Published in final edited form as:

Test (Madr). 2007 May ; 16(1): 145–171. doi:10.1007/s11749-006-0006-x.

Change-point diagnostics in competing risks models: Two posterior predictive p -value approaches

Chen-Pin Wang and

Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, San Antonio, USA

Malay Ghosh

Department of Statistics, University of Florida, Gainesville, USA

Chen-Pin Wang: WangC3@uthscsa.edu

Abstract

This paper presents a Bayesian diagnostic procedure for examining change-point assumption in the competing risks model framework. It considers the family of distributions arising from the cause-specific model as reported by Chiang (Introduction to stochastic processes in biostatistics. Wiley, New York, 1968) upon which change-points are added to accommodate possible distributional heterogeneity. Model departure, due to misspecification of change-points associated with either the overall survival distribution or cause-specific probabilities, is quantified in terms of a sequence of cumulative-sum statistics between each pair of adjacent change-points assumed. When assessing the asymptotic behavior of each sequence of cumulative-sum statistics using its posterior predictive p -values, see Rubin (Ann Stat 12:1151–1172, 1984) and partial posterior predictive p -values as reported by Bayarri and Berger (J Am Stat Assoc 95:1127–1142, 2000), we show that both types of p -values attain their greatest departure from 0.5 at the change-point that is missed in the assumed model, from which a diagnostic procedure is formalized. Statistical power of these two types of p -values is discussed.

Keywords

Change-point; Competing risks; Posterior predictive p -values

1 Introduction

In statistical modeling, change-points are used to mark the time points at which the parameters of the population distribution undergo sharp change. Change-point problems have been addressed in the framework of hypothesis testing (Hinkley 1971; Hinkley and Hinkley 1970; Pettitt 1980; James et al. 1987; Raftery and Akman 1986) and model estimation (Smith 1975; Raftery and Akman 1986; Barry and Hartigan 1993; Stephens 1994). Particularly, in the Bayesian paradigm, change-points are regarded as either model

parameters that are subject to posterior sampling (Smith 1975; Raftery and Akman 1986; Barry and Hartigan 1993; Stephens 1994), or a model's special characteristics to be sought via model selection procedures (e.g., Raftery and Akman 1986; Denison et al. 2002). Due to computational or analytical complexity, these formal procedures however can be hard tailored to suit situations that involve multiple (nested) processes, each having a distinct set of multiple change-points. One such situation, the subject matter of this paper, is the analysis of competing risks models where the underlying distributions are governed by multiple change-points. Specifically, we consider Chiang's (1968) cause-specific models where both the underlying survival distribution and the cause-specific failure probabilities are subject to multiple change-points. The main objective of the present paper is to identify the most likely change-points of these competing risks models using the Bayesian posterior predictive check approach.

Following Guttman (1967) and Rubin (1984), a series of studies on *posterior predictive p-values* (PPP) and their applications to model validation have been proposed (Meng 1994; Gelman et al. 1996a, 1996b; Glickman and Stern 1998). In essence, PPP of a statistic T is the tail area of the expected posterior predictive distribution of T with respect to its realization. PPP has been praised for its ease of implementation (see Chen et al. 2000; Robert 1994) as well as possible accommodation of improper objective priors when the posterior is proper.

However, PPP generally is less variable than uniform $(0, 1)$, the conventional reference distribution for gauging p -values under the null (Meng 1994). Specifically, when used as a formal test for evaluating the fit of the null model to a univariate T , PPP can be conservative (less likely to reject the null than it should) due to the so-called "double use of the data." This is because T is first included in part for computing the posterior distribution of the parameters needed for generating the posterior predictive samples and is subsequently used in evaluating the tail area of the target posterior predictive distribution (see Bayarri and Berger 2000). Thus, it is useful to consider adjusted PPP that conforms more closely to uniform $(0, 1)$. This has been made possible by Bayarri and Berger (2000), known as the partial posterior predictive p -value (PPPP) or the conditional posterior predictive p -value (CPPP). As in the spirit of PPP, both PPPP and CPPP involve calculating the posterior predictive distribution of some statistic T . However, while CPPP is based on the posterior of the model parameters ψ given some statistics $\mathbf{U}$, typically being the conditional MLE of ψ given T , PPPP is based on the posterior after removing the information about ψ contained in the observed T .

The present work is primarily a comparison of PPPP with PPP in identifying change-points in the cause-specific competing risks model considered in Sect. 2. We note that if the assumed model is correct, then both the means of PPP and PPPP converge (as the sample size approaches to infinity) in probability to 0.5. Hence the discrepancy, either $|E(\text{PPP}) - 0.5|$ or $|E(\text{PPPP}) - 0.5|$, is a useful measure of model departure. Naturally, this discrepancy measure can be used to assess the relative fit of two statistics, say T_1 and T_2 , by comparing the extent of their departure (see (3.6) and (3.7) in Sect. 3). We say that T_2 is more affected than T_1 by model misspecification if such a departure index for T_2 exceeds that for T_1 . This technique is employed here to examine the change-point assumption associated with both

the overall survival distribution and cause-specific probabilities via evaluating the relative fits of their corresponding sequence(s) of cumulative-sum statistics as described in Sect. 4. We formalize a diagnostic procedure based on our theorems derived for the single change-point situation, where the asymptotic means of PPP's (or PPPP's) associated with these cumulative-sum statistics are piecewise monotone so that their maximum departure from 0.5 is attained at the change-point missed in the assumed model. When verifying our theorems using simulations under various scenarios, we find that PPPP is often more powerful than PPP in revealing the absolute departure, but not necessarily more effective for change-point detection which relies instead on the increment of PPP (or PPPP) departure between consecutive time points (see Sect. 5). It is shown that PPP can be superior to PPPP as the magnitude of the increments associated with PPP are greater than those of PPPP, specially in the neighborhood of some change-points. This result however is not consistent throughout the finite sample situations. Furthermore, we propose a forward model selection procedure as described in Sect. 4.3 to appropriate the use of these posterior predictive diagnostics for models involving multiple change-points.

2 Change-point competing risks models

In the general competing risks setting, subjects are exposed to multiple distinct risks, often referred to as “competing risks,” that jointly govern the event occurrence. We assume the causes of events to be mutually exclusive; therefore, each subject's event occurrence is associated with a single cause. We also assume the event occurrence to be nonrecursive. For data of this kind, one often elicits its underlying distribution from the family of multivariate survival distributions. To ensure model identifiability (Tsiatis 1975), we adopt the cause-specific model (Chiang 1968). This cause-specific modeling (CSM) requires specification of the overall survival distribution and joint cause-specific failure probabilities. It however does not impose an explicit dependence structure among the risks.

Suppose that there exist K competing risks labeled as $1, \dots, K$. We consider the situation where a random sample of size n is recruited at time x_0 , and each study subject i is followed up to $\min\{x_{W_i}, x_J\}$, where W_i denotes the discrete waiting time of event occurrence. Let the event occurrence history of each subject i be coded by $\mathbf{Y}_i = (Y_i^{(1)}, \dots, Y_i^{(W_i)})$ and the cause of event by $C_i^{(W_i)}$. If the event occurs during $[x_{w_i-1}, x_{w_i}]$, i.e., $W_i = w_i$, and $w_i \leq J$, then $Y_i^{(j)} = 0$ for $j < w_i$, $Y_i^{(w_i)} = 1$, and $1 \leq C_i^{(w_i)} \leq K$. If $w_i > J$ (meaning that no event occurs during $[x_0, x_J]$), then $Y_i^{(j)} = 0$ for $1 \leq j \leq J$, and both w_i and $C_i^{(w_i)}$ are censored. Thus under the noninformative censoring assumption, the CSM-based joint likelihood of $\mathbf{Y}_i = \mathbf{y}_i$ and $C_i^{(W_i)} = c_i^{(w_i)}$ is given by

$$\{p(y_i^{(1)}, \dots, y_i^{(w_i)})p(c_i^{(w_i)} | y_i^{(w_i)} = 1)\}^{\mathbf{1}_{[w_i \leq J]}} \{p(y_i^{(1)}, \dots, y_i^{(J)})\}^{\mathbf{1}_{[w_i > J]}}, \quad (2.1)$$

where $p(\cdot)$ is the generic symbol for a probability density function, and $\mathbf{1}_{[\cdot]}$ is the usual indicator function.

