, 2005) (in this case, there is no $\boldsymbol{\xi}_i$). The types of marginal likelihoods just described (and others not listed) are easily applied using our method. Even so, in what follows, we will focus on the following specification (which is suitable for changes in mean and variance for data supported on $\mathbb{R}$). Let $\boldsymbol{\theta}_i = (\mu_i, \sigma_i^2)^\top \in \boldsymbol{\Theta}_i = \mathbb{R} \times (0, +\infty)$ and

$$\begin{aligned} \mathcal{L}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\theta}_i) &= \prod_{t \in S_{i,j}} \text{Normal}(y_{i,t} \mid \mu_i, \sigma_i^2), \\ G_i(\boldsymbol{\theta}_i) &= \text{Normal}(\mu_i \mid \mu_{i,0}, (\kappa_{i,0})^{-1}\sigma_i^2)\text{Inv-Gamma}(\sigma_i^2 \mid \alpha_{i,0}, \beta_{i,0}). \end{aligned} \tag{4}$$

Here, $\mu_{i,0} \in \mathbb{R}$ and $\kappa_{i,0}, \alpha_{i,0}, \beta_{i,0} > 0$ are fixed hyperparameters, Normal$(\,\cdot \mid \mu, \sigma^2)$ denotes a normal density with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 > 0$, and Inv-Gamma$(\,\cdot \mid \alpha, \beta)$ denotes an inverse Gamma density with shape $\alpha > 0$ and scale $\beta > 0$. It is well known that the jth *data factor* induced by (4) is given by

$$\mathcal{F}_i(\boldsymbol{y}_{i,j}) = t_{n_{i,j}}(\boldsymbol{y}_{i,j} \mid 2\alpha_{i,0}, \mu_{i,0}\mathbf{1}_{n_{i,j}}, (\alpha_{i,0})^{-1}\beta_{i,0}\{\mathbf{I}_{n_{i,j}} + (\kappa_{i,0})^{-1}\mathbf{J}_{n_{i,j}}\}), \tag{5}$$

where $n_{i,j}$ is the cardinality of $S_{i,j}$, $\mathbf{1}_p \in \mathbb{R}^p$ is the vector with entries equal 1, $\mathbf{I}_p \in \mathbb{R}^{p \times p}$ is the identity matrix and $\mathbf{J}_p \in \mathbb{R}^{p \times p}$ is the matrix with all entries equal 1. Also, $t_p(\,\cdot \mid \nu, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the p-dimensional Student's t-density with degrees of freedom $\nu > 0$, location vector $\boldsymbol{\mu} \in \mathbb{R}^p$ and scale matrix $\boldsymbol{\Sigma} \in \mathbb{S}^{p \times p}$, where $\mathbb{S}^{p \times p}$ denotes the space of positive-definite matrices.

The set of hyperparameters $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})^\top$ play a crucial role in determining what constitutes a change point. For example, setting $\alpha_{i,0}$ close to one will result in (5) having thick tails so that a change point would necessarily need to be far from the center. Conversely, with a large value of $\alpha_{i,0}$ (5) approximates a normal distribution and

points not far from the center can still be change points. Consequently, thought must be dedicated to assigning values to the marginal likelihood parameters. In Section 3.2 we discuss an empirical Bayes method that produces reasonable values for them in the absence of prior information.

3 The Joint Prior Distribution on a Collection of Partitions

We now describe our approach of formulating a joint model for a sequence of partitions. As mentioned, the partition ρ_i for the i-th time series $\boldsymbol{y}_i$ has a one-to-one correspondence with $\boldsymbol{c}_i$. Thus, any prior distribution for $\mathbf{C} = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_L)$, say $\pi(\mathbf{C})$, uniquely determines a prior for $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_L)$. We start by describing our joint model as an extension of the change point PPM and then we connect it to $\pi(\mathbf{C})$, which is what we ultimately use in our approach as it facilitates computation.

In our setup, rather than consider a single probability parameter p_i for the i-th series, we define $\tilde{\boldsymbol{p}}_i = (p_{i,1}, \ldots, p_{i,n-1})^\top$ with $p_{i,t} \in [0,1]$ and extend the cohesion in (1) to

$$c^\star(S_{i,j}\,;\,\tilde{\boldsymbol{p}}_i) = \begin{cases} p_{i,\tau_{i,j}} \prod_{t=\tau_{i,j-1}+1}^{\tau_{i,j}-1}(1-p_{i,t}), & \text{if} \quad \tau_{i,j} < n, \\ \prod_{t=\tau_{i,j-1}+1}^{\tau_{i,j}-1}(1-p_{i,t}), & \text{if} \quad \tau_{i,j} = n. \end{cases} \tag{6}$$

Using the cohesion (6) for contiguous partitions still results in $\sum_{\rho_i \in \mathcal{C}_n} \prod_{j=1}^{k_i} c^\star(S_{i,j}\,;\,\tilde{\boldsymbol{p}}_i) = 1$. Therefore, the partition probabilities become

$$\mathbb{P}(\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\} \mid \tilde{\boldsymbol{p}}_i) = \prod_{j=1}^{k_i} c^\star(S_{i,j}\,;\,\tilde{\boldsymbol{p}}_i) = \prod_{t \in \boldsymbol{T}_i} p_{i,t} \prod_{t \notin \boldsymbol{T}_i} (1-p_{i,t}),$$

where $\boldsymbol{T}_i = \{\tau_{i,1}, \ldots, \tau_{i,k_i-1}\}$. Including all L partitions, the joint partition model becomes

$$\begin{aligned} \mathbb{P}(\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\} : i = 1, \ldots, L \mid \tilde{\boldsymbol{p}}_1, \ldots, \tilde{\boldsymbol{p}}_L) \quad &= \textstyle\prod_{i=1}^{L} \mathbb{P}(\rho_i = \{S_{i,1}, \ldots, S_{i,k_i}\} \mid \tilde{\boldsymbol{p}}_i) \\ &= \textstyle\prod_{i=1}^{L} \left\{ \prod_{t \in \boldsymbol{T}_i} p_{i,t} \prod_{t \notin \boldsymbol{T}_i} (1-p_{i,t}) \right\}. \end{aligned}$$

Next we induce correlation in the sequence $\rho_1, \ldots, \rho_L$ by jointly modeling the L-dimensional vector of probabilities $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{L,t})^\top$ (which are supported on the space $(0,1)^L$). To specify a multivariate distribution for $\boldsymbol{p}_t$, we consider the bijective transformation $\text{logit}(\boldsymbol{p}_t) = \left(\log\left(\frac{p_{1,t}}{1-p_{1,t}}\right), \ldots, \log\left(\frac{p_{L,t}}{1-p_{L,t}}\right)\right)^\top$ which is defined on the Euclidean space $\mathbb{R}^L$, and model it with a multivariate Student's-t distribution. The reason for selecting a Student's-t distribution instead of, for example, a multivariate normal distribution is that extreme probabilities (near 0 or 1) are more achievable due to the thicker tails of the Student's-t. In summary, the proposed model for partitions $\boldsymbol{\rho}$ can be formulated using the following hierarchical structure:

$$\begin{aligned} c_{i,t} \mid p_{i,t} &\overset{\text{ind}}{\sim} p_{i,t}^{c_{i,t}} (1-p_{i,t})^{1-c_{i,t}}, \\ \text{logit}(\boldsymbol{p}_t) &\overset{\text{iid}}{\sim} t_L(\nu_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0), \end{aligned} \tag{7}$$

with $\boldsymbol{\mu}_0 = (\mu_1^0, \ldots, \mu_L^0)^\top$, $\boldsymbol{\Sigma}_0 = (\sigma^0_{l_1,l_2} : l_1, l_2 \in \{1, \ldots, L\})$. Note that, specifying adequate values for ν_0, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ in (7) must be done with caution as $p_{i,t}$ are not invariant to their selection. In the absence of information regarding these parameters, we provide an empirical-based approach for selecting them in Section 3.2. The full likelihood of our extension to the change point PPM is simply the product of (4) across the L series. In other words, we are assuming conditional independence between series given all the partition structures. In what follows, we will refer to the model comprised of the full likelihood, (6), and (7) as the correlated change point product partition model or simply, CCP-PPM.

A correlated change point PPM based on (2) is available from our construction if $p_{i,t} = p_i$ for all $t \in \boldsymbol{T}_i$ and the "global" $(p_1, \ldots, p_L)$ are modeled as in (7). We consider this special case of our approach (which we denote as CCP-PPM0) in the simulation study of Section 4. In addition, the properties listed in Section 3.1 hold for the CCP-PPM0 as well as the CCP-PPM. This is because the propositions in Section 3.1 correspond to functionals of $c_{i,t}$ that have been averaged over change point probabilities and also because $\text{logit}(\boldsymbol{p}_t)$ in (7) are assumed to follow an iid model. It is important to stress however, that our extension (*i.e.*, introducing $p_{i,t}$) provides more flexibility for modeling simultaneous change point configurations $\boldsymbol{\rho}$ by allowing us to correlate probabilities of a change point at each time point. As a consequence, the change point indicators $c_{i,t}$ are assumed conditionally independent with their own probability $p_{i,t} \in (0, 1)$ of detecting a change ($c_{i,t} = 1$). This added flexibility permits borrowing of strength when estimating the $\boldsymbol{p}_t$ vectors. The benefits of doing this are highlighted in the simulation study of Section 4. We next discuss some properties that hold for both the CCP-PPM and the CCP-PPM0.

