A Bayesian multiple partition model for multiparametric change point detection

Ricardo Cunha Pedroso

Belo Horizonte/MG, Brasil

February 2020

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação em Estatística

A Bayesian multiple partition model for multiparametric change point detection

Ricardo Cunha Pedroso

Master's thesis presented to the Department of Statistics of the Universidade Federal de Minas Gerais (UFMG), as a partial requirement to obtain the Master's degree in Statistics.

Supervisor: Prof. Rosangela Helena Loschi, D.Sc.

Belo Horizonte/MG, Brasil

February 2020

Ficha catalográfica elaborada pela bibliotecária Belkiz Inez Rezende Costa CRB 6ª Região nº 1510

Pedroso, Ricardo Cunha.

P372b A Bayesian multiple partition model for multiparametric change points detection / Ricardo Cunha Pedroso — Belo Horizonte, 2020.
75 f. il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de Minas Gerais – Departamento de Estatísitca.

Orientadora: Rosângela Helena Loschi.

1. Estatística - Teses. 2. Ponto de mudança – Teses. 3 Modelo partição produto – Teses. 4. Amostrador de Gibbs - Teses. I. Orientadora. II.Título.

CDU 519.2 (043)

UNIVERSIDADE FEDERAL DE MINAS GERAIS

PROGRAMA DE PÓS-GRADUAÇÃO EM ESTATÍSTICA

UFMG

ATA DA DEFESA DA DISSERTAÇÃO DO ALUNO

RICARDO CUNHA PEDROSO

Realizou-se, no dia 14 de fevereiro de 2020, às 14:00 horas, Sala 2040 ICEx, da Universidade Federal de Minas Gerais, a 249ª defesa de dissertação, intitulada *A Bayesian multiple partition model for multiparametric change points detection*, apresentada por RICARDO CUNHA PEDROSO, número de registro 2018657571, graduado no curso de CIÊNCIAS ATUARIAIS, como requisito parcial para a obtenção do grau de Mestre em ESTATÍSTICA, à seguinte Comissão Examinadora: Prof(a). Rosângela Helena Loschi - Orientador (DEST/UFMG), Prof(a). Fabio Nogueira Demarqui (DEST/UFMG), Prof(a). Fernando Ferraz do Nascimento (UFPI).

A Comissão considerou a dissertação:

(X) Aprovada

( ) Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.
Belo Horizonte, 14 de fevereiro de 2020.

Prof(a). Rosângela Helena Loschi (Doutora)

Prof(a). Fabio Nogueira Demarqui (Doutor)

Prof(a). Fernando Ferraz do Nascimento (Doutor)

Acknowledgements

I am grateful to the Statistics department of UFMG for its supportive environment that helped me to conclude a master's course in a way that I am very proud of it. I am indebted with professors Fábio Nogueira Demarqui, Cristiano de Carvalho Santos and Wagner Barreto de Souza for their several important technical and professional advices throughout my course.

I am specially grateful to my supervisor Rosangela Helena Loschi for introducing me in a research area that was completely new for me. I also thank her for her unrelenting dedication to my learning in our weekly meetings, besides many other activities she is responsible for. I would also like to thank her for the many delightful Italian coffees we shared.

My two years of full dedication to my master's course were financially supported by the Brazilian research agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Resumo

O objetivo principal em um problema de múltiplos pontos de mudança em uma sequência de observações é a estimação do número e localização desses pontos de mudança. Além disso, é de interesse estimar o regime de cada agrupamento de observações que formam a partição induzida pelos pontos de mudança. O Modelo de Partição de Produto (MPP), introduzido por Hartigan [1990], foi aplicado pela primeira vez a problemas de múltiplos pontos de mudança por Barry and Hartigan [1992; 1993]. O MPP é um modelo estatístico eficiente para detecção de múltiplos pontos de mudança. Barry and Hartigan [1993] aplicaram o MPP para detectar mudanças na média de sequências univariadas de observações Normais, com o número de mudanças e suas localizações desconhecidas, e assumindo uma variação constante desconhecida. Loschi et al. [1999] propôs uma extensão do MPP para detectar mudanças na média e variância das sequências univariadas de observações Normais, mas sem especificar quais parâmetros foram alterados. Se o MPP é aplicado para identificar mudanças em dois ou mais parâmetros, um dos maiores desafios é identificar qual parâmetro sofreu cada mudança. A distribuição *a posteriori* para a partição aleatória indica apenas os instantes quando as mudanças ocorreram, mas não qual parâmetro sofreu a mudança. Alterações em diferentes parâmetros podem ocorrer em momentos diferentes. Nesta dissertação, propomos um novo modelo que permite identificar quais parâmetros sofreram as mudanças. Introduzimos um MPP com múltiplas partições para identificar múltiplas mudanças em múltiplos parâmetros, em dados observados sequencialmente. O modelo proposto estende trabalhos anteriores assumindo que diferentes parâmetros do modelo podem sofrer quantidades distintas de mudanças, que podem ocorrer em diferentes instantes. Isso é obtido considerando-se uma partição aleatória diferente associada a cada parâmetro. O modelo é definido em um contexto geral e assumindo partições independentes. Apresentamos um esquema geral do amostrador de Gibbs para simular da distribuição *a posteriori* com base na estratégia de amostragem por blocos proposta por Liu [1994]. Aplicamos o modelo proposto a dados Normais sujeitos a mudanças na média e na variância. Avaliamos o desempenho do modelo proposto por meio de um estudo de simulação de Monte Carlo e também considerando aplicações de dados reais. Seu desempenho é comparado aos métodos de Barry and Hartigan [1993] e Loschi et al. [1999]. Esses estudos mostram que o modelo proposto é competitivo e enriquece a análise de problemas de pontos de mudança.

Key-words: Ponto de mudança, Modelo Partição Produto, Amostrador de Gibbs, Multipartição, Agrupamento.

Abstract

The main goal in a multiple change point problem in a sequence of observations is the estimation of the number and location of the change points. Additionally, it is of interest to estimate the regime structure of each different cluster of observations that form the partition induced by the change points. The Product Partition Model (PPM), introduced Hartigan [1990], was first applied to multiple change point problems by Barry and Hartigan [1992; 1993]. It is an efficient statistical model to multiple change points detection. Barry and Hartigan [1993] applied the PPM detect changes in the mean of univariate sequences of Normal observations, with the number of changes and their locations both unkown, and assuming an unknown constant variance. Loschi et al. [1999] proposed an extension of the PPM to detect changes in the mean and variance of univariate sequences of Normal observations, but without specifying which of the parameters have changed. If the PPM is applied to identify changes in two or more parameters, one of the greatest challenges is to identify which parameter has changed. The posterior distribution for the random partition only indicates the instants when the changes occured, but does not which parameter has changed. Changes in different parameters may occur at different times. In this thesis, we propose a new model that permits to identify the parameter or parameters that have changed. We introduce a multipartition PPM to detect multiple changes in multiple parameters in sequentially observed data. The proposed model extends previous works by assuming that different model parameters may experience distinct numbers of changes that may occur at different points of time. That is attained by considering a different random partition associated with each different parameter. The model is defined in a general context and assuming independent partitions. We present a general Gibbs scheme to generate from the posteriors, based on the blocking strategy proposed by Liu [1994]. We apply the proposed model to Normal data subject to changes in the mean and variance. We evaluate the performance of the proposed model through a Monte Carlo simulation study and also considering real data applications. Its performance is compared with Barry and Hartigan [1993] and Loschi et al. [1999] methods. These studies show that the proposed model is competitive and enriches the analysis of change point problems.

Key-words: Change point, Product Partition Model, Gibbs sampling, Multipartition, Clustering.

List of Figures

List of Tables

Contents

1 Introduction

A change point may be defined as a structural change in the regime of a sequence of observations ordered, for example, by time or position. The detection of change points is a relevant topic of research in many areas, such as analysis of historical environmental measurements, analysis of financial time series, genetics and other areas. The main goal in a multiple change point problem is the estimation of the number and location of these change points. Additionally, it is of interest to estimate the regime structure of each different cluster of observations that form the partition induced by the change points.

The Product Partition Model (PPM) introduced by Hartigan [1990], and its extension to change point problems proposed by Barry and Hartigan [1992], is an efficient statistical model to multiple change points detection. Barry and Hartigan [1993] applied the PPM detect changes in the mean of univariate sequences of Normal observations, with the number of changes and their locations both unkown, and assuming an unknown constant variance. One of the greatest contributions of their work is the proposed Gibbs sampler scheme to sample from the random partition that provides information about the change points locations. Loschi et al. [1999] proposed an extension of the PPM to detect changes in the mean and variance of univariate sequences of Normal observations, but without specifying which of the parameters experienced the change. The models proposed by Barry and Hartigan [1993] and Loschi et al. [1999] are described and applied all over this text, and we refer to these models as BH93 and L99 models, respectively. Other important applications of the PPM to change point problems are cited next. Loschi and Cruz [2002] and Loschi et al. [2003] analyse the influence of some prior specifications for the parameters of the L99 model. In Loschi and Cruz [2005], the PPM is extended to generate the posterior probability that each instant of time is a change point. Fearnhead [2006] proposed an efficient algorithm to simulate from a class of change point models that includes the PPM. Loschi et al. [2010] applied the PPM to identify multiple change points in linear regression models and also modified the original algorithm proposed by Barry and Hartigan [1993] in order to obtain samples from the posteriors of the regression parameters. Nyamundanda et al. [2015] improved the PPM for the purpose of detecting multiple change points in both the mean and covariance structures of multivariate correlated sequences of Gaussian data.

The PPM has also been applied in many other distinct statistical problems. Quintana and Iglesias [2003] presented a theoretic formulation of the PPM that may be used in different decision problems such as estimation or hypothesis testing and clustering methods simultaneously. Also, they proved that the Dirichlet process (Ferguson [1973]) is a specific configuration of the PPM. Jordan et al. [2007] proposed the application of the PPM to a prediction problem with categorical predictor variables and to a meta-analysis model. Hegarty and Barry [2008] used the PPM in a regional mapping of disease risk. Demarqui et al. [2008] applied the PPM to estimate the time grid in picewise exponential

models. Monteiro et al. [2011] replaced the assumption of constant parameter values within each cluster to the assumption of correlated different parameter values. Müller et al. [2011] developed a PPM for clustering with covariates, such that the probability distribution of the random partition may depend on the covariates. Page et al. [2016] extended the PPM to a spatial setting and modeled the partitioning of spatial locations into spatially dependent clusters. García and Gutiérrez-Peña [2019] proposed a nonparametric extension of the PPM, such that they associate a random measure instead of a parametric distribution to each cluster of the random partition, and do not impose any specific form to the observations model. An extension of the PPM to spatio-temporal clustering was proposed by Teixeira et al. [2019].

Many distinct approachs have been proposed for change point problems, some of them are cited next. Chib [1998] formulated a change point model in terms of a latent discrete state variable that evolves according to a discrete-time Markov process and indicates the regime of each particular observation. Fearnhead [2005] proposed a regression model that is a combination of independent linear regression models on disjoint segments, such that both the number, the position and the parameters of the regressions are to be estimated. Fearnhead and Liu [2007] introduced a particle filter simulation algorithm to online change point problems, that is, observations are obtained incrementally over time, and new inferences are required each time that an observation is made. Fearnhead and Rigaill [2019] presented a penalized cost approach to change points detection that is robust to the presence of outliers, based on alternative loss functions that are less sensitive to outliers. Peluso et al. [2019] proposed a Bayesian semiparametric multiple change point model, in terms of the Markov process formulation proposed by Chib [1998], such that different groups of structural parameters follow separate change point processes. Mira and Petrone [1996], Martínez et al. [2014] and Haynes et al. [2017] are examples of nonparametric approachs for change point problems. Dierckx and Teugels [2010], Nascimento and Moura e Silva [2017] and Lattanzi and Leonelli [2019] are examples of change point approachs for the identification of extreme regimes. A spatio-temporal change point model can be found in Majumdar et al. [2005].

If the PPM is applied to identify multiple change points in two or more parameters, one of the greatest challenges is to identify which parameter experienced the change. The posterior distribution for the random partition only indicates the instants when the changes occured, but does not which parameter have changed. Only one or all parameters may change in a specific instant, and changes in different parameters can occur in different instants.

Our main goal is to propose a model that permits to identify the parameter or parameters that have changed. We introduce a multiple partition PPM to detect multiple structural changes in multiple parameters in sequentially observed data. We assume that different groups of parameters may experience multiple changes at different times, with both the number and location of these changes unknown. The proposed model extends previous PPM such that it allows to identify in which group of structural parameters the change occurred. The proposed multiple random partitions PPM assumes

that different partitions are associated to different parameters. We refer to this model as a Bayesian multipartition change point model (BMCP).

In Chapter 2, we review some existing product partition models and present the tools that form the basic background required to the development of this thesis. In Section 2.1, we describe the PPM. In Section 2.2, we describe the convenience of the PPM to multiple change points detection, as proposed by Barry and Hartigan [1992]. The BH93 model is presented in Section 2.3. This model is one of the main backgrounds of this work, since we adopt their Gibbs sampler strategy to sample from the random partitions in the proposed model. In Section 2.3.2, we develop an alternative and more informative implementation of the BH93 model. In this approach, the parameters and partitions are estimated through a Gibbs sampler strategy that makes use of the blocking technique proposed by Liu [1994]. This blocking technique is an adaptation of the conventional Gibbs sampler that plays an important role to the estimation procedures used in this thesis. The L99 model is described in Section 2.4. In Section 2.5, we present some properties of the model that permit some prior specifications to be more or less informative, based on our expectation about the number of changes. Finally, in Section 2.6, we discuss some features of the alternative implementation of the BH93 model, proposed in Section 2.3.2. We also present a motivation to the new model to be proposed in Chapter 3. We analyse a simulated data set in which both BH93 and L99 models fail to correctly estimate the parameters and partitions.

In Chapter 3, we introduce the new Bayesian multipartition change point model, which is the central contribution of this thesis. We denote it as BMCP model. In Section 3.1, we formulate the BMCP model, and in Sections 3.2 and 3.3 we present the general configuration of the prior and posterior distributions for parameters and partitions. In Section 3.4, we present a general Gibbs scheme to estimate the parameters and partitions of the BMCP model. In Section 3.5, we apply the BMCP model to Normal data subject to changes in the mean and variance, considering two random partitions, one for each of these two groups of parameters.

Chapter 4 presents the Monte Carlo simulation studies that evaluate the proposed BMCP model considering simulated data sets. Different scenes are considered, with changes in all the parameters or only in one of them. We present Monte Carlo simulation results that compare the performance of the BMCP, L99 and BH93 models. In all the scenes, Normal data subject to changes in mean and variance are considered.

In Chapter 5, we apply the BMCP model in two real data sets. Chapter 6 summarizes the contributions of this thesis and point out some suggestions of future related works. Appendix A describes the probability densitiy functions used along this thesis. In Appendix B we present summarized theoretical descriptions of the Markov Chain Monte Carlo (MCMC) methods.

2 Product Partition Model for multiple change point detection

The Product Partition Model (PPM) is a model for cluster detection introduced by Hartigan [1990]. The clusters are represented by a random partition ρ, and the main goal is to infer about ρ. This chapter presents the PPM and some applications of this model to detection of multiple change points. We refer to many of the results described here along the future chapters.

2.1 The Product Partition Model

Let $\boldsymbol{S}$ denotes a set of n objects indexed by 1 to n and S_j denotes a nonempty subset of $\boldsymbol{S}$. $\boldsymbol{X} = (X_1, \ldots, X_n)$ represents the observations associated to each object of $\boldsymbol{S}$ and $\boldsymbol{X}_{S_j}$ the combined observations $\{X_i, i \in S_j\}$. To define the PPM, Hartigan [1990] considers a partition ρ of $\boldsymbol{S}$ represented by the subsets $S_1, S_2, \ldots, S_b$ with the property that each object in $\boldsymbol{S}$ lies in exactly one of these subsets, also called components of the partition. Define $\mathcal{P}$ as the set of all possible partitions of $\boldsymbol{S}$. The PPM assumes a prior product distribution for the random partition $\rho \in \mathcal{P}$ given by

$$p(\rho) = P\left(\rho = \{S_1, \ldots, S_b\}\right) = \frac{\prod_{j=1}^{b} c(S_j)}{\sum_{\rho \in \mathcal{P}} \prod_{S_j \in \rho} c(S_j)}. \tag{2.1}$$

where $c(S_j)$ is a non-negative cohesion assigned to the subset S_j of $\boldsymbol{S}$, which represents the prior knowledge about the similarity level of the objects in S_j. The PPM also assumes that, given $\rho = \{S_1, \ldots, S_b\}$, the clusters of observations $\boldsymbol{X}_{S_1}, \ldots, \boldsymbol{X}_{S_b}$ are independent, such that

$$f\left(\boldsymbol{X} | \rho = \{S_1, \ldots, S_b\}\right) = \prod_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}). \tag{2.2}$$

where $f_{S_j}(\boldsymbol{X}_{S_j})$ denotes the joint distribution of the observations in $\boldsymbol{X}_{S_j}$, $j = 1, \ldots, b$.

As defined in Barry and Hartigan [1992], any joint probability distribution for observations and partition $(\boldsymbol{X}, \rho)$ that satisfies the product form in Eq. (2.1) and the independence condition for observations given the partition in Eq. (2.2) is called a product partition model. As a consequence of these assumptions, the posterior distribution of ρ given the observations $\boldsymbol{X}$ is also a product partition model with cohesions $c(S_j) f_{S_j}(\boldsymbol{X}_{S_j})$, as shown next. Using the Bayes' theorem, it follows that

$$p(\rho|\boldsymbol{X}) = P(\rho = \{S_1, \ldots, S_b\}|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\rho = \{S_1, \ldots, S_b\})P(\rho = \{S_1, \ldots, S_b\})}{\sum\limits_{\rho \in \mathcal{P}} f(\boldsymbol{X}|\rho)p(\rho)}$$

$$= \frac{\prod\limits_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}) \times \dfrac{\prod\limits_{j=1}^{b} c(S_j)}{\sum\limits_{\rho \in \mathcal{P}} \prod\limits_{j=1}^{b} c(S_j)}}{\sum\limits_{\rho \in \mathcal{P}} \left(\prod\limits_{S_j \in \rho} f_{S_j}(\boldsymbol{X}_{S_j}) \times \dfrac{\prod\limits_{S_j \in \rho} c(S_j)}{\sum\limits_{\rho \in \mathcal{P}} \prod\limits_{S_j \in \rho} c(S_j)} \right)} = \frac{\prod\limits_{j=1}^{b} c(S_j) f_{S_j}(\boldsymbol{X}_{S_j})}{\sum\limits_{\rho \in \mathcal{P}} \prod\limits_{S_j \in \rho} c(S_j) f_{S_j}(\boldsymbol{X}_{S_j})}. \tag{2.3}$$

To calculate the marginal distribution of each X_i, $i = 1, \ldots, n$, a central definition is the relevance probability $r(S_j)$, which is the probability that each subset S_j is a component of the random partition ρ, that is,

$$r(S_j) = P(S_j \in \rho) = \sum_{\rho \in \mathcal{P}|S_j \in \rho} p(\rho). \tag{2.4}$$

Considering expression (2.4), the marginal density $f(X_i)$, $i = 1, \ldots, n$, is given by

$$\begin{aligned} f(X_i) &= \sum_{\rho \in \mathcal{P}} f(X_i|\rho)p(\rho) = \sum_{S_j|i \in S_j} \sum_{\rho \in \mathcal{P}|S_j \in \rho} f(X_i|\rho)p(\rho) \\ &= \sum_{S_j|i \in S_j} \sum_{\rho \in \mathcal{P}|S_j \in \rho} f_{S_j}(X_i)p(\rho) = \sum_{S_j|i \in S_j} f_{S_j}(X_i)r(S_j), \end{aligned} \tag{2.5}$$

which means that $f(X_i)$ is the average of the possible component densities containing X_i, wheighted by their respective relevance probabilities. The component density $f_{S_j}(\boldsymbol{X}_{S_j})$ for the observations in each subset S_j is called data factor.

The PPM can consider parametric probabilistic models for the observations. To that end, it is assumed that the observations $X_1, \ldots, X_n$ are independent, given the parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, with independent conditional marginal densities $f(X_i|\theta_i)$, $i = 1, \ldots, n$. Given a partition $\rho = \{S_1, \ldots, S_b\}$, $\theta_i = \theta_{S_j}$ for every $i \in S_j$, $j = 1, \ldots, b$, and the parameters $\theta_{S_1}, \ldots, \theta_{S_b}$ are independent with densities $f_{S_j}(\theta_{S_j})$, $j = 1, \ldots, b$, that is,

$$f\left(\boldsymbol{\theta}|\rho = \{S_1, \ldots, S_b\}\right) = \prod_{j=1}^{b} f_{S_j}(\theta_{S_j}), \tag{2.6}$$

where $f_{S_j}(\theta_{S_j})$ is called the block prior density, for $j = 1, \ldots, b$. The data factor $f_{S_j}(\boldsymbol{X}_{S_j})$ is the predictive distribution for the observations in cluster S_j and is given by

$$f_{S_j}(\boldsymbol{X}_{S_j}) = \int \left(\prod_{i \in S_j} f(X_i|\theta_{S_j}) \right) f_{S_j}(\theta_{S_j}) d\theta_{S_j}. \tag{2.7}$$

The conditional distribution on partition and parameters given the observations is also a product partition model, such that

$$
\begin{aligned}
f\left(\boldsymbol{\theta}, \rho|\boldsymbol{X}\right) = \frac{f(\boldsymbol{X}|\boldsymbol{\theta},\rho)f(\boldsymbol{\theta}|\rho)p(\rho)}{\sum\limits_{\rho\in\mathcal{P}} f(\boldsymbol{X}|\rho)p(\rho)} &= \frac{\prod\limits_{S_j\in\rho}\left(\prod\limits_{i\in S_j} f(X_i|\theta_{S_j})\right)\times\prod\limits_{S_j\in\rho} f_{S_j}(\theta_{S_j})\times\frac{\prod\limits_{S_j\in\rho} c(S_j)}{\sum\limits_{\rho\in\mathcal{P}}\prod\limits_{S_j\in\rho} c(S_j)}}{\sum\limits_{\rho\in\mathcal{P}}\left(\prod\limits_{S_j\in\rho} f_{S_j}(\boldsymbol{X}_{S_j})\times\frac{\prod\limits_{S_j\in\rho} c(S_j)}{\sum\limits_{\rho\in\mathcal{P}}\prod\limits_{S_j\in\rho} c(S_j)}\right)} \\
&= \prod_{S_j\in\rho}\frac{\left(\prod\limits_{i\in S_j} f(X_i|\theta_{S_j})\right) f_{S_j}(\theta_{S_j})}{f_{S_j}(\boldsymbol{X}_{S_j})}\times\frac{\prod\limits_{S_j\in\rho} c(S_j)f_{S_j}(\boldsymbol{X}_{S_j})}{\sum\limits_{\rho\in\mathcal{P}}\prod\limits_{S_j\in\rho} c(S_j)f_{S_j}(\boldsymbol{X}_{S_j})} \\
&= \prod_{S_j\in\rho} f_{S_j}(\theta_{S_j}|\boldsymbol{X}_{S_j})\times p(\rho|\boldsymbol{X}),
\end{aligned}
\tag{2.8}
$$

where $f_{S_j}(\theta_{S_j}|\boldsymbol{X}_{S_j}) = \left(\prod\limits_{i\in S_j} f(X_i|\theta_{S_j})\right) f_{S_j}(\theta_{S_j}) \Big/ f_{S_j}(\boldsymbol{X}_{S_j})$ is the posterior distribution for the common parameter θ_{S_j} indexing the distribution of the observations in cluster S_j, based only on data information in the cluster S_j, and $c(S_j)f_{S_j}(\boldsymbol{X}_{S_j})$ are the posterior cohesions of each $S_j \in \rho$. It follows that the posterior density of θ_i, $i = 1, \ldots, n$, is the average of the posterior distributions of each block $S_j|i \in S_j$, given the observations in S_j, weighted by the posterior relevance probability $r(S_j|\boldsymbol{X})$, which is the probability that $S_j \in \rho$, given $\boldsymbol{X}$. Formally, the posterior density of θ_i is given by

$$
\begin{aligned}
f\left(\theta_i|\boldsymbol{X}\right) &= \sum_{\rho\in\mathcal{P}} f(\theta_i|\boldsymbol{X},\rho)p(\rho|\boldsymbol{X}) = \sum_{S_j|i\in S_j}\ \sum_{\rho\in\mathcal{P}|S_j\in\rho} f(\theta_i|\boldsymbol{X},\rho)p(\rho|\boldsymbol{X}) \\
&= \sum_{S_j|i\in S_j}\ \sum_{\rho\in\mathcal{P}|S_j\in\rho} f_{S_j}(\theta_i|\boldsymbol{X}_{S_j})p(\rho|\boldsymbol{X}) \\
&= \sum_{S_j|i\in S_j} f_{S_j}(\theta_i|\boldsymbol{X}_{S_j}) \sum_{\rho\in\mathcal{P}|S_j\in\rho} p(\rho|\boldsymbol{X}) = \sum_{S_j|i\in S_j} f_{S_j}(\theta_i|\boldsymbol{X}_{S_j})r(S_j|\boldsymbol{X}),
\end{aligned}
\tag{2.9}
$$

where

$$
r(S_j|\boldsymbol{X}) = P(S_j \in \rho|\boldsymbol{X}) = \sum_{\rho\in\mathcal{P}|S_j\in\rho} p(\rho|\boldsymbol{X}) \tag{2.10}
$$

Thus, point estimates for θ_i, $i = 1, \ldots, n$, using the square loss function, can be computed by first conditioning on the partition and then averaging over all partitions, such that

$$
E\left(\theta_i|\boldsymbol{X}\right) = \sum_{\rho\in\mathcal{P}} E(\theta_i|\boldsymbol{X},\rho)p(\rho|\boldsymbol{X}) = \sum_{S_j|i\in S_j} E_{S_j}(\theta_i|\boldsymbol{X}_{S_j})r(S_j|\boldsymbol{X}), \tag{2.11}
$$

where $E_{S_j}(\theta_i|\boldsymbol{X}_{S_j})$ denotes the posterior expectation of θ_i when $\theta_i \in S_j \in \rho$.

A recursive method to exact computation of prior and posterior relevances is available in Hartigan [1990]. A similar recursive method, suitable to change point models, was introduced by Yao [1984]. It is detailed stated in Barry and Hartigan [1992]. Although the relevances play an essential role in statistical inference concerning the PPM, their exact calculation demands a high computational effort. Sampling strategies that do not depend on exact computation of relevances were proposed by Barry and Hartigan [1993] and are used in this work.

2.2 The Product Partition Model as a Bayesian multiple change point model

A multiple change point problem consists of estimating changes in one or in a group of structural parameters in the probability model of a sequence of observations, such that the number of changes and their respective locations are both unknown. In the cases when the changes are observed over a group of parameters (for example, location and scale paramenters), the changes are assigned to the group of parameters as a whole, regardless of which parameters have changed.