As a generalization of (2.1), change-points are posited upon such CSM to characterize temporal change experienced during the process. Specifically, among the change-point competing risks models (CCRM) discussed below, we assume that $[x_j, x_{j+1})$ contains a change-point of $\mathbf{Y}_i$ if the parameters indexing $p(Y_i^{(1)}, \dots, Y_i^{(j)})$ differ from those of $p(Y_i^{(j+1)}, \dots, Y_i^{(J)})$. Also, a change-point of $\mathbf{C}$ is posited in $[x_j, x_{j+1})$ if $p(C_i^{(j')} | Y_i^{(j')}=1) \neq p(C_i^{(j'+1)} | Y_i^{(j'+1)}=1)$.

Let $\Pr(Y_i^{(1)}=y_i^{(1)}, \dots, Y_i^{(J)}=y_i^{(J)})$ be parameterized in terms of a multinomial distribution, say

$$\Pr(Y_i^{(j)}=1)=\theta^{(j)} \quad (2.2)$$

for $1 \leq j \leq J$ and $1 \leq i \leq n$, where $\sum_{j=1}^{J+1} \theta^{(j)}=1$ with $\theta^{(J+1)} = \Pr(W_i > J)$. For $\mathbf{C}$, we posit

$$\Pr(C_i^{(j)}=k | Y_i^{(j)}=1)=\rho_k^{(j)} \quad (2.3)$$

for $1 \leq k \leq K$, $1 \leq j \leq J$, and $1 \leq i \leq n$, where $\sum_{k=1}^K \rho_k^{(j)}=1$ for all j .

Suppose that there is no change in the distribution of $\mathbf{Y}$ nor that of $\mathbf{C}$ within $[x_0, x_J)$. Then it is equivalent to assume that

$$\theta^{(1)} = \dots = \theta^{(J)} \quad (2.4)$$

and

$$\rho_k^{(1)} = \dots = \rho_k^{(J)} \quad (2.5)$$

for $1 \leq k \leq K$.

The CCRM of our primary interest contains at least one change-point associated with $\mathbf{Y}$ or $\mathbf{C}$ falling within (x_0, x_J) , or both. We begin with formalizing CCRM with one change-point for $\mathbf{Y}$ and one for $\mathbf{C}$. Suppose that the change-points associated with $\mathbf{Y}$ and $\mathbf{C}$ fall into $[x_{\nu_y}, x_{\nu_y+1})$ and $[x_{\nu_c}, x_{\nu_c+1})$, respectively, for $1 \leq \nu_y < J$ and $1 \leq \nu_c < J$. Then

$$\theta^{(1)} = \dots = \theta^{(\nu_y)} \text{ and } \theta^{(\nu_y+1)} = \dots = \theta^{(J)} \quad (2.6)$$

with $\theta^{(\nu_y)} \neq \theta^{(\nu_y+1)}$, and

$$\rho_k^{(1)} = \dots = \rho_k^{(\nu_c)} \text{ and } \rho_k^{(\nu_c+1)} = \dots = \rho_k^{(J)} \quad (2.7)$$

with $\rho_k^{(\nu_c)} \neq \rho_k^{(\nu_c+1)}$ for at least one $k \in \{1, \dots, K\}$.

To write the likelihood function for the above CCRM, we let $\boldsymbol{\theta} = (\theta_1, \theta_2)$ and $\boldsymbol{\rho} = (\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$, where $\theta_1 = \theta^{(1)}$, $\theta_2 = \theta^{(\nu_y+1)}$, $\boldsymbol{\rho}_1 = (\rho_{11}, \dots, \rho_{1K}) = (\rho_1^{(1)}, \dots, \rho_K^{(1)})$ and $\boldsymbol{\rho}_2 = (\rho_{21}, \dots, \rho_{2K}) = (\rho_1^{(\nu_c+1)}, \dots, \rho_K^{(\nu_c+1)})$. Also, we define a catalog of notation below:

- $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_n)$, where $\mathbf{h}_i = (y_i^{(1)}, \dots, y_i^{(w_i)}, c_i^{(w_i)})$.
- d_{jk} : the total number of cause- k failures observed within $[x_0, x_j]$.
- $d_{j\cdot}$: the total number of failures observed within $[x_0, x_j]$.
- s_j : the total number of survival intervals observed within $[x_0, x_j]$.

Suppose that $x_{\tilde{\nu}_y}$ and $x_{\tilde{\nu}_c}$ are the change-points assumed by the model (which may not necessarily be the true change-points, x_{ν_y} and x_{ν_c} , unless the model is correct). Then the joint likelihood of $(\mathbf{Y}, \mathbf{C})$ is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\rho} | \mathbf{h}) = L^y(\boldsymbol{\theta} | \mathbf{h}) L^c(\boldsymbol{\rho} | \mathbf{h}), \quad (2.8)$$

where

$$L^y(\boldsymbol{\theta} | \mathbf{h}) \propto \begin{cases} \{\theta_1^{d_{\tilde{\nu}_y}} \theta_2^{(d_{J\cdot} - d_{\tilde{\nu}_y})}\} \{1 - \tilde{\nu}_y \theta_1 - (J - \tilde{\nu}_y) \theta_2\}^{(n - d_{J\cdot})} & \text{if } \tilde{\nu}_y \leq J, \\ \theta_1^{d_{J\cdot}} \{1 - J \theta_1\}^{(n - d_{J\cdot})} & \text{if } \tilde{\nu}_y > J, \end{cases}$$

and

$$L^c(\boldsymbol{\rho} | \mathbf{h}) \propto \begin{cases} \prod_{k=1}^K \{(\rho_{1k})^{d_{\tilde{\nu}_c k}} (\rho_{2k})^{(d_{JK} - d_{\tilde{\nu}_c k})}\} & \text{if } \tilde{\nu}_c \leq J, \\ \prod_{k=1}^K \{(\rho_{1k})^{d_{JK}}\} & \text{if } \tilde{\nu}_c > J. \end{cases}$$

By adding additional x'_{ν} s in (2.2)–(2.7), we can construct CCRM with multiple change-points.

Bayesian analysis of CCRM requires in addition specification of priors for model parameters and estimation of the associated posteriors. We defer these technical details to Sect. 5. Instead, we discuss next posterior predictive p-value diagnostics employed herein for checking CCRM assumption regarding change-points.

3 Posterior predictive checks

Let T denote a statistic used for model diagnostics. Also throughout this paper, all statistics annotated with $*$ are referred to their posterior predictive counterparts. Thus T^* denotes the posterior predictand of T based on the assumed model. Then the Bayesian goodness-of-fit of T can be assessed by its posterior predictive p -value,

$$\text{PPP}(T; \mathbf{U}) = 1 - F^*(T; \mathbf{U}), \quad (3.1)$$

where

$$F^*(T; \mathbf{U}) = \int \Pr(T^* \leq T | \boldsymbol{\psi}, T) \pi(\boldsymbol{\psi} | \mathbf{U}) d\boldsymbol{\psi}, \quad (3.2)$$

$\mathbf{U}$ is a function of $\mathbf{h}$ (as defined in Sect. 2), and $\pi(\boldsymbol{\psi} | \mathbf{U})$ is the posterior of $\boldsymbol{\psi}$ given $\mathbf{U}$. Such a fit index can be viewed as a measure of compatibility between the assumed model and the observed data, where T is elicited to suit the purpose of model diagnostics.

For a given T , it is natural to consider the comparison of (3.1) derived from various $\mathbf{U}$'s. Below we discuss three versions of commonly used $\mathbf{U}$ and their resulting p -values.

Originally, Guttman (1967) and Rubin (1984) considered

$$\int \Pr(T^* \leq T | \boldsymbol{\psi}, T) \pi(\boldsymbol{\psi} | \mathbf{h}) d\boldsymbol{\psi}, \quad (3.3)$$

where $\mathbf{h}$ is as defined previously, the entire dataset in our case. This, due to the likelihood factorization theorem, is essentially equivalent to $\int \Pr(T^* \leq T | \boldsymbol{\psi}, T) \pi(\boldsymbol{\psi} | \hat{\boldsymbol{\psi}}_{SS}) d\boldsymbol{\psi}$, or putting $\mathbf{U} = \hat{\boldsymbol{\psi}}_{SS}$ into (3.1), where $\hat{\boldsymbol{\psi}}_{SS}$ denotes the sufficient statistic(s) for $\boldsymbol{\psi}$ and thus contains all the relevant data associated with $\boldsymbol{\psi}$.