3.1 Properties of the Joint Model on Partitions

In this section, we provide some interesting properties that are consequences of modeling the change point indicators with (7). The proofs of all propositions are provided in the online supplementary material (Quinlan et al., 2022).

Proposition 1. *Under the assumptions in* (7)*, each of the* $\boldsymbol{c}_i$*,* $i = 1, \ldots, L$*, follows a change point PPM based on Yao's cohesion with probability parameter*

$$\phi_i = \int_{\mathbb{R}} \left\{ \frac{\exp(z)}{1 + \exp(z)} \right\} t_1(z \mid \nu_0, \mu_i^0, \sigma_{i,i}^0) \mathrm{d}z. \tag{8}$$

A consequence of Proposition 1 is that the correlation in the change point probabilities from our model only exists across the L series for a fixed t. Within a series, the probability of a change point at time t_1 is independent of time t_2. The next proposition provides an interesting result regarding the number of expected change points based on the CCP-PPM.

Proposition 2. *Under the assumptions in* (7)*, the number of change points* $(k_i - 1)$ *for the* i*th series satisfies* $(k_i - 1) \sim \text{Binomial}(n - 1, \phi_i)$*, where* ϕ_i *is given by* (8)*. Additionally, for series* i *and* s*, the distribution for* $(k_i - 1, k_s - 1)^\top$ *is a mixture of a*

product of two Poisson-Binomial distributions (Wang, 1993) with $\text{Cov}(k_i - 1, k_s - 1) = (n-1)(\varphi_{i,s} - \phi_i\phi_s)$*, where*

$$\varphi_{i,s} = \int_{\mathbb{R}^2} \left[\frac{\exp(z_i)\exp(z_s)}{\{1+\exp(z_i)\}\{1+\exp(z_s)\}} \right] t_2(\boldsymbol{z}_A \mid \nu_0, \boldsymbol{\mu}_{A,0}, \boldsymbol{\Sigma}_{A,0}) \mathrm{d}\boldsymbol{z}_A. \tag{9}$$

Here, $\boldsymbol{z}_A = (z_i, z_s)^\top$*,* $\boldsymbol{\mu}_{A,0} = (\mu_i^0, \mu_s^0)^\top$ *and* $\boldsymbol{\Sigma}_{A,0} = (\sigma^0_{l_1,l_2} : l_1, l_2 \in \{i, s\})$*.*

Note that $\text{E}(k_i - 1) = (n-1)\phi_i$ easily follows from Proposition 2. In the top plot of Figure 2 we display $corr(k_i - 1, k_s - 1)$ by numerically approximating $\varphi_{i,s}$ and using $\mu_i^0 \in \{-10, -7, -4\}$ and $\sigma^0_{i,i} \in \{0.5, 1, 2\}$. As expected the number of change points in two series with high correlation between $\text{logit}(p_{i,t})$ and $\text{logit}(p_{s,t})$ will be similar. For the values of μ_i^0 and $\sigma^0_{i,i}$, used in Figure 2, the expected number of change points in each series ranges between 6.19 and 12.29 *a priori*.

The last proposition derives the conditional probabilities of change point indicators. These are of particular interest as they illustrate how the probability of a change point across series varies as one series experiences a change point.

Proposition 3. *Under the assumptions in (7), the probability of a change point occurring at time* t *in the* i*th series given that one occurred at time* t *in the* s*th series is*

$$\mathbb{P}(c_{i,t} = 1 \mid c_{s,t} = 1) = \frac{\varphi_{i,s}}{\phi_s}. \tag{10}$$

Here ϕ_s *and* $\varphi_{i,s}$ *are given by (8) and (9), respectively.*

The bottom row of Figure 2 provides values for $\mathbb{P}(c_{i,t} = 1 \mid c_{s,t} = 1)$. The integral in $\mathbb{P}(c_{i,t} = 1 \mid c_{s,t} = 1)$ was approximated using the statistical software `R` (R Core Team 2022). As expected, the higher the correlation between $\text{logit}(p_{i,t})$ and $\text{logit}(p_{s,t})$ the higher the conditional probabilities *a priori*, for the values of μ_i^0 and $\sigma^0_{i,i}$ considered.

3.2 Selection of Tuning Parameters

Like most change point methods, the posterior probability of classifying a point as a change point can be sensitive to "tuning" parameters. For the CCP-PPM these correspond to the marginal likelihood and prior distribution parameters. In some cases, the practitioner can inform the procedure regarding a change point, which guides tuning parameter selection. Without this information, it is appealing to have a procedure that produces *default* values for the tuning parameters. Therefore, we describe an empirical Bayes approach to selecting values for $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})^\top$ of the marginal likelihood and $(\nu_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ of the prior. The approach we describe is geared towards situations in which the magnitudes of change points are relatively small and is based on moment matching.

First note that from (5) we have

$$0 \leq \text{Corr}(y_{i,r}, y_{i,s}) \leq \frac{1}{1+\kappa_{i,0}},$$

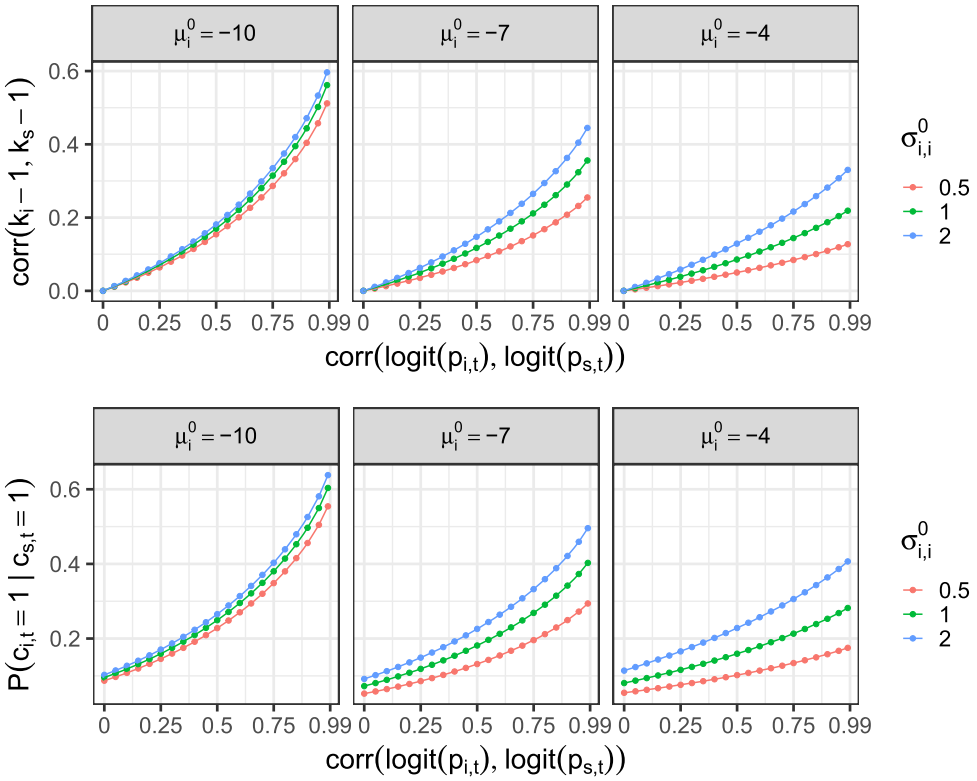

Figure 2: The top row displays $\text{Corr}(k_i - 1, k_s - 1)$ for the a small collection of values for μ_i^0, σ_i^0, and increasing sequence of correlations between $\text{logit}(p_{i,t})$ and $\text{logit}(p_{s,t})$. The integral in (9) was approximated numerically. The bottom row displays values of $\mathbb{P}(c_{i,t} = 1 \mid c_{s,t} = 1)$ for the same values of μ_i^0 and σ_i^0 and correlations between $\text{logit}(p_{i,t})$ and $\text{logit}(p_{s,t})$.

for all $r, s \in \{1, \ldots, n\}$ and $r \neq s$. Under the scenario of no change points, $\text{Corr}(y_{i,r}, y_{i,s})$ is equal to the upper bound $(1 + \kappa_{i,0})^{-1}$, which is constant as a function of (r, s). Thus, a reasonable value for $\kappa_{i,0}$ could be empirically selected using $\text{Corr}(y_{i,r}, y_{i,s})$. Since $(1 + \kappa_{i,0})^{-1}$ is a constant function of (r, s) any estimated ℓ-lag autocorrelation, say $\widehat{c}_{i,\ell}$, of $\boldsymbol{y}_i$ that is positive could be used. We opt to use the smallest $\ell \geq 1$ such that $\widehat{c}_{i,\ell} > 0$. Then set $\kappa_{i,0} = (\widehat{c}_{i,\ell})^{-1}(1 - \widehat{c}_{i,\ell})$. From there, moment estimators based on the first two moments and Mardia's kurtosis coefficient (Mardia, 1970) can be used to provide values for $\mu_{i,0}$, $\alpha_{i,0}$, and $\beta_{i,0}$. To see this let $(d, m, s)^\top$ denote Mardia's kurtosis coefficient and the first two moments. Next set $d = 2\alpha_{i,0}$, $m = \mu_{i,0}$ and $s = (\alpha_{i,0}\kappa_{i,0})^{-1/2}\{\beta_{i,0}(\kappa_{i,0} + 1)\}^{1/2}$. Then, we have that $\mu_{i,0} = \widehat{m}_i$, $\alpha_{i,0} = 0.5\widehat{d}_i$ and $\beta_{i,0} = 0.5\widehat{d}_i(1 - \widehat{c}_{i,\ell})\widehat{s}_i^2$ where $(\widehat{d}_i, \widehat{m}_i, \widehat{s}_i)^\top$ denote the set of moment estimators of $(d, m, s)^\top$.