A PPM for change point identification is proposed by Barry and Hartigan [1992] and Barry and Hartigan [1993]. They consider $\boldsymbol{X} = (X_1, \ldots, X_n)$ a sequence of observations at consecutive points in time that undergoes structural changes at unknown times. Suppose $X_1, \ldots, X_n$ independent given the sequences of unkown structural parameters $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, with independent conditional marginal densities $f(X_i|\theta_i)$, $i = 1, \ldots, n$. The process is modeled by supposing that $\boldsymbol{S}$ is partitioned into contiguous subsequences (clusters or blocks) $S_1, \ldots, S_b$ that induces a partition of $\boldsymbol{\theta}$ in equal parameter values $\theta_{S_1}, \ldots, \theta_{S_b}$. That is, there exists a partition $\rho = \{i_0, i_1, \ldots, i_b\}$ of the set of indexes $I = \{1, 2, \ldots, n\}$, denoted by

$$\rho = \{i_0, i_1, \ldots, i_b\}, \quad 0 = i_0 < i_1 < \cdots < i_b = n,$$

such that, given ρ, there exist the common parameters $\theta_{S_1}, \ldots, \theta_{S_b}$ such that

$$\theta_i = \theta_{S_j} \quad \text{for} \quad i \in S_j = \{i_{j-1} + 1, i_{j-1} + 2, \ldots, i_j\}, \quad j = 1, 2, \ldots, b.$$

The points $i_1, \ldots, i_b$ are the end points of the blocks $S_1, \ldots, S_b$ of ρ. The first point of each block is called a change point. If the partition ρ is known, observations whose indexes belong to the same block have the same independent distribution. Note that $\rho = \{i_0, i_1, \ldots, i_b\}$ is equivalent to $\rho = \{S_1, \ldots, S_b\}$ as denoted in Section 2.1, but now with the constraint that $S_1, \ldots, S_b$ are contiguos blocks. A PPM for change point identification is thus defined if we consider the likelihood in Eq. (2.2) and the prior specifications in Eqs. (2.1) and (2.6), setting cohesion zero for all subsets of I which are not composed by contiguous indexes.

The number of end points (or blocks) $B = b$ is a random variable ranging from 1 to n, and has prior and posterior distributions given, respectively, by

$$P(B=b)=\sum_{\rho\in\mathcal{B}}p(\rho)=\sum_{\rho\in\mathcal{B}}\frac{\prod_{j=1}^{b}c(S_j)}{\sum_{\rho\in\mathcal{P}}\prod_{S_j\in\rho}c(S_j)},\quad b\in I \tag{2.12}$$

and

$$P(B=b|\boldsymbol{X})=\sum_{\rho\in\mathcal{B}}p(\rho|\boldsymbol{X})=\sum_{\rho\in\mathcal{B}}\frac{\prod_{j=1}^{b}c(S_j)f_{S_j}(\boldsymbol{X}_{S_j})}{\sum_{\rho\in\mathcal{P}}\prod_{S_j\in\rho}c(S_j)f_{S_j}(\boldsymbol{X}_{S_j})},\quad b\in I, \tag{2.13}$$

where $\mathcal{B}$ is the set of all partitions of I with b contiguous blocks.

One of the greatest challenges in the PPM is to sample from the posterior distribution of ρ. Barry and Hartigan [1993] proposed an interesting approach to sample from this posterior. This approach is discussed in the next section.

2.3 The BH93 model for detection of multiple change points

This section presents an outline of the Bayesian approach proposed by Barry and Hartigan [1993] for multiple change point detection in Normal means. This approach is based in a PPM and proceeds with a MCMC scheme to estimate the parameters and partitions. An implementation of this approach is available in the R package **bcp** (Erdman and Emerson [2007]). In addition, we present an alternative estimation procedure for this Bayesian approach, based on a partially colapsed Gibbs sampler strategy, as defined by Van Dyk and Park [2008]. This sampling strategy will be considered to estimate the Bayesian multipartition change point model proposed in Chapter 3.

2.3.1 The BH93 model

The BH93 model is applied to the case where the observations $X_1,\ldots,X_n$ are independent given the sequence of parameters $\mu_1,\ldots,\mu_n,\sigma^2$, with $X_i\sim N(\mu_i,\sigma^2)$, $i=1,\ldots,n$. The prior distribution for the block parameter μ_{S_j} is the Normal distribution $N(\mu_0,\sigma_0^2/n_j)$, where $n_j=|S_j|$ and $|S_j|$ is the cardinality of the set S_j, $j=1,\ldots,b$. The block prior cohesions follow the parametric approach suggested by Yao [1984], which considers the probability p that a change occurs at any instant in the sequence, so that

$$c(S_j)=\begin{cases}(1-p)^{n_j-1}p & \text{if } \ j=1,2,\ldots,b-1,\\ (1-p)^{n_j-1} & \text{if } \ j=b.\end{cases} \tag{2.14}$$

Assuming the cohesions in Eq. (2.14) and the prior distribution in Eq. (2.1), the prior distribution of ρ, given p, is given by

$$p(\rho|p) = p^{b-1}(1-p)^{n-b}, \tag{2.15}$$

where b is the number of blocks in ρ.

The block density of the observations $\boldsymbol{X}_{S_j}$, given μ_0, σ_0^2 and σ^2, is

$$f_{S_j}(\boldsymbol{X}_{S_j}|\mu_0, \sigma_0^2, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n_j/2}} \left(\frac{\sigma^2}{\sigma_0^2+\sigma^2}\right)^{1/2} \exp\left[-\frac{\sum_{i\in S_j}(X_i - \overline{X}_{S_j})^2}{2\sigma^2} - \frac{n_j(\overline{X}_{S_j}-\mu_0)^2}{2(\sigma_0^2+\sigma^2)}\right], \tag{2.16}$$

with $\overline{X}_{S_j} = \sum_{i\in S_j} X_i/n_j$. Barry and Hartigan [1993] assume the following independent priors for the parameters p, μ_0, σ^2 and $w = \sigma^2/(\sigma_0^2+\sigma^2)$:

$$\begin{aligned}
&f(\mu_0) \propto 1, && -\infty < \mu_0 < \infty,\\
&f(\sigma^2) \propto 1/\sigma^2, && 0 < \sigma^2 < \infty,\\
&f(p) \propto 1/p_0, && 0 < p < p_0,\\
&f(w) \propto 1/w_0, && 0 < w < w_0.
\end{aligned}$$

The authors suggest using $p_0 = 0.2$ and $w_0 = 0.2$. Under this model structure, the posterior distribution for the partition ρ is $p(\rho|X) \propto f(\boldsymbol{X}|\rho)p(\rho)$, where

$$\begin{aligned}
f(\boldsymbol{X}|\rho) &\propto \int_0^{w_0}\int_{-\infty}^{\infty}\int_0^{\infty} f(\boldsymbol{X}|\rho,\sigma^2,\mu_0,w)\frac{1}{\sigma^2}\frac{1}{w_0}\, d\sigma^2\, d\mu_0\, dw\\
&\propto \int_0^{w_0}\int_{-\infty}^{\infty}\int_0^{\infty} \left(\prod_{S_j\in\rho} f_{S_j}(\boldsymbol{X}_{S_j})\right)\frac{1}{\sigma^2}\frac{1}{w_0}\, d\sigma^2\, d\mu_0\, dw\\
&\propto \int_0^{w_0} \frac{w^{(b-1)/2}}{[W-Vw]^{(n-1)/2}}\, dw,
\end{aligned} \tag{2.17}$$

b is the number of blocks in ρ, $\overline{X} = \sum_{i=1}^{n} X_i/n$, $V = \sum_{S_j\in\rho} n_j(\overline{X}_{S_j}-\mu_0)^2$, $W = \sum_{S_j\in\rho}\sum_{i\in S_j}(X_i - \overline{X}_{S_j})^2$, and

$$\begin{aligned}
p(\rho) = \int_0^{p_0} p(\rho|p)f(p)\, dp &= \int_0^{p_0}\left(\prod_{j=1}^{b} c(S_j)\right)\frac{1}{p_0}\, dp\\
&= \int_0^{p_0}\left(\prod_{j=1}^{b-1}(1-p)^{n_j}p\right)(1-p)^{n_b}\frac{1}{p_0}\, dp\\
&\propto \int_0^{p_0} p^{b-1}(1-p)^{n-b}\, dp.
\end{aligned} \tag{2.18}$$

Thus, it follows that

$$p(\rho|\boldsymbol{X}) \propto \int_0^{w_0} \frac{w^{(b-1)/2}}{[W-Vw]^{(n-1)/2}}\, dw \int_0^{p_0} p^{b-1}(1-p)^{n-b}\, dp. \tag{2.19}$$

The assumptions considered by the BH93 model determine the posterior means of $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ is such that

$$E(\mu_i|\boldsymbol{X}) = \sum_{\rho \in \mathcal{P}} \left[(1-w^*)\overline{X}_{S_j} + w^*\overline{X}\right] p(\rho|\boldsymbol{X}),\ i \in S_j \in \rho, \tag{2.20}$$

with

$$w^* = \frac{\displaystyle\int_0^{w_0} \frac{w^{(b+1)/2}}{[W+Vw]^{(n-1)/2}}\, dw}{\displaystyle\int_0^{w_0} \frac{w^{(b-1)/2}}{[W+Vw]^{(n-1)/2}}\, dw}.$$

It also follows that the posterior mean of σ^2 is given by

$$E(\sigma^2|\boldsymbol{X}) = \sum_{\rho \in \mathcal{P}} \frac{1}{n-3} \frac{\displaystyle\int_0^{w_0} \frac{w^{(b-1)/2}}{[W+Vw]^{(n-3)/2}} dw}{\displaystyle\int_0^{w_0} \frac{w^{(b-1)/2}}{[W+Vw]^{(n-1)/2}} dw}\, p(\rho|\boldsymbol{X}). \tag{2.21}$$

Although exact computation of Eqs. (2.20) and (2.21) is possible, the sum over all partitions has a high computational cost. A MCMC implementation was proposed to achieve an approximation to these quantities. A fixed dimension representation of ρ is considered. Let $\boldsymbol{U} = (U_1, \ldots, U_{n-1})$ be a vector of auxiliary variables such that for $i = 1, \ldots, n-1$,

$$U_i = \begin{cases} 1 & \text{if } \mu_i = \mu_{i+1}, \\ 0 & \text{if } \mu_i \neq \mu_{i+1}. \end{cases}$$

Instant n is always an end point, that is equivalent to $U_n = 0$. Thus, $\boldsymbol{U}$ is a vector of random variables assuming values in $\{0, 1\}$, indicating whether or not a change point occurred at each time i, $i = 1, \ldots, n-1$. Using the Gibbs sampler described in Section B.2, in Appendix B, a sample from $p(\rho|\boldsymbol{X})$ can be obtained by successively sampling from

$$U_i^{(t)} \sim\ p\left(U_i|\boldsymbol{X}, U_1^{(t)}, \ldots, U_{i-1}^{(t)}, U_{i+1}^{(t-1)}, \ldots, U_{n-1}^{(t-1)}\right),\ \text{for } i = 1, \ldots, n-1.$$

As detailed explained in Gamerman and Lopes [2006], this sampling can be made based on a random quantity u generated from an U(0,1) distribution, as shown next. Define

$$p_i = P\left(U_i = 1|\boldsymbol{X}, U_k, k \neq i\right) = 1 - P\left(U_i = 0|\boldsymbol{X}, U_k, k \neq i\right).$$

Thus,

$$U_i^{(t)} = \mathbf{1}\left[u \leq p_i\right] = \mathbf{1}\left[\frac{u}{1-u} \leq \frac{p_i}{1-p_i}\right], \tag{2.22}$$

where $\mathbf{1}[A]$ represents the indicator function of event A and the ratio $\dfrac{p_i}{1-p_i}$ follows from Eq. (2.19) as

$$\frac{p_i}{1-p_i} = \frac{\displaystyle\int_0^{w_0} \frac{w^{b/2}}{[W_1 - V_1 w]^{(n-1)/2}}\, dw \int_0^{p_0} p^b (1-p)^{n-b-1}\, dp}{\displaystyle\int_0^{w_0} \frac{w^{(b-1)/2}}{[W_0 - V_0 w]^{(n-1)/2}}\, dw \int_0^{p_0} p^{b-1} (1-p)^{n-b}\, dp}, \tag{2.23}$$

where W_0, V_0, W_1 and V_1 are equivalent to W and V defined for Eq. (2.17). Initialize all $U_i^{(0)} = 0$. After each iteraction t, the posterior means in Eqs. (2.20) and (2.21) are computed and their average give an approximation for $\boldsymbol{\mu}$ and σ^2.

2.3.2 An alternative approach to the BH93 model

In this section, we propose an alternative and more refined estimation procedure for the BH93 model. Estimation of the partition and parameters of the BH93 model can be made through an MCMC implementation that simulate from the posterior joint distribution of $(\boldsymbol{\mu}, \sigma^2, \rho, \mu_0, w, p|\boldsymbol{X})$. It eliminates the use of numerical integrations and refines the inference about the parameters, providing samples of their posterior marginal distributions. The Gibbs sampler scheme described next is built following the conventional methodology described in Section B.2, in Appendix B. It considers the equivalent configuration of ρ as a binary vector $\boldsymbol{U} \in \{0,1\}^{n-1}$.

i. Initialize $\boldsymbol{\mu}^{(0)}, \sigma^{2(0)}, \boldsymbol{U}^{(0)}, \mu_0^{(0)}, w^{(0)}, p^{(0)}$ (initialize all $U_i^{(0)} = 0$).

ii. Set $t = 1$.

iii. Sequentially generate:

1. $p^{(t)} \sim f(p \mid \sigma^{2(t-1)}, \mu_0^{(t-1)}, w^{(t-1)}, \boldsymbol{U}^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{X})$
2. $\sigma^{2(t)} \sim f(\sigma^2 \mid p^{(t)}, \mu_0^{(t-1)}, w^{(t-1)}, \boldsymbol{U}^{(t-1)}, \boldsymbol{\mu}^{(t-1)} \boldsymbol{X})$
3. $\mu_0^{(t)} \sim f(\mu_0 \mid p^{(t)}, \sigma^{2(t)}, w^{(t-1)}, \boldsymbol{U}^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{X})$
4. $w^{(t)} \sim f(w \mid p^{(t)}, \sigma^{2(t)}, \mu_0^{(t)}, \boldsymbol{U}^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{X})$
5. $U_i^{(t)} \sim p(U_i \mid p^{(t)}, \sigma^{2(t)}, \mu_0^{(t)}, w^{(t)}, U_1^{(t)}, \ldots, U_{i-1}^{(t)}, U_{i+1}^{(t-1)}, \ldots, U_{n-1}^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{X})$, for $i = 1, \ldots, n-1$.
6. $\mu_i^{(t)} \sim f(\mu_{S_j | i \in S_j} \mid p^{(t)}, \sigma^{2(t)}, \mu_0^{(t)}, w^{(t)}, \boldsymbol{U}^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

iv. Set $t = t + 1$ and return to step 3 until convergence is reached.

This Gibbs sampler is not feasible to run. The conditional distribution in step iii.5 is degenerate in $U_i = 0$ or $U_i = 1$, because it depends on parameter vector $\boldsymbol{\mu}$, that implies the partition $\boldsymbol{U}$. The definition of the PPM implies that given the parameters, the partition is exactly specified. To solve this problem, we replace this distribution in step iii.5 with a conditional distribution that integrates out the parameter vector $\boldsymbol{\mu}$, but preserving the convergence of the samples to the desired target distribution. This modified Gibbs sampler can be defined as a partially colapsed Gibbs sampler, as proposed by Van Dyk and Park [2008]. A similar Gibbs sampler strategy was used by Dobigeon et al. [2007], and deeply discussed by Park and Van Dyk [2009].

The modified Gibbs sampler, with the parameter $\boldsymbol{\mu}$ integrated out of the conditional distribution in step iii.5 is feasible to implement. This partially colapsed Gibbs sampler is equivalent to use the blocking technique proposed by (Liu [1994]) to combine the sample distrubutions of $\boldsymbol{U}$ and $\boldsymbol{\mu}$. As pointed out by Van Dyk and Park [2008], the partially colapsed Gibbs sampler technique can be viewed as a generalization of blocking.

Define $\boldsymbol{U}_{(-i)} = (U_1, \ldots, U_{i-1}, U_{i+1}, \ldots, U_{n-1})$. The conditional distribution in step iii.5, that is given by

$$\begin{aligned} & p(U_i|\boldsymbol{X}, p, \sigma^2, \mu_0, w, \boldsymbol{U}_{(-i)}, \boldsymbol{\mu}) \\ \propto\ & f(\boldsymbol{X}|\sigma^2, \boldsymbol{U}, \boldsymbol{\mu}) f(\boldsymbol{\mu}|\mu_0, w, \sigma^2, \boldsymbol{U}) p(\boldsymbol{U}|p) \\ \propto\ & \prod_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}|\mu_{S_j}, \sigma^2) \times \prod_{j=1}^{b} f_{S_j}(\mu_{S_j}|\mu_0, w, \sigma^2) \times \prod_{j=1}^{b} c(S_j) \\ \propto\ & \prod_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}|\mu_{S_j}, \sigma^2) f_{S_j}(\mu_{S_j}|\mu_0, w, \sigma^2) c(S_j), \end{aligned} \tag{2.24}$$

is replaced by

$$\begin{aligned} & p(U_i \Big| \boldsymbol{X}, p, \sigma^2, \mu_0, w, \boldsymbol{U}_{(-i)}) \\ \propto\ & \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \prod_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}|\mu_{S_j}, \sigma^2) f_{S_j}(\mu_{S_j}|\mu_0, w, \sigma^2) c(S_j) d\mu_{S_1} \ldots d\mu_{S_b} \\ \propto\ & \prod_{j=1}^{b} \int_{-\infty}^{\infty} f_{S_j}(\boldsymbol{X}_{S_j}|\mu_{S_j}, \sigma^2) f_{S_j}(\mu_{S_j}|\mu_0, w, \sigma^2) c(S_j) d\mu_{S_j} \\ \propto\ & \prod_{j=1}^{b} f_{S_j}(\boldsymbol{X}_{S_j}|\mu_0, w, \sigma_0^2) c(S_j), \end{aligned} \tag{2.25}$$

with $f_{S_j}(\boldsymbol{X}_{S_j})$ as defined in Eq. (2.16). The same procedure used in Eq. (2.22) to sample from the partition may be applied considering this marginal conditional distribution, such that when $i \in S_j$,

$$\frac{p_i}{1-p_i} = \frac{f_{S_j}(\boldsymbol{X}_{S_j}|\mu_0, w, \sigma_0^2) c(S_j)}{f_{S_{ji}}(\boldsymbol{X}_{S_{ji}}|\mu_0, w, \sigma_0^2) c(S_{ji}) \times f_{S_{ij}}(\boldsymbol{X}_{S_{ij}}|\mu_0, w, \sigma_0^2) c(S_{ij})} \tag{2.26}$$

where $S_{ji} = \{i_{j-1}+1, i_{j-1}+2, \ldots, i-1, i\}$ and $S_{ij} = \{i+1, i+2, \ldots, i_j-1, i_j\}$ form a partition of the block S_j with two blocks such that i is the end point of S_{ji} (equivalent to $U_i = 0$). The proposed Gibbs scheme to the BH93 model is outlined next. The definition of these conditional distributions are described in Section A.1, in Appendix A.

i. Initialize $\boldsymbol{\mu}^{(0)}, \sigma^{2(0)}, \boldsymbol{U}^{(0)}, \mu_0^{(0)}, w^{(0)}, p^{(0)}$ (initialize all $U_i^{(0)} = 0$).

ii. Set $t = 1$.

iii. Sequentially generate:

1. $p^{(t)} \sim$ Beta $(b, n-b+1, [0, p_0])$
2. $\sigma^{2(t)} \sim \text{IG}\left(\frac{n+b}{2}, \frac{\sum_{i=1}^{n}(X_i-\mu_i)^2 + \frac{w}{1-w}\sum_{i=1}^{n}(\mu_i-\mu_0)^2}{2}\right)$
3. $\mu_0^{(t)} \sim \text{N}\left(\frac{\sum_{i=1}^{n}\mu_i}{n}, \frac{\sigma^2(1-w)}{nw}\right)$
4. $w^{(t)} \sim f(w|b, n_1, \ldots, n_b, \sigma^2, \boldsymbol{\mu}, \mu_0, w_0)$
5. $U_i^{(t)} \sim p(U_i|\boldsymbol{X}, p, \sigma^2, \mu_0, w, \boldsymbol{U}_{(-i)})$, for $i = 1, \ldots, n-1$.
6. $\mu_i^{(t)}|i \in S_j \sim \text{N}\left((1-w)\overline{X}_{S_j} + w\mu_0, \frac{\sigma^2(1-w)}{n_j}\right)$, for $i = 1, \ldots, n$.

iv. Set $t = t+1$ and return to step iii until convergence is reached.

The iteration indicators (t) and $(t-1)$ are not displayed to simplify the expressions of the conditional distributions, but in each step it must be used the most recently sampled value of each variable that is not sampled in that step. The quantity b represents the number of blocks of the current sampled partition. The Beta distribution in step iii.1 is truncated in $[0, p_0]$. The w conditional distribution is not a known model. Thus, a Metropolis-Hastings (M-H) algorithm was proposed to this sample. The general M-H algorithm is briefly described in Section B.1, in Appendix B. The samples from $\boldsymbol{U}^{(t)}$ in step iii.5 consider a conditional distribution with the parameter vector $\boldsymbol{\mu}$ integrated out, and $\boldsymbol{\mu}$ is sampled in the next step. It is a blocking strategy tha combines the samples of $\boldsymbol{U}$ and $\boldsymbol{\mu}$.

2.4 The L99 model for multiple change points detection

The L99 model was first presented in Loschi et al. [1999], but it is also detailed described in Loschi and Cruz [2002] and Loschi et al. [2003]. This model is an extension of the PPM to multiple change point detection in both mean and variance of sequences of Normal observations. It assumes

that given the sequence of unkown parameters $\theta_1 = (\mu_1, \sigma_1^2), \ldots, \theta_n = (\mu_n, \sigma_n^2)$, the observations $X_1, \ldots, X_n$ are independent with $X_i|\mu_i, \sigma_i^2 \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$.

It is also assumed that given a partition $\rho = \{S_1, \ldots, S_b\}$, there exists $\theta_{S_j} = (\mu_{S_j}, \sigma_{S_j}^2)$, $j = 1, \ldots, b$, such that $(\mu_i, \sigma_i^2) = (\mu_{S_j}, \sigma_{S_j}^2)$ for $i \in S_j$, $i = 1, \ldots, n$. The joint block prior distribution of $(\mu_{S_j}, \sigma_{S_j}^2)$ is given by

$$\mu_{S_j}|\sigma_{S_j}^2 \sim N(m, v\sigma_{S_j}^2) \quad \text{and} \quad \sigma_{S_j}^2 \sim IG(a/2, d/2), \tag{2.27}$$

that is called Normal-Inverse-Gamma distribution, which is a conjugate prior for the Normal model with mean and variance unknown. Similar to the BH93 model, the block prior cohesions follow the parametric model described in Eq. (2.14).

2.5 Prior specifications for the random partitions

The change point models discussed in this thesis consider the parametric cohesion model introduced by Yao [1984], described in Eq. (2.14), to determine the prior probability distribution of the random partition ρ. As presented by Loschi et al. [2003] and Loschi and Cruz [2005], and later discussed by Monteiro et al. [2011] and Peluso et al. [2019], the prior distribution of the parameter p in Eq. (2.14) may induce higher or lower informative levels for this prior, based on prior knowledge about the number of changes in each particular application. The results presented in this section support the specification of the prior distributions of the random partitions in the applications developed throughout this thesis.

Consider the distribution $p(\rho|p)$ in Eq. (2.15). If we specify the prior distribution of the parameter p to be a $Beta(\alpha, \beta)$, we have that

$$\begin{aligned} p(\rho) &= \int_0^1 p(\rho|p) f(p)\, dp = \int_0^1 p^{b-1}(1-p)^{n-b} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1}\, dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{\alpha+b-2}(1-p)^{n+\beta-b-1}\, dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+b-1)\Gamma(n+\beta-b)}{\Gamma(\alpha+\beta+n-1)}. \end{aligned} \tag{2.28}$$

Thus, the probability distribution of the random number of blocks B of ρ is given by

$$\begin{aligned} P(B=b) &= \sum_{\rho \in \mathcal{B}} p(\rho) = \sum_{\rho \in \mathcal{B}} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+b-1)\Gamma(n+\beta-b)}{\Gamma(\alpha+\beta+n-1)} \\ &= \binom{n-1}{b-1} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+b-1)\Gamma(n+\beta-b)}{\Gamma(\alpha+\beta+n-1)}, \end{aligned} \tag{2.29}$$

for $b = 1, \ldots, n$, where $\mathcal{B}$ is the set of all partitions of $I = \{1, 2, \ldots, n\}$ with b contiguous blocks. The random number of blocks B and the random number of change points, denoted by N, in a partition ρ are related by $N = B - 1$. Thus,

$$P(N = c) = P(B = c + 1) = \binom{n-1}{c} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+c)\Gamma(n-1+\beta-c)}{\Gamma(\alpha+\beta+n-1)}, \tag{2.30}$$

for $c = 0, 1, \ldots, n-1$. Eq. (2.30) determines that $N \sim \textit{Beta-Binomial}(n-1, \alpha, \beta)$, such that

$$E(N) = (n-1)\frac{\alpha}{\alpha+\beta} \tag{2.31}$$

and

$$Var(N) = (n-1)\frac{\alpha\beta(\alpha+\beta+n-1)}{(\alpha+\beta)^2(\alpha+\beta+1)} \tag{2.32}$$

The hyperparameters α and β may be determined in order to insert in the model some prior beliefs about the number of changes that have occurred in the application, and how confident we are about this beliefs. If we believe that $N = c$ regime changes have occurred, we can incorporate this expectation in the model by choosing the parameters α and β such that $E(N) = c$, and also regulate our uncertainty about this expectation by the variance of N determined by the choice of α and β. We are not restricted to an integer expected number of changes. For example, if we believe that two or three changes have occurred, we can set c as real number in the open interval $(2, 3)$.

If we fix $E(N) = c$, it follows from (2.31) that the parameters α and β are related by $\beta = \alpha(n-1-c)/c$. Denote $k = (n-1-c)/c$. Thus, we have that

$$Var(N) = c\ \frac{k}{1+k}\ \frac{\alpha(1+k)+n-1}{\alpha(1+k)+1}, \tag{2.33}$$

such that $\dfrac{\mathrm{d}}{\mathrm{d}\alpha} Var(N) = c\ \dfrac{k}{[\alpha(1+k)+1]^2}(2-n) < 0$ for $n \geq 3$ and $\lim\limits_{\alpha\to\infty} Var(N) = c(n-1-c)/(n-1)$. Thus, $Var(N)$ decreases as α increases, and we can consider more or less informative levels for the prior knowledge about the number of changes by choosing higher or lower values of α, respectively, with β following from $\beta = \alpha(n-1-c)/c$. Table 1 presents some examples of highly informative possible specifications of the N prior to scenes with $n = 100$ or $n = 200$ observations and expected number of changes varing in 1 to 9. We empirically choose $\alpha = 50$, that is a value large enough such that $Var(N)$ is next to $\lim\limits_{\alpha\to\infty} Var(N)$, and the β values in Table 1 follow from the relation $\beta = \alpha(n-1-c)/c$.

	$n = 100$			$n = 200$		
$E(N)$	β	$Var(N)$	$\lim_{\alpha\to\infty} Var(N)$	β	$Var(N)$	$\lim_{\alpha\to\infty} Var(N)$
1	4,900	1.01	0.99	9,900	1.01	0.99
2	2,425	2.04	1.96	4,925	2.06	1.98
3	1,600	3.08	2.91	3,267	3.13	2.95
4	1,188	4.14	3.84	2,438	4.23	3.92
5	940	5.22	4.75	1,940	5.36	4.87
6	775	6.31	5.64	1,608	6.51	5.82
7	657	7.41	6.51	1,371	7.69	6.75
8	569	8.52	7.35	1,194	8.9	7.68
9	500	9.64	8.18	1,056	10.13	8.59

Table 1: Examples of prior specifications of N prior to scenes with 100 or 200 observations and expected number of changes varing in 1 to 9, with $\alpha = 50$.