In lieu of (3.3), Bayarri and Berger (2000) proposed two alternatives: the partial posterior predictive p -values (PPPP),

$$\int \Pr(T^* \leq T | \boldsymbol{\psi}, T) \pi(\boldsymbol{\psi} | \hat{\boldsymbol{\psi}}_{SS} \setminus T) d\boldsymbol{\psi}, \quad (3.4)$$

and the conditional posterior predictive p -values (CPPP),

$$\int \Pr(T^* \leq T | \boldsymbol{\psi}, T) \pi(\boldsymbol{\psi} | \hat{\boldsymbol{\psi}}_T) d\boldsymbol{\psi}, \quad (3.5)$$

where $\pi(\boldsymbol{\psi} | \hat{\boldsymbol{\psi}}_{SS} \setminus T) \propto \pi(\boldsymbol{\psi}) p(\hat{\boldsymbol{\psi}}_{SS} | \boldsymbol{\psi}) / p(T | \boldsymbol{\psi})$ corresponds to the posterior distribution of $\boldsymbol{\psi}$ based on all the relevant information about $\boldsymbol{\psi}$ but T (see (9) in Bayarri and Berger 2000), while $\hat{\boldsymbol{\psi}}_T$ is the MLE of $\boldsymbol{\psi}$ with respect to $p(\mathbf{h} | \boldsymbol{\psi}, T)$, the conditional likelihood of the data given T (see (13) in Bayarri and Berger 2000). In particular, Bayarri and Berger (2000) recommended (3.4) over (3.5) for discrete distributions.

The choice of $\mathbf{U}$ in (3.1) is critical in terms of concluding the absolute fit of a scalar T , since it anchors the reference distribution with respect to which the goodness-of-fit of T is assessed. Bayarri and Berger (2000) have shown that both (3.4) and (3.5) are generally superior to (3.3), since the information about $\boldsymbol{\psi}$ contained in T is used only once, and thus the resulting inference is not susceptible of being smoothed towards the null. Nevertheless, (3.3), (3.4), and (3.5) are asymptotically equivalent when a special T is selected. For instance, if T is ancillary under the null hypothesis, then all three emerge as the frequentist p -value.

Later, in Sects. 4 and 5, we give another instance where, in CCRM analyses, the diagnostics due to (3.3) concur those due to (3.4) when model fits are examined with relative fit across multiple T statistics quantified by (3.3) or (3.4). That is, for $T_1 \leq T_2$, (3.3) draws conclusion based on the sign of

$$\left| E \left[\int \Pr(T_1^* \leq T_1 | \psi, T_1) \pi(\psi | \hat{\psi}_{SS}) d\psi \right] - 0.5 \right| - \left| E \left[\int \Pr(T_2^* \leq T_2 | \psi, T_2) \pi(\psi | \hat{\psi}_{SS}) d\psi \right] - 0.5 \right|, \quad (3.6)$$

while (3.4) is based on the sign of

$$\left| E \left[\int \Pr(T_1^* \leq T_1 | \psi, T_1) \pi(\psi | \hat{\psi}_{SS} \setminus T_1) d\psi \right] - 0.5 \right| - \left| E \left[\int \Pr(T_2^* \leq T_2 | \psi, T_2) \pi(\psi | \hat{\psi}_{SS} \setminus T_2) d\psi \right] - 0.5 \right|. \quad (3.7)$$

Both these indices can be interpreted as the average impact of model misspecification on the prediction of T_2 compared to that on the prediction of T_1 . Therefore, despite different $\mathbf{U}$'s are used in (3.3) and (3.4), they arrive at the same conclusion as long as (3.6) and (3.7) have the same sign.

In general, statistical power of these different types of posterior predictive p -values (for detecting a certain model misspecification) can be evaluated based on both absolute and relative scales. Based on the absolute scale, one posterior predictive p -value is considered to be more powerful than the other if its departure from 0.5 is the greatest for all diagnostic statistics under all θ and $\mathbf{h}$. While differences in the departure among the diagnostic statistics can be useful power indices when they are more direct measures (compared with the absolute departure) for the nature of model mis-specification. For diagnostics based on (3.6) and (3.7), greater the discrepancy of the departure greater the power to differentiate the fit between T_1 and T_2 .

For change-point diagnostics in CCRM, we consider two sequences of T 's, one for $\mathbf{Y}$ and one for $\mathbf{C}$, and rely on (3.6) and (3.7) to guide the elicitation of their change-points. These diagnostic statistics are discussed next.

4 Application of PPP and PPPP to change-point diagnostics for CCRM

4.1 The PPP approach

In the light of Page (1955) and Pettitt (1980), we consider two types of cumulative-sum statistics for change-points diagnostics for CCRM and evaluate their associated PPP's. Without loss of generality, we shall describe our procedure for the situation where there is only one change-point associated with either $\mathbf{Y}$ or $\mathbf{C}$, or both $\mathbf{Y}$ and $\mathbf{C}$, within $(x_0, x_J]$, and the null model assumes $x_{v_Y} \sim x_J$ and $x_{v_C} \sim x_J$ (i.e., no change in the distribution occurs during $(x_0, x_J]$).

Let $\pi(\theta, \rho) = \pi(\theta)\pi(\rho)$ and $\pi(\theta, \rho | \hat{\theta}_{SS}, \hat{\rho}_{SS}) = \pi_{\theta}^y(\theta) \pi_{\rho}^c(\rho) \equiv \pi(\theta | \hat{\theta}_{SS}) \pi(\rho | \hat{\rho}_{SS})$, where $\hat{\theta}_{SS} = d_J$ and $\hat{\rho}_{SS} = (d_1, \dots, d_K)$. To verify whether the assumption $x_{v_Y} \sim x_J$ (i.e., no change-point associated with $\mathbf{Y}$ during $(x_0, x_J]$) is adequate, we evaluate (3.3) for

$\mathbf{T}=(T_1, \dots, T_{J-1})' \equiv \left(\frac{d_{1.}}{d_{J.}}, \dots, \frac{d_{(J-1).}}{d_{J.}} \right)'$ and $U = \hat{\theta}_{SS} = d_{J.}$. Here T_j corresponds to the realized failure proportion accrued over $(x_0, x_j]$, and $\hat{\theta}_{SS} = d_{J.}$ is the total number of failures observed during $(x_0, x_J]$. Thus, calculating

$$PPP^y(j) \equiv PPP^y\left(\frac{d_{j.}}{d_{J.}}; \hat{\theta}_{SS}; \theta\right) = \int \Pr\left(\frac{d_{j.}^*}{d_{J.}^*} \leq \frac{d_{j.}}{d_{J.}} \middle| \theta, \frac{d_{j.}}{d_{J.}}\right) \pi_\theta^y(\psi | d_{J.}) d\psi \quad (4.1)$$

for $1 \leq j < J$ yields a sequence of PPP's to inform the fit of these cumulative-sum statistics.

As in (4.1), the change-point assumption for $\mathbf{C}$ (i.e., whether $x_{v_c} = x_j$) is examined by the posterior predictive p -values of cumulative cause-specific failure probabilities:

$$PPP^c(j, k) \equiv PPP^c(d_{jk}; (\hat{\rho}_{SS}, \mathbf{Y}); \boldsymbol{\rho}) = \int \Pr(d_{jk}^* \leq d_{jk} \middle| \boldsymbol{\rho}, d_{jk}, \mathbf{Y}) \pi_{\rho, k}^c(\psi \middle| \hat{\rho}_{SS}) d\boldsymbol{\rho} \quad (4.2)$$

for $1 \leq j < J$ and $1 \leq k < K$, where $\hat{\boldsymbol{\rho}} = (d_{J1}, \dots, d_{JK})'$ and $\pi_{\rho, k}^c(\rho_k | \hat{\rho}_{SS}) = \pi(\rho_k | \hat{\rho}_{SS})$.

To identify the true change-points x_{v_y} and x_{v_c} based on (4.1) and (4.2), respectively, we examine the patterns of the means and quantiles associated with each sequence of posterior predictive p -values. These diagnostic quantities are derived from the empirical distributions of (4.1) and (4.2) as described in Sect. 5.2.

We point out twofold practical advantages of the proposed PPP diagnostic technique. First, this technique can be extended to examining CCRM's with multiple change-points. For such a situation, we carry out PPP diagnostics in conjunction with a forward model selection procedure as described in Sect. 4.3. Second, the proposed PPP's can effectively distinguish change-points from those regular-points in the way that if there is one true change-point missed in the interval where PPP's are assessed, then the average statistical power for detecting so attains its local maximum at that change-point. Technical details of our assertion are summarized in the theorems below.