Now, we focus on $(\nu_0, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)^\top$. Although $\text{E}(\boldsymbol{p}_t)$ and $\text{Var}(\boldsymbol{p}_t)$ do not exist in a closed

form, a first-order Taylor expansion provides approximations to them. If $\nu_0 > 2$, then

$$\begin{aligned}
\mathrm{E}(\boldsymbol{p}_t) &\approx \left(\frac{\exp(\mu_1^0)}{1+\exp(\mu_1^0)}, \ldots, \frac{\exp(\mu_L^0)}{1+\exp(\mu_L^0)} \right)^\top, \\
\mathrm{Var}(\boldsymbol{p}_t) &\approx \left(\frac{\nu_0}{\nu_0 - 2} \right) \mathbf{J}(\boldsymbol{\mu}_0) \boldsymbol{\Sigma}_0 \mathbf{J}(\boldsymbol{\mu}_0), \\
\mathbf{J}(\boldsymbol{\mu}_0) &= \mathrm{diag} \left(\frac{\exp(\mu_1^0)}{\{1+\exp(\mu_1^0)\}^2}, \ldots, \frac{\exp(\mu_L^0)}{\{1+\exp(\mu_L^0)\}^2} \right).
\end{aligned}$$

After choosing prior guesses for $\mathrm{E}(\boldsymbol{p}_t)$ and $\mathrm{Var}(\boldsymbol{p}_t)$, say $\boldsymbol{m}_0 = (m_1^0, \ldots, m_L^0)^\top$ and $\mathbf{S}_0$ respectively, we set

$$\begin{aligned}
\boldsymbol{\mu}_0 &= \left(\log \left(\frac{m_1^0}{1-m_1^0} \right), \ldots, \log \left(\frac{m_L^0}{1-m_L^0} \right) \right)^\top, \\
\boldsymbol{\Sigma}_0 &= \left(\frac{\nu_0 - 2}{\nu_0} \right) \mathbf{D}(\boldsymbol{m}_0)^{-1} \mathbf{S}_0 \mathbf{D}(\boldsymbol{m}_0)^{-1}, \\
\mathbf{D}(\boldsymbol{m}_0)^{-1} &= \mathrm{diag} \left(\frac{1}{m_1^0(1-m_1^0)}, \ldots, \frac{1}{m_L^0(1-m_L^0)} \right).
\end{aligned}$$

We recommend setting $\nu_0 = 3$, which is the least integer such that the approximations described above exist. In the case that no prior information is available to guide specifying $\boldsymbol{m}_0$ and $\mathbf{S}_0$, the following empirical approach can be employed. Set $m_1^0 = \cdots = m_L^0 = n^{-1}$. For $\mathbf{S}_0$, a compound symmetry covariance matrix $\sigma_0^2\{(1-r_0)\mathbf{I}_L + r_0 \mathbf{J}_L\}$ can be used, where $\sigma_0^2 = m_1^0(1-m_1^0)/n = n^{-3}(n-1)$ and $r_0 = 0.5$.

3.3 Posterior Sampling

The joint posterior distribution of $\boldsymbol{p}_t$ and $\mathbf{C}$ is not analytically tractable. Therefore we resort to sampling from it using Markov Chain Monte Carlo (MCMC) methods. The MCMC algorithm we construct is very straightforward to implement and is an extension of the Gibbs sampler described in Loschi et al. (2003) with the main difference being that we employ Metropolis steps to update the $p_{i,t}$. This MCMC strategy is adopted due to its simplicity. On the one hand, the method uses the change point indicators that easily represent the partition structure observed in the data series. On the other hand, we need to specify the *data factors* or marginal likelihoods. Note that, as we mentioned earlier, the *data factors* can be obtained by specifying a probability distribution (a likelihood) conditional on the parameters of interest within a particular block (block-specific parameters) and prior distribution for these parameters. Although we generally consider a conjugate pair of those distributions, this is not a restriction. The choices of the likelihood function and the prior distribution can be general.

Before model fitting, we recommend scaling each series to have mean zero and standard deviation one. The full conditional distributions used in our algorithm are described next. Notice that there is no updating step for $\boldsymbol{\theta}$ as it has been analytically integrated out of the likelihood.

- For each $i \in \{1, \ldots, L\}$ and $t \in \{1, \ldots, n-1\}$ update $p_{i,t}$ according to its full conditional density $\pi(p_{i,t} \mid \cdots)$, which is proportional to

$$\frac{1}{p_{i,t}}\left(\frac{p_{i,t}}{1-p_{i,t}}\right)^{c_{i,t}}\left\{1+\frac{(\text{logit}(\boldsymbol{p}_t)-\boldsymbol{\mu}_0)'\boldsymbol{\Sigma}_0^{-1}(\text{logit}(\boldsymbol{p}_t)-\boldsymbol{\mu}_0)}{\nu_0}\right\}^{-\frac{\nu_0+L}{2}} 1(p_{i,t} \in (0,1)).$$

  Here, $1(\,\cdot \in S)$ is the indicator function of the set S. To update $p_{i,t}$, we employ a random walk Metropolis step with a normal centered at the previous iteration's value as a candidate density. The standard deviation of the normal candidate density is set to 0.005 which produces an acceptance rate in the general range of 0.2 and 0.5

- For each $i \in \{1, \ldots, L\}$, $t \in \{1, \ldots, n-1\}$ and $a \in \{0, 1\}$, define the set of change point indicators $\boldsymbol{c}^{(a)} = (c_1^{(a)}, \ldots, c_{n-1}^{(a)})^\top$ such that

$$c_s^{(a)} = \begin{cases} c_{i,s}, & \text{if } s \neq t \\ a, & \text{if } s = t. \end{cases}$$

  Using $\boldsymbol{c}^{(a)}$, we construct the corresponding set of cluster labels $\boldsymbol{e}^{(a)} = (e_1^{(a)}, \ldots, e_n^{(a)})^\top$. Then, after computing

$$\varpi_{i,t} = \frac{\mathbb{P}(c_{i,t}=1 \mid \cdots)}{\mathbb{P}(c_{i,t}=0 \mid \cdots)} = \frac{\mathcal{F}_i(e_t^{(1)} \mid \boldsymbol{\xi}_i)\mathcal{F}_i(e_{t+1}^{(1)} \mid \boldsymbol{\xi}_i)}{\mathcal{F}_i(e_t^{(0)} \mid \boldsymbol{\xi}_i)}\left(\frac{p_{i,t}}{1-p_{i,t}}\right),$$

  where $\mathcal{F}_i(j \mid \boldsymbol{\xi}_i) = \mathcal{F}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\xi}_i)$, $c_{i,t}$ can be updated using a Bernoulli distribution with probability parameter

$$\mathbb{P}(c_{i,t}=1 \mid \cdots) = \frac{\varpi_{i,t}}{1+\varpi_{i,t}}.$$

Now, an MCMC algorithm can be obtained by cycling through each of the full conditionals individually. If a model is proposed so that $\boldsymbol{\xi}_i$ is available, it is relatively straightforward to update $\boldsymbol{\xi}_i$ in the Gibbs sampler using a Metropolis step. The update is based on the following full conditional of $\boldsymbol{\xi}_i$ for each $i \in \{1, \ldots, L\}$

$$\pi(\boldsymbol{\xi}_i \mid \cdots) \propto \left\{\prod_{j=1}^{k_i} \mathcal{F}_i(\boldsymbol{y}_{i,j} \mid \boldsymbol{\xi}_i)\right\} h_i(\boldsymbol{\xi}_i),$$

where h_i is a prior density for $\boldsymbol{\xi}_i$.

4 Simulation Study

We conduct a numerical experiment to study the CCP-PPM's ability to detect multiple change points. The experiment is based on generating data sets containing change points whose times are dependent across series, mimicking the *contagion* idea. We consider change points that result from simultaneous changes in the mean and variance

of a normal distribution. Data sets are generated using four scenarios, with each one producing data sets containing $L = 4$ series of $n = 100$ observations. The scenarios are detailed next.

The first scenario referred as *data type 1* uses the CCP-PPM as a data generating mechanism. We set $\nu_0 = 3$, $\boldsymbol{\mu}_0 = (-6, -6, -6, -6)^\top$, and $\boldsymbol{\Sigma}_0$ to a compound symmetric matrix with variance 10 and correlation 0.9. Based on these values, (7) is used to create partitions. Once partitions are formed, we use a Normal$((\ell - 1), 1)$ for $\ell = 1, \ldots, 4$ to generate cluster specific means. Cluster specific variances were all generated using an Inv-Gamma$(10, 1)$ distribution. Once cluster specific parameters were generated, observations were generated using a Normal$(0, 1)$. An example of this type of data is displayed in the top row of Figure 3.