Additionally, fixing $\alpha = \beta = 1$ impplies that $N \sim Beta\text{-}Binomial(n-1, 1, 1)$, which is equivalent to N follow a Uniform discrete distribution over $\{0, 1, \ldots, n-1\}$, with $E(N) = (n-1)/2$ and $Var(N) = (n-1)(n+1)/12$. In this case, we assume prior distributions for p and N corresponding to the noninformative Bayes–Laplace prior distributions (Laplace [1774]).

A similar analysis can be made for the prior specification of the parameter p in the BH93 model, that is, $p \sim U(0, p_0)$, where p_0 is a fixed value in $(0, 1)$. In this case, we have the prior probability distribution of N given by

$$P(N = c) = \binom{n-1}{c} \int_0^{p_0} p^c (1-p)^{n-c-1} \frac{1}{p_0}\, dp, \tag{2.34}$$

for $c = 0, \ldots, n-1$, with

$$E(N) = E(E(N|p)) = E((n-1)p) = (n-1)\frac{p_0}{2}. \tag{2.35}$$

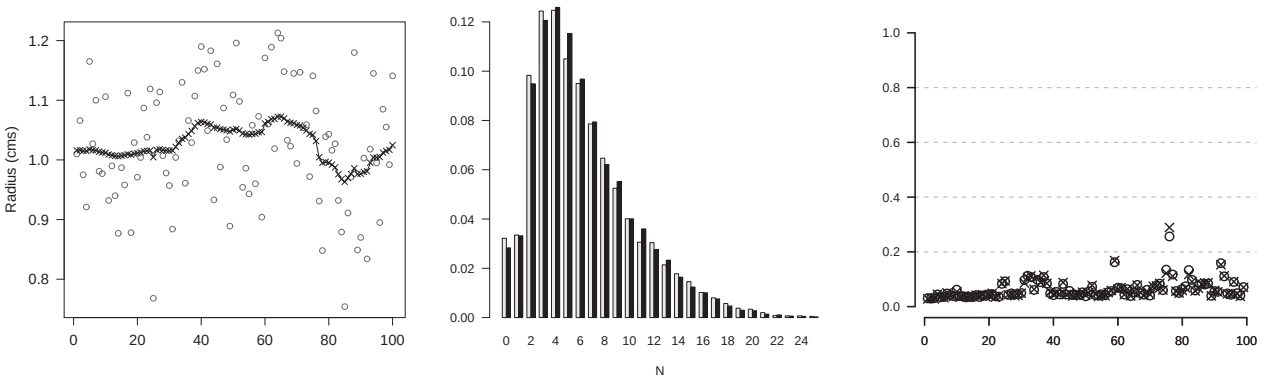
The model properties presented in this Section support some prior specifications considered in the applications presented in the Section 2.6, in the Monte Carlo simulation study presented in Chapter 4, and in the real data applications presented in Chapter 5.

2.6 Applications of the L99 and BH93 models

This section has two main purposes. First, we analyze the Lombard's dataset using the R package **bcp** and the algorithm proposed in Section 2.3.2, to compare their results. After that, we analyze a simulated data set experiencing changes in mean and variance at different times, using the L99 and BH93 models, to evaluate their limitations in this conditions and motivate the new model. The applications of the L99 model presented in this section consider the hyperparameters m, v, a and d fixed as $(m, v, a, d) = (0, 2, 2, 2)$, that implies in a flat configuration for the distribution in (2.27), and a

$Beta(\alpha, \beta)$ prior for the parameter p of the cohesions, with α and β to be specified in each application, based on the model properties discussed in Section 2.5. For the BH93 model, we follow the suggestion proposed in the original paper, that is, $p_0 = w_0 = 0.2$

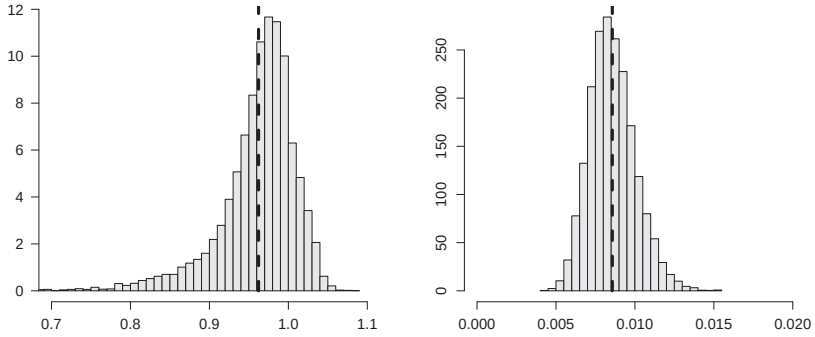
First, we apply the Gibbs scheme described in Section 2.3.2 and compare it with the R package **bcp**. We consider the $n = 100$ sequential observations of the Lombard's dataset analyzed by Barry and Hartigan [1993]. We generate 10,000 samples after a burn-in period of 4,000 discarded samples. The results are displayed in Figure 1.



(a) Posterior means of μ_i, $i = 1, \ldots, n$.

(b) Number of change points.

(c) Posterior probabilities of the i^{th} instant to be an end point.



(d) μ_{85}

(e) σ^2

Figure 1: BH93 model estimates to Lombard's dataset: (a) Lombard's data (gray circles) and posterior means estimates using bcp (black ×) and alternative approach (black solid line). (b) Estimated posterior distribution of N using bcp (gray bars) and alternative approach (black bars). (c) Estimated posterior probability of each instant to be a change point using bcp (×) and alternative approach (○). (d) Estimated marginal posterior density of μ_{85} using the alternative approach (histogram) and posterior point estimate using bcp (black dashed line). (e) Estimated marginal posterior density of σ^2 using alternative approach (histogram), and posterior point estimate informed in the original paper as 0.00857 (dashed line).

The point estimates in Figure 1a are the posterior sample means of $\mu_1, \ldots, \mu_{100}$. The posterior distribution of the number of change points N, in Figure 1b, is obtained by noticing that

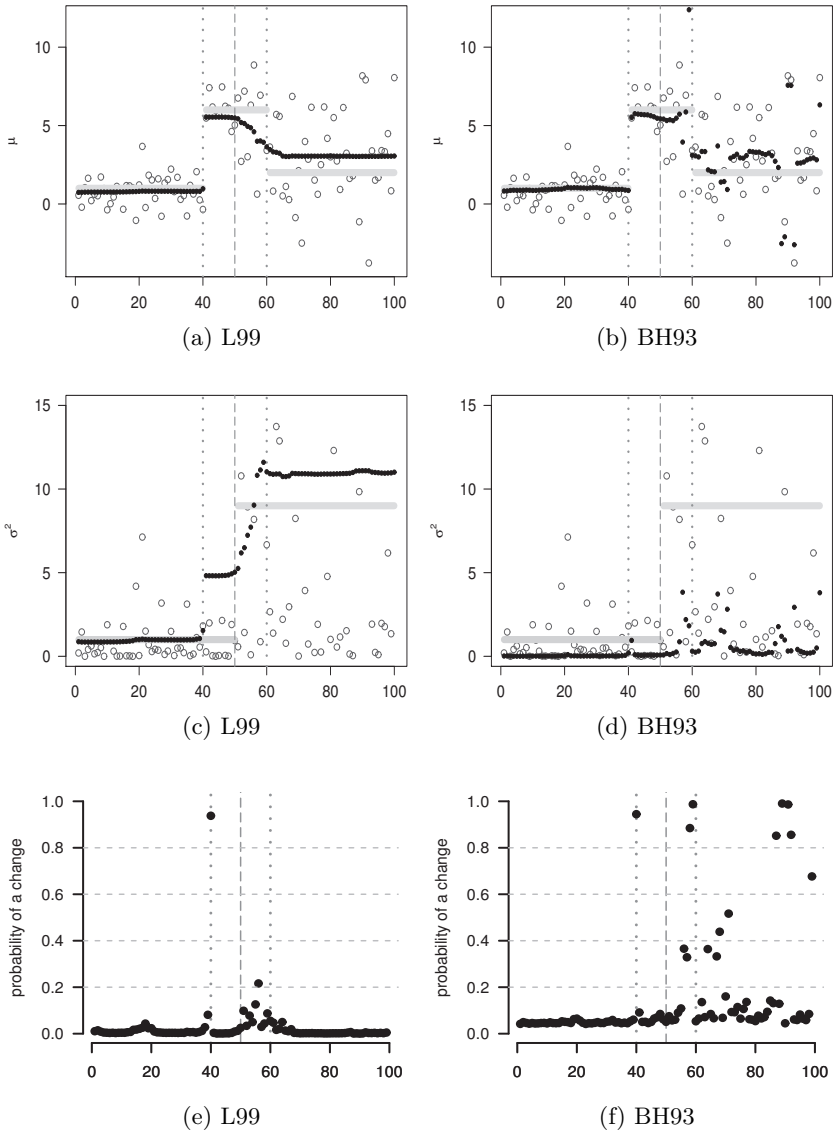
$$N = \sum_{i=1}^{n-1} (1 - U_i), \tag{2.36}$$

and the posterior probability of the i^{th} instant to be an end point in Figure 1c (and also in Figures 2e and 2f) are estimated by the proportion of the generated samples in which $U_i = 0$, $i = 1, \ldots, n-1$, as proposed by Loschi and Cruz [2005]. These estimated posterior probabilities provide useful information about the true partition, in addition to the estimated posterior probabilities of ρ, which are often too flat, not providing sufficient evidence to make decision about the most likely partition.

As shown in Figures 1a, 1b and 1c, the same results are obtained with the **bcp** package and the proposed partially colapsed Gibbs sampler, which confirms it as a useful sampling strategy. Moreover, the partially colapsed Gibbs sampler provides estimates of the marginal posterior densities of each μ_i, $i = 1, \ldots, n$, and σ^2 (Figures 1d and 1e), while the **bcp** package procedure returns only point estimates of these parameters, which may be an inappropriate information in the case of an asymmetric or bimodal posterior density, that are usual shapes for the posterior sample distributions of the parameters around the change points.

In another application, we apply the BH93 and L99 models to a simulated sequence of independent Normal data $\boldsymbol{X} = (X_1, \ldots, X_{100})$ where changes occur in the mean at instants 41 and 61, and the true means are 1, 6 and 2 for the 1^{st}, 2^{nd} and 3^{rd} clusters, respectively. Only one variance change is assumed, at instant 51. The variance changes from 1 in the cluster $(X_1, \ldots, X_{50})$ to 9 in the cluster $(X_{51}, \ldots, X_{100})$. We use the R package **bcp** to apply the BH93 model, with $p_0 = w_0 = 0.2$, as suggested in the original paper. We developed an implementation of the L99 model in R language, and considered the prior configuration presented in Section 2.4 to the block parameters. The *Beta* prior of the cohesion parameter p is specified as $p \sim Beta(50, 1600)$, following the informative prior specifications presented in Table 1 to an expected number of changes $E(N) = 3$ in $n = 100$ observations. Figure 2 summarizes the results of this application. In this application, the BH93 model does not provide good point estimates for the parameters $\boldsymbol{\mu}$ from the instant when the variance changes to the end of the sequence, and gives high probability of being a change point to a considerable number of false change points, as shown in Figures 2b and 2f. The parameters $\boldsymbol{\sigma}$ are underestimated by the BH93 model, as shown in Figure 2d. It is a default feature of the variance estimates provided by the BH93 model, as can be noticed in all the applications presented in this thesis.

Although the L99 model provides slightly better estimates for the means, both models fail to identify the true change points. Under the L99 model, the posterior estimates for the variance are clearly affected by the changes in the mean, and also the change in the variance affect the mean estimates. Moreover, the L99 and BH93 approachs do not provide information about in which parameter the change occurred. As an attempt to account for these problems, in Chapter 3 we introduce the Bayesian multipartition change point model which allows different groups of parameters to be separately partitioned. Thus, we may identify the changes and the parameter in which the changes occurred.



(a) L99 (b) BH93

(c) L99 (d) BH93

(e) L99 (f) BH93

Figure 2: L99 and BH93 model estimates to simulated Normal data: (a-b) Simulated data (gray circles), posterior mean estimates (black dots) and true mean values (gray solid horizontal lines). (c-d) The squared error (gray circles), posterior variance estimates (black dots) and the true variance values (gray solid horizontal lines). (e-f) Estimated posterior probability of each instant to be a change point. In Figures (a-f), the vertical gray lines indicate the true change points in the mean (dotted line) and variance (dashed line).

3 Multipartition model for detection of multiple change points

The PPM for multiple change point identification in sequences of observations generated by multiparametric models (Loschi et al. [1999], Loschi et al. [2010]) permits to identify when regime changes occurred but does not indicate which structural parameter experienced the change. The model proposed in this chapter extend the PPM to a multiparametric change point application such that each group of structural parameters experiences its own change point process. We propose a model with more than one partition, one for each group of structural parameters. As a result, in addition to detecting the regime changes, it is also known which structural parameter has changed.

3.1 Model assumptions and likelihood

Consider the sequence of n observations $\boldsymbol{X} = (X_1, \ldots, X_n)$ that is independent, given the d sequences of unkown structural parameters $\boldsymbol{\theta}_k = (\theta_{k,1}, \ldots, \theta_{k,n})$, $k = 1, \ldots, d$, and with conditional marginal densities $f(X_i|\theta_{1,i}, \ldots, \theta_{d,i})$, $i = 1, \ldots, n$. Suppose that each $\boldsymbol{\theta}_k$ undergoes sudden multiple changes at unknown times. Suppose that $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ are each one partitioned into contiguous blocks. Thus, there exist d partitions ρ_k, $k = 1, \ldots, d$, of the set of indexes $I = \{1, \ldots, n\}$, denoted by

$$\rho_k = \{i_{k,0}, i_{k,1}, \ldots, i_{k,b_k}\}, \quad 0 = i_{k,0} < i_{k,1} < \cdots < i_{k,b_k} = n.$$

Assume that for each given ρ_k, there exist the common parameters $\theta_{S_{k,1}}, \ldots, \theta_{S_{k,b_k}}$ such that

$$\theta_{k,i} = \theta_{S_{k,j_k}} \quad \text{for} \quad i \in S_{k,j_k} = \{i_{k,j_k-1} + 1, i_{k,j_k-1} + 2, \ldots, i_{k,j_k}\}, \quad i = 1, \ldots, n, \quad j_k = 1, \ldots, b_k.$$

The points i_{k,j_k} are the end points of the blocks S_{k,j_k} of the partitions ρ_k, for $j_k = 1, \ldots, b_k$, and define the change points of $\boldsymbol{\theta}_k$, $k = 1, \ldots, d$. The first point of each block is said to be a change point. The assumptions above determine that if the partition ρ_k is known, observations whose indexes i belongs to the same block of ρ_k have the same value for $\theta_{k,i}$. The number of end points or blocks $B_k = b_k$, in each parameter $\boldsymbol{\theta}_k$, $k = 1, \ldots, d$, is a random variable ranging from 1 to n.

As a consequence of each sequence of parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ to be blocked through an specific process represented by a specific partition, we are not constrained to blocks in matching positions, neither the same number of blocks between the partitions. For example, in a observational model with two groups of structural parameters (that is, $d = 2$), a subsequence of observations setted by ρ_1 with equal parameter $\theta_{S_{1,j_1}}$ may not be all setted by ρ_2 with equal parameter $\theta_{S_{2,j_2}}$. For a better view of this specific blocking feature of the model, we itemize in Table 2 all possible combinations of partitions ρ_1 and ρ_2 for a sequence of $n = 3$ observations, with $(b_1, b_2) \in \{1, 2, 3\} \times \{1, 2, 3\}$.

$(\boldsymbol{b_1}, \boldsymbol{b_2})$	$\begin{pmatrix} \theta_{1,1} & \theta_{1,2} & \theta_{1,3} \\ \theta_{2,1} & \theta_{2,2} & \theta_{2,3} \end{pmatrix}$
$(1,1)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} & \theta_{S_{1,1}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} & \theta_{S_{2,1}} \end{pmatrix}$
$(1,2)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} & \theta_{S_{1,1}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} & \theta_{S_{1,1}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} & \theta_{S_{2,2}} \end{pmatrix}$
$(1,3)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} & \theta_{S_{1,1}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \mid & \theta_{S_{2,3}} \end{pmatrix}$
$(2,1)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} & \theta_{S_{2,1}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} & \theta_{S_{2,1}} \end{pmatrix}$
$(2,2)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} & \theta_{S_{2,2}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} & \theta_{S_{2,2}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \end{pmatrix}$
$(2,3)$	$\begin{pmatrix} \theta_{S_{1,1}} & \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \mid & \theta_{S_{2,3}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} & \theta_{S_{1,2}} \\ \theta_{S_{2,1}} & \theta_{S_{2,2}} \mid & \theta_{S_{2,3}} \end{pmatrix}$
$(3,1)$	$\begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \mid & \theta_{S_{1,3}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} & \theta_{S_{2,1}} \end{pmatrix}$
$(3,2)$	$\begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \mid & \theta_{S_{1,3}} \\ \theta_{S_{2,1}} & \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \end{pmatrix} \begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \mid & \theta_{S_{1,3}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} & \theta_{S_{2,2}} \end{pmatrix}$
$(3,3)$	$\begin{pmatrix} \theta_{S_{1,1}} \mid & \theta_{S_{1,2}} \mid & \theta_{S_{1,3}} \\ \theta_{S_{2,1}} \mid & \theta_{S_{2,2}} \mid & \theta_{S_{2,3}} \end{pmatrix}$

Table 2: All possible combinations of partitions ρ_1 and ρ_2 for a sequence of $n = 3$ observations.

For notation simplicity, denote by $\boldsymbol{X}_{S_{1,j_1} \cdots S_{k,j_k}}$ the subsequence of observations which indexes belong to $S_{1,j_1} \cap S_{2,j_2} \cdots \cap S_{k,j_k}$, $k = 1, \ldots, d$. Assume that $\{X_i, i \in S_{1,j_1} \cdots S_{d,j_d}\}$ are independent and identicaly distributed, given $(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_d}, \rho_1, \ldots, \rho_d)$, with conditional marginal density $f(X_i|\theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}})$. Under these assumptions, the likelihood function of $(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_d}, \rho_1, \ldots, \rho_d)$ is given by

$$f(\boldsymbol{X}|\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d) = \prod_{j_1=1}^{b_1} \prod_{j_2 : \bigcap_{l=1}^{2} S_{l,j_l} \neq \emptyset} \cdots \prod_{j_d : \bigcap_{l=1}^{d} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}}|\theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}}), \quad (3.1)$$

where

$$f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}}|\theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}}) = \prod_{i \in \bigcap_{l=1}^{d} S_{l,j_l}} f(X_i|\theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}}). \quad (3.2)$$

To exemplify, consider $n = 3$ and $d = 2$, and assume $(b_1, b_2) = (2,2)$. If ρ_1 and ρ_2 are such that $S_{1,1} = \{1\}$, $S_{1,2} = \{2,3\}$, $S_{2,1} = \{1,2\}$ and $S_{2,2} = \{3\}$, which corresponds to 5^{th} line, 4^{th} column of Table 2, it follows that

$$\begin{aligned} & f(X_1, X_2, X_3 | \theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{2,1}, \theta_{2,2}, \theta_{2,3}, \rho_1 = \{0,1,3\}, \rho_2 = \{0,2,3\}) \\ = & \prod_{j_1=1}^{2} \prod_{j_2 : \bigcap_{l=1}^{2} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} S_{2,j_2}} | \theta_{S_{1,j_1}}, \theta_{S_{2,j_2}}) \\ = & f(X_1 | \theta_{S_{1,1}}, \theta_{S_{2,1}}) \times f(X_2 | \theta_{S_{1,2}}, \theta_{S_{2,1}}) \times f(X_3 | \theta_{S_{1,2}}, \theta_{S_{2,2}}). \end{aligned}$$

Another possible configuration is $S_{1,1} = \{1,2\}$, $S_{1,2} = \{3\}$, $S_{2,1} = \{1,2\}$ and $S_{2,2} = \{3\}$, that is the case in 5^{th} line and 1^{st} column of Table 2. In this case, the likelihood is given by

$$\begin{aligned} & f(X_1, X_2, X_3 | \theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{2,1}, \theta_{2,2}, \theta_{2,3}, \rho_1 = \{0,2,3\}, \rho_2 = \{0,2,3\}) \\ = & \prod_{j_1=1}^{2} \prod_{j_2 : \bigcap_{l=1}^{2} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} S_{2,j_2}} | \theta_{S_{1,j_1}}, \theta_{S_{2,j_2}}) \\ = & f(X_1 | \theta_{S_{1,1}}, \theta_{S_{2,1}}) \times f(X_2 | \theta_{S_{1,1}}, \theta_{S_{2,1}}) \times f(X_3 | \theta_{S_{1,2}}, \theta_{S_{2,2}}). \end{aligned}$$

It is relevant to note that the order of the indexes in the products of Eq. (3.1) is exchangeable. Let $\{k_1, \ldots, k_d\}$ denote any permutation of $\{1, \ldots, d\}$. Eq. (3.1) can be generalized by

$$\begin{aligned} & f(\boldsymbol{X} | \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d) \\ = & \prod_{j_{k_1}=1}^{b_{k_1}} \prod_{j_{k_2} : S_{k_1,j_{k_1}} \cap S_{k_2,j_{k_2}} \neq \emptyset} \cdots \prod_{j_{k_d} : S_{k_1,j_{k_1}} \cap \cdots \cap S_{k_d,j_{k_d}} \neq \emptyset} f(\boldsymbol{X}_{S_{k_1,j_{k_1}} \cdots S_{k_d,j_{k_d}}} | \theta_{S_{k_1,j_{k_1}}}, \ldots, \theta_{S_{k_d,j_{k_d}}}), \end{aligned} \tag{3.3}$$

where

$$f(\boldsymbol{X}_{S_{k_1,j_{k_1}} \cdots S_{k_d,j_{k_d}}} | \theta_{S_{k_1,j_{k_1}}}, \ldots, \theta_{S_{k_d,j_{k_d}}}) = \prod_{i \in \bigcap_{l=1}^{d} S_{k_l,j_l}} f(X_i | \theta_{S_{k_1,j_{k_1}}}, \ldots, \theta_{S_{k_d,j_{k_d}}}), \tag{3.4}$$

and $\boldsymbol{X}_{S_{k_1,j_{k_1}} \cdots S_{k_l,j_{k_l}}}$ denotes the subsequence of observations which indexes belong to $S_{k_1,j_{k_1}} \cap \cdots \cap S_{k_l,j_{k_l}}$, $l = 1, \ldots, d$.

3.2 Prior distributions

To specify a PPM, it is necessary to set product prior distributions for the partitions ρ_k and prior distributions for $\boldsymbol{\theta}_k$ given ρ_k, for $k = 1, ..., d$. Given the partitions ρ_k, $k = 1, ..., d$, we assume

that (i) $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ are independent and (ii) $\theta_{S_{k,1}}, \ldots, \theta_{S_{k,b_k}}$, $k = 1, \ldots, d$, are independent, such that the joint prior distribution of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ given $\rho_1, \ldots, \rho_d$ is

$$\begin{aligned} f(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d | \rho_1, \ldots, \rho_d) &= \prod_{k=1}^{d} f(\boldsymbol{\theta}_k | \rho_1, \ldots, \rho_d) = \prod_{k=1}^{d} f(\boldsymbol{\theta}_k | \rho_k) \\ &= \prod_{k=1}^{d} \prod_{j_k=1}^{b_k} f_{S_{k,j_k}}(\theta_{S_{k,j_k}}), \end{aligned} \tag{3.5}$$

where $f_{S_{k,j_k}}(\theta_{S_{k,j_k}})$ is the block prior density, for $k = 1, \ldots, d$, $j_k = 1, \ldots, b_k$. To simplify the notation, we set $f_{S_{k,j_k}}(\theta_{S_{k,j_k}}) = f_k(\theta_{S_{k,j_k}})$.

We assume that each random partition ρ_k, $k = 1, ..., d$, are independent and that each partition ρ_k has a product partition distribution, denoted as $p(\rho_k)$, given by

$$p(\rho_k) = P(\rho_k = \{S_{k,1}, \ldots, S_{k,b_k}\}) = \frac{\prod_{j_k=1}^{b_k} c_k(S_{k,j_k})}{\sum_{\rho_k \in \mathcal{P}} \prod_{S_{k,j_k} \in \rho_k} c_k(S_{k,j_k})}, \quad k = 1, \ldots, d, \tag{3.6}$$

where the cohesions $c_k(S_{k,j_k})$, *a priori*, measure how strongly we believe that the components of $\boldsymbol{\theta}_k$ in S_{k,j_k} are a cluster, $k = 1, \ldots, d$.

3.3 Joint posterior distribution of structural parameters and partitions

Under the proposed model assumptions in (3.1), (3.5) and (3.6), the posterior distribution for parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ and partitions $\rho_1, \ldots, \rho_d$ becomes

$$f(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d | \boldsymbol{X}) = \frac{f(\boldsymbol{X} | \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d) f(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d | \rho_1, \ldots, \rho_d) p(\rho_1) \cdots p(\rho_d)}{f(\boldsymbol{X})}$$

$$\propto \prod_{j_1=1}^{b_1} \prod_{j_2 : \bigcap_{l=1}^{2} S_{l,j_l} \neq \emptyset}^{} \cdots \prod_{j_d : \bigcap_{l=1}^{d} S_{l,j_l} \neq \emptyset}^{} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}} | \theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}}) \times \prod_{k=1}^{d} \prod_{j_k=1}^{b_k} f_k(\theta_{S_{k,j_k}}) c_k(S_{k,j_k}), \tag{3.7}$$

where the proportionality is supressing the term $\left[f(\boldsymbol{X}) \prod_{k=1}^{d} \sum_{\rho_k \in \mathcal{P}} \prod_{S_{k,j_k} \in \rho_k} c_k(S_{k,j_k}) \right]^{-1}$.

3.4 Sampling scheme

Inference about the partitions $\rho_1, \ldots, \rho_d$ and structural parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d$ can be made through a partially colapsed Gibbs sampler based on the same blocking strategy described in Section 2.3.2, but now adapted to consider d groups of structural parameters and d respective random partitions.