Theorem 4.1 *Let x_{v_y} denote the true change-point of $\mathbf{Y}$ in a CCRM. Suppose that $\mathbf{Y}$ is fitted by a CCRM assuming that $x_{v_y} \sim x_J$. Then*

- i. (Meng 1994) $E(PPP^y(j)) = 0.5$ for $1 \leq j < J$ as $n \rightarrow \infty$ if $x_{v_y} = x_J$. If $x_{v_c} < x_J$, then
- ii. $|E(PPP^y(j)) - 0.5|$ is asymptotically monotone decreasing with $|x_j - x_{v_c}|$, and
- iii. for $0 < \alpha < 0.5$ and $|x_j - x_{v_y}| < |x_{j'} - x_{v_y}|$,

$$\lim_{n \rightarrow \infty} \Pr(PPP^y(j) \in A_\alpha) \geq \lim_{n \rightarrow \infty} \Pr(PPP^y(j') \in A_\alpha) \quad (4.3)$$

when $\theta_1 < \theta_2$ and $A_\alpha = (0, \alpha]$, or $\theta_1 > \theta_2$ and $A_\alpha = [1 - \alpha, 1)$.

Proof Rewrite $PPP^y(j)$ as

$$\int \int \Pr \left(\frac{d_{j.}^*}{d_{j.}^*} \leq \frac{d_{j.}}{d_{j.}} \middle| \psi, d_{j.}^*, d_{j.}, d_{j.} \right) p(d_{j.}^* | \psi) \pi_{\theta}^y(\psi) dd_{j.}^* d\psi.$$

Note that $d_{j.}^* |_{[d_{j.}^*, d_{j.}, d_{j.}]} \sim d_{j.}^* |_{d_{j.}^*}$ and $d_{j.}^* |_{d_{j.}^*} \sim \text{Bin}(d_{j.}^*, w_j)$, where $w_j = jJ^{-1}$. Applying the

Central Limit Theorem (CLT) to $\frac{d_{j.}^*}{d_{j.}^*}$ gives

$$\text{PPP}^y(j) = \int \int \Phi \left(\frac{\sqrt{d_{j.}^*} (\frac{d_{j.}}{d_{j.}^*} - w_j)}{\sqrt{w_j(1-w_j)}} + o_p(1) \right) p(d_{j.}^* | \psi) \pi_{\theta}^y(\psi) dd_{j.}^* d\psi = \int \Phi \left(\frac{\sqrt{d_{j.}^*} (\frac{d_{j.}}{d_{j.}^*} - w_j)}{\sqrt{w_j(1-w_j)}} + o_p(1) \right) p(d_{j.}^* | d_{j.}) dd_{j.}^*.$$

Let

$$\begin{aligned} \kappa_j &= \begin{cases} jJ^{-1} & \text{if } x_j \in [x_1, x_{\nu_y}], \\ \nu_y J^{-1} & \text{if } x_j \in (x_{\nu_y}, x_J], \end{cases} \\ \tau_j &= \begin{cases} 0 & \text{if } x_j \in [x_1, x_{\nu_y}], \\ (j - \nu_y)J^{-1} & \text{if } x_j \in (x_{\nu_y}, x_J], \end{cases} \\ \tilde{\theta} &= \kappa_{\nu_y} \theta_1 + (1 - \kappa_{\nu_y}) \theta_2, \\ \tilde{w}_j &= \kappa_j \theta_1 \tilde{\theta}^{-1} + \tau_j \theta_2 \tilde{\theta}^{-1}. \end{aligned}$$

Then, as $d_J \rightarrow \infty$

$$\frac{|\frac{d_{j.}}{d_{j.}^*} - w_j|}{\sqrt{w_j(1-w_j)}} \xrightarrow{p} \frac{|\tilde{w}_j - w_j|}{\sqrt{w_j(1-w_j)}} = \begin{cases} \sqrt{\frac{w_j}{1-w_j}} |1 - \theta_1 \tilde{\theta}^{-1}| & \text{if } x_j \in (x_0, x_{\nu_y}], \\ \sqrt{\frac{1-w_j}{w_j}} |\theta_1 - \theta_2| \tilde{\theta}^{-1} \kappa_{\nu_y} & \text{if } x_j \in (x_{\nu_y}, x_J], \end{cases}$$

which decreases with $|x_j - x_{\nu_y}|$. Since $d_{j.}/d_{j.} = O_p(1/\sqrt{d_{j.}})$ and is uniformly integrable, the lemma below implies (ii), that is:

Lemma 4.1 If $\Pr(U_1 \leq u | Z) < \Pr(U_2 \leq u | Z)$ then $\Pr(U_1 < u) < \Pr(U_2 < u)$.

To verify result (iii), according to Lemma 4.1, it is sufficient to consider

$$\begin{aligned} & \int \Pr \left(\Phi \left(\frac{\sqrt{d_{j.}^*} (\frac{d_{j.}}{d_{j.}^*} - w_j)}{\sqrt{w_j(1-w_j)}} \right) \in A_{\alpha} | d_{j.}^*, d_{j.}, d_{j.} \right) p(d_{j.} | d_{j.}) dd_{j.} \\ &= \int \Pr \left(\frac{\sqrt{d_{j.}^*} (\frac{d_{j.}}{d_{j.}^*} - w_j)}{\sqrt{w_j(1-w_j)}} \in \Phi^{-1}(A_{\alpha}) | d_{j.}^*, d_{j.}, d_{j.} \right) p(d_{j.} | d_{j.}) dd_{j.}. \end{aligned} \quad (4.4)$$

By CLT

$$\frac{\sqrt{d_{j.}}(\frac{d_{j.}}{d_{j.}} - \tilde{w}_j)}{\sqrt{\tilde{w}_j(1 - \tilde{w}_j)}} \Big|_{d_{j.}} \xrightarrow{d} N(0, 1).$$

Therefore, (4.4) becomes

$$\Phi \left(\sqrt{\frac{d_{j.}}{\tilde{w}_j(1 - \tilde{w}_j)}} |w_j - \tilde{w}_j| + \sqrt{\frac{d_{j.} w_j(1 - w_j)}{d_{j.}^* \tilde{w}_j(1 - \tilde{w}_j)}} z_\alpha + O_p(1) \right).$$

Note that the $O_p(1)$ term in the above equation is equal to $N(0, 1) + o_p(1)$ and hence is of the same order as the second term if $d_{j.}/d_{j.}^* = O(1)$. Therefore, for large $d_{j.}$, the above quantity is predominated by the leading term:

$$\frac{\sqrt{d_{j.}} |\tilde{w}_j - w_j|}{\sqrt{\tilde{w}_j(1 - \tilde{w}_j)}} = \begin{cases} \sqrt{d_{j.}} \sqrt{\frac{w_j}{1 - \tilde{w}_j}} \frac{|1 - \theta_1 \tilde{\theta}^{-1}|}{\sqrt{\theta_1 \tilde{\theta}^{-1}}} & \text{for } x_j \in (x_0, x_{\nu_y}], \\ \sqrt{d_{j.}} \sqrt{\frac{1 - w_j}{\tilde{w}_j}} \sqrt{\frac{\tilde{\theta}}{\theta_2}} & \text{for } x_j \in (x_{\nu_y}, x_J]. \end{cases}$$

Result (iii) follows, since the above quantity decreases with $|x_j - x_{\nu_y}|$ as both w_j and $\tilde{w}_j$ increase with j .

Theorem 4.2 Let x_{ν_c} denote the true change-point of $\mathbf{C}$ in a CCRM. Suppose that $\mathbf{C}$ is fitted by a CCRM assuming that $x_{\nu_c} \approx x_J$. Then

- i. (Meng 1994) $E(\text{PPP}^c(j, k)) = 0.5$ for $1 \leq j \leq J$ as $n \rightarrow \infty$ if $x_{\nu_c} \approx x_J$. If $x_{\nu_c} < x_J$, then
- ii. $|E(\text{PPP}^c(j, k)) - 0.5|$ is asymptotically monotone decreasing with $|x_j - x_{\nu_c}|$, and
- iii. For $0 < \alpha < 0.5$ and $|x_j - x_{\nu_y}| < |x_{j'} - x_{\nu_c}|$,

$$\lim_{n \rightarrow \infty} \Pr(\text{PPP}^c(j, k) \in A_\alpha) \geq \lim_{n \rightarrow \infty} \Pr(\text{PPP}^c(j', k) \in A_\alpha) \quad (4.5)$$

when $\rho_{1k} < \rho_{2k}$ and $A_\alpha = (0, \alpha]$, or $\rho_{1k} > \rho_{2k}$ and $A_\alpha = [1 - \alpha, 1)$.