The next two scenarios set change point locations at 25, 50, and 75 for each series. As a result, the change points of the four series are highly dependent. Under this setting, four clusters of 25 observations for each series are obtained. Given this type of partition, we produce observations in two ways. The first one, which we refer to as *data type 2*, generates observations using a normal distribution with the following cluster-specific means and variances:

- $\boldsymbol{\mu}_1^\star = (-1, 0, 1, 2)^\top$ and $\boldsymbol{\sigma}_1^\star = (0.1, 0.25, 0.5, 0.75)^\top$,
- $\boldsymbol{\mu}_2^\star = (2, 1, 0, -2)^\top$ and $\boldsymbol{\sigma}_2^\star = (0.1, 0.25, 0.5, 0.75)^\top$.
- $\boldsymbol{\mu}_3^\star = (0, 1, 2, 3)^\top$ and $\boldsymbol{\sigma}_3^\star = (0.75, 0.5, 0.25, 0.1)^\top$.
- $\boldsymbol{\mu}_4^\star = (-2, -1, 0, 1)^\top$ and $\boldsymbol{\sigma}_4^\star = (0.75, 0.5, 0.25, 0.1)^\top$.

The second scenario, which we refer to as *data type 3*, generates observations using a normal distribution and cluster-specific means and variances given by:

- $\boldsymbol{\mu}_1^\star = (-0.25, 0, 0.25, 0.5)^\top$ and $\boldsymbol{\sigma}_1^\star = (0.1, 0.25, 1, 0.25)^\top$,
- $\boldsymbol{\mu}_2^\star = (-0.25, 0, 0.25, 0.5)^\top$ and $\boldsymbol{\sigma}_2^\star = (0.1, 2, 0.5, 1)^\top$.
- $\boldsymbol{\mu}_3^\star = (0.25, 0, -0.25, 0.5)^\top$ and $\boldsymbol{\sigma}_3^\star = (0.5, 0.1, 1, 0.75)^\top$.
- $\boldsymbol{\mu}_4^\star = (0.25, 0, -0.25, 0.5)^\top$ and $\boldsymbol{\sigma}_4^\star = (0.25, 2, 0.75, 1)^\top$.

This scenario is included because it provides insight to how the CCP-PPM approach performs for data similar to that which we consider in Section 5. Examples of synthetic data sets created from these two scenarios are provided in the second and third rows of Figure 3. It is important to note that in the scenario *data type 2*, the means are the primary driver of change points, while in the scenario *data type 3* the variances, or volatility, is the primary driver of change points.

The last scenario which is referred to as *data type 4* employs the same mean and variance vectors as in *data type 2*, but change point locations for the second and third series are at time points 20, 50, and 70 rather than 25, 50 and 75. As result, one change

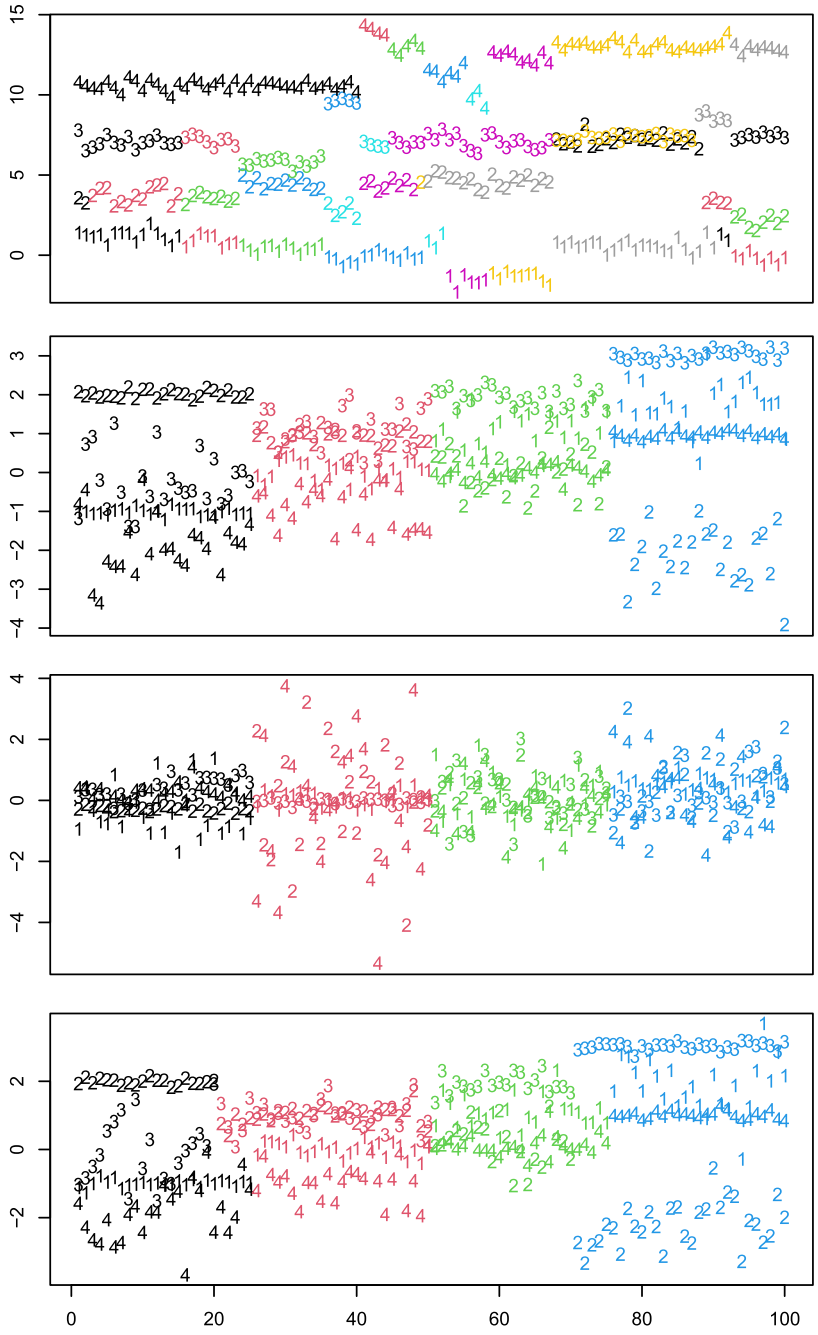

Figure 3: Example of the types of data sets used in the simulation study. Rows are ordered by data type. Points that are labeled as **1** coming from the first series and points labeled **2** come from the second, etc. The different colors identify the clusters formed by change points.

point is common across all four series while the other two change points are only shared across two time series. An example of data from *data type 4* can be found in the fourth row of Figure 3.

We simulated one hundred data sets for each scenario. Then, we fit the CCP-PPM by collecting 2000 MCMC iterates after discarding the initial 10,000 as burn-in and thinning by 10. Values for the CCP-PPM tuning's parameters were selected using the procedure in Section 3.2. All computing was carried out using the `ccp_ppm` function found in the `ppmSuite` (Page and Quinlan 2022) `R` package that is available on `CRAN`. To illustrate the benefit of introducing $p_{i,t}$ in the model rather than p_i, we also fit the CPP-PPM0 using the same MCMC specifications as with the CCP-PPM. In addition, tuning parameters for the CPP-PPM0 were selected in the same way and computing was carried out using the `cpp_ppm` function.

We compare the CCP-PPM to other methods in two ways. First, in order to see the utility of modeling the collection of partitions jointly rather than independently, for each data set we fit the following methods to each of the $L = 4$ series independently.

- The PPM-based method of Wang and Emerson (2015). The `bcp` package (Erdman and Emerson, 2007) found in the `R` statistical software (R Core Team, 2022) is used to implement this method. Default prior parameter values were used. We referred to this method as the *Wang* method.
- The method developed in Barry and Hartigan (1992) and Loschi and Cruz (2002). This method is our most natural competitor and is implemented using the `ppmSuite R` package. we use $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})^\top = (0, 1, 2, 1)^\top$ and $p_i \overset{\text{iid}}{\sim} \text{Beta}(1, 20)$. This method is referred to as the *Loschi* method.

Second, to see how the CCP-PPM performs compared to methods developed to find multivariate change points (i.e., change points in $\mathbb{R}^L$) we compared the CCP-PPM to four methods developed to detect change points in L-space. To identify change points from the CPP-PPM in L-space, we classify each time point that exhibits a change in at least one of the $L = 4$ time series as a "multivariate" change point. In Figure 5, we refer to this procedure as "CCP-PPM_Mult". We consider the following multivarate change point methods in our numerical experiment:

- The method of Matteson and James (2014). This method is implemented using the `R` package `ecp` (James et al., 2019). We referred to this method as the *Matteson* method. Default tuning values are used.
- The method of Arlot et al. (2019). This method is implemented using the `R` package `kcpRS` (Cabrieto and Meers 2019) and is based on the idea of kernel change point detection on running statistics (multivariate). We refer to this method as the *kcpRS* method and default tuning values are used.
- The method of Grundy et al. (2020). This method is implemented using the `R` package `geomcp` (Grundy, 2020) and which finds multivariate change points via

two geometric mappings. We refer to this method as the *geomcp* method and default tuning parameters are used.

- The method of Jin et al. (2022). The R-code that accompanies Jin et al. (2022) was used to fit the method. We refer to this method as the *Jin* method and default tuning parameters are used.