Let $\boldsymbol{\delta}$ represent a vector with the possibly existing hyperparameters in our Bayesian model. We have a model with the random quantities $(\boldsymbol{X}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d, \boldsymbol{\delta})$, and we are interested in sample from $f(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d, \boldsymbol{\delta} | \boldsymbol{X})$. To sample from the conditional distributions of $\rho_1, \ldots, \rho_d$, we consider the same fixed dimension representation of the partitions as the proposed in the BH93 model, that is, we consider the random vector $\boldsymbol{U}_k = (U_{k,1}, \ldots, U_{k,n-1})$, such that

$$U_{k,i} = \begin{cases} 1 & \text{if } \theta_{k,i} = \theta_{k,i+1}, \\ 0 & \text{if } \theta_{k,i} \neq \theta_{k,i+1}, \end{cases}$$

for $k = 1, \ldots, d$, $i = 1, \ldots, n-1$. The component $U_{k,i}$ indicates whether or not a change point occurred at time $i+1$ in the group of structural parameters $\boldsymbol{\theta}_k$. Thus, $\boldsymbol{U}_k = (U_{k,1}, \ldots, U_{k,n-1})$ determines the random partition ρ_k, for $k = 1, \ldots, d$, and the following equivalence is verified:

$$(\boldsymbol{X}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d, \boldsymbol{\delta}) \Leftrightarrow (\boldsymbol{X}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \boldsymbol{U_1}, \ldots, \boldsymbol{U_d}, \boldsymbol{\delta}).$$

In order to implement the BMCP model, the partially colapsed Gibbs sampler scheme described next may be considered to simulate from $f(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d, \rho_1, \ldots, \rho_d, \boldsymbol{\delta} | \boldsymbol{X})$.

i. Initialize $\boldsymbol{\delta}^{(0)}, \boldsymbol{\theta}_1^{(0)}, \ldots, \boldsymbol{\theta}_d^{(0)}, \boldsymbol{U}_1^{(0)}, \ldots, \boldsymbol{U}_d^{(0)}$ ($U_{k,i}^{(0)}$ may be initialized as all 0).

ii. Set $t = 1$.

iii. Sequentially generate:

0. $\boldsymbol{\delta}^{(t)} \sim f(\boldsymbol{\delta} \mid \boldsymbol{U}_1^{(t-1)}, \ldots, \boldsymbol{U}_d^{(t-1)}, \boldsymbol{\theta}_1^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$

1.1. $U_{1,i}^{(t)} \sim p(U_{1,i} \mid \boldsymbol{\delta}^{(t)}, U_{1,1}^{(t)}, \ldots, U_{1,i-1}^{(t)}, U_{1,i+1}^{(t-1)}, \ldots, U_{1,n-1}^{(t-1)}, \boldsymbol{U}_2^{(t-1)}, \ldots, \boldsymbol{U}_d^{(t-1)}, \boldsymbol{\theta}_2^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$,
for $i = 1, \ldots, n-1$.

1.2. $\theta_{1,i}^{(t)} \sim f(\theta_{S_{1,j_1} | i \in S_{1,j_1}} \mid \boldsymbol{\delta}^{(t)}, \boldsymbol{U}_1^{(t)}, \boldsymbol{U}_2^{(t-1)}, \ldots, \boldsymbol{U}_d^{(t-1)}, \boldsymbol{\theta}_2^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

$\vdots$

k.1. $U_{k,i}^{(t)} \sim$
$p(U_{k,i} \mid \boldsymbol{\delta}^{(t)}, \boldsymbol{U}_1^{(t)}, \ldots, \boldsymbol{U}_{k-1}^{(t)}, \boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_{k-1}^{(t)}, U_{k,1}^{(t)}, \ldots, U_{k,i-1}^{(t)}, U_{k,i+1}^{(t-1)}, \ldots, U_{k,n-1}^{(t-1)}, \boldsymbol{U}_{k+1}^{(t-1)}, \ldots, \boldsymbol{U}_d^{(t-1)}, \boldsymbol{\theta}_{k+1}^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$,
for $i = 1, \ldots, n-1$.

k.2. $\theta_{k,i}^{(t)} \sim f(\theta_{S_{k,j_k} | i \in S_{k,j_k}} \mid \boldsymbol{\delta}^{(t)}, \boldsymbol{U}_1^{(t)}, \ldots, \boldsymbol{U}_k^{(t)}, \boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_{k-1}^{(t)}, \boldsymbol{U}_{k+1}^{(t-1)}, \ldots, \boldsymbol{U}_d^{(t-1)}, \boldsymbol{\theta}_{k+1}^{(t-1)}, \ldots, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

$\vdots$

d.1. $U_{d,i}^{(t)} \sim p(U_{d,i} \mid \boldsymbol{\delta}^{(t)}, \boldsymbol{U}_1^{(t)}, \ldots, \boldsymbol{U}_{d-1}^{(t)}, \boldsymbol{\theta}_1^{(t)}, \ldots, \boldsymbol{\theta}_{d-1}^{(t)}, U_{d,1}^{(t)}, \ldots, U_{d,i-1}^{(t)}, U_{d,i+1}^{(t-1)}, \ldots, U_{d,n-1}^{(t-1)}, \boldsymbol{\theta}_d^{(t-1)}, \boldsymbol{X})$,
for $i = 1, \ldots, n-1$.

d.2. $\theta_{d,i}^{(t)} \sim f(\theta_{S_{d,j_d} | i \in S_{d,j_d}} \mid \boldsymbol{\delta}^{(t)}, \boldsymbol{U}_1^{(t)}, \ldots, \boldsymbol{U}_d^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

iv. Set $t = t + 1$ and return to step iii until convergence is reached.

The $\boldsymbol{\delta}$ sample in step iii.0 represents a group of samples of each hyperparameter that may belong to the model. The structural parameters vectors $\boldsymbol{\theta}_k$, $k = 1, \ldots, d$, are integrated out of their respective conditional distribution of $\boldsymbol{U}_k$, in step iii.k.1, which is a necessary procedure, as discussed in Section 2.3.2. Each pair of ordered samples (iii.k.1,iii.k.2) is a blocking step that combines the samples of $(\boldsymbol{U}_k, \boldsymbol{\theta}_k)$.

Denote $\boldsymbol{U}_{k,(-i)} = (U_{k,1}, \ldots, U_{k,i-1}, U_{k,i+1}, \ldots, U_{k,n-1})$. The samples from the conditional distributions in steps iii.k.1, $k = 1, \ldots, d$, follow the same procedure described in Eq. (2.22), considering now

$$\frac{p_{k,i}}{1 - p_{k,i}} = \frac{p(U_{k,i} = 1 \mid \boldsymbol{U}_1, \ldots, \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k,(-i)}, \boldsymbol{U}_{k+1}, \ldots, \boldsymbol{U}_d, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k+1}, \ldots, \boldsymbol{\theta}_d, \boldsymbol{\delta}, \boldsymbol{X})}{p(U_{k,i} = 0 \mid \boldsymbol{U}_1, \ldots, \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k,(-i)}, \boldsymbol{U}_{k+1}, \ldots, \boldsymbol{U}_d, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k+1}, \ldots, \boldsymbol{\theta}_d, \boldsymbol{\delta}, \boldsymbol{X})}, \quad (3.8)$$

It follows from Eq. (3.7) that

$$p(U_{k,i} \mid \boldsymbol{U}_1, \ldots, \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k,(-i)}, \boldsymbol{U}_{k+1}, \ldots, \boldsymbol{U}_d, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k+1}, \ldots, \boldsymbol{\theta}_d, \boldsymbol{\delta}, \boldsymbol{X})$$

$$= \int \cdots \int p(U_{k,i} \mid \boldsymbol{U}_1, \ldots, \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k,(-i)}, \boldsymbol{U}_{k+1}, \ldots, \boldsymbol{U}_d, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k+1}, \ldots, \boldsymbol{\theta}_d, \boldsymbol{\delta}, \boldsymbol{X})\ d\theta_{k,1} \ldots d\theta_{k,n}$$

$$\propto \int \cdots \int \prod_{j_1=1}^{b_1} \prod_{j_2 : \bigcap_{l=1}^{2} S_{l,j_l} \neq \emptyset} \cdots \prod_{j_d : \bigcap_{l=1}^{d} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}} | \theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}})$$

$$\times \prod_{j_1=1}^{b_1} f_1(\theta_{S_{1,j_1}}) c_1(S_{1,j_1}) \times \cdots \times \prod_{j_k=1}^{b_k} f_k(\theta_{S_{k,j_k}}) c_k(S_{k,j_k}) \times \cdots \times \prod_{j_d=1}^{b_d} f_d(\theta_{S_{d,j_d}}) c_d(S_{d,j_d})\ d\theta_{S_{k,j_1}} \ldots d\theta_{S_{k,j_{b_k}}}$$

$$\propto \int \cdots \int \prod_{j_k=1}^{b_k} \prod_{j_1 : \bigcap_{l \in \{k,1\}} S_{l,j_l} \neq \emptyset} \cdots \prod_{j_d : \bigcap_{l \in \{k,1,\ldots,d\}} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}} | \theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}})$$

$$\times \prod_{j_1=1}^{b_1} f_1(\theta_{S_{1,j_1}}) c_1(S_{1,j_1}) \times \cdots \times \prod_{j_k=1}^{b_k} f_k(\theta_{S_{k,j_k}}) c_k(S_{k,j_k}) \times \cdots \times \prod_{j_d=1}^{b_d} f_d(\theta_{S_{d,j_d}}) c_d(S_{d,j_d})\ d\theta_{S_{k,j_1}} \ldots d\theta_{S_{k,j_{b_k}}}$$

$$\propto \prod_{j_k=1}^{b_k} \int \prod_{j_1 : \bigcap_{l \in \{k,1\}} S_{l,j_l} \neq \emptyset} \cdots \prod_{j_d : \bigcap_{l \in \{k,1,\ldots,d\}} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}} | \theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}})$$

$$f_k(\theta_{S_{k,j_k}}) c_k(S_{k,j_k})\ d\theta_{S_{k,j_k}}$$

$$\propto \prod_{j_k=1}^{b_k} c_k(S_{k,j_k})\ \ f(\boldsymbol{X}_{S_{k,j_k}} | \theta_{S_{l,j_l}} : S_{l,j_l} \cap S_{k,j_k} \neq \emptyset), \tag{3.9}$$

where

$$f(\boldsymbol{X}_{S_{k,j_k}} | \theta_{S_{l,j_l}} : S_{k,j_k} \cap S_{l,j_l} \neq \emptyset)$$

$$= \int \prod_{j_1 : \bigcap_{l \in \{k,1\}} S_{l,j_l} \neq \emptyset} \cdots \prod_{j_d : \bigcap_{l \in \{k,1,\ldots,d\}} S_{l,j_l} \neq \emptyset} f(\boldsymbol{X}_{S_{1,j_1} \cdots S_{d,j_d}} | \theta_{S_{1,j_1}}, \ldots, \theta_{S_{d,j_d}}) f_k(\theta_{S_{k,j_k}})\ d\theta_{S_{k,j_k}}, \tag{3.10}$$

and $\{k, 1, \ldots, d\}$ is a simplified notation to $\{k, 1, \ldots, k-1, k+1, \ldots, d\}$.

Denote $S_{ki,j_k} = \{i_{k,j_{k-1}} + 1, i_{k,j_{k-1}} + 2, \ldots, i-1, i\}$ and $S_{ik,j_k} = \{i+1, i+2, \ldots, i_{k,j_k} - 1, i_{k,j_k}\}$. Therefore, for $i \in S_{k,j_k}$,

$$U_{k,i}^{(t)} = \mathbf{1}\left[u \leq p_{k,i}\right] = \mathbf{1}\left[\frac{u}{1-u} \leq \frac{p_{k,i}}{1-p_{k,i}}\right], \tag{3.11}$$

with

$$\frac{p_{k,i}}{1-p_{k,i}} = \frac{f(\boldsymbol{X}_{S_{k,j_k}}|\theta_{S_{l,j_l}} : S_{k,j_k} \cap S_{l,j_l} \neq \emptyset)c_k(S_{k,j_k})}{f(\boldsymbol{X}_{S_{ki,j_k}}|\theta_{S_{l,j_l}} : S_{ki,j_k} \cap S_{l,j_l} \neq \emptyset)c_{ki}(S_{ki,j_k}) \quad \times f(\boldsymbol{X}_{S_{ik,j_k}}|\theta_{S_{l,j_l}} : S_{ik,j_k} \cap S_{l,j_l} \neq \emptyset)c_{ik}(S_{ik,j_k})}. \tag{3.12}$$

The indexes subsets $S_{1,1}, \ldots, S_{1,b_1}, S_{2,1}, \ldots, S_{2,b_2}, \ldots, S_{d,1}, \ldots, S_{d,b_d}$, to be considered in the products configuration in the third line of Eq.(3.9) are directly determined by the given quantities $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_{k-1}, \boldsymbol{U}_{k,(-i)}, \boldsymbol{U}_{k+1}, \ldots, \boldsymbol{U}_d$ and $U_{k,i} = 0$ or $U_{k,i} = 1$. Also in the third line, we exchange the products order, setting the first product to be over the blocks of the respective partition k being sampled in step iii.k.1. This exchange permits the multiple integral becomes a product of single integrals. All terms that leave the Eq.(3.9) through the proportionality assumption are common to the numerator and denominator of the ratio in the Eq.(3.8), not affecting its final value.

In the next section, we apply the BMCP model and the sampling procedures presented before in this chapter to multiple change points detection in the mean and variance of sequentially observed Normal data.

3.5 The BMCP model for Normal data

Consider $X_1, \ldots, X_n$ are independent given the $d = 2$ sequences of unkown structural parameters $\boldsymbol{\theta}_1 = (\theta_{1,1}, \ldots, \theta_{1,n}) = (\mu_1, \ldots, \mu_n)$ and $\boldsymbol{\theta}_2 = (\theta_{2,1}, \ldots, \theta_{2,n}) = (\sigma_1^2, \ldots, \sigma_n^2)$, with conditional marginal densities $f(X_i|\mu_i, \sigma_i^2)$ as the density of a $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$.

Suppose $f(X_i|\mu_i, \sigma_i^2), i = 1, \ldots, n$, experience changes in $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)$ and $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_n^2)$, at unknown times, and there exist partitions ρ_1 and ρ_2 of $I = \{1, 2, \ldots, n\}$ such that, given ρ_1, there exist the common parameters $\mu_{S_{1,1}}, \ldots, \mu_{S_{1,b_1}}$ such that

$$\mu_i = \mu_{S_{1,j_1}} \quad \text{for} \quad i \in S_{1,j_1} = \{i_{1,j_1-1} + 1, i_{1,j_1-1} + 2, \ldots, i_{1,j_1}\}, \quad i = 1, \ldots, n, \quad j_1 = 1, 2, \ldots, b_1,$$

and, given ρ_2, there exist the common parameters $\sigma^2_{S_{2,1}}, \ldots, \sigma^2_{S_{2,b_2}}$ such that

$$\sigma_i^2 = \sigma^2_{S_{2,j_2}} \quad \text{for} \quad i \in S_{2,j_2} = \{i_{2,j_2-1} + 1, i_{2,j_2-1} + 2, \ldots, i_{2,j_2}\}, \quad i = 1, \ldots, n, \quad j_2 = 1, 2, \ldots, b_2.$$

To simplify the notation, denote $S_{1,j_1} = S_{j_1}$, $S_{2,j_2} = S_{j_2}$, $S_{1,j_1} \cap S_{2,j_2} = S_{j_1}S_{j_2}$, and denote by $\boldsymbol{X}_{S_{j_1}S_{j_2}}$ the sequence of observations which indexes belong to $S_{j_1}S_{j_2}$. Also, denote $\mu_{S_{1,j_1}} = \mu_{S_{j_1}}$ and $\sigma^2_{S_{2,j_2}} = \sigma^2_{S_{j_2}}$.

Assume that $\{X_i, i \in S_{j_1}S_{j_2}\}$ are independent and identicaly distributed with conditional marginal density $f(X_i|\mu_{S_{j_1}}, \sigma^2_{S_{j_2}})$. Thus, the likelihood function of $(\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho_1, \rho_2)$ is given by

$$f(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2) = \prod_{j_1=1}^{b_1} \prod_{j_2:S_{j_1}S_{j_2}\neq\emptyset} f(\boldsymbol{X}_{S_{j_1}S_{j_2}}|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}), \tag{3.13}$$

where

$$f(\boldsymbol{X}_{S_{j_1}S_{j_2}}|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}) = \prod_{i\in S_{j_1}S_{j_2}} f(X_i|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}). \tag{3.14}$$

Given (ρ_1,ρ_2), we assume $(\boldsymbol{\mu},\boldsymbol{\sigma})$ independent, with joint prior distribution given by

$$\begin{aligned} f(\boldsymbol{\mu},\boldsymbol{\sigma}|\rho_1,\rho_2) &= f(\boldsymbol{\mu}|\rho_1,\rho_2)f(\boldsymbol{\sigma}|\rho_1,\rho_2) = f(\boldsymbol{\mu}|\rho_1)f(\boldsymbol{\sigma}|\rho_2) \\ &= \prod_{j_1=1}^{b_1} f_{S_{1,j_1}}(\mu_{S_{j_1}}) \prod_{j_2=1}^{b_2} f_{S_{2,j_2}}(\sigma^2_{S_{j_2}}), \end{aligned} \tag{3.15}$$

where $f_{S_{1,j_1}}(\mu_{S_{j_1}})$ and $f_{S_{2,j_2}}(\sigma^2_{S_{j_2}})$ are the block prior densities, that we propose to be

$$\mu_{S_{j_1}} \sim N(\mu_0,\sigma_0^2), \quad j_1 = 1,...,b_1, \tag{3.16}$$

and

$$\sigma^2_{S_{j_2}} \sim IG(a/2,d/2), \quad j_2 = 1,...,b_2. \tag{3.17}$$

We assume the random partitions ρ_1 and ρ_2 are distributed according to independent product partition distributions, determined by block prior cohesions as the described in Eq. (2.14), with distinct parameters p_1 and p_2, respectively.

The posterior distribution for parameters and partitions is given by

$$\begin{aligned} f(\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2|\boldsymbol{X}) &= \frac{f(\boldsymbol{X}|\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2)f(\boldsymbol{\mu},\boldsymbol{\sigma}|\rho_1,\rho_2)p(\rho_1|p_1)f(p_1)p(\rho_2|p_2)f(p_2)}{f(\boldsymbol{X})} \\ &\propto \prod_{j_1=1}^{b_1} \prod_{j_2:S_{j_1}S_{j_2}\neq\emptyset} f(\boldsymbol{X}_{S_{j_1}S_{j_2}}|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}) \times \prod_{j_1=1}^{b_1} f_{S_{1,j_1}}(\mu_{S_{j_1}}) \times \prod_{j_2=1}^{b_2} f_{S_{2,j_2}}(\sigma^2_{S_{j_2}}) \\ &\times p_1^{\alpha_1+b_1-2}(1-p_1)^{n+\beta_1-b_1-1} \times p_2^{\alpha_2+b_2-2}(1-p_2)^{n+\beta_2-b_2-1}. \end{aligned} \tag{3.18}$$

Next, we present the partially colapsed Gibbs sampler scheme described in Section 3.4, now specified to the current Normal BMCP model, that is formed by the random quantities $(\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2)$, such that

$$(\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2) \Leftrightarrow (\boldsymbol{X},\boldsymbol{\mu},\boldsymbol{\sigma},\boldsymbol{U}_1,\boldsymbol{U}_2,p_1,p_2),$$

with $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ as defined in Section 3.4.

In order to implement the Normal BMCP model, the partially colapsed Gibbs sampler scheme described next may be considered to simulate from $f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{U}_1, \boldsymbol{U}_2, p_1, p_2|\boldsymbol{X})$.

i. Initialize $\boldsymbol{\mu}^{(0)}, \boldsymbol{\sigma}^{(0)}, \boldsymbol{U}_1^{(0)}, \boldsymbol{U}_2^{(0)}, p_1^{(0)}, p_2^{(0)}$ (initialize all $U_{k,i}^{(0)}$ to be 0).

ii. Set $t = 1$.

iii. Sequentially generate:

0.1. $p_1 \sim f(p_1 \mid \boldsymbol{U}_1^{(t-1)}, \boldsymbol{U}_2^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, p_2^{(t-1)}, \boldsymbol{X})$

0.1. $p_2 \sim f(p_2 \mid \boldsymbol{U}_1^{(t-1)}, \boldsymbol{U}_2^{(t-1)}, \boldsymbol{\mu}^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, p_1^{(t)}, \boldsymbol{X})$

1.1. $U_{1,i}^{(t)} \sim p(U_{1,i} \mid U_{1,1}^{(t)}, \ldots, U_{1,i-1}^{(t)}, U_{1,i+1}^{(t-1)}, \ldots, U_{1,n-1}^{(t-1)}, \boldsymbol{U}_2^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, p_1^{(t)}, p_2^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n-1$.

1.2. $\mu_i^{(t)} \sim f(\mu_{S_{j_1}|i\in S_{1,j_1}} \mid \boldsymbol{U}_1^{(t)}, \boldsymbol{U}_2^{(t-1)}, \boldsymbol{\sigma}^{(t-1)}, p_1^{(t)}, p_2^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

2.1. $U_{2,i}^{(t)} \sim p(U_{2,i} \mid \boldsymbol{U}_1^{(t)}, U_{2,1}^{(t)}, \ldots, U_{2,i-1}^{(t)}, U_{2,i+1}^{(t-1)}, \ldots, U_{2,n-1}^{(t-1)}, \boldsymbol{\mu}^{(t)}, p_1^{(t)}, p_2^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n-1$.

2.2. $\sigma_i^{2(t)} \sim f(\sigma^2_{S_{j_2}|i\in S_{2,j_2}} \mid \boldsymbol{U}_1^{(t)}, \boldsymbol{U}_2^{(t)}, \boldsymbol{\mu}^{(t)}, p_1^{(t)}, p_2^{(t)}, \boldsymbol{X})$, for $i = 1, \ldots, n$.

iv. Set $t = t + 1$ and return to step iii until convergence is reached.

Denote $n_{j_1} = |S_{1,j_1}|$, $n_{j_2} = |S_{2,j_2}|$, $n_{j_1 j_2} = |S_{1,j_1} \cap S_{2,j_2}|$ and $\overline{X}_{S_{j_1} S_{j_2}} = \sum_{i\in S_{j_1} S_{j_2}} X_i / n_{j_1 j_2}$. Also, to simplify notation, denote $j_2|S_{j_1} = \{j_2 : S_{1,j_1} \cap S_{2,j_2} \neq \emptyset\}$ and $j_1|S_{j_2} = \{j_1 : S_{1,j_1} \cap S_{2,j_2} \neq \emptyset\}$. To sample from the conditional distribution in step iii.1.1, the Eq. (3.10) is given by

$$f(\boldsymbol{X}_{S_{1,j_1}}|\sigma^2_{S_{2,j_2}} : S_{1,j_1} \cap S_{2,j_2} \neq \emptyset) = \int \prod_{j_2:S_{j_1}} f(\boldsymbol{X}_{S_{j_1} S_{j_2}}|\mu_{S_{j_1}}, \sigma^2_{S_{j_2}}) \times f_{S_{1,j_1}}(\mu_{S_{j_1}})\, d\mu_{S_{j_1}}$$

$$= \int \prod_{j_2:S_{j_1}} (2\pi\sigma^2_{S_{j_2}})^{-n_{j_1 j_2}/2} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \sum_{i\in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2\right\} \times (2\pi\sigma_0^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma_0^2}(\mu_{S_{j_1}} - \mu_0)^2\right\} d\mu_{S_{j_1}}$$

$$= (2\pi)^{-(n_r+1)/2} (\sigma_0^2)^{-1/2} \prod_{j_2:S_{j_1}} (\sigma^2_{S_{j_2}})^{-n_{j_1 j_2}/2}$$

$$\times \int \prod_{j_2:S_{j_1}} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \sum_{i\in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2\right\} \times \exp\left\{-\frac{1}{2\sigma_0^2}(\mu_{S_{j_1}} - \mu_0)^2\right\} d\mu_{S_{j_1}}. \tag{3.19}$$

The integral in Eq. (3.19) may be computed as

$$\int \prod_{j_2:S_{j_1}} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \sum_{i\in S_{j_1}S_{j_2}} (X_i-\mu_{S_{j_1}})^2\right\} \times \exp\left\{-\frac{1}{2\sigma_0^2}(\mu_{S_{j_1}}-\mu_0)^2\right\}\, d\mu_{S_{j_1}}$$

$$= \int \exp\left\{-\frac{1}{2}\left[\mu^2_{S_{j_1}}\left(\sum_{j_2:S_{j_1}} \frac{n_{j_1j_2}}{\sigma^2_{S_{j_2}}}+\frac{1}{\sigma_0^2}\right) - 2\mu_{S_{j_1}}\left(\sum_{j_2:S_{j_1}} \frac{n_{j_1j_2}\overline{X}_{S_{j_1}S_{j_2}}}{\sigma^2_{S_{j_2}}}+\frac{\mu_0}{\sigma_0^2}\right) + \sum_{j_2:S_{j_1}} \frac{\sum_{i\in S_{j_1}S_{j_2}} X_i^2}{\sigma^2_{S_{j_2}}}+\frac{\mu_0^2}{\sigma_0^2}\right]\right\} d\mu_{S_{j_1}}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{j_2:S_{j_1}} \frac{\sum_{i\in S_{j_1}S_{j_2}} X_i^2}{\sigma^2_{S_{j_2}}}+\frac{\mu_0^2}{\sigma_0^2}-\frac{\boldsymbol{Q_2^2}}{\boldsymbol{Q_1}}\right]\right\} \int \exp\left\{-\frac{\boldsymbol{Q_1}}{2}\left[\mu^2_{S_{j_1}} - 2\mu_{S_{j_1}}\frac{\boldsymbol{Q_2}}{\boldsymbol{Q_1}}+\left(\frac{\boldsymbol{Q_2}}{\boldsymbol{Q_1}}\right)^2\right]\right\} d\mu_{S_{j_1}}$$

$$= \exp\left\{-\frac{1}{2}\left[\sum_{j_2:S_{j_1}} \frac{\sum_{i\in S_{j_1}S_{j_2}} X_i^2}{\sigma^2_{S_{j_2}}}+\frac{\mu_0^2}{\sigma_0^2}-\frac{\boldsymbol{Q_2^2}}{\boldsymbol{Q_1}}\right]\right\} \left(\frac{\boldsymbol{Q_1}}{2\pi}\right)^{-1/2},$$

where $\boldsymbol{Q_1} = \sum_{j_2:S_{j_1}} \frac{n_{j_1j_2}}{\sigma^2_{S_{j_2}}}+\frac{1}{\sigma_0^2}$ and $\boldsymbol{Q_2} = \sum_{j_2:S_{j_1}} \frac{n_{j_1j_2}\overline{X}_{S_{j_1}S_{j_2}}}{\sigma^2_{S_{j_2}}}+\frac{\mu_0}{\sigma_0^2}$. Thus, the density in Eq. (3.19) becomes

$$\begin{aligned}
&f(\boldsymbol{X}_{S_{1,j_1}}|\sigma^2_{S_{2,j_2}} : S_{1,j_1}\cap S_{2,j_2}\neq\emptyset)\\
=&(2\pi)^{-(n_{j_1}+1)/2}(\sigma_0^2)^{-1/2}\prod_{j_2:S_{j_1}}(\sigma^2_{S_{j_2}})^{-n_{j_1j_2}/2}\times\exp\left\{-\frac{1}{2}\left[\sum_{j_2:S_{j_1}}\frac{\sum_{i\in S_{j_1}S_{j_2}}X_i^2}{\sigma^2_{S_{j_2}}}+\frac{\mu_0^2}{\sigma_0^2}-\frac{\boldsymbol{Q_2^2}}{\boldsymbol{Q_1}}\right]\right\}\left(\frac{\boldsymbol{Q_1}}{2\pi}\right)^{-1/2}\\
=&(2\pi)^{-n_{j_1}/2}(\sigma_0^2\boldsymbol{Q_1})^{-1/2}\prod_{j_2:S_{j_1}}(\sigma^2_{S_{j_2}})^{-n_{j_1j_2}/2}\times\exp\left\{-\frac{1}{2}\left[\sum_{j_2:S_{j_1}}\frac{\sum_{i\in S_{j_1}S_{j_2}}X_i^2}{\sigma^2_{S_{j_2}}}+\frac{\mu_0^2}{\sigma_0^2}-\frac{\boldsymbol{Q_2^2}}{\boldsymbol{Q_1}}\right]\right\},
\end{aligned} \tag{3.20}$$

that does not depend on $\boldsymbol{\mu}$.