Proof Following the argument used in the proof of Theorem 4.1, we rewrite $\text{PPP}^c(j; k)$ as

$$\int \Pr \left(\frac{\sqrt{d_{j.}}(\frac{d_{jk}^*}{d_{j.}} - \psi)}{\sqrt{\psi(1 - \psi)}} \leq \frac{\sqrt{d_{j.}}(\frac{d_{jk}^*}{d_{j.}} - \psi)}{\sqrt{\psi(1 - \psi)}} \Big| \psi, d_{jk}, \mathbf{Y} \right) \pi_{\rho, k}^c(\psi) d\psi.$$

Applying the CLT to $d_{jk}^*|_{d_{j.}}$, one obtains

$$\text{PPP}^c(j; k) = \int \Phi \left(\frac{\sqrt{d_{j.}}(\frac{d_{jk}^*}{d_{j.}} - \psi)}{\sqrt{\psi(1 - \psi)}} + o_p(1) \Big| \psi, d_{jk}, \mathbf{Y} \right) \pi_{\rho, k}^c(\psi) d\psi.$$

Recall the definition of ρ_1 and ρ_2 in Sect. 2. Also, let

$$\kappa_j = \begin{cases} 1 & \text{for } x_j \in [x_1, x_{\nu_c}], \\ d_{\nu_c}/d_j & \text{for } x_j \in (x_{\nu_c}, x_J], \end{cases} \quad \hat{\rho}_k = \kappa_J \rho_{1k} + (1 - \kappa_J) \rho_{2k}, \quad \tilde{\rho}_{jk} = \kappa_j \rho_{1k} + (1 - \kappa_j) \rho_{2k}.$$

Then according to Bayesian version CLT (see (Berger 1985, pp. 224–225)),

$$\text{PPP}^c(j; k) = \Phi \left(\sqrt{d_j} \cdot \left\{ \frac{(\frac{d_{jk}}{d_j} - \hat{\rho}_k)}{\sqrt{\hat{\rho}_k(1 - \hat{\rho}_k)}} + O_p(1/d_J) \right\} + o_p(1) \right) = \Phi \left(\sqrt{d_j} \cdot \left\{ \frac{(\frac{d_{jk}}{d_j} - \hat{\rho}_k)}{\sqrt{\hat{\rho}_k(1 - \hat{\rho}_k)}} \right\} + o_p(1) \right).$$

Since under $x_{\nu_c} < x_J$, $d_{jk}|d_j \sim \text{Bin}(d_j, \tilde{\rho}_{jk})$, CLT implies

$$\frac{d_{jk}}{d_j} = \tilde{\rho}_{jk} + O_p(1/\sqrt{d_j}).$$

Thus,

$$\text{PPP}^c(j; k) = \Phi \left(\sqrt{d_j} \cdot \left\{ \frac{(\tilde{\rho}_{jk} - \hat{\rho}_k)}{\sqrt{\hat{\rho}_k(1 - \hat{\rho}_k)}} \right\} + O_p(1) \right).$$

Note that the second term $O_p(1)$ above is more precisely $N(0, 1) + o_p(1)$, which is negligible compared with the first term when d_j is large. Hence, result (ii) follows from Lemma 4.1 and from

$$\sqrt{d_j} \cdot \frac{|\hat{\rho}_k - \tilde{\rho}_{jk}|}{|\rho_{1k} - \rho_{2k}|} = \begin{cases} \sqrt{d_j} \cdot (1 - \kappa_J) & \text{if } x_j \in (x_0, x_{\nu_c}], \\ d_{\nu_c} \cdot \left\{ \frac{1}{\sqrt{d_j}} - \frac{\sqrt{d_j}}{d_J} \right\} & \text{if } x_j \in (x_{\nu_c}, x_J], \end{cases}$$

which is monotone decreasing with $|x_j - x_{\nu_c}|$.

As in the proof for $\text{PPP}^v(j)$, to prove (iii), we evaluate

$$\begin{aligned}
& \int \Pr \left(\Phi \left(\frac{\sqrt{d_{j.}}(d_{jk} - \psi)}{\sqrt{\psi(1-\psi)}} \mid \psi, d_{jk}, \mathbf{Y} \right) \in A_\alpha \right) \pi_{\rho,k}^c(\psi) d\psi \\
&= \int \Pr \left(\frac{\sqrt{d_{j.}}(d_{jk} - \psi)}{\sqrt{\psi(1-\psi)}} \in \Phi^{-1}(A_\alpha) \mid \psi, d_{jk}, \mathbf{Y} \right) \pi_{\rho,k}^c(\psi) d\psi = \int \Phi \left(\frac{\sqrt{d_{j.}}(\psi - \tilde{\rho}_{jk})}{\sqrt{\tilde{\rho}_{jk}(1-\tilde{\rho}_{jk})}} \right. \\
&\quad \left. + z_\alpha \sqrt{\frac{\psi(1-\psi)}{\tilde{\rho}_{jk}(1-\tilde{\rho}_{jk})}} + o_p(1) \mid \psi, \mathbf{Y}, \cdot \right) \pi_{\rho,k}^c(\psi) d\psi \\
&= \Phi \left(\frac{\sqrt{d_{j.}}|\hat{\rho}_k - \tilde{\rho}_{jk}|}{\sqrt{\tilde{\rho}_{jk}(1-\tilde{\rho}_{jk})}} + \sqrt{\frac{\hat{\rho}_k(1-\hat{\rho}_k)}{\tilde{\rho}_{jk}(1-\tilde{\rho}_{jk})}} z_\alpha + O_p(1) \right).
\end{aligned}$$

We note that $O_p(1)$ in the above equation is indeed $N(0, 1) + o_p(1)$, which together with the second term of order $O(1)$ is negligible compared with the first term for large enough $d_{j.}$.

Thus,

$$\sqrt{\frac{d_{j.}}{\tilde{\rho}_{jk}(1-\tilde{\rho}_{jk})}} \frac{|\hat{\rho}_k - \tilde{\rho}_{jk}|}{|\rho_{1k} - \rho_{2k}|} = \begin{cases} \frac{\sqrt{d_{j.}}(1-\kappa_{j.})}{\sqrt{\rho_{1k}(1-\rho_{1k})}} \text{ for } x_j \in (x_0, x_{\nu_c}], (d_{j.} - d_{j.})\kappa_{j.} \\ \sqrt{d_{\nu_c.}\rho_{1k} + (d_{j.} - d_{\nu_c.})\rho_{2k}} \sqrt{d_{\nu_c.}(1-\rho_{2k}) + (d_{j.} - d_{\nu_c.})(1-\rho_{1k})} \text{ for } x_j \in (x_{\nu_c}, x_J] \end{cases}$$

completes the proof, since d_j is monotone nondecreasing in j .

We note that if α is the “nominal” tolerance for type I error and A_α is the corresponding critical region of rejecting $x_{\nu_c} < x_J$, then $\Pr(\text{PPP}^T(\cdot) \in A_\alpha)$ evaluated under $x_{\nu_c} < x_J$ would be the power for detecting parameter difference between $(x_0, x_{\nu_c}]$ and $(x_{\nu_c}, x_J]$. Therefore, the results imply that when $x_{\nu_c} < x_J$, and yet the model assumes $x_{\nu_c} \sim x_J$, the asymptotic power for detecting change-point misspecification is maximized at x_{ν_c} for any $0 < \alpha < 0.5$. By the same token, x_j is expected to be the missed change-point associated with C if $\text{PPP}^c(j, k)$ achieves the greatest discriminatory power for some $k \in \{1, \dots, K\}$.

4.2 The PPPP approach

The cumulative-sum statistics considered in Sect. 4.1 are further examined by (3.4). As we show below, PPPP exhibits similar departure pattern as PPP in the sense of (3.6) and (3.7).

Following Bayarri and Berger’s (2000) prescription of PPPP, we compute

$$\text{PPPP}^y(j) \equiv \text{PPPP}(T_j; \hat{\theta}_{\text{ss}} \setminus d_{j.}; \theta) = \int \Pr \left(\frac{d_{j.}^*}{d_{j.}^*} \leq \frac{d_{j.}}{d_{j.}} \mid \theta, \frac{d_{j.}}{d_{j.}} \right) \pi_\theta^{y \setminus d_{j.}}(\psi) d\theta \quad (4.6)$$

and

$$\text{PPPP}^c(j, k) \equiv \text{PPPP}(T_j; (\hat{\rho}_{\text{SS}} \setminus d_{jk}, \mathbf{Y}); \rho) = \int \Pr(d_{jk}^* \leq d_{jk} | \rho, d_{jk}, \mathbf{Y}) \pi_{\rho, k}^{c \setminus d_{jk}}(\psi) d\psi, \quad (4.7)$$

where $\pi_{\theta}^{y \setminus d_j}(\theta) \equiv \pi(\theta | \hat{\theta}_{\text{SS}} \setminus d_j)$ and $\pi_{\rho, k}^{c \setminus d_{jk}}(\rho_k) \equiv \pi(\rho_k | \hat{\rho}_{\text{SS}} \setminus d_{jk}, \mathbf{Y})$.