After fitting the methods described above in each scenario, we classify any point as a change if its posterior probability of being a change is greater than 0.5. Then, we compute the overall misclassification rate, sensitivity (true positive rate), and area under the curve (AUC). These metrics give us information on the accuracy of identifying points as change points. In addition, we calculate the adjusted Rand index (Rand, 1971) (ARI) between the estimated partition based on change points and the true partition. This metric illustrates how well each method does at recovering the true partition of contiguous clusters.

The results of the simulation study are provided in Figures 4 and 5. From Figure 4 it appears that the CPP-PPM performs better than all other methods with regards to AUC and ARI regardless of how data were generated. With regards to missclassification rate and sensitivity, *Loschi* performs similarly to CPP-PPM for *data type 1* and *data type 3*. For *data type 2* and *data type 4* it seems that CPP-PPM has a slight advantage over the competing methods. This suggests that when change point times are correlated, using a model that incorporates this correlation is useful. In addition, since the CPP-PPM performs at least as well as CPP-PPM0 for all metrics and all data scenarios, it appears that the borrowing of strength when estimating p_{it} provides some benefit rather than considering a single p_i for each time series.

Figure 5 tells a similar story with regards to ARI and AUC. Mainly, that the CPP-PPM outperforms the other methods when it comes to estimating the partition and detecting change points in L-space regardless of type of data. In addition, it appears that the CPP-PPM has a lower misclassification rate and higher sensitivity compared to all the other methods regardless of datatype. In summary, it appears that the CCP-PPM overall performs the best at detecting the correct number and location of change points.

5 Finance Data Application

We now turn our attention to the application that motivated our proposal. As is commonly done in financial applications, we analyze returns rather than prices. Returns are defined as $R_t = (P_t - P_{t-1})/P_{t-1}$ for $t \in \{2, \ldots, n\}$, where P_t is the daily price. When analyzing the data set, we consider *contagion* both between mature markets and emerging ones and also between emerging markets. In addition to the USA market (a mature market), we include the most important Latin American markets (emerging markets), namely Argentinean, Brazilian, Chilean and Mexican markets. Consequently, we fit the CCP-PPM and the change point PPM of Loschi or simply *Loschi* method (which is perhaps our method's most natural competitor), treating each series independently.

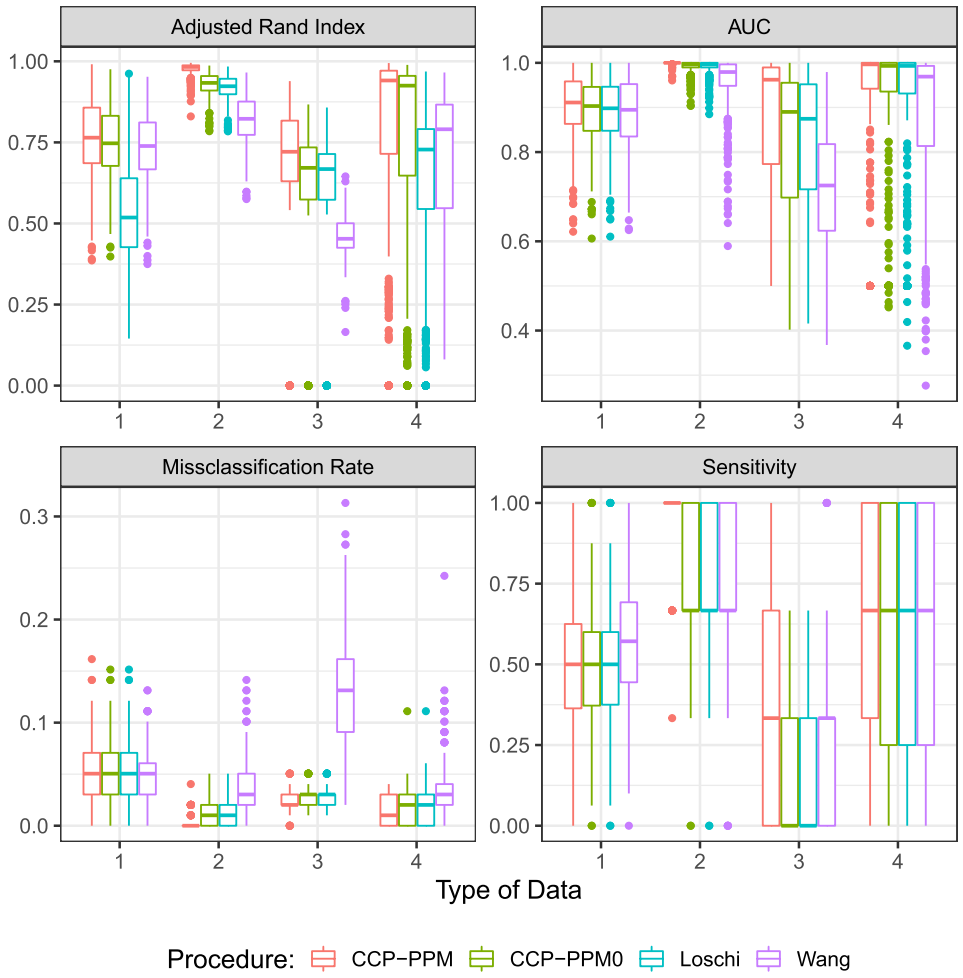

Figure 4: Results from the simulation study comparing CPP-PPM to methods that were fit to each of the $L = 4$ series independently. Each boxplot displays the results for each method based on the 100 data sets generated. For the adjusted Rand index, sensitivity (true positive rate), and AUC (area under the curve) higher values indicate superior performance.

We considered the return series of each countries main stock indexes, namely, the MERVAL (Índice de Mercado de Valores de Buenos Aires) of Argentina, the IBOVESPA (Índice da Bolsa de Valores do Estado de São Paulo) of Brazil, the IPSA (Índice de Precios Selectivos de Acciones) of Chile, the IPyC (Índice de Precios y Cotizaciones) of Mexico, and the Dow Jones (Dow Jones Industrial Average) of USA. The stock returns were recorded daily from October 31, 1995 to October 31, 2000.

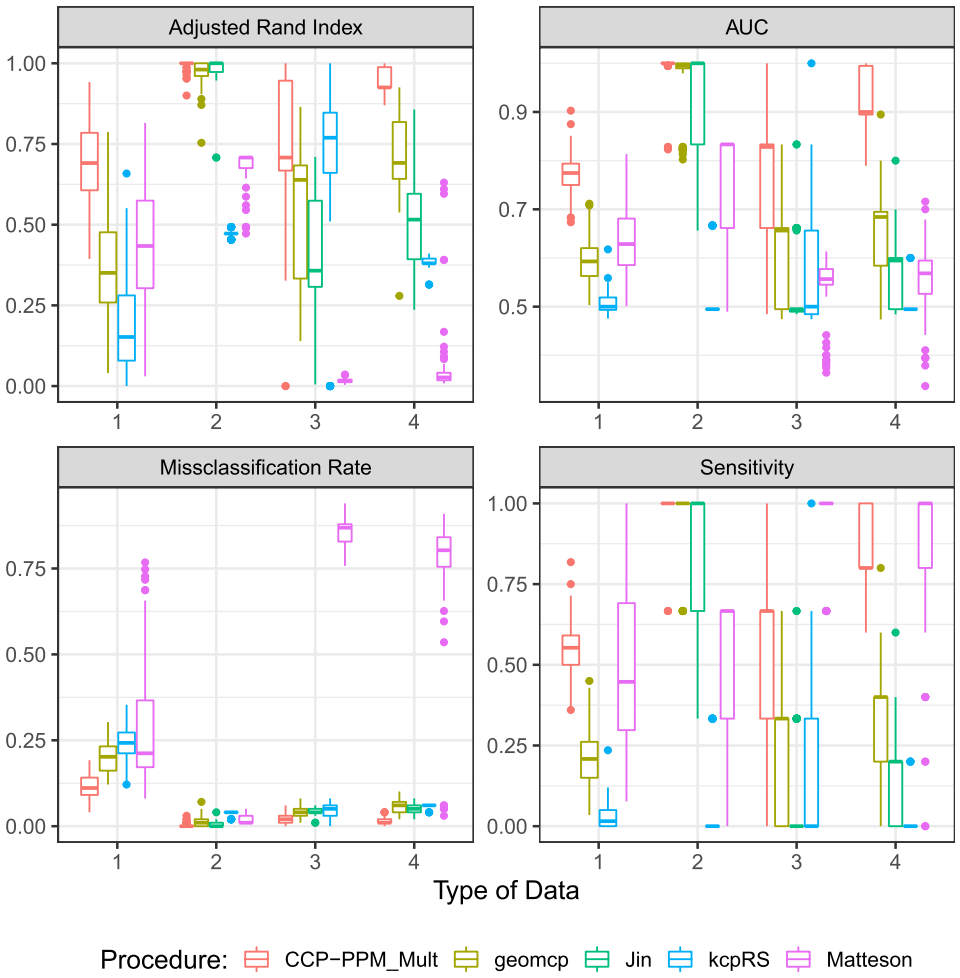

Figure 5: Results from the simulation study comparing the CPP-PPM to methods that find multivariate change points. Each boxplot displays the results for each method based on the 100 data sets generated. For the adjusted Rand index, sensitivity (true positive rate), and AUC (area under the curve) higher values indicate superior performance.