To sample from the conditional distribution in step iii.2.1, the Eq. (3.10) is given by

$$f(\boldsymbol{X}_{S_{2,j_2}}|\mu_{S_{1,j_1}} : S_{2,j_2} \cap S_{1,j_1} \neq \emptyset) = \int \prod_{j_1:S_{j_2}} f(\boldsymbol{X}_{M_r L_s}|\mu_{S_{j_1}}, \sigma^2_{S_{j_2}}) \times f_{S_{2,j_2}}(\sigma^2_{S_{j_2}})\, d\sigma^2_{S_{j_2}}$$

$$= \int \prod_{j_1:S_{j_2}} (2\pi\sigma^2_{S_{j_2}})^{-n_{j_1 j_2}/2} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \sum_{i \in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2\right\} \times \frac{(a/2)^{d/2}}{\Gamma(d/2)} (\sigma^2_{S_{j_2}})^{-(d+2)/2} \exp\left\{-\frac{a}{2\sigma^2_{S_{j_2}}}\right\} d\sigma^2_{S_{j_2}}$$

$$= (2\pi)^{-n_{j_2}/2} \frac{(a/2)^{d/2}}{\Gamma(d/2)} \int (\sigma^2_{S_{j_2}})^{-n_{j_2}/2} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \sum_{j_1:S_{j_2}} \sum_{i \in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2\right\} \times (\sigma^2_{S_{j_2}})^{-(d+2)/2} \exp\left\{-\frac{a}{2\sigma^2_{S_{j_2}}}\right\} d\sigma^2_{S_{j_2}}$$

$$= (2\pi)^{-n_{j_2}/2} \frac{(a/2)^{d/2}}{\Gamma(d/2)} \int (\sigma^2_{S_{j_2}})^{-(n_{j_2}+d+2)/2} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \left[\sum_{j_1:S_{j_2}} \sum_{i \in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2 + a\right]\right\} d\sigma^2_{S_{j_2}}. \tag{3.21}$$

The integral in Eq. (3.21) may be computed as

$$\int (\sigma^2_{S_{j_2}})^{-(n_{j_2}+d+2)/2} \exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}} \left[\sum_{j_1:S_{j_2}} \sum_{i \in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2 + a\right]\right\} d\sigma^2_{S_{j_2}}$$

$$= \int (\sigma^2_{S_{j_2}})^{-(\boldsymbol{D}+2)/2} \exp\left\{-\frac{\boldsymbol{A}}{2\sigma^2_{S_{j_2}}}\right\} d\sigma^2_{S_{j_2}}$$

$$= \frac{\Gamma(\boldsymbol{D}/2)}{(\boldsymbol{A}/2)^{\boldsymbol{D}/2}} \int \frac{(\boldsymbol{A}/2)^{\boldsymbol{D}/2}}{\Gamma(\boldsymbol{D}/2)} (\sigma^2_{S_{j_2}})^{-(\boldsymbol{D}+2)/2} \exp\left\{-\frac{\boldsymbol{A}}{2\sigma^2_{S_{j_2}}}\right\} d\sigma^2_{S_{j_2}}$$

$$= \frac{\Gamma(\boldsymbol{D}/2)}{(\boldsymbol{A}/2)^{\boldsymbol{D}/2}}$$

where $\boldsymbol{D} = n_{j_2} + d$ and $\boldsymbol{A} = \sum_{j_1:S_{j_2}} \sum_{i \in S_{j_1} S_{j_2}} (X_i - \mu_{S_{j_1}})^2 + a$. Thus, the density in Eq. (3.21) becomes

$$f(\boldsymbol{X}_{S_{2,j_2}}|\mu_{S_{1,j_1}} : S_{2,j_2} \cap S_{1,j_1} \neq \emptyset) = (2\pi)^{-n_{j_2}/2} \frac{(a/2)^{d/2}}{\Gamma(d/2)} \frac{\Gamma(\boldsymbol{D}/2)}{(\boldsymbol{A}/2)^{\boldsymbol{D}/2}}, \tag{3.22}$$

that does not depend on $\boldsymbol{\sigma}$. Eqs. (3.20) and (3.22) are used to sample from partitions ρ_1 and ρ_2 according to Eqs. (3.11) and (3.12), respectively.

The conditional distributions of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, p_1 and p_2 are presented next.

$$f(\boldsymbol{\mu}|\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2,\boldsymbol{X}) \propto f(\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2|\boldsymbol{X})$$

$$\propto \prod_{j_1=1}^{b_1}\prod_{j_2:S_{j_1}} f(\boldsymbol{X}_{S_{j_1}S_{j_2}}|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}) \times \prod_{j_1=1}^{b_1} f_{S_{1,j_1}}(\mu_{S_{j_1}})$$

$$\propto \prod_{j_1=1}^{b_1}\prod_{j_2:S_{j_1}} (2\pi\sigma^2_{S_{j_2}})^{-n_{j_1j_2}/2}\exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}}\sum_{i\in S_{j_1}S_{j_2}}(X_i-\mu_{S_{j_1}})^2\right\} \times \prod_{j_1=1}^{b_1}(2\pi\sigma_0^2)^{-1/2}\exp\left\{-\frac{1}{2\sigma_0^2}(\mu_{S_{j_1}}-\mu_0)^2\right\}$$

$$\propto \prod_{j_1=1}^{b_1}\exp\left\{-\frac{1}{2}\left[\sum_{j_2:S_{j_1}}\sum_{i\in S_{j_1}S_{j_2}}\frac{(X_i-\mu_{S_{j_1}})^2}{\sigma^2_{S_{j_2}}}+\frac{(\mu_{S_{j_1}}-\mu_0)^2}{\sigma_0^2}\right]\right\}$$

$$\propto \prod_{j_1=1}^{b_1}\exp\left\{-\frac{1}{2}\left[\mu^2_{S_{j_1}}\left(\sum_{j_2:S_{j_1}}\frac{n_{j_1j_2}}{\sigma^2_{S_{j_2}}}+\frac{1}{\sigma_0^2}\right)-2\mu_{S_{j_1}}\left(\sum_{j_2:S_{j_1}}\frac{n_{j_1j_2}\overline{X}_{S_{j_1}S_{j_2}}}{\sigma^2_{S_{j_2}}}+\frac{\mu_0}{\sigma_0^2}\right)\right]\right\}$$

$$\propto \prod_{j_1=1}^{b_1}\exp\left\{-\frac{\boldsymbol{M_1}}{2}\left[\mu^2_{S_{j_1}}-2\mu_{S_{j_1}}\frac{\boldsymbol{M_2}}{\boldsymbol{M_1}}\right]\right\},$$

where $\boldsymbol{M_1}=\sum_{j_2:S_{j_1}}\frac{n_{j_1j_2}}{\sigma^2_{S_{j_2}}}+\frac{1}{\sigma_0^2}$ and $\boldsymbol{M_2}=\sum_{j_2:S_{j_1}}\frac{n_{j_1j_2}\overline{X}_{S_{j_1}S_{j_2}}}{\sigma^2_{S_{j_2}}}+\frac{\mu_0}{\sigma_0^2}$. Thus, the sample distribution in step iii.1.2 is given by

$$\mu_{S_{j_1}}|\boldsymbol{\sigma},\rho_1,\rho_2,\boldsymbol{X} \sim N\left(\frac{\boldsymbol{M_2}}{\boldsymbol{M_1}},\boldsymbol{M_1}^{-1}\right), \quad \text{for } i\in S_{1,j_1}, \quad j_1=1,\ldots,b_1. \tag{3.23}$$

$$f(\boldsymbol{\sigma}|\boldsymbol{\mu},\rho_1,\rho_2,p_1,p_2,\boldsymbol{X}) \propto f(\boldsymbol{\mu},\boldsymbol{\sigma},\rho_1,\rho_2,p_1,p_2|\boldsymbol{X})$$

$$\propto \prod_{j_2=1}^{b_2}\prod_{j_1:S_{j_2}} f(\boldsymbol{X}_{S_{j_1}S_{j_2}}|\mu_{S_{j_1}},\sigma^2_{S_{j_2}}) \times \prod_{j_2=1}^{b_2} f_{S_{2,j_2}}(\sigma^2_{S_{j_2}})$$

$$\propto \prod_{j_2=1}^{b_2}\prod_{j_1:S_{j_2}} (2\pi\sigma^2_{S_{j_2}})^{-n_{j_1j_2}/2}\exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}}\sum_{i\in S_{j_1}S_{j_2}}(X_i-\mu_{S_{j_1}})^2\right\} \times \prod_{j_2=1}^{b_2}\frac{(a/2)^{d/2}}{\Gamma(d/2)}(\sigma^2_{S_{j_2}})^{-(d+2)/2}\exp\left\{-\frac{a}{2\sigma^2_{S_{j_2}}}\right\}$$

$$\propto \prod_{j_2=1}^{b_2}(\sigma^2_{S_{j_2}})^{-(n_{j_2}+d+2)/2}\exp\left\{-\frac{1}{2\sigma^2_{S_{j_2}}}\left[\sum_{j_1:S_{j_2}}\sum_{i\in S_{j_1}S_{j_2}}(X_i-\mu_{S_{j_1}})^2+a\right]\right\}$$

$$\propto \prod_{j_2=1}^{b_2}(\sigma^2_{S_{j_2}})^{-(\boldsymbol{D}+2)/2}\exp\left\{-\frac{\boldsymbol{A}}{2\sigma^2_{S_{j_2}}}\right\},$$

with $\boldsymbol{A}$ and $\boldsymbol{D}$ as defined for Eq. (3.22). Thus, the sample distribution in step iii.2.2 is given by

$$\sigma^2_{S_{j_2}} | \boldsymbol{\mu}, \rho_1, \rho_2, \boldsymbol{X} \sim IG\left(\boldsymbol{A}/2, \boldsymbol{D}/2\right), \quad \text{for } i \in S_{2,j_2}, \quad j_2 = 1, \ldots, b_2. \tag{3.24}$$

$$f(p_1|\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho_1, \rho_2, p_2, \boldsymbol{X}) \propto f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho_1, \rho_2, p_1, p_2|\boldsymbol{X}) \propto p_1^{\alpha_1+b_1-2}(1-p_1)^{n+\beta_1-b_1-1}$$

and

$$f(p_2|\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho_1, \rho_2, p_1, \boldsymbol{X}) \propto f(\boldsymbol{\mu}, \boldsymbol{\sigma}, \rho_1, \rho_2, p_1, p_2|\boldsymbol{X}) \propto p_2^{\alpha_2+b_2-2}(1-p_2)^{n+\beta_2-b_2-1}.$$

Thus, the sample distributions in steps 0.1 and 0.2 are given by

$$p_1|\rho_1, \boldsymbol{X} \sim Beta(\alpha_1 + b_1 - 1, n + \beta_1 - b_1) \tag{3.25}$$

and

$$p_2|\rho_2, \boldsymbol{X} \sim Beta(\alpha_2 + b_2 - 1, n + \beta_2 - b_2). \tag{3.26}$$

The performance of the BMCP model introduced in this section will be evaluated in Chapter 4 through simulation studies.

4 Monte Carlo simulation study

In this chapter, we run a Monte Carlo simulation study to evaluate the performance of the BMCP model to identify multiple change points in Normal means and variances. We compare it with the L99 and BH93 models. The data are generated from Normal distributions according to the simulation schemes summarized in Table 3, that compiles the partitions, parameters and hyperparameters of each simulated scene. We concentrate in 7 scenes, with 100 or 200 observations, multiple changes in mean and/or variance, and the total number of changes varying from 1 to 9. For each scene we generate 100 data sets. For each data set, the posterior distribution is obtained through a MCMC scheme with 14,000 iterations, with the first 4,000 discarded as a burn-in period. We consider the BH93 model implementation available in the R package **bcp**. The L99 and BMCP models were implemented in C++ and connected to R through the **Rcpp** package (Eddelbuettel and François [2011]).

To analyze the data, we consider the prior distributions mentioned in previous chapters. The hyperparameters of the block prior distributions of the BMCP model, specified in 3.16 and 3.17, are fixed as $(\mu_0, \sigma_0^2) = (0, 100)$ and $(a, d) = (2, 2)$, which are flat proposals for both the block mean and the block variance. The hyperparameters of the joint block prior distribution of the L99 model and the hyperparameters of the BH93 model are fixed as described in Section 2.6. The hyperparameters of the prior distributions of the random partitions, $\alpha_1, \beta_1, \alpha_2$ and β_2 for the BMCP model, and α and β for the L99 model, are specified based on the model properties discussed in Section 2.5. For the scenes 1 to 6, they are stated based on the true number of changes in each partition, following what is shown in Table 1. For the BMCP model, this procedure may be considered to each partition separately, due to the independence assumption between the partitions determined in Section 3.2. For the scenes where there is no change in mean or variance, that is, $N = 0$, we considered the $Beta$ hyperparameters of the case when $N = 1$ because there is no possible configuration of these hyperparameters that implies $E(N) = 0$. For the scene 7, we assume the Bayes-Laplace prior $Beta(1, 1)$ for parameters p_1, p_2 and p of the BMCP and L99 models, as mentioned in Section 2.5.

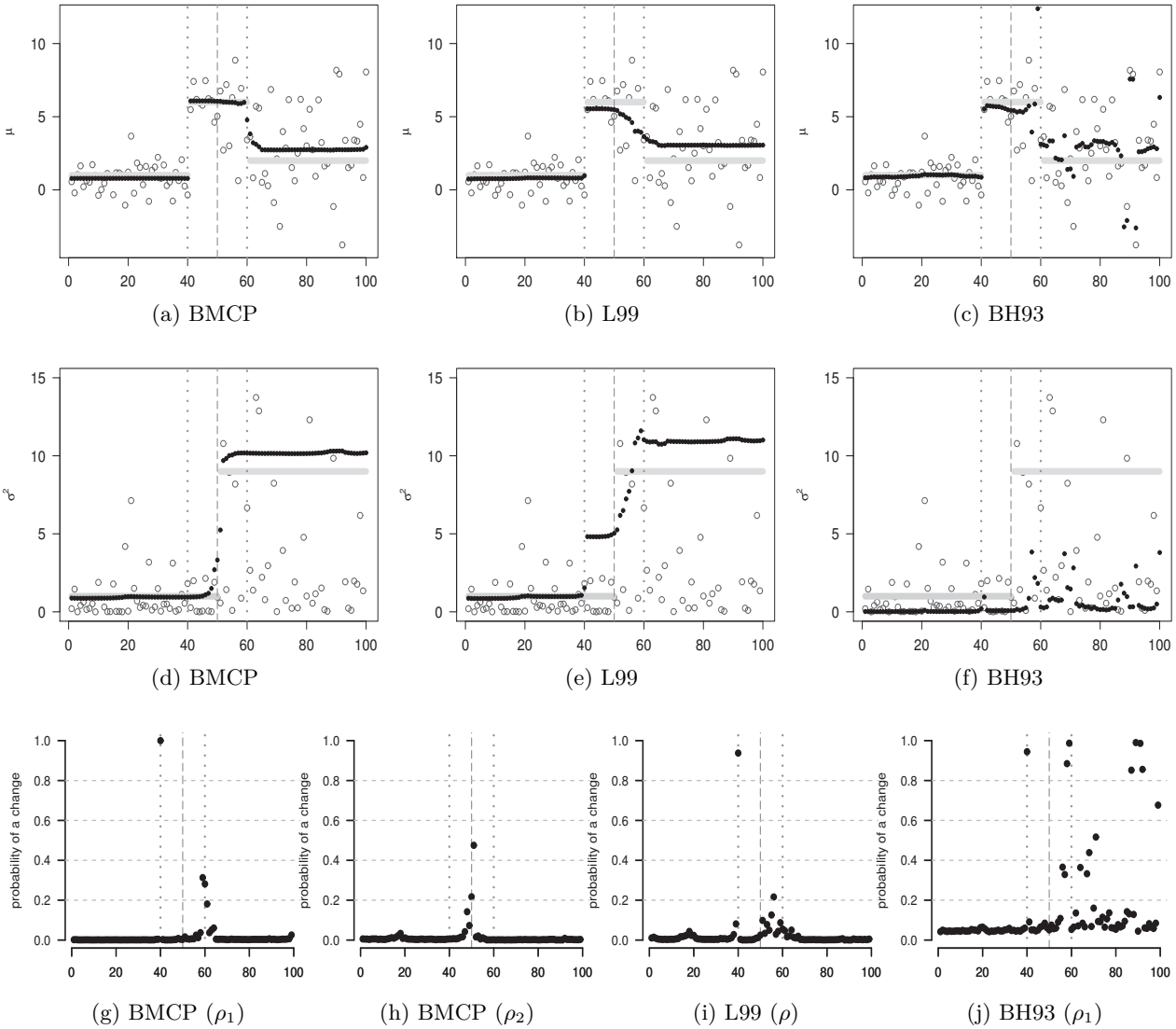
The representation of the partitions described in this chapter ignore the first element, previously defined as 0. We represent the partition of the L99 model as ρ. Note that it is equivalent to the union of the elements of the partitions ρ_1 and ρ_2 of the BMCP model. For example, in the scene of the Section 2.6 we have that $\rho = \{40, 50, 60, 100\}$. The partition of the BH93 model is represented by ρ_1 because this model objective is to identify only the changes in the mean. The 90% precision intervals in the graphics that show the posterior mean and variance estimates are the quantiles of 5% and 95% of the Monte Carlo replications. In the graphics of the posterior mean and variance estimates and in the graphics that show the estimated probability of a change, the horizontal axis display the observations indexes $i = 1, \ldots, n$, and the vertical gray lines indicate the true end points, relative to changes in the mean (dotted line), variance (dashed line) and both the mean and the variance (solid line). In the graphics of the prior and posterior distributions of the number of changes N, the vertical solid gray line indicates the true N in the respective partition.

Scene	n	**True parameters** μ **and** σ	**True partitions** ρ_1, ρ_2 **and** ρ	**Prior distributions of** p_1, p_2 **and** p
1	100	$\mu_i = \begin{cases} 0, & i = 1, \dots, 60 \\ 2, & i = 61, \dots, 100 \end{cases}$ $\sigma_i^2 = 1, \ i = 1, \dots, 100$	$\rho_1 = \{60, 100\}$ $\rho_2 = \{100\}$ $\rho = \{60, 100\}$	$p_1 \sim Beta(50, 4900)$ $p_2 \sim Beta(50, 4900)$ $p \sim Beta(50, 4900)$
2	100	$\mu_i = \begin{cases} 0, & i = 1, \dots, 10 \\ 3, & i = 11, \dots, 20 \\ 0, & i = 21, \dots, 30 \\ 3, & i = 31, \dots, 40 \\ 0, & i = 41, \dots, 50 \\ 3, & i = 51, \dots, 60 \\ 0, & i = 61, \dots, 70 \\ 3, & i = 71, \dots, 80 \\ 0, & i = 81, \dots, 90 \\ 3, & i = 91, \dots, 100 \end{cases}$ $\sigma_i^2 = 1, \ i = 1, \dots, 100$	$\rho_1 = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$ $\rho_2 = \{100\}$ $\rho = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$	$p_1 \sim Beta(50, 500)$ $p_2 \sim Beta(50, 4900)$ $p \sim Beta(50, 500)$
3	200	$\mu_i = 1, \ i = 1, \dots, 200$ $\sigma_i^2 = \begin{cases} 0, & i = 1, \dots, 50 \\ 4, & i = 51, \dots, 100 \\ 0, & i = 101, \dots, 150 \\ 9, & i = 151, \dots, 200 \end{cases}$	$\rho_1 = \{200\}$ $\rho_2 = \{50, 100, 150, 200\}$ $\rho = \{50, 100, 150, 200\}$	$p_1 \sim Beta(50, 9900)$ $p_2 \sim Beta(50, 3267)$ $p \sim Beta(50, 3267)$
4	200	$\mu_i = \begin{cases} 0, & i = 1, \dots, 60 \\ 2, & i = 61, \dots, 140 \\ 0, & i = 141, \dots, 200 \end{cases}$ $\sigma_i^2 = \begin{cases} 1, & i = 1, \dots, 60 \\ 4, & i = 61, \dots, 140 \\ 1, & i = 141, \dots, 200 \end{cases}$	$\rho_1 = \{60, 140, 200\}$ $\rho_2 = \{60, 140, 200\}$ $\rho = \{60, 140, 200\}$	$p_1 \sim Beta(50, 4925)$ $p_2 \sim Beta(50, 4925)$ $p \sim Beta(50, 4925)$
5	200	$\mu_i = \begin{cases} 0, & i = 1, \dots, 40 \\ 3, & i = 41, \dots, 120 \\ 0, & i = 121, \dots, 200 \end{cases}$ $\sigma_i^2 = \begin{cases} 1, & i = 1, \dots, 80 \\ 4, & i = 81, \dots, 160 \\ 2, & i = 161, \dots, 200 \end{cases}$	$\rho_1 = \{40, 120, 200\}$ $\rho_2 = \{80, 160, 200\}$ $\rho = \{40, 80, 120, 160, 200\}$	$p_1 \sim Beta(50, 4925)$ $p_2 \sim Beta(50, 4925)$ $p \sim Beta(50, 2438)$
6	200	$\mu_i = \begin{cases} 0, & i = 1, \dots, 40 \\ 2, & i = 41, \dots, 80 \\ 4, & i = 81, \dots, 120 \\ 2, & i = 121, \dots, 160 \\ 0, & i = 161, \dots, 200 \end{cases}$ $\sigma_i^2 = \begin{cases} 1, & i = 1, \dots, 100 \\ 2, & i = 100, \dots, 200 \end{cases}$	$\rho_1 = \{40, 80, 120, 160, 200\}$ $\rho_2 = \{100, 200\}$ $\rho = \{40, 80, 100, 120, 160, 200\}$	$p_1 \sim Beta(50, 2438)$ $p_2 \sim Beta(50, 9900)$ $p \sim Beta(50, 1940)$
7	200	$\mu_i = \begin{cases} 0.25, & i = 1, \dots, 50 \\ 1.5, & i = 51, \dots, 100 \\ 0.1, & i = 101, \dots, 150 \\ 1, & i = 151, \dots, 200 \end{cases}$ $\sigma_i^2 = \begin{cases} 0.3, & i = 1, \dots, 65 \\ 1.2, & i = 66, \dots, 115 \\ 0.15, & i = 116, \dots, 165 \\ 1, & i = 166, \dots, 200 \end{cases}$	$\rho_1 = \{50, 100, 150, 200\}$ $\rho_2 = \{65, 115, 165, 200\}$ $\rho = \{50, 65, 100, 115, 150, 165, 200\}$	$p_1 \sim Beta(1, 1)$ $p_2 \sim Beta(1, 1)$ $p \sim Beta(1, 1)$

Table 3: Partitions, parameters and hyperparameters considered in the simulation scenes.

4.1 BMCP applied to data in Section 2.6

Before analyzing the Monte Carlo simulation results, we present in Figure 3 the results of the application of the BMCP model to the sample considered in Section 2.6. Following the notation described in Section 3.5, this scene may be summarized by $\rho_1 = \{40, 60, 100\}$, $(\mu_{S_1}, \mu_{S_2}, \mu_{S_3}) = (1, 6, 2)$, $\rho_2 = \{50, 100\}$ and $(\sigma^2_{S_1}, \sigma^2_{S_2}) = (1, 9)$. We can see in Figures 3a-3f, that the BMCP model provides improved estimates for the mean and variance when compared to the other models. Also, as shown in Figures 3g-3j, it gives distinctly larger probabilities of a change to the true change points in mean and variance, correctly identifying the two partitions, which is not true to the L99 and BH93 models.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

Figure 3: BMCP, L99 and BH93 model estimates to simulated Normal data: (a-c) Simulated data (gray circles), posterior mean estimates (black dots) and true mean values (gray solid horizontal lines). (d-f) The squared error (gray circles), posterior variance estimates (black dots) and the true variance values (gray solid horizontal lines). (g-j) Estimated posterior probability of each instant to be a change point. In Figures (a-j), the vertical gray lines indicate the true change points in the mean (dotted line) and variance (dashed line).

4.2 Scenes 1 and 2: mean changes with constant variance

In this section, we evaluate the performance of the BMCP, L99 and BH93 models in two constant variance scenes.