According to (Robins et al. 2000), the asymptotic result (i) of Theorems 4.1 and 4.2 hold true for (4.6) and (4.7), respectively. Since under the null model, d_{jk}^* given d_j^* is independent of model parameter, substituting π_{θ}^y by $\pi_{\theta}^{y \setminus d_j}$ in the proof of Theorem 4.1 proves that the asymptotic results (ii) and (iii) in Theorem 4.1 also hold true for (4.6). Results (ii) and (iii) of Theorem 4.2 hold true for (4.7), and a proof is given below.

Proof Recall that, for each $1 \leq k \leq K$, $\{T_j^* = d_{jk}^* : 1 \leq j \leq J\}$ and $\pi_{\rho, k}^{c \setminus d_{jk}}$ as given above. According to the derivation established in the proof of Theorem 4.1, one only needs to

compute $\sqrt{d_j} |\hat{\rho}_{k \setminus j} - \tilde{\rho}_{jk}| / \sqrt{\hat{\rho}_{k \setminus j}(1 - \hat{\rho}_{k \setminus j})}$, where

$$\hat{\rho}_{k \setminus j} = \begin{cases} \frac{d_{\nu_c} - d_j}{d_j - d_j} \rho_{1k} + \frac{d_j - d_{\nu_c}}{d_j - d_j} \rho_{2k} & \text{for } x_j \in (x_0, x_{\nu_c}], \\ \rho_{2k} & \text{for } x_j \in (x_{\nu_c}, x_J] \end{cases}$$

is the asymptotic mean of $\rho_{1k} | \mathbf{h} \setminus d_{jk}$. Then

$$\frac{\sqrt{d_j} |\hat{\rho}_{k \setminus j} - \tilde{\rho}_{jk}|}{\sqrt{\hat{\rho}_{k \setminus j}(1 - \hat{\rho}_{k \setminus j})}} = \begin{cases} \frac{\sqrt{d_j} |\rho_{2k} - \rho_{1k}| (d_j - d_{\nu_c})}{\sqrt{\{(d_{\nu_c} - d_j) \rho_{1k} + (d_j - d_{\nu_c}) \rho_{2k}\} \{d_j - d_{\nu_c} \rho_{1k} - d_j(1 - \rho_{1k}) - (d_j - d_{\nu_c}) \rho_{2k}\}}} & \text{for } x_j \in (x_0, x_{\nu_c}], \\ \frac{d_{\nu_c} |\rho_{2k} - \rho_{1k}|}{\sqrt{d_j \rho_{2k}(1 - \rho_{2k})}} & \text{for } x_j \in (x_{\nu_c}, x_J] \end{cases}$$

implies that $|E(\text{PPPP}^c(j; k)) - 0.5|$ decreases with $|x_j - x_{\nu_c}|$, which concludes result (ii).

Similarly, to prove (iii) for $\text{PPPP}^c(j; k)$, it is sufficient to show that

$\sqrt{\xi_j} \times |\hat{\rho}_{k \setminus j} - \tilde{\rho}_{jk}| / \sqrt{\tilde{\rho}_{jk}(1 - \tilde{\rho}_{jk})}$ decreases with $|x_j - x_{\nu_c}|$. This is shown in the calculation below.

For $x_j \in (x_0, x_{\nu_c}]$,

$$\frac{\sqrt{d_j} |\hat{\rho}_{k \setminus j} - \tilde{\rho}_{jk}|}{\sqrt{\tilde{\rho}_{jk}(1 - \tilde{\rho}_{jk})}} = \frac{\sqrt{d_j} (d_j - d_{\nu_c}) |\rho_{2k} - \rho_{1k}|}{(d_j - d_j) \sqrt{\rho_{1k}(1 - \rho_{1k})}},$$

which is nondecreasing with j .

For $x_j \in (x_{\nu_c}, x_J]$, write

$$\frac{\sqrt{d_j}|\hat{\rho}_{k\setminus j} - \tilde{\rho}_{jk}|}{\sqrt{\tilde{\rho}_{jk}(1 - \tilde{\rho}_{jk})}} = \begin{cases} \frac{d_{\nu_c}|\rho_{2k} - \rho_{1k}|}{\sqrt{\rho_{2k} - \frac{d_{\nu_c}}{d_j}(\rho_{2k} - \rho_{1k})} \sqrt{d_j(1 - \rho_{2k}) + d_{\nu_c}(\rho_{2k} - \rho_{1k})}} & \text{if } \rho_{1k} \leq \rho_{2k}, \\ \frac{d_{\nu_c}|\rho_{2k} - \rho_{1k}|}{\sqrt{(d_j - d_{\nu_c})\rho_{2k} + d_{\nu_c}\rho_{1k}} \sqrt{1 - \rho_{2k} - \frac{d_{\nu_c}}{d_j}(\rho_{1k} - \rho_{2k})}} & \text{if } \rho_{1k} \geq \rho_{2k}, \end{cases}$$

which is nonincreasing with j .

Therefore, the diagnostics based on PPPP agree with those of PPP to the extent that both their means converge to 0.5 when $x_v = x_J$, and the magnitudes of their departure from 0.5 decrease with $|x_j - x_v|$ when $x_v < x_J$. Consequently, the graphical diagnostic technique as described in Sect. 4.1 can be used here to examine the distributions of (4.6) and (4.7), and thus to find the true change-points.

For finite sample situations, diagnostics derived from PPPP indeed differ from those due to PPP in terms of their absolute departure from 0.5 as well as the increment of their departure between consecutive time points. As we show in Sect. 5.2, compared with PPP, PPPP often has a greater absolute departure from 0.5 and thus is more powerful for revealing the magnitude of the departure. In terms of the power of detecting change-point misspecification, the increment of the departure is another useful index; greater the increment of the departure greater the power for detecting change-points. In that sense, PPP can be more powerful than PPPP, since the increments of adjacent PPP departure around certain change-point neighborhood exceed those of PPPP (see simulation results in Sect. 5.2).

4.3 Extension to CCRMs with multiple change-points

The technique described in Sects. 4.1, 4.2 can be modified easily to examine CCRMs with multiple change-points. We consider a forward model selection procedure beginning with PPP (or PPPP) diagnostics for a model with one change-point, say x_v (since the model involves more than one change-point). From PPP (or PPPP) diagnostics of this single change-point model, one should identify potential change-point(s) falling into at least one of $[x_0, x_v]$ or $[x_v, x_J]$. If the fit is inadequate in only one of the two intervals, add a new change-point in that interval based on the departure pattern and the results derived in Sects. 4.1, 4.2. While if the fit is inadequate in both intervals, then identify a new change-point in each interval to replace x_v . It is followed by refitting the data using the model with the new set of revised change-points, and evaluation of the model fit. This iteration of model refinement and model validation continues until adequate fit is achieved. During this process, we consider increasing the number of change-points from u to $(u + 1)$, unless no fit is adequate after exhausting all the models with u change-points or less.

Note that since the cumulative-sum diagnostic statistics within each pair of adjacent change-points assumed are formed separately and thus memoryless, the departure due to misfit in one such interval will not affect the mean departure in other intervals. That is, to verify whether there exists a change-point x_{v_y} associated with $\mathbf{Y}$ falling into $(x_{\tilde{v}_y}^\ell, x_{\tilde{v}_y}^{\ell+1}]$, we

evaluate PPP and PPPP for $\mathbf{T}^\ell \equiv \left(\frac{d_{(\tilde{\nu}_y^\ell+1)} - d_{\tilde{\nu}_y^\ell}}{d_{\tilde{\nu}_y^{\ell+1}} - d_{\tilde{\nu}_y^\ell}}, \dots, \frac{d_{(\tilde{\nu}_y^{\ell+1}-1)} - d_{\tilde{\nu}_y^\ell}}{d_{\tilde{\nu}_y^{\ell+1}} - d_{\tilde{\nu}_y^\ell}} \right)'$ and

$U = \hat{\theta}_{SS} = d_{\tilde{\nu}_y^{\ell+1}} - d_{\tilde{\nu}_y^\ell}$, where the j th component of $\mathbf{T}^\ell$ corresponds to the observed failure proportion accrued over $(x_{\tilde{\nu}_y^\ell}, x_{\tilde{\nu}_y^\ell+j}]$. Similarly, to verify whether there is a change-point x_{ν_c} associated with $\mathbf{C}$ falling into $(x_{\tilde{\nu}_c^\ell}, x_{\tilde{\nu}_c^{\ell+1}}]$, we consider

$\mathbf{T}_k^\ell \equiv (d_{(\tilde{\nu}_c^{\ell+1})k} - d_{\tilde{\nu}_c^\ell k}, \dots, d_{(\tilde{\nu}_c^{\ell+1}-1)k} - d_{\tilde{\nu}_c^\ell k})'$ for $1 \leq k \leq K$ and $\mathbf{U} = (\hat{\rho}_{SS}, \mathbf{y})$, where the j th component of $\mathbf{T}_k^\ell$ is the observed failure proportion due to cause k during $(x_{\tilde{\nu}_c^\ell}, x_{\tilde{\nu}_c^\ell+j}]$ and $\hat{\rho}_{SS} = (d_{\tilde{\nu}_c^{\ell+1}k} - d_{\tilde{\nu}_c^\ell k}, d_{\tilde{\nu}_c^{\ell+1}} - d_{\tilde{\nu}_c^\ell})$.