We employ the procedure described in Section 3.2 to produce values for the tuning parameters in (5) and (7). This resulted in values for $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})^\top$ that are listed in Table 1. These tuning parameter values were used for both the CCP-PPM and that of *Loschi* method.

To specify values for $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma_0}$, we first set $m_1^0 = \cdots = m_5^0 = n^{-1}$ (in the application $n = 1309$) and used a compound symmetry matrix for $\mathbf{S}_0$ with variance $\sigma_0^2 = m_1^0(1 - m_1^0)/n = n^{-3}(n-1)$ and correlation 0.5. This resulted in $\boldsymbol{\mu}_0 = (-7.1762)\mathbf{1}_5$ and $\boldsymbol{\Sigma}_0$

Series	$\mu_{i,0}$	$\kappa_{i,0}$	$\alpha_{i,0}$	$\beta_{i,0}$
USA	0.009	188.924	2.212	1.233
Mexico	−0.014	9.075	1.596	0.574
Argentina	0.009	10.186	1.328	0.402
Chile	−0.022	2.533	1.864	0.656
Brazil	0.022	11.010	1.349	0.423

Table 1: Values of $(\mu_{i,0}, \kappa_{i,0}, \alpha_{i,0}, \beta_{i,0})^\top$ for each country's series. The values are the result of applying the empirical procedure described in 3.2 to each country's returns.

being a compound symmetric matrix with variance 0.334 and correlation 0.5. We set $\nu_0 = 3$. For the *Loschi* method $p_i \sim \text{Beta}(a, b)$ and we set $a = 1304.5$ and $b = 681209.9$. These values were selected based on setting the mean number of clusters *a priori* to 3.5 with a variance of 2.5. Both methods were fit by collecting 1000 MCMC samples after discarding the first 10,000 as burn-in and thinning by 5 (*i.e.*, 15,000 total samples were collected). The CCP-PPM was fit using the `ccp_ppm` function while *Loschi*'s method was fit using the `icp_ppm` function both of which are available in the `ppmSuite`-package that can be found on `CRAN`.

There are two approaches that could used to estimate change points. The first classifies points as change points if their posterior probability of being a change point is greater than some pre-specified value. The second classifies points as change points based on a partition estimate. We report both as both require input from the user (pre-specified probability cut-off for the first approach and a loss function in the second approach).

We first explore the *a posteriori* dependence between partitions from the five markets. To do this, at each MCMC iteration we computed the ARI for all possible pairs of partitions (which is 10 in this application). The CCP-PPM produced slightly more similar partitions across countries than the *Loschi* method. The overall average pairwise ARI for CCP-PPM turned out to be 0.51 compared to 0.48 from the *Loschi* method.

Next we explore the posterior change point probabilities which are displayed in the first column of Figure 6. The black points correspond to the CCP-PPM and the red to *Loschi*. For both, change point probabilities were estimated using the posterior means of $c_{i,t}$. It seems that there is a general agreement between the two methods regarding the location of potential change points. However, the CCP-PPM seems to produce probabilities that are closer to one for these points compared to the *Loschi* method. In fact, the *Loschi* method never records a change point probability greater than 0.75. Similarly, both methods agree on the general location of points that have a small chance of being a change point, although the CCP-PPM seems to push these probabilities closer to zero compared to the *Loschi* method.

Figure 6, second and third columns, shows the partition estimates under the CCP-PPM (second column) and *Loschi* method (third column). Partition estimates were obtained using the `salso` (Dahl et al., 2021) `R` package and the generalization of the

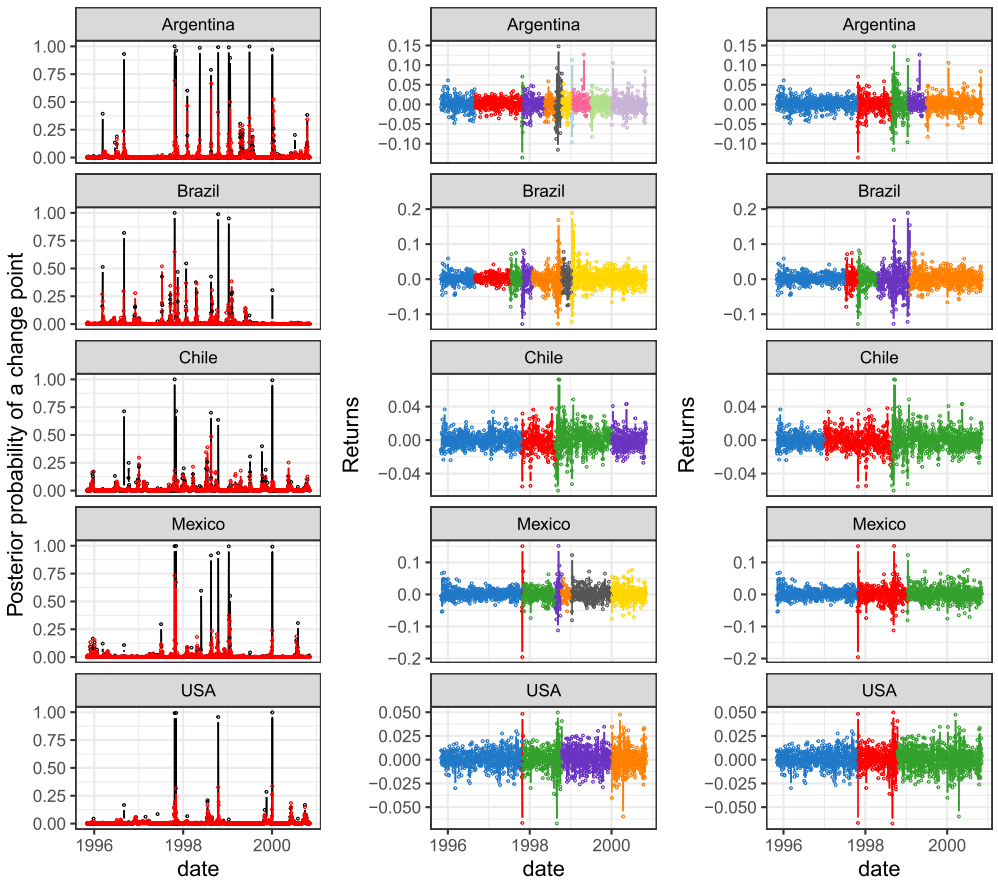

Figure 6: The left plot displays the posterior probability of each point being a change point with black points corresponding to the CCP-PPM method and red points to the *Loschi* method. The middle plot displays the partition estimate of each series for the CCP-PPM and the right plot that for *Loschi*. Both sets of partitions were estimated using the `salso` package in `R`.

Variation of Information loss function (Meilă, 2007). Since in our case it seems natural to penalize change point false positives more than false negatives, we set the false positive penalty parameter of the `salso` function to $a = 25$ (see Dahl et al., 2022, for more details). We note briefly that setting $a = 25$ was driven primarily by *Loschi*'s method. If $a > 25$, then *Loschi*'s method tended to smooth over some change points and for $a < 25$ it tended to produce more change points than what would be desired. However, the CCP-PPM was reasonably robust to a's value. This is a consequence of the change point probabilities from the *Loschi* method being more central (i.e., closer to 0.5) than those from the CCP-PPM.

Apparent differences between the estimated partitions exist, and they illustrate how

the CCP-PPM takes into account the dependency between the index series or the *contagion* phenomenon. For example, the CCP-PPM method identifies a shared partition for Brazil, Chile, Mexico, and the USA at the end of 1997. It is important to stress that in July 1997, the Thai government ran out of foreign currency, forcing it to float the Thai baht which is a factor in starting the 1997 Asian financial crisis, or Asian Flu. This crisis spread internationally, affecting some Asian stock markets such as Indonesia, South Korea, Hong Kong, Laos, Malaysia, Philippines, Brunei, mainland China, Singapore, Taiwan, and Vietnam. According to Harrigan (2000), the Asian crises' overall effect on the United States were small. However, as mentioned by Stallings (1998), the Asian Flu hit the Latin American markets in October 1997, when bond spreads widened abruptly implying more risk. In the Argentinean market, our approach identified the second cluster at the end of 1996. Note that, at the end of 1995, the (real) gross domestic product (GDP) in this country fell by 2.5 percent. However, by the end of 1996, it rebounded by 5.5 percent (International Monetary Found, 2003), possibly affecting the performance of the MERVAL index.

Another cluster or change point our method identifies is related to the Russian crisis or Russian Cold in August 1998. This crisis started when the Russian government and the Russian Central Bank devalued the ruble and defaulted on its debt. It is important to stress that although most countries experienced changes in their stock market returns series at the end of 1998, the Argentinean market experienced a change in mid-1998. Moreover, the IBOVESPA index (Brazil) experienced a change at the beginning of 1999, just after the Russian crisis. This crisis was known as the Samba effect and was produced when the Minas Gerais State Governor, Itamar Franco, stopped paying Minas Gerais debt to other states, generating unleashing capital flights. Note that a small cluster is detected by the CCP-PPM in the indexes of Argentina, Mexico, and USA towards the end of 1998, but something the *Loschi* method misses. These clusters provide some evidence that the *contagion* phenomena from a mature market to emerging ones is being captured by the CCP-PPM, which includes dependency between partitions.