The results for the scene 1, with constant variance and one mean change, are summarized in Figure 4 and Tables 4 and 5. The three models provide appropriate mean estimates, with similar precision. The precision interval in the BH93 model are not so regular as the other models. It may be a consequence of the different prior specification for parameter p in this model, that is not so informative as the considered for the L99 and BMCP models. The underestimated variance in Figure 4f and the less regular estimates for the mean provided by the BH93 model are common to all scenes analysed in this chapter. The BMCP and L99 models provide suitable posterior variance estimates. As can be seen in Figures 4g-4j, all the three models correctly identify the true mean change point, estimating high probabilities to the true change point to be a change point. The BMCP also provides the information that there is no change in the variance. The results in Table 4 support that all the three models correctly identify the true partitions, such that the estimated posterior mode are the true partitions. Based on Figures 4k-4n and Table 5, we can conclude that the BMCP and L99 models provide posterior distributions for the number of changes concentrated in the true values, with both the estimated posterior expectation and posterior mode providing adequate estimates for the true N values of each partition. In the case of the BH93 model, only the mode provides a correct estimative for the number of mean changes.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{60, 100\}$	1	0.469828
$\rho_1 = \{59, 100\}$	1	0.156611
$\rho_1 = \{61, 100\}$	1	0.132427
$\rho_1 = \{58, 100\}$	1	0.046720
$\rho_1 = \{62, 100\}$	1	0.035384

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{100\}$	0	0.645319
$\rho_2 = \{99, 100\}$	1	0.005375
$\rho_2 = \{98, 100\}$	1	0.005245
$\rho_2 = \{93, 100\}$	1	0.004245
$\rho_2 = \{97, 100\}$	1	0.003793

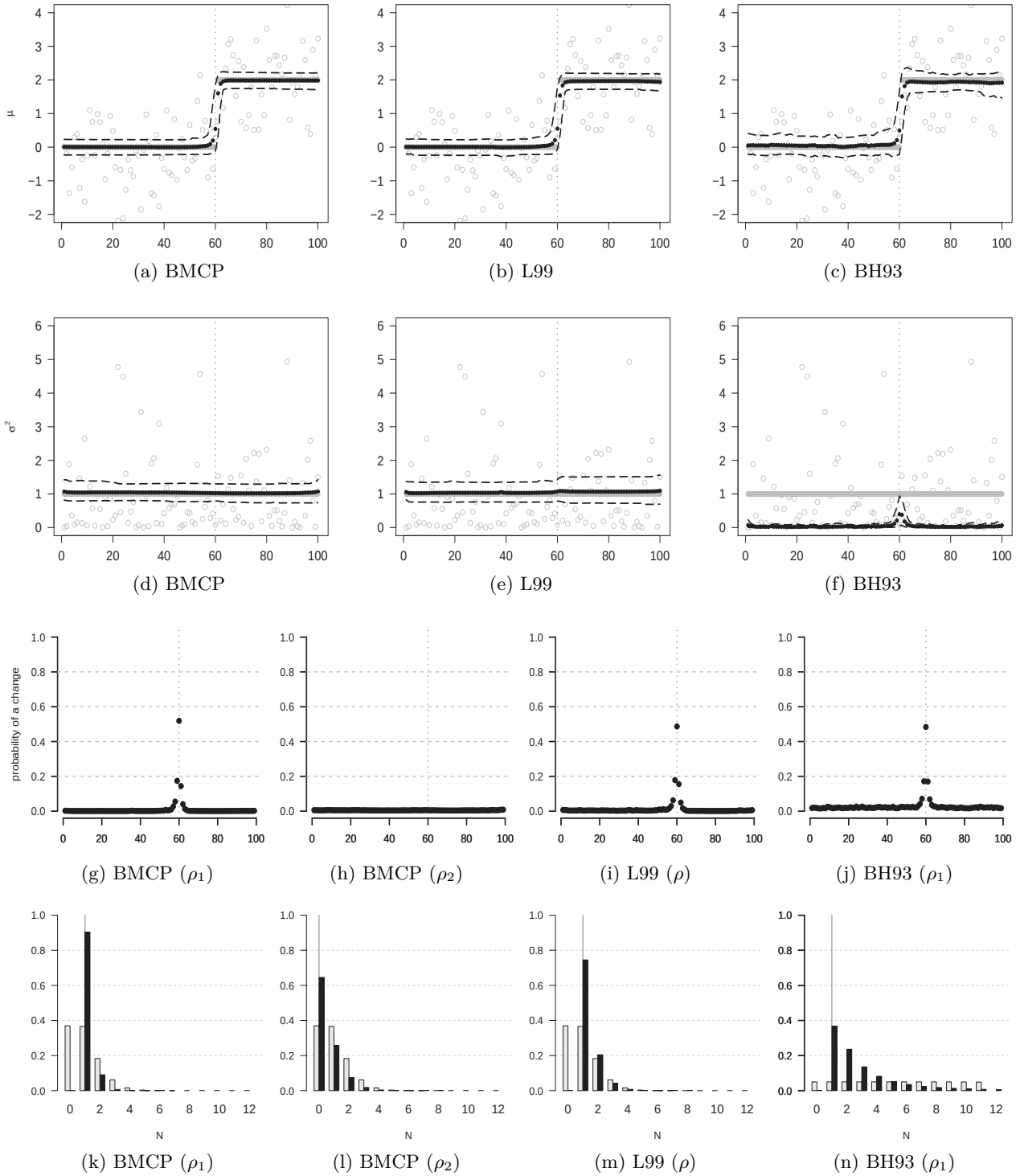
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{60, 100\}$	1	0.369748
$\rho = \{59, 100\}$	1	0.130202
$\rho = \{61, 100\}$	1	0.117891
$\rho = \{58, 100\}$	1	0.040332
$\rho = \{62, 100\}$	1	0.035497

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{60, 100\}$	1	0.183917
$\rho_1 = \{61, 100\}$	1	0.064499
$\rho_1 = \{59, 100\}$	1	0.063816
$\rho_1 = \{58, 100\}$	1	0.018677
$\rho_1 = \{62, 100\}$	1	0.015206

Table 4: Top five most likely partitions based on the average estimated posterior probability for the scene 1.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	1	1.1	0.1	1	0.9
BMCP	ρ_2	0	0.5	0.6	0	0.65
L99	ρ	1	1.3	0.4	1	0.75
BH93	ρ_1	1	2.9	7.2	1	0.37

Table 5: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 1.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 4: Graphical summary for the MC simulation study of the scene 1. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

The results for the scene 2, with constant variance and nine changes in the mean, are summarized in Figure 5 and Tables 6 and 7. The three models provide appropriate mean estimates, with similar precision, but the BMCP and the BH93 models provide more accurate estimates. Only the BMCP model provides correct variance estimates, because the posterior variance estimates of the L99 model are affected by the mean changes (Figures 5d and 5e). As can be seen in Figures 5g-5j, all the three models correctly identify the true changes, with the additional information of the BMCP model that the changes ocurred only in the mean. The results in Table 6 show that the average posterior distributions of the partitions give larger probabilities to the true partitions, but in this case we have a smooth probability function, and the probability of a change available in Figures 5g-5j is a better evidence in favor of the true partitions. Based on Figures 5k-5n and Table 7, all the models overestimate the true number of mean changes, with the BMCP model providing the better estimates. The absence of changes in variance is correctly identified by the BMCP model in all the results related to scene 2.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.03256
$\rho_1 = \{10, 20, 31, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00661
$\rho_1 = \{10, 19, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00594
$\rho_1 = \{10, 20, 29, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00462
$\rho_1 = \{10, 20, 30, 40, 50, 60, 71, 80, 90, 100\}$	9	0.00447

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{100\}$	0	0.64037
$\rho_2 = \{99, 100\}$	1	0.00470
$\rho_2 = \{98, 100\}$	1	0.00386
$\rho_2 = \{1, 100\}$	1	0.00354
$\rho_2 = \{97, 100\}$	1	0.00327

L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00176
$\rho = \{10, 20, 31, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00031
$\rho = \{10, 20, 30, 40, 48, 50, 60, 70, 80, 90, 100\}$	10	0.00030
$\rho = \{10, 20, 29, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00029
$\rho = \{10, 19, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00019

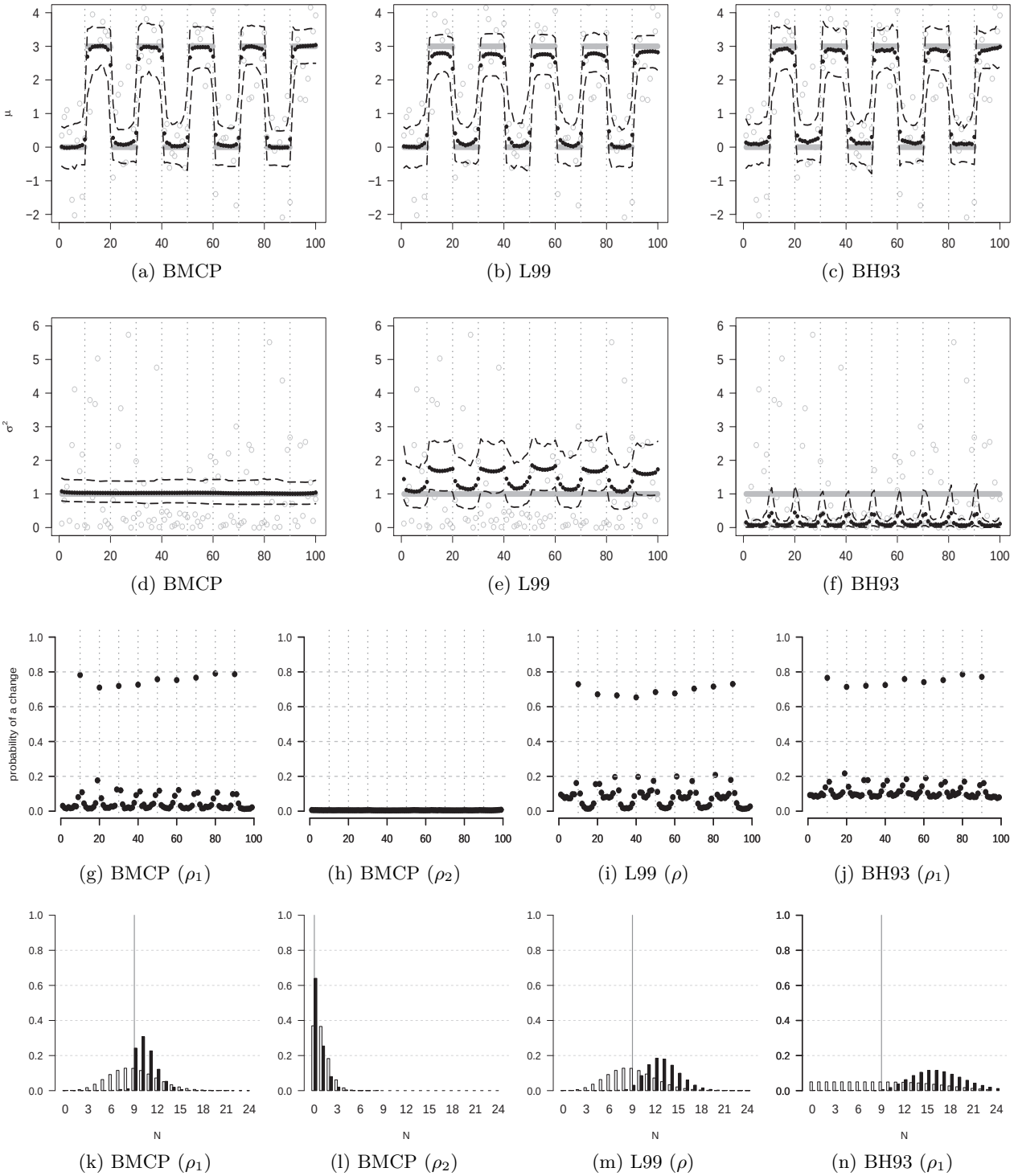
BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00086
$\rho = \{10, 20, 31, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00025
$\rho = \{10, 19, 30, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00009
$\rho = \{10, 20, 30, 40, 50, 60, 70, 81, 90, 100\}$	9	0.00008
$\rho = \{10, 20, 29, 40, 50, 60, 70, 80, 90, 100\}$	9	0.00007

Table 6: Top five most likely partitions based on the average estimated posterior probability for the scene 2.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	9	10.4	2	10	0.31
BMCP	ρ_2	0	0.5	0.6	0	0.64
L99	ρ	9	12.9	4.8	12	0.18
BH93	ρ_1	9	16.4	11.5	16	0.11

Table 7: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 2.

The two scenes with changes in mean and constant variance show the ability of the BMCP model to correctly identify the changes in this situation. Unlike the L99 model, the BMCP model point out that the changes occurred only in the mean and the variance estimates are not affected by the mean changes.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 5: Graphical summary for the MC simulation study of the scene 2. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

4.3 Scene 3: variance changes with constant mean

We evaluate now the performance of the models in a constant mean scene. The results are summarized in Figure 6 and Tables 8 and 9. The three models provide appropriate mean estimates, with the BMCP presenting the highest precision (Figures 6a-6c). The BMCP and L99 models provide suitable posterior variance estimates, with similar precisions (Figures 6d-6f). The results in Table 8 show that the most likely partitions estimated by the L99 model consider only one (the most severe) change of the three changes. The average posterior distributions of the partitions give larger probabilities to the true partitions only for the BMCP and BH93 models, but we have a smooth estimated posterior probability distribution for ρ_2, and the estimated probability of a change available in Figures 6g-6j is thus a better evidence in favor of the true partitions. The distinctly higher probability of a change presented in Figures 6h and 6i for the true change points indicates that both the BMCP and the L99 models correctly recognize the changes in the behavior of the sequence of observations. In the case of the BMCP model, Figures 6g and 6h correctly indicate that this changes occur only in the variance. Based on Figures 6k-6n and Table 9, we can conclude that the BMCP model overestimates the number of mean changes, but correctly indicates the absence of changes in the variance. The L99 model provides posterior distributions for the number of changes concentrated in the true value $N = 3$, with the estimated posterior expectation and posterior mode providing adequate estimates for the number of changes in ρ. The BH93 model provides an incorrect estimation for the number and location of the mean changes, which is a consequence that its assumption of constant variance is strongly broken in this scene.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{200\}$	0	0.888989
$\rho_1 = \{198, 200\}$	1	0.003847
$\rho_1 = \{199, 200\}$	1	0.003257
$\rho_1 = \{195, 200\}$	1	0.002525
$\rho_1 = \{193, 200\}$	1	0.001660

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{50, 100, 150, 200\}$	3	0.001550
$\rho_2 = \{150, 200\}$	1	0.001408
$\rho_2 = \{51, 100, 150, 200\}$	3	0.001130
$\rho_2 = \{151, 200\}$	1	0.001066
$\rho_2 = \{51, 100, 151, 200\}$	3	0.000973

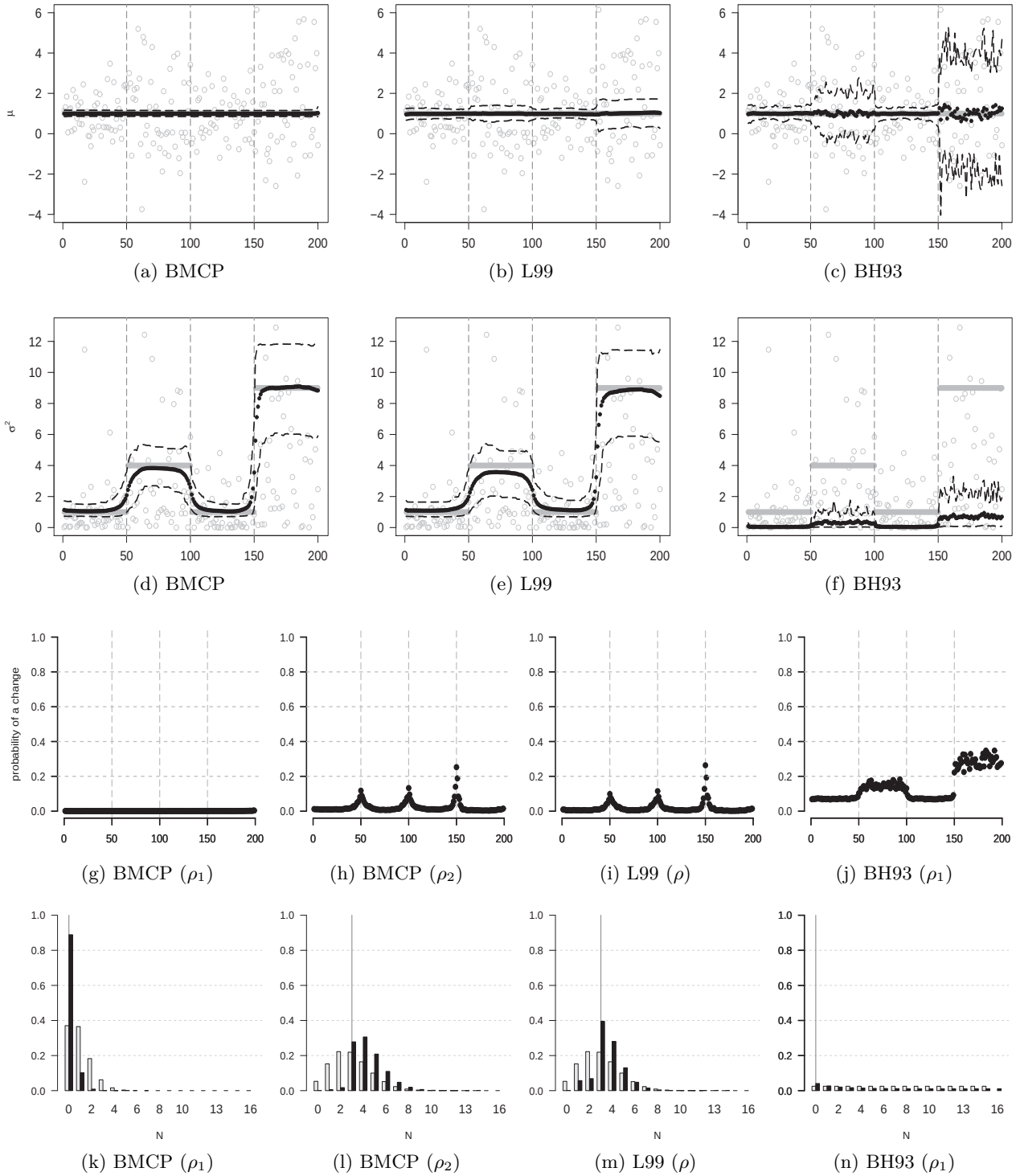
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{150, 200\}$	1	0.018335
$\rho = \{151, 200\}$	1	0.012897
$\rho = \{149, 200\}$	1	0.007634
$\rho = \{152, 200\}$	1	0.004169
$\rho = \{148, 200\}$	1	0.004023

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{200\}$	0	0.040672
$\rho_1 = \{198, 200\}$	1	0.001044
$\rho_1 = \{181, 200\}$	1	0.000663
$\rho_1 = \{184, 186, 200\}$	2	0.000635
$\rho_1 = \{183, 200\}$	1	0.000634

Table 8: Top five most likely partitions based on the average estimated posterior probability for the scene 3.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	0	0.1	0.1	0	0.89
BMCP	ρ_2	3	4.4	1.9	4	0.31
L99	ρ	3	3.6	1.6	3	0.4
BH93	ρ_1	0	28.6	234.7	0	0.04

Table 9: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 3.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 6: Graphical summary for the MC simulation study of the scene 3. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

4.4 Scene 4: mean and variance changes at the same time

In this section, we evaluate the performance of the models in a scene where two changes occur in both the mean and variance at the same time. The Monte Carlo simulation results are summarized in Figure 7 and Tables 10 and 11. The three models provide good mean estimates, with the L99 model providing the most accurate estimates (Figures 7a-7c). The same is true for the variance estimates (Figures 7d-7f), except for the BH93 model that underestimates the variance, as is the case in all the scenes we analyse in this chapter. The better performance of the L99 model in the mean and variance estimation in this scene is expected because, as described in Section 2.4, this model is formulated with the assumption that both the mean and the variance parameters change at the same time, which is the case in this scene. Comparing Figures 7g-7j, the L99 model provides the largest average probabilities of a change for the true changes. The BMCP model provides distinctly larger probabilities of change for the true changes in mean and variance, with these probabilities being smaller for the changes in the variance. Based on Figure 7j, we can say that the BH93 model correctly identifies the changes in mean. Figures 7k-7m and Table 11 show that the L99 and the BMCP models provide posterior distributions for the number of changes highly concentrated in the true value $N = 2$. The results in Table 10 show that the average posterior distributions of all the partitions have their mode equal to the true partition

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{60, 140, 200\}$	2	0.050056
$\rho_1 = \{61, 139, 200\}$	2	0.028129
$\rho_1 = \{61, 140, 200\}$	2	0.026343
$\rho_1 = \{60, 139, 200\}$	2	0.023128
$\rho_1 = \{62, 139, 200\}$	2	0.020103

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{60, 140, 200\}$	2	0.006304
$\rho_2 = \{59, 140, 200\}$	2	0.005154
$\rho_2 = \{60, 139, 200\}$	2	0.004889
$\rho_2 = \{60, 141, 200\}$	2	0.004874
$\rho_2 = \{61, 140, 200\}$	2	0.004494

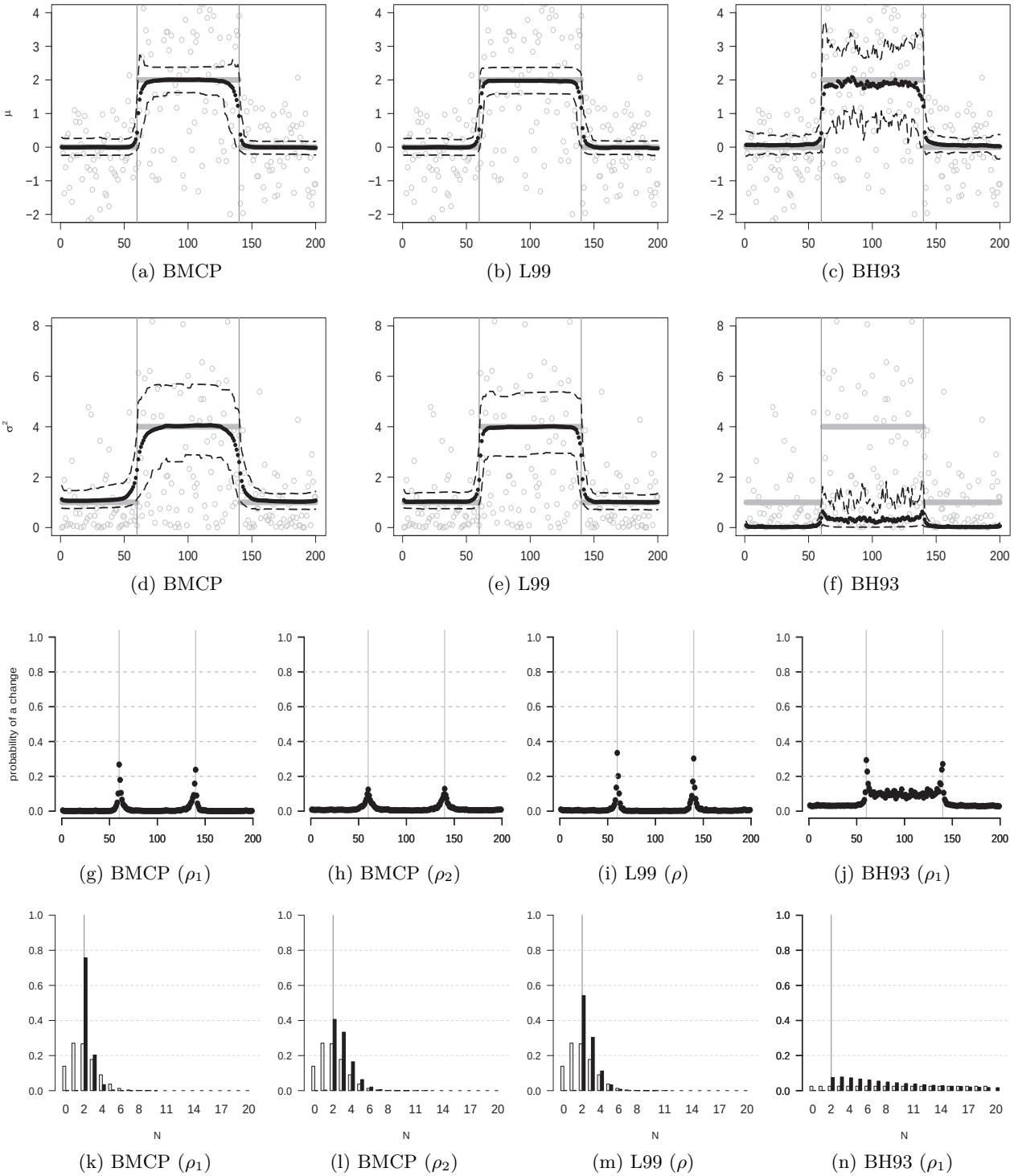
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{60, 140, 200\}$	2	0.057594
$\rho = \{61, 140, 200\}$	2	0.028148
$\rho = \{61, 139, 200\}$	2	0.027855
$\rho = \{60, 139, 200\}$	2	0.026327
$\rho = \{60, 141, 200\}$	2	0.024905

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{60, 140, 200\}$	2	0.004427
$\rho_1 = \{61, 140, 200\}$	2	0.002886
$\rho_1 = \{61, 139, 200\}$	2	0.002721
$\rho_1 = \{60, 139, 200\}$	2	0.002465
$\rho_1 = \{60, 141, 200\}$	2	0.002370

Table 10: Top five most likely partitions based on the average estimated posterior probability for the scene 4.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	2	2.3	0.3	2	0.76
BMCP	ρ_2	2	3	1.2	2	0.41
L99	ρ	2	2.7	0.8	2	0.54
BH93	ρ_1	2	12.8	102.2	3	0.08

Table 11: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 4.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 7: Graphical summary for the MC simulation study of the scene 4. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

4.5 Scenes 5, 6 and 7: mean and variance changes at different times

In this section, we evaluate the performance of the BMCP, L99 and BH93 models in simulated scenes where changes occur in the mean and in the variance at different times.

The scene 5, with two changes in mean and two changes in variance, is summarized in Figure 8 and Tables 12 and 13. All the models provide appropriate mean estimates, with the BMCP model providing the most accurate estimates (Figures 8a-8c). The BMCP model also provides the better estimates for the variance (Figures 8d-8f). As shown in Figure 8e, the variance estimates of the L99 model are clearly affected by the changes in the mean, as it is also the case in the constant variance scene 1, in Section 4.2, which does not occur in the BMCP variance estimates (Figure 8d). As can be seen in Figures 8h and 8i, both the BMCP and the L99 models do not identify the last change in variance. Figures 8g-8i are a good illustration of the most relevant output of the BMCP model that is not provided by the L99 model: both the BMCP and the L99 models provide closer values to the probabilities for the true changes in the mean and in the variance, but the BMCP model, unlike the L99 model, provide this information separately for the mean and the variance, such that we know in which parameter the change occurred. Figures 8k-8n show that the L99 and the BMCP models provide posterior distributions for the number of changes highly concentrated in the true values $N = 2$ (two changes in mean and two changes in variance) and $N = 4$ (four total changes) for the L99 model.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{40, 120, 200\}$	2	0.209763
$\rho_1 = \{40, 121, 200\}$	2	0.089323
$\rho_1 = \{40, 119, 200\}$	2	0.084880
$\rho_1 = \{40, 118, 200\}$	2	0.047314
$\rho_1 = \{40, 122, 200\}$	2	0.033951

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{80, 200\}$	1	0.026173
$\rho_2 = \{81, 200\}$	1	0.024465
$\rho_2 = \{79, 200\}$	1	0.018039
$\rho_2 = \{82, 200\}$	1	0.014761
$\rho_2 = \{78, 200\}$	1	0.010790

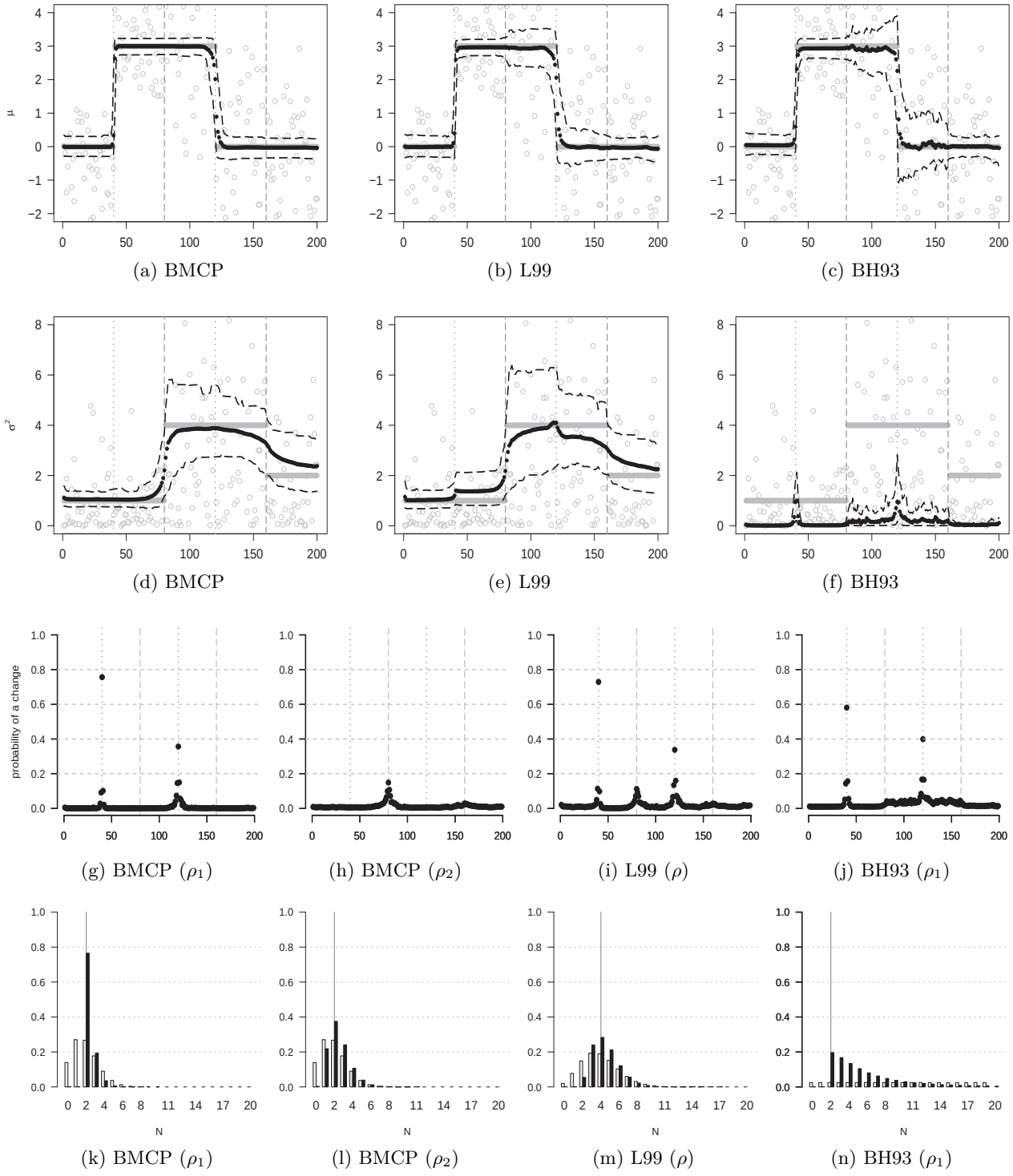
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{40, 120, 200\}$	2	0.012472
$\rho = \{40, 80, 120, 200\}$	3	0.006363
$\rho = \{40, 119, 200\}$	2	0.005487
$\rho = \{40, 81, 120, 200\}$	3	0.005333
$\rho = \{40, 121, 200\}$	2	0.004157