5 Numerical analysis

To justify the results derived in Sect. 4, we examine the distributions of (4.1), (4.2), (4.6), and (4.7), using two simulated datasets arising from CCRMs as described in Sect. 5.2. Detailed Bayesian computation in these CCRM analyses are given in Sect. 5.1.

5.1 Bayesian computation

The derivation below is based on (2.8) while its extension to general CCRM is straightforward. For the ease of exposition, we consider priors that arise from the Dirichlet family of which diffuseness is appropriated by tuning the hyper shape parameters. Note that under the usual regularity conditions, including model identifiability and posterior propriety, our conclusions about asymptotic posterior predictive diagnostics derived in Sect. 4 can however be extended to other priors that meet the noninformative criterion required in Bernstein–von Mises theorem.

Let the Dirichlet form prior for (2.8) be posited as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\rho}) = \pi(\boldsymbol{\theta}^\dagger, \boldsymbol{\rho}) \propto \left\{ \prod_{u=1}^2 (\theta_u^\dagger)^{\alpha_u-1} \right\} (1 - \theta_1^\dagger - \theta_2^\dagger)^{\alpha_0-1} \left\{ \prod_{u=1}^2 \times \prod_{k=1}^K (\rho_{uk})^{\gamma_{uk}-1} \right\} \quad (5.1)$$

for $\{0 < \alpha_u < \varepsilon : 0 \leq u \leq 2\}$ and $\{0 < \gamma_{uk} < \varepsilon : 1 \leq u \leq 2, 1 \leq k \leq K\}$, where $\boldsymbol{\theta}^\dagger = (\theta_1^\dagger, \theta_2^\dagger)$ with $\theta_1^\dagger = \tilde{\nu}_y \theta_1$ and $\theta_2^\dagger = (\mathbf{J} - \tilde{\nu}_y) \theta_2$, and a suitable positive ε is specified in (5.1) to control the degree of (un)certainty about $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$.

The posterior used for calculating PPP thus is

$$\pi(\boldsymbol{\theta}^\dagger, \boldsymbol{\rho} | \hat{\theta}_{SS}, \hat{\rho}_{SS}) = \pi_{\boldsymbol{\theta}^\dagger}^y(\boldsymbol{\theta}^\dagger | \hat{\theta}_{SS}) \pi_{\boldsymbol{\rho}}^c(\boldsymbol{\rho} | \hat{\rho}_{SS}), \quad (5.2)$$

where

$$\pi_{\boldsymbol{\theta}^\dagger}^y(\boldsymbol{\theta}^\dagger) \equiv \pi^y(\boldsymbol{\theta}^\dagger | \hat{\theta}_{SS}) \propto (\theta_1^\dagger)^{(d_{\tilde{\nu}_y} + \alpha_1 - 1)} (\theta_2^\dagger)^{(d_{\mathbf{J}} - d_{\tilde{\nu}_y} + \alpha_2 - 1)} (1 - \theta_1^\dagger - \theta_2^\dagger)^{n - d_{\mathbf{J}} + \alpha_0 - 1}$$

and

$$\pi_{\rho}^c(\boldsymbol{\rho}) \equiv \pi^c(\boldsymbol{\rho}|\hat{\rho}_{SS}) \propto \prod_{k=1}^K \left\{ \rho_{1k}^{(d_{\tilde{v}_{ck}} + \gamma_{1k} - 1)} \rho_{2k}^{(d_{Jk} - d_{\tilde{v}_{ck}} + \gamma_{2k} - 1)} \right\}.$$

Since the cumulative-sum statistics for diagnostics are $\{d_j : j < \tilde{v}_y\}$, $\{(d_j - d_{\tilde{v}_y}) : j > \tilde{v}_y\}$, $\{d_{jk} : j < \tilde{v}_c, k = 1, \dots, K\}$, and $\{(d_{jk} - d_{\tilde{v}_{ck}}) : j > \tilde{v}_c, k = 1, \dots, K\}$, therefore all the posteriors for their associated PPP calculation follow Dirichlet distributions.

To obtain the posteriors for deriving $PPPP^y(j)$ and $PPPP^c(j, k)$, we first obtain the following conditional likelihoods:

$$\begin{aligned} p(\mathbf{Y}|d_j, \boldsymbol{\theta}) &\propto \left[\frac{\{(\tilde{\nu}_y - j)\theta_1\}^{(d_{\tilde{v}_y} - d_j)} \{(J - \tilde{\nu}_y)\theta_2\}^{(d_J - d_{\tilde{v}_y})}}{\{1 - j\theta_1\}^{(n - d_j)}} \times \{1 - \tilde{\nu}_y\theta_1 - (J - \tilde{\nu}_y)\theta_2\}^{(n - d_J)} \right]^{1_{[j < \tilde{v}_y]}} \\ &\times \left[\frac{\{\tilde{\nu}_y\theta_1\}^{d_{\tilde{v}_y}} \{(j - \tilde{\nu}_y)\theta_2\}^{(d_j - d_{\tilde{v}_y})}}{\{\tilde{\nu}_y\theta_1 + (j - \tilde{\nu}_y)\theta_2\}^{d_j}} \right]^{1_{[j \geq \tilde{v}_y]}} \\ &\times \left[\frac{\{(J - j)\theta_2\}^{(d_J - d_j)} \{1 - \tilde{\nu}_y\theta_1 - (J - \tilde{\nu}_y)\theta_2\}^{(n - d_J)}}{\{1 - \tilde{\nu}_y\theta_1 - (j - \nu_y)\theta_2\}^{(n - d_j)}} \right]^{1_{[j \geq \tilde{v}_y]}} \end{aligned}$$

and

$$\begin{aligned} p(\mathbf{C}|d_{jk}, \mathbf{Y}, \boldsymbol{\rho}) &\propto \left[\{\rho_{1k}\}^{(d_{\tilde{v}_{ck}} - d_{jk})} \{1 - \rho_{1k}\}^{(d_{\tilde{v}_c} - d_j) - (d_{\tilde{v}_{ck}} - d_{jk})} \right]^{1_{[j < \tilde{v}_c]}} \\ &\times \left[\rho_{2k}^{(d_{Jk} - d_{\tilde{v}_{ck}})} (1 - \rho_{2k})^{(d_J - d_{\tilde{v}_c}) - (d_{Jk} - d_{\tilde{v}_{ck}})} \right]^{1_{[j < \tilde{v}_c]}} \\ &\times \left[\{\rho_{2k}\}^{(d_{Jk} - d_{jk})} \{1 - \rho_{2k}\}^{(d_J - d_j) - (d_{Jk} - d_{jk})} \right]^{1_{[j \geq \tilde{v}_c]}}. \end{aligned}$$

Since, under (5.1)

$$\pi_{\theta}^{y \setminus d_j}(\boldsymbol{\theta}|\hat{\theta}_{SS} \setminus d_j) \propto p(\mathbf{Y}|d_j, \boldsymbol{\theta}) \left\{ \prod_{u=1}^2 (\theta_u^\dagger)^{\alpha_u - 1} \right\} (1 - \theta_1^\dagger - \theta_2^\dagger)^{\alpha_0 - 1}$$

does not yield a standard form posterior for computing $PPPP^y(j)$, the posteriors of $\boldsymbol{\theta}$ are obtained using Metropolis–Gibbs sampling. The posteriors of ρ_{1k} and ρ_{2k} used for calculating $PPPP^c(j, k)$ have explicit beta distributions:

$$\pi_{\rho}^{c \setminus d_{jk}}(\rho_{1k}, \rho_{2k}|\hat{\rho}_{SS} \setminus d_{jk}, d_j) \propto \left\{ \rho_{1k}^{(d_{\tilde{v}_{ck}} - d_{jk} + \gamma_{1k} - 1)} (1 - \rho_{1k})^{(d_{\tilde{v}_c} - d_j) - (d_{\tilde{v}_{ck}} - d_{jk})} \right\}$$

for $j < \tilde{v}_c$, and

$$\pi_{\rho}^{c \setminus d_{jk}}(\rho_{2k}|\hat{\rho}_{SS} \setminus d_{jk}, d_j) \propto \left\{ \rho_{2k}^{(d_{Jk} - d_{jk} + \gamma_{2k} - 1)} (1 - \rho_{2k})^{(d_J - d_j) - (d_{Jk} - d_{jk})} \right\}$$

for $j < \tilde{v}_c$

As our diagnostic procedure involves examining the distributional property of (4.1), (4.2), (4.6), and (4.7), we employ the bootstrap technique to obtain their empirical distributions by repeating the steps below.