Finally, the CCP-PPM also detected a cluster after the year 2000. The corresponding change point can be explained by the dot-com bubble, caused by excessive speculation of some internet companies in the late 1990s. On January 14, 2000, the Dow Jones Industrial Average reached its dot-com bubble peak. This cluster is observed in the Argentinean, Chilean, Mexican, and USA markets. Possibly, the dot-com bubble may have affected other economies over the Latin American region, evidencing some *contagion* effect. In this case, *Loschi* method did not detect a change point at or near the above-mentioned date.

As mentioned, the estimated partitions in Figure 6 depend on the value $a = 25$. To show a more complete picture of both methods performance, we provide Figure 7. In this figure, all points with a posterior probability of being a change point less than 0.4 are colored gray. The left column corresponds to the CCP-PPM fit while the right the *Loschi* method. The CCP-PPM fit has more power in detecting change points compared to *Loschi* method, without inflating the false-positive rate. The points highlighted by the CCP-PPM fit are at least plausibly change points, and those associated with more pronounced volatility generally have a larger probability of being a change point, which is a desirable characteristic.

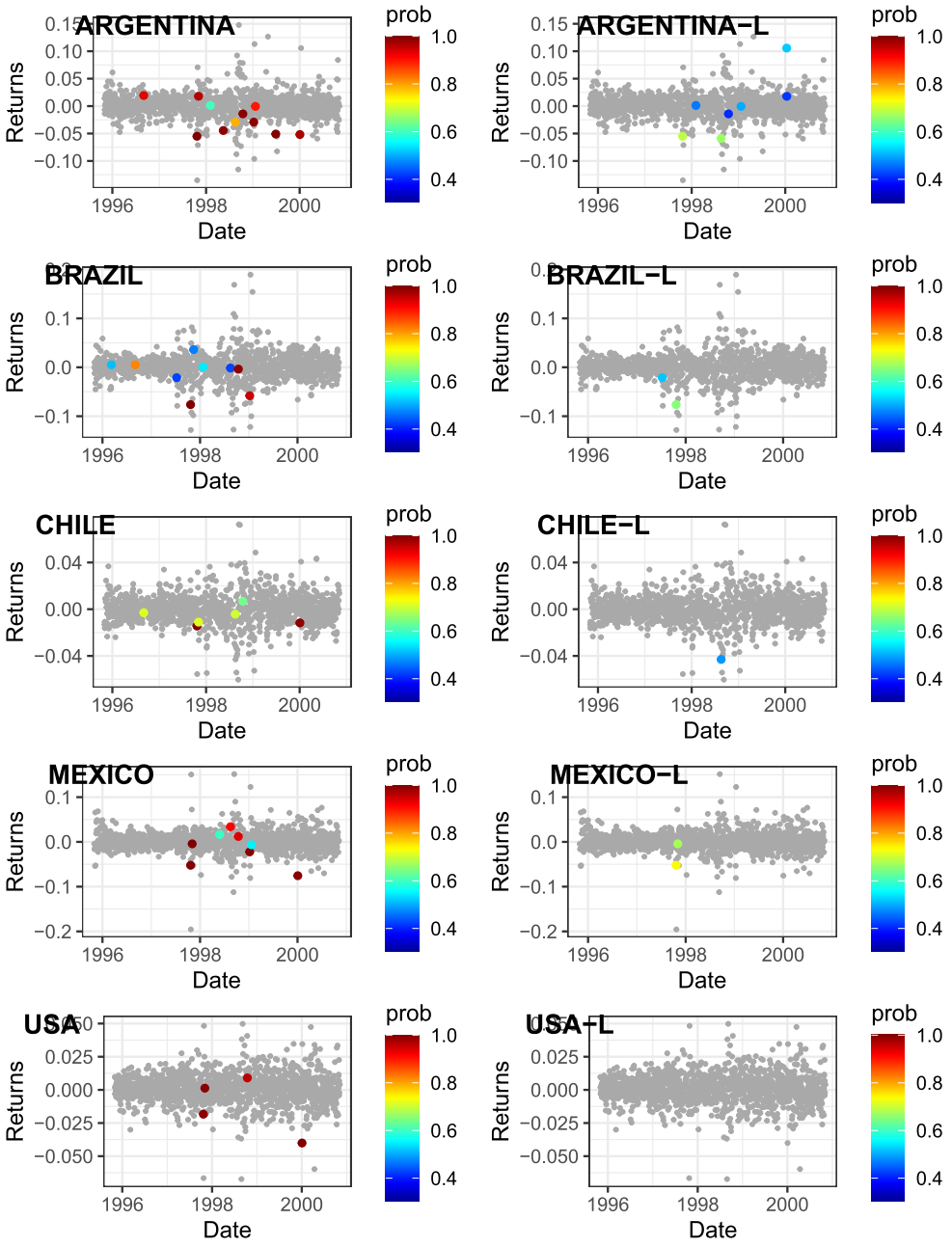

Figure 7: Change point posterior probabilities. Gray points correspond to locations whose posterior probability of being a change point was less than 0.4. The left column displays results under the CCP-PPM and the right column under the *Loschi* method.

6 Discussion

In this paper we developed a new change point detection model for L time series with n observations in an arbitrary space, which undergo sudden changes in their distributional parameters. By making dependent the vector of latent change point probabilities $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{L,t})^\top$ at a specific time $t \in \{1, \ldots, n-1\}$, the corresponding L partitions with contiguous clusters are encouraged to be correlated. We provide some theoretical results that help to better understand the main features of our model, a useful procedure to guide the specification of all parameters that are involved in, and simple pseudo-code to perform posterior inference via MCMC methods. Through a small simulation study, we compared the ability of our model with other compelling approaches to detect highly dependent changes under different scenarios, showing an improvement in detecting change points. Additionally, we applied our method to the returns of emerging Latin American and US markets, obtaining exciting results about possible *contagion effect* between the economies of these countries based on the dependence between change point locations.

In terms of extending the proposed model with the aim of making it more flexible, several directions can be pursued. For instance, the assumption that the vectors of change point probabilities $\boldsymbol{p}_t = (p_{1,t}, \ldots, p_{L,t})^\top$ are independent and identically distributed through time t can be relaxed. One possible approach would be to model $\boldsymbol{p}_t$ with a stationary process in $(0, 1)^L$. Another interesting direction would be to incorporate time-dependent covariates in the marginal likelihood function to describe abrupt changes in a regression curve. A similar situation, but more complex, is to incorporate covariates in the distribution for contiguous partitions. Finally, the computational cost involved in the MCMC algorithm for posterior inference increases rapidly as the length and number of time series grow. It would be appealing to develop strategies that mitigate the so-called "curse of dimensionality". These are all topics of future research.

Supplementary Material

Supplementary Material: Joint Random Partition Models for Multivariate Change Point Analysis
(DOI: 10.1214/22-BA1344SUPP; .pdf). The Supplementary Material contains all the proofs for propositions detailed in Section 3.1.

References

Arellano-Valle, R., Castro, L. M., and Loschi, R. (2013). "Change point detection in the skew-normal model parameters." *Communications in Statistics. Theory and Methods*, 42: 603–618. MR3211938. doi: https://doi.org/10.1080/03610926.2011.611321. 21

Arlot, S., Celisse, A., and Harchaoui, Z. (2019). "A kernel multiple change-point algorithm via model selection." *Journal of Machine Learning Research*, 20: 1–56. MR4048973. 36

Barry, D. and Hartigan, J. A. (1992). "Product partition models for change point problems." *Annals of Statistics*, 20(1): 260–279. MR1150343. doi: https://doi.org/10.1214/aos/1176348521. 22, 25, 26, 36

Cabrieto, J. and Meers, K. (2019). *kcpRS: Kernel change point detection on the running statistics*. R package version 1.0.0. 36

Cheon, S. and Kim, J. (2010). "Multiple change-point detection of multivariate mean vectors with the Bayesian approach." *Computational Statistics & Data Analysis*, 54(2): 406–415. MR2756435. doi: https://doi.org/10.1016/j.csda.2009.09.003. 22

Dahl, D. B., Johnson, D. J., and Mueller, P. (2022). "Search algorithms and loss functions for Bayesian clustering." *Journal of Computational and Graphical Statistics*, DOI: 10.1080/10618600.2022.2069779. MR4513380. doi: https://doi.org/10.1080/10618600.2022.2069779. 41

Dahl, D. B., Johnson, D. J., and Müller, P. (2021). *salso: Search Algorithms and Loss Functions for Bayesian Clustering*. R package version 0.2.23. MR4513380. doi: https://doi.org/10.1080/10618600.2022.2069779. 40

De Blasi, P., Favaro, S., Lijoi, A., Mena, R. H., Prünster, I., and Ruggiero, M. (2015). "Are Gibbs-type priors the most natural generalization of the Dirichlet process?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2): 212–229. 24

de P. Filleti, J., Hotta, L. K., and Zevallos, M. (2008). "Analysis of contagion in emerging markets." *Journal of Data Science*, 6: 601–626. 21

Erdman, C. and Emerson, J. W. (2007). "bcp: An R package for performing a Bayesian analysis of change point problems." *Journal of Statistical Software*, 23(3): 1–13. 22, 36

Fan, Z. and Mackey, L. (2017). "Empirical Bayesian analysis of simultaneous changepoints in multiple data sequences." *Annals of Applied Statistics*, 11: 2200–2221. MR3743294. doi: https://doi.org/10.1214/17-AOAS1075. 24, 25