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{40, 120, 200\}$	2	0.045443
$\rho_1 = \{40, 119, 200\}$	2	0.018004
$\rho_1 = \{40, 121, 200\}$	2	0.016670
$\rho_1 = \{40, 118, 200\}$	2	0.010768
$\rho_1 = \{39, 120, 200\}$	2	0.010651

Table 12: Top five most likely partitions based on the average estimated posterior probability for the scene 5.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	2	2.3	0.3	2	0.77
BMCP	ρ_2	2	2.4	1.4	2	0.37
L99	ρ	4	4.4	2.2	4	0.28
BH93	ρ_1	2	6.1	24.9	2	0.2

Table 13: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 5.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 8: Graphical summary for the MC simulation study of the scene 5. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

The same conclusions made about the scene 5 can be extended to the scene 6, displayed next in Figure 9 and Tables 14 and 15, with four changes in mean and one change in variance. Moreover, the BMCP has a better performance in the identification of the variance change than the L99 model, as can be seen in Figures 9h and 9i and in the average posterior probabilities $p(\rho_2|\boldsymbol{X})$ and $p(\rho|\boldsymbol{X})$ in Table 14. Thus, this scene is an example where the BMCP model not only identifies the changes and parameters that experienced these changes, but also identifies a variance change that is not identified by the L99 model, indicating that the multipartition model we propose in this thesis is a powerfull strategy to change point detection in situations where different parameters change at different times.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{40, 80, 134, 200\}$	3	0.004045
$\rho_1 = \{40, 80, 120, 160, 200\}$	4	0.003690
$\rho_1 = \{40, 80, 121, 160, 200\}$	4	0.003576
$\rho_1 = \{40, 80, 120, 200\}$	3	0.003243
$\rho_1 = \{39, 80, 121, 200\}$	3	0.002842

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{100, 200\}$	1	0.111871
$\rho_2 = \{101, 200\}$	1	0.072110
$\rho_2 = \{99, 200\}$	1	0.069860
$\rho_2 = \{102, 200\}$	1	0.050833
$\rho_2 = \{98, 200\}$	1	0.045440

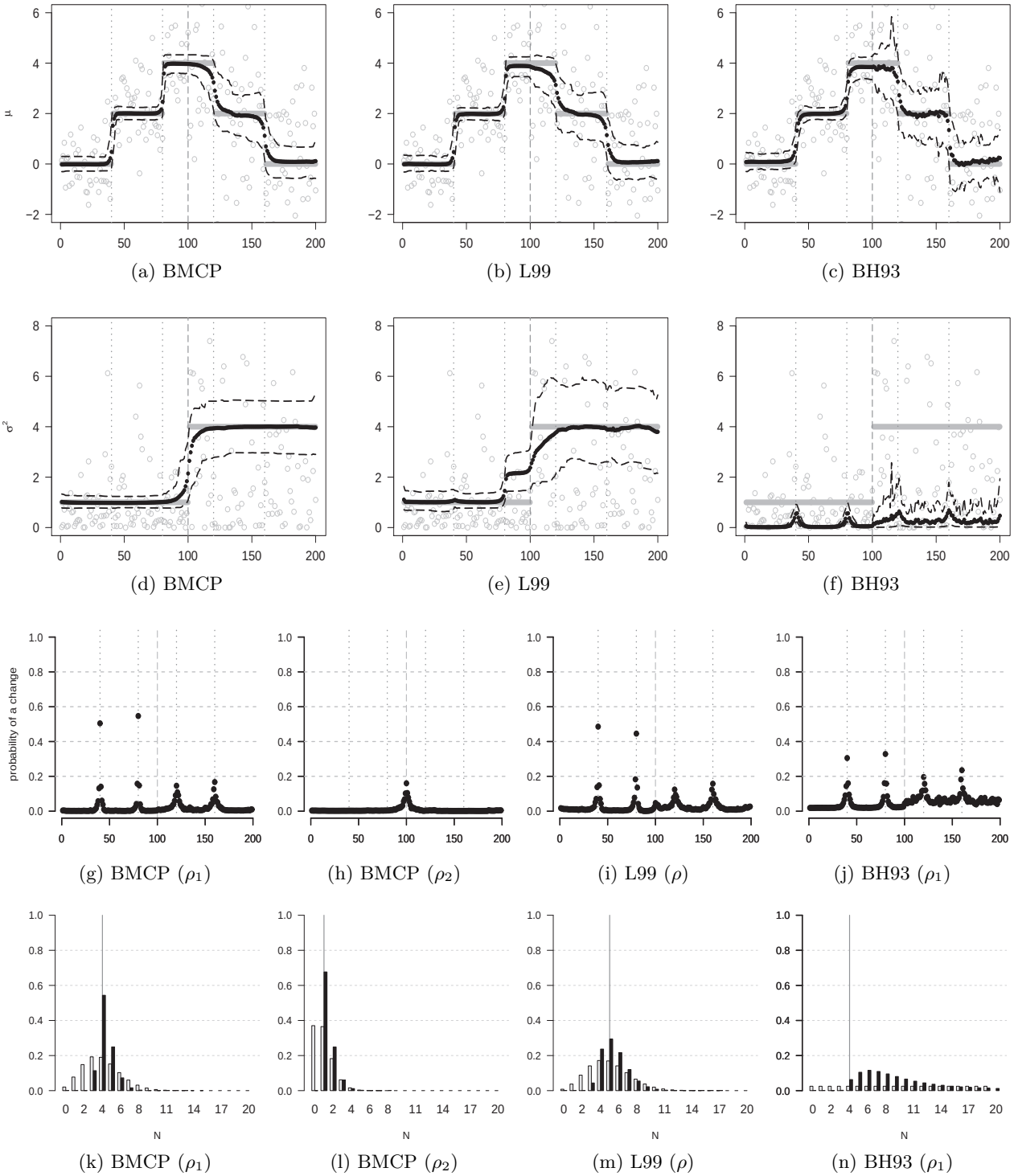
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{40, 81, 121, 200\}$	3	0.001413
$\rho = \{40, 80, 121, 200\}$	3	0.001376
$\rho = \{39, 80, 121, 200\}$	3	0.001099
$\rho = \{40, 80, 120, 200\}$	3	0.000856
$\rho = \{40, 79, 120, 200\}$	3	0.000595

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{40, 80, 122, 160, 200\}$	4	0.000163
$\rho_1 = \{40, 81, 120, 158, 200\}$	4	0.000143
$\rho_1 = \{40, 80, 121, 160, 200\}$	4	0.000141
$\rho_1 = \{40, 80, 121, 167, 200\}$	4	0.000135
$\rho_1 = \{40, 79, 121, 167, 200\}$	4	0.000118

Table 14: Top five most likely partitions based on the average estimated posterior probability for the scene 6.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	4	4.3	0.8	4	0.54
BMCP	ρ_2	1	1.4	0.5	1	0.68
L99	ρ	5	5.4	2.1	5	0.29
BH93	ρ_1	4	10.7	44.5	6	0.12

Table 15: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 6.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 9: Graphical summary for the MC simulation study of the scene 6. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

The scene 7, with three changes in mean and three changes in variance, considers the non-informative $Beta(1,1)$ prior for parameters p_1 and p_2. This scene was inspired in a simulated scene proposed by Peluso et al. [2019] to evaluate their multiparametric change point model in a situation where a variance change occur a few instants after each mean change. Our results are displayed next in Figure 10 and Tables 16 and 17. Based on Table 16 and Figures 10g-10i, the BMCP model correctly identifies the true change points in mean and variance, while the L99 model does not identify two changes in the variance. We can see in Table 17 that the BMCP model provides correct estimates for the number of changes in the mean and in the variance through the estimated posterior mode, while the L99 model correctly estimates the total number of changes through the estimated posterior expectation.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{50, 100, 150, 200\}$	3	0.034641
$\rho_1 = \{50, 99, 150, 200\}$	3	0.020887
$\rho_1 = \{50, 101, 150, 200\}$	3	0.015268
$\rho_1 = \{50, 98, 150, 200\}$	3	0.012232
$\rho_1 = \{51, 100, 150, 200\}$	3	0.010630

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{65, 115, 165, 200\}$	3	0.000981
$\rho_2 = \{65, 114, 165, 200\}$	3	0.000878
$\rho_2 = \{65, 115, 166, 200\}$	3	0.000877
$\rho_2 = \{64, 115, 165, 200\}$	3	0.000755
$\rho_2 = \{64, 114, 165, 200\}$	3	0.000754

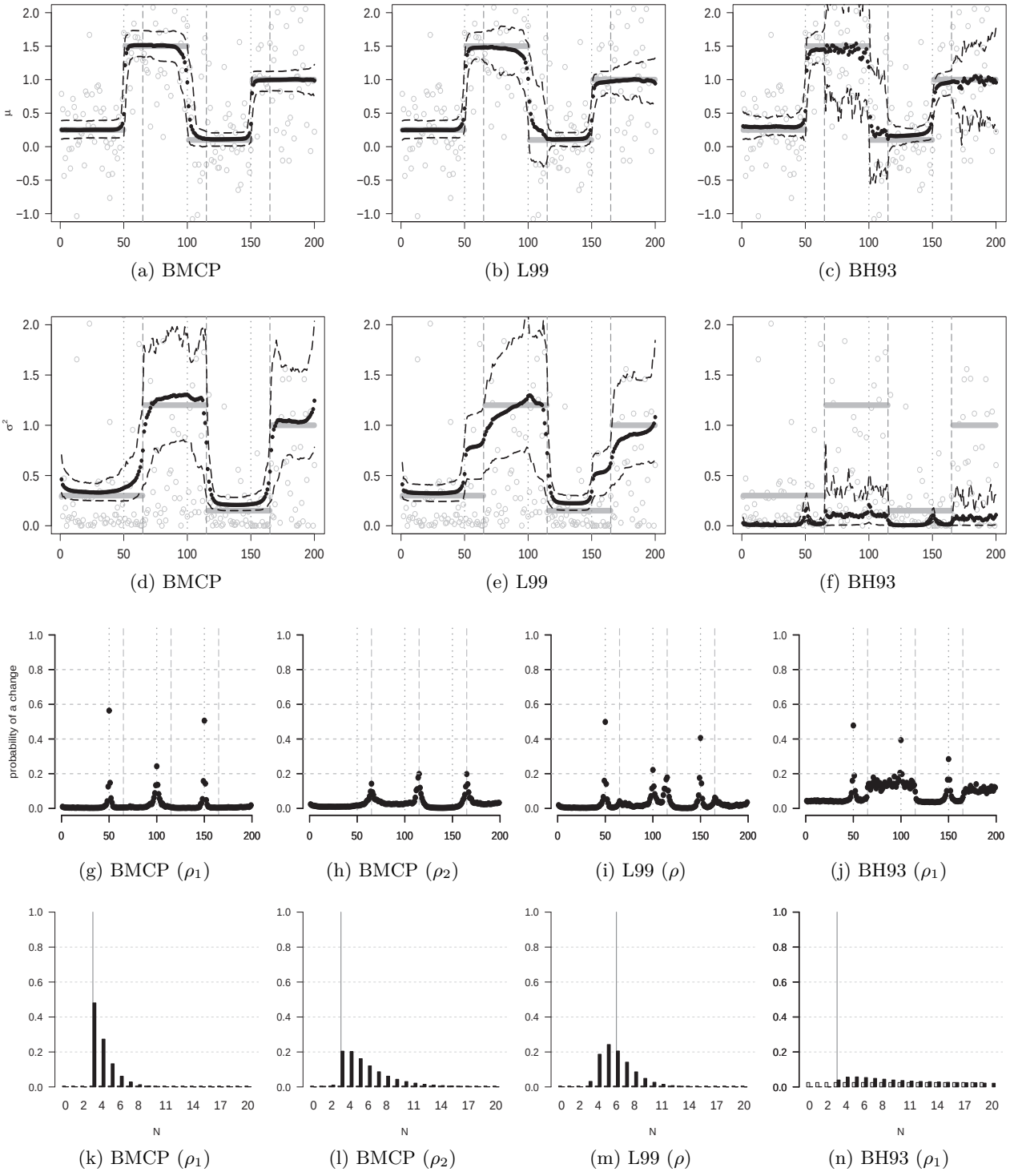
L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{50, 100, 114, 150, 200\}$	4	0.001616
$\rho = \{50, 100, 115, 150, 200\}$	4	0.001074
$\rho = \{50, 99, 114, 150, 200\}$	4	0.000931
$\rho = \{50, 100, 112, 150, 200\}$	4	0.000797
$\rho = \{50, 100, 113, 150, 200\}$	4	0.000759

BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{50, 100, 150, 200\}$	3	0.001361
$\rho_1 = \{50, 99, 150, 200\}$	3	0.001051
$\rho_1 = \{49, 100, 150, 200\}$	3	0.000939
$\rho_1 = \{50, 100, 149, 200\}$	3	0.000723
$\rho_1 = \{50, 100, 151, 200\}$	3	0.000619

Table 16: Most likely partitions based on the average estimated posterior probability for the scene 7.

Model	Partition	True N	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	3	4	1.7	3	0.48
BMCP	ρ_2	3	5.8	9.3	3	0.2
L99	ρ	6	6	3.8	5	0.24
BH93	ρ_1	3	17.5	134.6	5	0.06

Table 17: Summary of the average estimates of the posterior probability of the number of changes N for each model and respective partitions for the scene 7.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 10: Graphical summary for the MC simulation study of the scene 7. Figures (a-c): average posterior means estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample example (gray circles) and the true mean values (solid gray horizontal lines). Figures (d-f): average posterior variance estimates (black dots) and 90% precision interval (dashed lines) with a simulated sample squared error (gray circles) and the true variance values (solid gray horizontal lines). Figures (g-j): average estimates of the posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and average estimated posterior (black bars) probability distributions of the number of change points.

5 Real data applications

In this chapter, we apply the BMCP model to two real data sets: the Us EX-post real interest rate and the Mexican Peso/US Dollar exchange rate. We also compare BMCP estimates with that obtained applying the L99 and BH93 models. In both the cases, we consider the noninformative Bayes-Laplace prior $Beta(1, 1)$ for the cohesion parameters p_1, p_2 and p of the BMCP and L99 models, as mentioned in Section 2.5, which assumes that, on average, 50% of the instants will be a change point.

5.1 Real data application 1: US Ex-post real interest rate

The US ex-post real interest rate we analyse in this section is available at the RealInt data set of the R package **bcp**. This data set is a quarterly time series from 1961/1 to 1986/3 of the three-month treasury bill rate deflated by the Consumer Price Index (CPI) inflation rate. It has a total of $n = 103$ observations and is presented in Figure 11. The presence of structural changes in the mean and variance of this time series was analysed by Garcia and Perron [1996] and Bai and Perron [2003].

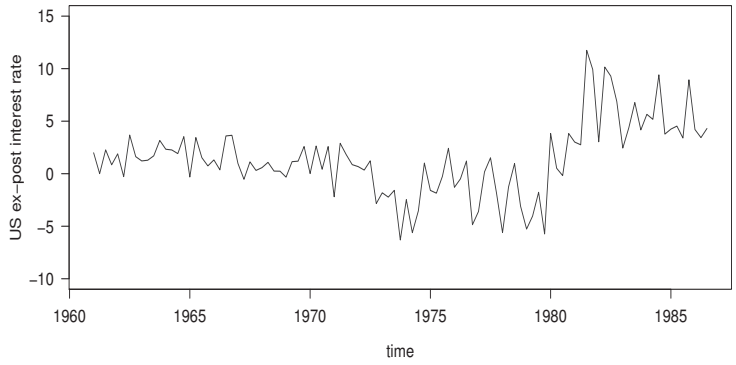


Figure 11: US ex-post real interest rate quarterly time series from 1961/1 to 1986/3.

We apply the BMCP to possibly detect changes in mean and variance of this data set. We also apply the L99 and BH93 models, to compare the results. We consider the same prior specifications of the simulation study presented in Chapter 4: $(\mu_0, \sigma_0^2, a, d) = (0, 100, 2, 2)$ for the BMCP model, $(m, v, a, d) = (0, 2, 2, 2)$ for the L99 model and $w_0 = p_0 = 0.2$ for the BH93 model. The posterior distributions are obtained through a MCMC scheme with 14,000 iterations, with the first 4,000 discarded as a burn-in period. The simulations took 3.2, 3.3 and 6.7 seconds for the BH93, L99 and BMCP models, respectively. The results are summarized in Figure 12 and Tables 18 and 19. In Figures 12a-12j, the vertical lines indicate the end points of the ρ_1 blocks (red dotted lines) and of the ρ_2 blocks (red dashed lines) of the most probable partitions ρ_1 and ρ_2 according to the BMCP model. Combining the most probable partitions estimated by the BMCP model, presented in Table 18, with the instants with the highest probabilities of a change in Figures 12g and 12h, we can say that the BMCP model indicates two changes in the mean, at instants 48 and 80 (one time next the end points 47 and 79), and one change in the variance, at instant 52 (one time next the end point 51). The posterior modes

for the number of changes in mean and variance estimated by the BMCP (see Table 19) also support this statement. Based on Figures 12a-12f, the BMCP and L99 models provide similar posterior point estimates for the means and the variances, with similar precision, but the L99 model estimates for the variance are affected by the two mean changes, just as observed in the Monte Carlo study in Chapter 4, in the scenes with the mean and the variance changes occurring at different times. The BH93 model mean estimates become unstable after the variance change indicated by the BMCP model, and the variance has lower estimates than BMCP and L99 models, maintaining the same behavior observed in the Monte Carlo study.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{47, 79, 103\}$	2	0.2067
$\rho_1 = \{47, 76, 103\}$	2	0.0996
$\rho_1 = \{47, 78, 103\}$	2	0.0221
$\rho_1 = \{47, 80, 103\}$	2	0.0218
$\rho_1 = \{47, 76, 82, 103\}$	3	0.0191

BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{51, 103\}$	1	0.1300
$\rho_2 = \{50, 103\}$	1	0.0452
$\rho_2 = \{49, 103\}$	1	0.0254
$\rho_2 = \{48, 103\}$	1	0.0136
$\rho_2 = \{39, 103\}$	1	0.0113

L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{47, 79, 103\}$	2	0.0867
$\rho = \{47, 76, 103\}$	2	0.0704
$\rho = \{47, 76, 82, 103\}$	3	0.0238
$\rho = \{46, 79, 103\}$	2	0.0165
$\rho = \{46, 76, 103\}$	2	0.0113

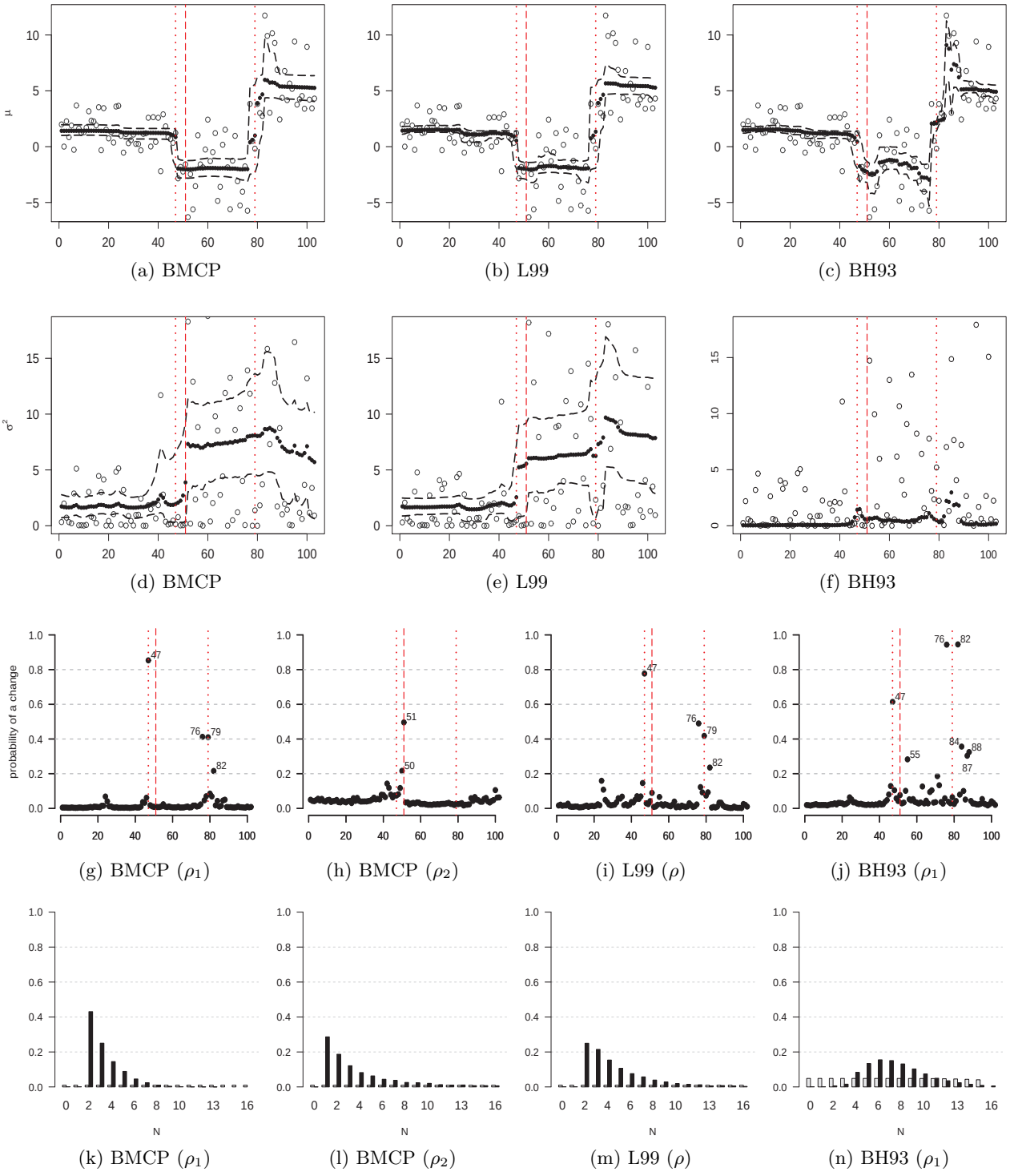
BH93	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{47, 76, 82, 88, 103\}$	4	0.0116
$\rho_1 = \{47, 76, 82, 87, 103\}$	4	0.0115
$\rho_1 = \{47, 76, 82, 84, 103\}$	4	0.0083
$\rho_1 = \{46, 76, 82, 87, 103\}$	4	0.0031
$\rho_1 = \{47, 76, 82, 103\}$	3	0.0030

Table 18: Top five most likely partitions based on the estimated posterior probability.

Model	Partition	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	3.2	2.2	2	0.43
BMCP	ρ_2	5.2	55.5	1	0.29
L99	ρ	4.7	9.1	2	0.25
BH93	ρ_1	7.6	8	6	0.16

Table 19: Summary of the estimates of the posterior probability of the number of changes N for each model and respective partitions, for real data application 1.

Our results for the BMCP and L99 models applied to the US Ex-post real interest rate data set are much similar to the results obtained by the model in Garcia and Perron [1996]. They used the Markov switching model proposed by Hamilton [1989], allowing a maximum number of three possible regimes affecting both the mean and the variance. They estimate two changes, in 1973 and 1981 (they do not specify the quarter of the year), while the BMCP model estimates two change points in the mean, in 1972/4 and 1980/4, that are the observations one time next the end points 47 and 79. Garcia and Perron [1996] results also indicate that the variance for the second and third blocks are equivalent, similar to the BMCP indication of a unique change in the variance next to the first change in the mean. The L99 estimates also coincide with the BMCP and the Garcia and Perron [1996] results. The results of Bai and Perron [2003] contrast with BMCP and Garcia and Perron [1996] applications because they found three regime changes. It is relevant to note that the two changes in the mean indicated by the BMCP model match two of the three changes indicated by Bai and Perron [2003]. The other change indicated by Bai and Perron [2003] occur in 1966/4.



(a) BMCP (b) L99 (c) BH93

(d) BMCP (e) L99 (f) BH93

(g) BMCP (ρ_1) (h) BMCP (ρ_2) (i) L99 (ρ) (j) BH93 (ρ_1)

(k) BMCP (ρ_1) (l) BMCP (ρ_2) (m) L99 (ρ) (n) BH93 (ρ_1)

Figure 12: Graphical summary for the application of the BMCP, L99 and BH93 models for the US Ex-post Real Interest Rate data set. Figures (a-c): posterior means estimates (black dots), 90% precision interval (dashed lines) and observation values (black circles). Figures (d-f): posterior variance estimates (black dots), 90% precision interval (dashed lines) and the square of the difference between the observation values and the respective posterior mean estimate (black circles). Figures (g-j): estimated posterior probability of each instant to be a change point. Figures (k-n): prior (gray bars) and estimated posterior (black bars) probability distributions of the number of change points.

5.2 Real data application 2: Mexican Peso/US Dollar exchange rate

The Mexican Peso/US Dollar exchange rate we analyse in this section is available at www.federalreserve.gov, and is presented in Figure 13. This data set is composed by daily records of Mexican Peso/US Dollar exchange rate from January 2007 to December 2012, a total of $n = 1,510$ observations. The presence of regime changes in this time series was analysed by Martínez et al. [2014] fitting a nonparametric change point detection model.

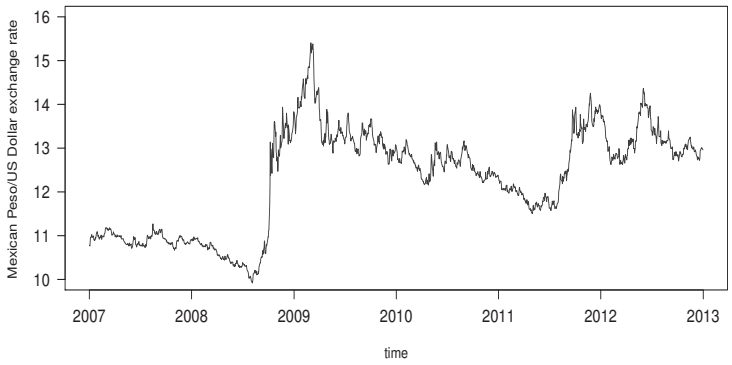


Figure 13: Daily records of Mexican Peso/US Dollar exchange rate from January 2007 to December 2012.