García, E. C. and Gutiérrez-Peña, E. (2019). "Nonparametric product partition models for multiple change-points analysis." *Communications in Statistics - Simulation and Computation*, 48(7): 1922–1947. MR3981008. doi: https://doi.org/10.1080/03610918.2018.1429616. 22

Grundy, T. (2020). *changepoint.geo: geometrically inspired multivariate changepoint detection*. R package version 1.0.1. 36

Grundy, T., Killick, R., and Mihaylov, G. (2020). "High-dimensional changepoint detection via a geometrically inspired mapping." *Statistics and Computing*, 30: 1155–1166. MR4108696. doi: https://doi.org/10.1007/s11222-020-09940-y. 36

Gupta, S. K., Gupta, N., and Singh, V. P. (2021). "Variable-sized cluster analysis for 3D pattern characterization of trends in precipitation and change-point detection." *Journal of Hydrologic Engineering*, 26(1): 04020056. 21

Harlé, F., Chatelain, F., Gouy-Pailler, C., and Achard, S. (2016). "Bayesian model for multiple mhange-points detection in multivariate time series." *IEEE Transactions on Signal Processing*, 64(16): 4351–4362. MR3528848. doi: https://doi.org/10.1109/TSP.2016.2566609. 24

Harrigan, J. (2000). "The impact of the Asia crisis on U.S industry: An almost-free lunch?" Federal Reserve Bank of New York, Economic Policy Review. 42

International Monetary Found (2003). "El papel del FMI en la Argentina, 1991-2002." 42

James, N. A., Zhang, W., and Matteson, D. S. (2019). "ecp: An R Package for nonparametric multiple change point analysis of multivariate data. R package version 3.1.2." 36

Jin, H., Yin, G., Yuan, B., and Jiang, F. (2022). "Bayesian hierarchical model for change point detection in multivariate sequences." *Technometrics*, 64(4): 177–186. MR4410912. doi: https://doi.org/10.1080/00401706.2021.1927848. 24, 37

Jones, C., Clayton, S., Ribalet, F., Armbrust, E. V., and Harchaoui, Z. (2021). "A kernel-based change detection method to map shifts in phytoplankton communities measured by flow cytometry." *Methods in Ecology and Evolution*, 12(9): 1687–1698. 21

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). "Optimal detection of changepoints with a linear computational cost." *Journal of the American Statistical Association*, 107(500): 1590–1598. MR3036418. doi: https://doi.org/10.1080/01621459.2012.737745. 24

Lijoi, A., Mena, R., and Prünster, I. (2007). "Bayesian nonparametric estimation of the probability of discovering new species." *Biometrika*, 4(94): 769–786. MR2416792. doi: https://doi.org/10.1093/biomet/asm061. 24

Loschi, R. and Cruz, F. (2002). "Analysis of the influence of some prior specifications in the identification of change points via product partition model." *Computational Statistics & Data Analysis*, 39: 477–501. MR1910023. doi: https://doi.org/10.1016/S0167-9473(01)00095-0. 22, 25, 36

Loschi, R. and Cruz, F. (2005). "Extension to the product partition model: computing the probability of a change." *Computational Statistics & Data Analysis*, 48(2): 255–268. MR2133587. doi: https://doi.org/10.1016/j.csda.2004.03.003. 22, 25, 27

Loschi, R., Cruz, F., and Arellano-Valle, R. (2005). "Multiple change point analysis for the regular exponential family using the product partition model." *Journal of Data Science*, 3(3): 305–330. 22

Loschi, R., Cruz, F., Iglesias, P., and Arellano-Valle, R. (2003). "A Gibbs sampling scheme to product partition model: An application to change-point problems." *Computers & Operations Research*, 30: 463–482. 22, 25, 26, 32

Loschi, R., Pontel, J., and Cruz, F. (2010). "Multiple change-point analysis for linear regression models." *Chilean Journal of Statistics*, 1(2): 93–112. MR2756121. 22

Lowell, J. F., Neu, C. R., and Tong, D. (1998). *Financial Crises and Contagion in Emerging Market Countries*. Santa Monica, CA: RAND Corporation. 21

Mardia, K. (1970). "Measures of multivariate skewness and kurtosis with applications." *Biometrika*, 57(3): 519–530. MR0397994. doi: https://doi.org/10.1093/biomet/57.3.519. 31

Martínez, A. F. and Mena, R. H. (2014). "On a nonparametric change point detection model in markovian regimes." *Bayesian Analysis*, 9(4): 823–858. MR3293958. doi: https://doi.org/10.1214/14-BA878. 22, 24, 27

Matteson, D. and James, N. (2014). "A nonparametric approach for multiple change point analysis of multivariate data." *Journal of the American Statistical Association*, 109(505): 334–345. MR3180567. doi: https://doi.org/10.1080/01621459.2013.849605. 24, 36

Meilă, M. (2007). "Comparing clusterings-an information based distance." *Journal of Multivariate Analysis*, 98: 873–895. MR2325412. doi: https://doi.org/10.1016/j.jmva.2006.11.013. 41

Nyamundanda, G., Hegarty, A., and Hayes, K. (2015). "Product partition latent variable model for multiple change-point detection in multivariate data." *Journal of Applied Statistics*, 42(11): 2321–2334. MR3399288. doi: https://doi.org/10.1080/02664763.2015.1029444. 24

Padilla, O. H. M., Yu, Y., Wang, D., and Rinaldo, A. (2021). "Optimal nonparametric change point analysis." *Electronic Journal of Statistics*, 15(1): 1154–1201. MR4255296. doi: https://doi.org/10.1214/21-ejs1809. 24

Page, G. L. and Quinlan, J. J. (2022). *ppmSuite: A collection of models that employ a product partition prior distribution on partitions*. R package version 0.2.1. 22, 36

Page, G. L., Quintana, F. A., and Dahl, D. B. (2022). "Dependent modeling of temporal sequences of random partitions." *Journal of Computational and Graphical Statistics*, 31(2): 614–627. MR4425090. doi: https://doi.org/10.1080/10618600.2021.1987255. 25

Pedroso, R. C., Loschi, R. H., and Quintana, F. A. (2021). "Multipartition model for multiple change point identification." arXiv:2107.11456v2. 22

Quinlan, José J., Page, Garritt L., and Castro, Luis M. (2022). "Supplementary Material for "Joint Random Partition Models for Multivariate Change Point Analysis"." *Bayesian Analysis*. doi: https://doi.org/10.1214/22-BA1344SUPP. 29

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ 22, 30, 36

Rand, W. M. (1971). "Objective criteria for the evaluation of clustering methods." *Journal of the American Statistical Association*, 66(336): 846–850. 37

Song, J. and Kang, J. (2021a). "Change point analysis in Bitcoin return series: a robust approach." *Communications for Statistical Applications and Methods*, 28(5): 511–520. 25

Song, J. and Kang, J. (2021b). "Test for parameter change in the presence of outliers: the density power divergence based approach." *Journal of Statistical Computation and Simulation*, 5: 1016–1039. MR4232468. doi: https://doi.org/10.1080/00949655.2020.1842407. 25

Stallings, B. (1998). "Impact of th Asian crisis on Latin America." United Nations Economic Commission for Latin America and the Caribbean. 42

Truong, C., Oudre, L., and Vayatis, N. (2020). "Selective review of offline change point detection methods." *Signal Processing*, 167: 107299. 21

Tveten, M., Eckley, I. A., and Fearnhead, P. (2022). "Scalable changepoint and anomaly detection in cross-correlated data with an application to condition monitoring." *Annals of Applied Statistics*, 16(2): 721–743. MR4438809. doi: https://doi.org/10.1214/21-aoas1508. 24

Valdés, R. (2000). "Emerging markets contagion: evidence and theory." Available at SSRN: https://ssrn.com/abstract=69093 or http://dx.doi.org/10.2139/ssrn.69093. 21

Wang, X. and Emerson, J. W. (2015). "Bayesian change point analysis of linear models on graphs." arXiv:1509.00817. 22, 36

Wang, Y. H. (1993). "On the number of successes in independent trials." *Statistica Sinica*, 3(2): 295–312. MR1243388. 30

Wood, K., Roberts, S., and Zohren, S. (2021). "Slow momentum with fast reversion: a trading strategy using deep learning and changepoint detection." *The Journal of Financial Data Science*, jfds.2021.1.081. MR4236146. doi: https://doi.org/10.1098/rsta.2020.0209. 21

Yao, Y.-C. (1984). "Estimation of a noisy discrete-time step function: Bayes and empirical Bayes approaches." *Annals of Statistics*, 12(4): 1434–1447. MR0760698. doi: https://doi.org/10.1214/aos/1176346802. 22, 26

Zanini, C. T. P., Müller, P., Ji, Y., and Quintana, F. A. (2019). "A Bayesian random partition model for sequential refinement and coagulation." *Biometrics*, 75(3): 988–999. MR4012103. doi: https://doi.org/10.1111/biom.13047. 25

Zhu, X., Li, Y., Liang, C., Chen, J., and Wu, D. (2013). "Copula based change point detection for financial contagion in Chinese banking." *Procedia Computer Science*, 17: 619–626. 25

Zhu, X., Xie, Y., Li, J., and Wu, D. (2015). "Change point detection for subprime crisis in American banking: from the perspective of risk dependence." *International Review of Economics and Finance*, 38: 18–28. 25