A priori, we assume the same hyperparameters specified in Section 5.1, excepted by $\sigma_0^2 = 10,000$, which reduced the number of short blocks (with less than ten observations, for example) that the BMCP model was identifying when we first assumed $\sigma_0^2 = 100$. Because of the higher number of observations in this application when compared to the application in Section 5.1 and to the simulation scenes discussed in Chapter 4, we increased the number of MCMC iterations to 140,000, with the first 40,000 discarded as a burn-in period. The simulations took 3.9, 41.8 and 246.4 minutes for the BH93, L99 and BMCP models, respectively. The BH93 model did not provide convenient estimates. The estimated posterior expected number of changes is very large, around 180, which is equivalent to occur, on average, one change at each eight observations. Because of this, we do not present the BH93 model results.

The results for the BMCP and L99 models are summarized in Figure 14 and Tables 20 and 21. In Figures 14a-14g, the vertical lines indicate the end points of the most probable partitions ρ_1 (red dotted lines) and ρ_2 (red dashed lines) according to the BMCP model, and the end points of the most probable partition ρ (blue solid lines) according to the L99 model. In this application, the BMCP and L99 models provide different results. According to the estimates of the posterior expected value and mode presented in Table 21, the number of changes in the mean obtained by the BMCP model is more than twelve times the total number of changes obtained by fitting the L99 model. The estimated posterior distribution of the ρ_1 and ρ_2 are too flat, not providing strong evidence about the respective change point locations. In this cases, the probability of a change in Figures 14e-14g may be a more useful result. For example, we could determine the true partition as the one composed by the instants with probability of a change greater than some probability threshold. On the other hand, the most probable partition ρ of the L99 model is estimated with probability 0.38.

Our results for the BMCP model applied to the Mexican Peso/US Dollar exchange rates are different from the results of the nonparametric change point model proposed by Martínez et al. [2014]. They estimmate a most probable partition with seven change points while the BMCP model estimates around forty changes in the mean and three changes in the variance (see Table 21). They also apply the L99 model to this data set, but we found different results. They estimate a most probable partition with four change points occurring with probability 0.08465, while we estimate three change points, with probability 0.3837. This difference may have been caused by some different prior specification.

BMCP	N	$p(\rho_1\|\boldsymbol{X})$
$\rho_1 = \{85, 144, 185, 278, 316, 348, 391, 421, 428, 448, 478, 514, 537, 555, 570, 655, 672, 701, 732, 798, 817,$ $843, 919, 940, 1009, 1070, 1157, 1182, 1190, 1228, 1268, 1276, 1325, 1347, 1353, 1383, 1406, 1436, 1510\}$	38	0.00030
$\rho_1 = \{87, 143, 187, 283, 317, 346, 392, 423, 430, 448, 478, 513, 539, 555, 569, 657, 672, 702, 730, 798, 818,$ $843, 919, 941, 1008, 1068, 1159, 1182, 1188, 1230, 1269, 1275, 1324, 1347, 1353, 1384, 1430, 1510\}$	37	0.00026
$\rho_1 = \{83, 147, 185, 284, 316, 350, 392, 422, 429, 447, 477, 514, 539, 554, 568, 654, 672, 703, 732, 798, 818, 842, 921,$ $934, 949, 1014, 1069, 1159, 1183, 1190, 1231, 1237, 1269, 1279, 1327, 1348, 1358, 1382, 1408, 1431, 1510\}$	40	0.00025
$\rho_1 = \{84, 144, 184, 274, 316, 347, 394, 422, 430, 447, 476, 513, 538, 554, 568, 655, 673, 701, 731, 795, 818, 842,$ $920, 940, 1010, 1070, 1158, 1182, 1190, 1231, 1236, 1268, 1276, 1325, 1348, 1358, 1382, 1408, 1433, 1510\}$	39	0.00022
$\rho_1 = \{81, 144, 184, 282, 318, 351, 393, 423, 432, 448, 477, 514, 537, 555, 569, 638, 643, 656, 671, 701, 732, 801, 819,$ $843, 919, 941, 1009, 1070, 1158, 1182, 1188, 1231, 1237, 1268, 1275, 1326, 1347, 1352, 1383, 1408, 1432, 1510\}$	41	0.00021

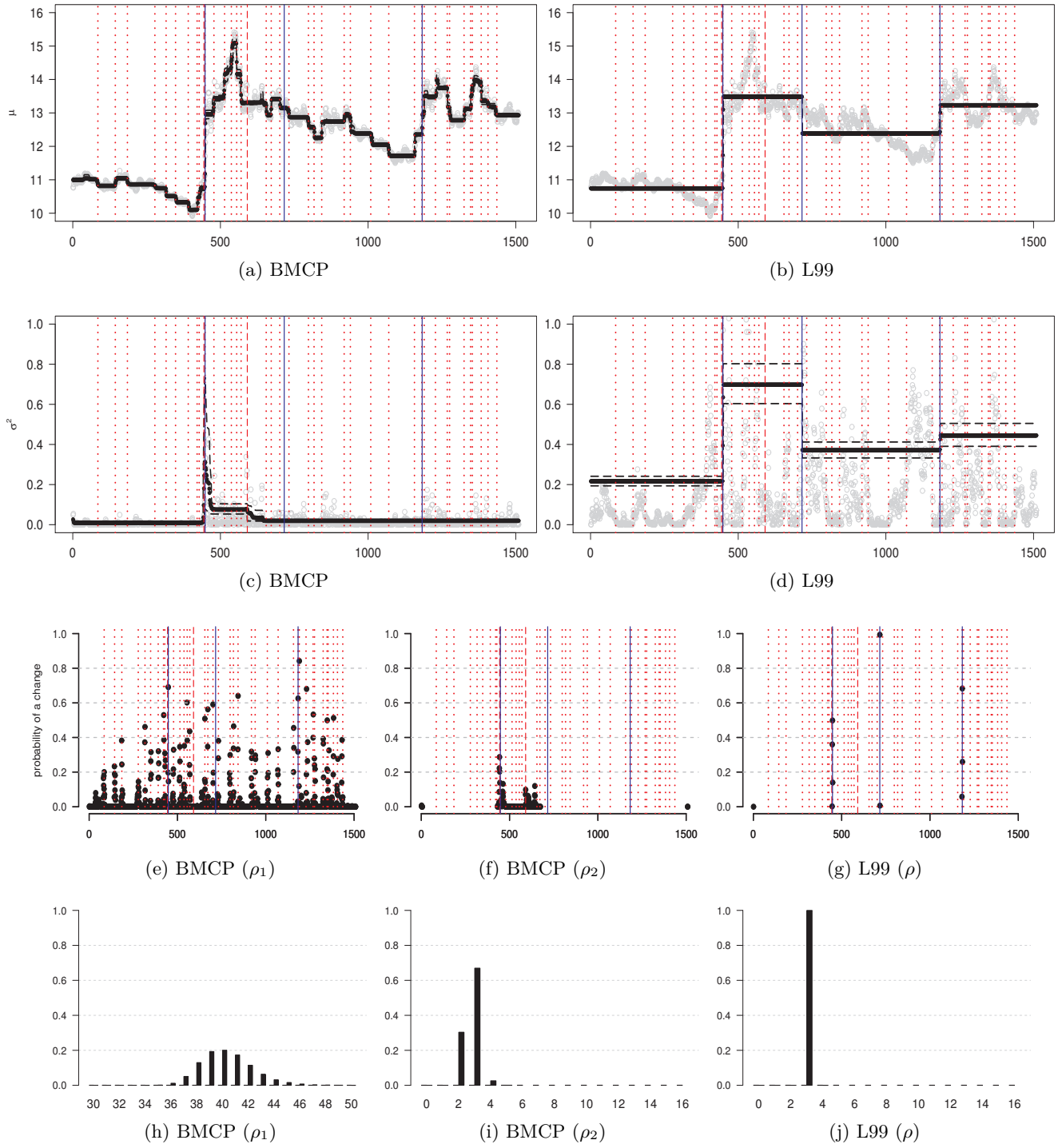
BMCP	N	$p(\rho_2\|\boldsymbol{X})$
$\rho_2 = \{444, 591, 1510\}$	2	0.01498
$\rho_2 = \{443, 590, 1510\}$	2	0.01388
$\rho_2 = \{443, 591, 1510\}$	2	0.01304
$\rho_2 = \{443, 604, 1510\}$	2	0.01048
$\rho_2 = \{442, 591, 1510\}$	2	0.00984

L99	N	$p(\rho\|\boldsymbol{X})$
$\rho = \{448, 716, 1183, 1510\}$	3	0.38370
$\rho = \{447, 716, 1183, 1510\}$	3	0.20571
$\rho = \{448, 716, 1184, 1510\}$	3	0.11439
$\rho = \{449, 716, 1183, 1510\}$	3	0.09291
$\rho = \{447, 716, 1184, 1510\}$	3	0.09025

Table 20: Top five most likely partitions for the BMCP and L99 models, based on the estimated posterior probability.

Model	Partition	$E(N\|X)$	$V(N\|X)$	$Mo(N\|X)$	$P(N = Mo\|X)$
BMCP	ρ_1	40.2	4.0369	40	0.2007
BMCP	ρ_2	2.7	0.2555	3	0.6699
L99	ρ	3	0.0001	3	0.9999

Table 21: Summary of the estimate of the posterior probability of the number of changes N for each model and respective partitions, for real data application 2.



(a) BMCP (b) L99

(c) BMCP (d) L99

(e) BMCP (ρ_1) (f) BMCP (ρ_2) (g) L99 (ρ)

(h) BMCP (ρ_1) (i) BMCP (ρ_2) (j) L99 (ρ)

Figure 14: Graphical summary for the application of the BMCP, L99 and BH93 models for the Mexican Peso/US Dollar data set. Figures (a-b): posterior means estimates (black dots), 90% precision interval (dashed lines) and observation values (black circles). Figures (c-d): posterior variance estimates (black dots), 90% precision interval (dashed lines) and the square of the difference between observation values and the respective mean estimate (black circles). Figures (e-g): estimated posterior probability of each instant to be a change point. Figures (h-j): prior (gray bars) and estimated posterior (black bars) probability distributions of the number of change points.

In the next chapter, we summarize the contributions of this thesis and point out some suggestions of future related works.

6 Conclusions

In this thesis, we introduce a new multipartition model to multiple change point detection in multiparametric models, referred to as BMCP model, such that we can identify separately the number and location of change points in different parameters. We developed an application of this model to Normal data and the results of the Monte Carlo simulation study presented in Chapter 4 show that the proposed model has similar ability to detect changes in the mean of Normal data when compared to the L99 and BH93 model, and has superior ability to identify changes in the variance of Normal data when compared to the L99 model. Moreover, the BMCP has a better performance than the L99 and BH93 models to identify the changes in the mean and in the variance of Normal data in scenes where the mean and the variance change at different times. Additionally to the L99 model, the BMCP model provides the information of which of these parameters experienced the change.

Some improvements to the model introduced in this thesis may be considered for future work. The influence of the prior specifications can be analysed. We can try other prior specifications for the block parameters, for example, the block variance prior may be a half-Cauchy (Gelman et al. [2006]) or a SBeta2 (Pérez et al. [2017]). The general change point model introduced in Chapter 3 may be applied to other multiparametric probability models, for example, the generalized exponential model, skew normal model and extreme value distributions. The independent *Beta* prior distributions assumed for the parameters p_k of the prior cohesions of each partition ρ_k, k=1, . . . , d, may be replaced by a joint correlated distribution for $(p_1, \ldots, p_d)$. Similar to the BMCP model, the change point model proposed by Peluso et al. [2019] also identifies changes separately in different parameters. This work suggests that we can have improved estimates if we make the parameters p_k time dependent. Another improvement may be the extension of the BMCP model to change point detection in multivariate sequences of observations.

The Normal data application of the BMCP model and the L99 model were implemented in C++ and will be submitted as a new R package to the Comprehensive R Archive Network (CRAN).

A Probability distributions

A.1 Basic distributions

This appendix presents the density functions and parameters description of the probability distributions considered in this thesis.

A.1.1 Univariate Normal

$X \sim N(\mu, \sigma^2)$ if its density function is given by

$$f(x|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right],\ \ x \in \mathbb{R}\ ,\ \ \mu \in \mathbb{R}\ ,\ \ \sigma^2 \in \mathbb{R}_+\ .$$

A.1.2 Inverse-Gamma

$X \sim IG(a, d)$ if its density function is given by

$$f(x|a,d) = \frac{a^d}{\Gamma(d)}\ x^{-(d+1)} e^{-a/x},\ \ x \in \mathbb{R}_+,\ \ a \in \mathbb{R}_+,\ \ d \in \mathbb{R}_+.$$

A.1.3 Beta

$X \sim Beta(\alpha, \beta)$ if its density function is given by

$$f(x|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ x^{\alpha-1}(1-x)^{\beta-1},\ \ x \in [0,1],\ \ \alpha \in \mathbb{R}_+,\ \ \beta \in \mathbb{R}_+.$$

A.1.4 Beta-Binomial

$X \sim Beta\text{-}Binomial(n, \alpha, \beta)$ if its density function is given by

$$f(x|\alpha,\beta) = \binom{n}{x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(n+\beta-x)}{\Gamma(\alpha+\beta+n)},\ \ x = 0,1,\ldots,n,\ \ n \in \mathbb{N},\ \ \alpha \in \mathbb{R}_+,\ \ \beta \in \mathbb{R}_+.$$

B Markov Chain Monte Carlo methods

This appendix briefly describes some Markov Chain Monte Carlo (MCMC) sampling methods used to evaluate the quantities of interest in this work: Metropolis-Hastings algorithm (M-H) and Gibbs sampler, which are powerful tools to simulate from complex high-dimensional distributions that can be evaluated but not easily sampled. Detailed and enlightening descriptions of these methods are available in Gamerman and Lopes [2006] or Liu [2008].

The partially collapsed Gibbs sampling method (PCG) proposed by Van Dyk and Park [2008] is also described here. It enables sampling from our proposed joint model for parameters and partitions, for which an ordinary Gibbs sampler is not feasible to run.

B.1 Metropolis-Hastings algorithm

Consider a distribution π from which a sample must be drawn. The MCMC sampling methods consist in construct a Markov chain for which the stationary distribution equals the target distribution π. It contributes to perform Bayesian inference when π is a posterior distribution for parameters $\pi(\theta|\boldsymbol{X})$. For a sufficiently large $t \in \mathbb{N}$, a realization $\theta^{(t)}$ from this chain will have approximate distribution $\pi(\theta|\boldsymbol{X})$.

The Metropolis-Hastings (M-H) method was introduced by Metropolis et al. [1953] and Hastings [1970]. A sample is generated from an arbitrary auxiliary distribution and it is accepted or rejected with some stated probability. Let $q(\cdot|\theta)$ represent a distribution that is easy to sample from and $\pi(\theta|\boldsymbol{X})$ the target distribution. The M-H algorithm can be specified by the following steps:

1. Initialize the iteration counter of the chain $t = 1$ and set an arbitrary initial value $\theta^{(0)}$.
2. Generate a value θ' from the density $q(\cdot|\theta^{(t-1)})$.
3. Compute the acceptance probability $\alpha(\theta^{(t-1)}, \theta') = \min\left\{1, \dfrac{\pi(\theta'|\boldsymbol{X})}{\pi(\theta^{(t-1)}|\boldsymbol{X})} \dfrac{q(\theta^{(t-1)}|\theta')}{q(\theta'|\theta^{(t-1)})}\right\}$.
4. Generate $u \sim \mathrm{U}(0,1)$. If $u \leq \alpha(\theta^{(t-1)}, \theta')$, accept $\theta^{(t)} = \theta'$, otherwise, $\theta^{(t)} = \theta^{(t-1)}$.
5. Set $t = t + 1$ and return to step 2 until convergence is reached.

B.2 Gibbs sampler

Assume $\pi(\boldsymbol{\theta}|\boldsymbol{X})$ is the target posterior distribution, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$, and the full conditional distributions $\pi(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d, \boldsymbol{X})$, $i = 1, \ldots, d$, are completely known and can be sampled from. The Gibbs sampler (Geman and Geman [1984]) is an MCMC method based on iterative simulation from full conditional distributions, as exemplified next:

1. Initialize the iteration counter of the chain $t = 1$ and set and set an arbitrary initial value $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})$.
2. Generate $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \ldots, \theta_d^{(t)})$ from $\theta^{(t-1)}$ through successive generation of

$$\begin{aligned}
\theta_1^{(t)} &\sim \pi(\theta_1|\theta_1^{(t-1)}, \ldots, \theta_d^{(t-1)}, \boldsymbol{X}),\\
\theta_2^{(t)} &\sim \pi(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}, \boldsymbol{X}),\\
&\vdots\\
\theta_d^{(t)} &\sim \pi(\theta_d|\theta_1^{(t)}, \ldots, \theta_{d-1}^{(t)}, \boldsymbol{X}).
\end{aligned}$$

3. Set $t = t + 1$ and return to step 2 until convergence is reached.

When convergence is reached, $\boldsymbol{\theta}^{(t)}$ is a sample from $\pi(\boldsymbol{\theta}|\boldsymbol{X})$. Clearly, a chain constructed via M-H or Gibbs sampler algorithms defines a Markov chain, since the sample in stage t is only dependent on the previous stage $t - 1$.

B.3 Partially collapsed Gibbs sampler

This section is a short presentation of thr Partially Colapsed Gibbs sampler (PCG), introduced by Van Dyk and Park [2008], which consists of a strategy developed to improve the convergence characteristics of a conventional Gibbs sampler.

To understand the PCG, it is necessary some insight into two sampling procedures related to the Gibbs sampler: grouping, also known as blocking, and collapsing, also known as marginalizing. Consider the Gibbs sampler scheme in Section B.2 with $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$. If we sample from some two components of $\boldsymbol{\theta}$ together, for example θ_2 and θ_3, this procedure is referred as grouping (it is possible if we are able to sample from $\pi(\theta_2|\theta_1)$ and then from $\pi(\theta_3|\theta_2, \theta_1)$). Collapsing consists of integrating out some random (irrelevant) parameters of the model. Both collapsing and grouping procedures usually result in more efficient Gibbs sampling schemes (Liu [1994]).

Based on grouping and collapsing, the PCG enables integrating out different components in different steps of the Gibbs algorithm without losing the convergence of the chain to the original target distribution.

To transform a Gibbs sampler into a PCG, three basic tools are used: marginalization, permutation and trimming. Park and Van Dyk [2009] use a simple and enlightening schematic example to illustrate how these three tools are used to transform a Gibbs sampler in a PCG sampler. This example is briefly illustrated here. It supposes a Gibbs sampler with target joint distribution $p(W, X, Y, Z)$ and also suppose it is possible to directly sample from $p(Y|X, Z)$ and $p(Z|X, Y)$, which are both conditional distributions of $\int p(W, X, Y, Z)dW$. They start with an ordinary Gibbs sampler.

Sampler 1.

1. Draw $W \sim p(W|X, Y, Z)$
2. Draw $X \sim p(X|W, Y, Z)$
3. Draw $Y \sim p(Y|W, X, Z)$
4. Draw $Z \sim p(Z|W, X, Y)$

The first tool is marginalization. It entails moving a group of unknowns components from being conditioned upon to being sampled in one or more steps of a Gibbs sampler. The marginalized group can differ among the steps. In this example, W is moved from being conditioned upon to being sampled in steps 3 and 4. Sampler 2 is a generalization of Sampler 1 that updates component W multiple times within each iteration. Thus, Sampler 2 has the same target distribution as Sampler 1.

Sampler 2.

1. Draw $W^\star \sim p(W|X, Y, Z)$
2. Draw $X \sim p(X|W, Y, Z)$
3. Draw $(W^\star, Y) \sim p(W, Y|X, Z)$
4. Draw $(W, Z) \sim p(W, Z|X, Y)$

The second tool is permutation. Drawing W three times in Sample 2 may be inefficient, but removing any two of the three draws affects the convergence of the chain to the target distribution. Only intermediate quantities whose values are not conditioned upon in subsequent steps may be removed. Permuting the steps of a Gibbs sampler does not alter its target distribution but can enable certain intermediate quantities to be removed. Thus, Sampler 3 has the same target distribution as Sampler 2.

Sampler 3.

1. Draw $(W^\star, Y) \sim p(W, Y|X, Z)$
2. Draw $(W^\star, Z) \sim p(W, Z|X, Y)$
3. Draw $W \sim p(W|X, Y, Z)$
4. Draw $X \sim p(X|W, Y, Z)$

The output of the iteration consists of the most recently sampled value of each quantity at the end of the iteration. The superscript * designates an intermediate quantity that is sampled but is not part of the output of an iteration. In Sampler 3, the intermediate draws of W in steps 1 and 2 are not used in the subsequent steps because a new value of W is sampled in step 3. Thus, they can be removed (or trimmed) from the sampler without affecting the stationary distribution of the chain. Therefore, Sampler 4 has the same stationary distribution as Sample 3.

Sampler 4.

1. Draw $(W^\star, Y) \sim p(W, Y|X, Z)$
2. Draw $(W^\star, Z) \sim p(W, Z|X, Y)$
3. Draw $W \sim p(W|X, Y, Z)$
4. Draw $X \sim p(X|W, Y, Z)$

Sampler 4 is not a Gibbs sampler, it is a PCG. Its conditional distributions are incompatible and permuting the order of the draws may alter the stationary distribution of the chain.

References

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18(1):1–22.

Barber, D., Cemgil, A. T., and Chiappa, S. (2011). *Bayesian time series models*. Cambridge University Press.

Barry, D. and Hartigan, J. A. (1992). Product partition models for change point problems. *The Annals of Statistics*, 20(1):260–279.

Barry, D. and Hartigan, J. A. (1993). A Bayesian analysis for change point problems. *Journal of the American Statistical Association*, 88(421):309–319.

Chib, S. (1998). Estimation and comparison of multiple change-point models. *Journal of econometrics*, 86(2):221–241.

Demarqui, F. N., Loschi, R. H., and Colosimo, E. A. (2008). Estimating the grid of time-points for the piecewise exponential model. *Lifetime Data Analysis*, 14(3):333–356.

Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *Environmetrics*, 21(7-8):661–686.

Dobigeon, N., Tourneret, J.-Y., and Scargle, J. D. (2007). Joint segmentation of multivariate astronomical time series: Bayesian sampling with a hierarchical model. *IEEE Transactions on Signal Processing*, 55(2):414–423.

Eddelbuettel, D. (2013). *Seamless R and C++ Integration with Rcpp*. Springer, New York. ISBN 978-1-4614-6867-7.

Eddelbuettel, D. and Balamuta, J. J. (2017). Extending extitR with extitC++: A Brief Introduction to extitRcpp. *PeerJ Preprints*, 5:e3188v1.

Eddelbuettel, D. and François, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software*, 40(8):1–18.

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.

Fearnhead, P. (2005). Exact Bayesian curve fitting and signal segmentation. *IEEE Transactions on Signal Processing*, 53(6):2160–2166.

Fearnhead, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and computing*, 16(2):203–213.

Fearnhead, P. and Liu, Z. (2007). On-line inference for multiple changepoint problems. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):589–605.

Fearnhead, P. and Liu, Z. (2011). Efficient Bayesian analysis of multiple changepoint models with dependence across segments. *Statistics and Computing*, 21(2):217–229.

Fearnhead, P. and Rigaill, G. (2019). Changepoint detection in the presence of outliers. *Journal of the American Statistical Association*, 114(525):169–183.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1(2):209–230.

Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. Chapman and Hall/CRC.

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.

Garcia, R. and Perron, P. (1996). An analysis of the real interest rate under regime shifts. *The Review of Economics and Statistics*, pages 111–125.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A. et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, 1(3):515–534.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721–741.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

Hartigan, J. A. (1990). Partition models. *Communications in statistics-Theory and methods*, 19(8):2745–2756.

Hastings, W. K. (1970). Monte carlo sampling methods using Markov chains and their applications.

Haynes, K., Fearnhead, P., and Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing*, 27(5):1293–1305.

Hegarty, A. and Barry, D. (2008). Bayesian disease mapping using product partition models. *Statistics in medicine*, 27(19):3868–3893.

Hlavac, M. (2018). *stargazer: Well-Formatted Regression and Summary Statistics Tables*. R package version 5.2.2.

Jordan, C., Livingstone, V., and Barry, D. (2007). Statistical modelling using product partition models. *Statistical Modelling*, 7(3):275–295.

Laplace, P.-S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des Sciences présentés par divers savans*, 6:621–656.

Lattanzi, C. and Leonelli, M. (2019). A changepoint approach for the identification of financial extreme regimes. *arXiv preprint arXiv:1902.09205*.

Liu, J. S. (1994). The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association*, 89(427):958–966.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Loschi, R. and Cruz, F. (2002). An analysis of the influence of some prior specifications in the identification of change points via product partition model. *Computational Statistics & Data Analysis*, 39(4):477–501.

Loschi, R., Iglesias, P., and Arellano-Valle, R. (1999). Bayesian detection of change points in the chilean stock market. In *Proceedings of the Section on Bayesian Statistical Science*, pages 160–165.

Loschi, R. H. and Cruz, F. R. (2005). Extension to the product partition model: computing the probability of a change. *Computational Statistics & Data Analysis*, 48(2):255–268.

Loschi, R. H., Cruz, F. R. B., Iglesias, P. L., and Arellano-Valle, R. B. (2003). A Gibbs sampling scheme to the product partition model: an application to change-point problems. *Computers & Operations Research*, 30(3):463–482.

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

Majumdar, A., Gelfand, A. E., and Banerjee, S. (2005). Spatio-temporal change-point modeling. *Journal of Statistical Planning and Inference*, 130(1-2):149–166.

Martínez, A. F., Mena, R. H., et al. (2014). On a nonparametric change point detection model in Markovian regimes. *Bayesian Analysis*, 9(4):823–858.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092.

Mira, A. and Petrone, S. (1996). Bayesian hierarchical nonparametric inference for change-point problems. *Bayesian Statistics*, 5:693–703.

Monteiro, J. V., Assunçao, R. M., Loschi, R. H., et al. (2011). Product partition models with correlated parameters. *Bayesian Analysis*, 6(4):691–726.

Müller, P., Quintana, F., and Rosner, G. L. (2011). A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*, 20(1):260–278.

Nascimento, F. F. d. and Moura e Silva, W. V. (2017). A Bayesian model for multiple change point to extremes, with application to environmental and financial data. *Journal of Applied Statistics*, 44(13):2410–2426.

Nemeth, C., Fearnhead, P., and Mihaylova, L. (2013). Sequential monte carlo methods for state and parameter estimation in abruptly changing environments. *IEEE Transactions on Signal Processing*, 62(5):1245–1255.

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.

Page, G. L., Quintana, F. A., et al. (2016). Spatial product partition models. *Bayesian Analysis*, 11(1):265–298.

Park, T. and Van Dyk, D. A. (2009). Partially collapsed Gibbs samplers: Illustrations and applications. *Journal of Computational and Graphical Statistics*, 18(2):283–305.

Peluso, S., Chib, S., Mira, A., et al. (2019). Semiparametric multivariate and multiple change-point modeling. *Bayesian Analysis*, 14(3):727–751.

Pérez, M.-E., Pericchi, L. R., Ramírez, I. C., et al. (2017). The scaled beta2 distribution as a robust prior for scales. *Bayesian Analysis*, 12(3):615–637.

Quintana, F. A. and Iglesias, P. L. (2003). Bayesian clustering and product partition models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):557–574.

R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Teixeira, L. V., Assunção, R. M., and Loschi, R. H. (2019). Bayesian space-time partitioning by sampling and pruning spanning trees. *Journal of Machine Learning Research*, 20(85):1–35.

Van Dyk, D. A. and Park, T. (2008). Partially collapsed Gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796.

Yao, Y.-C. (1984). Estimation of a noisy discrete-time step function: Bayes and empirical bayes approaches. *The Annals of Statistics*, pages 1434–1447.