- a. Generate $\mathbf{h}$ based on the empirical distribution of the observed data.
- b. For each $\mathbf{h}$ generated in (a), draw posterior samples of $\boldsymbol{\psi}$ given $\mathbf{U}$ based on the assumed model, where $\mathbf{U} = \boldsymbol{\psi}_{SS}$ for PPP and $\mathbf{U} = \mathbf{h}T$ for PPPP.
- c. Draw predicted samples $\mathbf{T}^*$ based on the posterior of $\boldsymbol{\psi}$ obtained in (b).
- d. Obtain PPP (and PPPP) estimates by comparing the empirical distribution of $\mathbf{T}^*$ with respect to the realization of $\mathbf{T}$. That is, for each T_j , one computes the proportion of posterior predictive samples of T_j that are no larger than T_j .
- e. Suppose that $\text{PPP}_r(j)$ and $\text{PPPP}_r(j)$ denote PPP and PPPP of T_j derived from the r th bootstrap sample ($r = 1, \dots, R$), respectively. Then, for a given $0 < q < 1$, calculate the q th quantile for PPP of each T_j , using $\inf\{\text{PPP}_r(j): (\sum_{r=1}^R \mathbf{1}_{[\text{PPP}_r(j) < q]})/R \geq q\}$ and $\inf\{\text{PPPP}_r(j): (\sum_{r=1}^R \mathbf{1}_{[\text{PPPP}_r(j) < q]})/R \geq q\}$ for PPPP.

In our examples, each bootstrap sample of $\mathbf{Y}$, say

$(d_{1.}^{\text{boot}}, d_{2.}^{\text{boot}} - d_{1.}^{\text{boot}}, \dots, d_{J.}^{\text{boot}} - d_{(J-1).}^{\text{boot}}, n - d_{J.}^{\text{boot}})$, is drawn from the multinomial distribution of size n with probabilities

$\left(\frac{d_{1.} + .5}{n + .5(J+1)}, \frac{d_{2.} - d_{1.} + .5}{n + .5(J+1)}, \dots, \frac{d_{J.} - d_{(J-1).} + .5}{n + .5(J+1)}, \frac{n - d_{J.} + .5}{n + .5(J+1)} \right)$, and a bootstrap sample of

$\mathbf{C}$ during $(x_{j-1}, x_j]$ is drawn from the multinomial distribution of size $(d_{j.}^{\text{boot}} - d_{(j-1).}^{\text{boot}})$ with

probabilities $\left(\frac{d_{j1} - d_{(j-1)1} + .5}{d_{j.} - d_{(j-1).} + .5K}, \dots, \frac{d_{jK} - d_{(j-1)K} + .5}{d_{j.} - d_{(j-1).} + .5K} \right)$.

5.2 Simulation studies

5.2.1 The simulated data—Our simulation study considers CCRMs involving both single change-point and two change-points. Specifications of these CCRM's are given in Tables 1 and 2, respectively. In the first example, the data are drawn from a CCRM with $v_y = 4$ and $v_c = 4$. Using this example, we show how the methods described in Sects. 4.1–4.2 help locate the true change-points when the data are fit by the incorrect models which assume $\tilde{v}_y = v_y$ and $\tilde{v}_c = v_c$. The CCRM assumed for the second example allows two change-points for both $\mathbf{Y}$ and $\mathbf{C}$: $v_y = 3, 5$ and $v_c = 3, 5$. This example demonstrates how to obtain correct change-points from two incorrectly specified CCRM's using the diagnostic procedure as described in Sect. 4.3: the initial model assumes no change-point (i.e., $\tilde{v}_y = 7$ and $\tilde{v}_c = 7$), and the one change-point model revised from the initial model based on PPP and PPPP diagnostics considers $\tilde{v}_y = 4$ and $\tilde{v}_c = 4$.

5.2.2 Simulation results—We first summarize the results based on the PPP approach. We then compare PPP and PPPP in terms of their statistical power of informing change-point misspecification.

Result (i) of Theorems 4.1 and 4.2 is justified by Figs. 1 and 2, respectively, which show no systematic departure of PPP when all change-points are correctly specified in CCRM. Results (ii) and (iii) of our theorems are justified by Figs. 3 and 4, which show that each missed change-point is effectively distinguished from its neighbor regular points; the PPP attains the greatest departure from 0.5 at the change-point, while such departure decreases monotone as the index of PPP moves farther from the index of the change-point.

To demonstrate the forward model selection procedure as described in Sect. 4.3, we examine CCRM's that are fit to data arising from Model 2 which contains two change-points for both $\mathbf{Y}$ and $\mathbf{C}$. We begin with the model with no change-point and obtain its PPP diagnostics for $\mathbf{Y}$. As the result (the left-hand column of Fig. 5) suggests that both x_3 and x_4 can be potential change-points, we refine the model by first adding x_4 as the single change-point for $\mathbf{Y}$ and re-examine the fit. Then according to the diagnostics shown in the right-hand column of Fig. 5, we replace x_4 by x_3 and x_5 as the new change-points added within $[x_0, x_4)$ and $[x_5, x_7)$, respectively. Consequently, we conclude x_3 and x_5 as the change-points of $\mathbf{Y}$, since the diagnostic, as shown in the right-hand column of Fig. 1, suggests adequate fit. Instead of x_4 , we have also considered adding x_3 as the first change-point. When examining this model using the proposed PPP diagnostics, we find that the fit is adequate for data falling in $[x_0, x_3)$, while x_5 should be added as a new change-point in $[x_4, x_7)$. Similarly, to obtain the change-points associated with $\mathbf{C}$, we first examine the assumption of $v_c = 7$ (see Fig. 6) and then consider $v_c = 3$ and $v_c = 4$. Finally, the diagnostics shown in Fig. 7 find x_3 and x_5 to be the change-points for $\mathbf{C}$.

The PPPP's counterparts associated with the PPP's shown in Figs. 4, 6, and 7 are presented in Figs. 8, 9, and 10, respectively. By contrasting PPP and PPPP using these figures, we find that both show similar departure patterns, but neither yields an admissible diagnostic for change-point mis-specification. More specifically, compared with PPP's, the means of PPPP's often exhibit greater departure from 0.5. However, since the increments of the departure associated with PPPP's are not necessarily greater than those associated with PPP's (e.g., the increment between $\text{PPPP}^c(4, 2)$ and $\text{PPPP}^c(5, 2)$ for $n = 500$ and $n = 1000$ in Fig. 8, the increment between $\text{PPPP}^c(4, 3)$ and $\text{PPPP}^c(5, 3)$ for $n = 500$ in Fig. 8, and that between $\text{PPPP}^c(5, 1)$ and $\text{PPPP}^c(6, 1)$ for $n = 500$ and $n = 1000$ in Fig. 10), the diagnostic power between PPP and PPPP remains inconclusive for the finite sample situation. Note that no comparison between $\text{PPP}^y(j)$'s and $\text{PPPP}^y(j)$'s is needed when $d_{J-} = n$ and $\tilde{v}_y = J$, since $d_{j-}^* \sim \text{Bin}(n, j/J)$, which does not depend on θ .

6 Conclusion

This paper presents a comparison between PPP and PPPP regarding their performance in detecting change-point mis-specification in the competing risks analysis framework. For the CCRM's considered in Sect. 2, we conclude that, asymptotically, diagnostics due to PPP is as effective as those due to PPPP, yet PPP is more computationally economic than PPPP. In

our simulations for finite sample situations, the power of the two diagnostics varies across different scenarios.

It is worth of noticing that both PPP and PPPP are consistent with those likelihood-ratio based tests (Page 1955; Hinkley and Hinkley 1970) to the extent that the maximum discriminatory power is expected at the change-point that is missed. In terms of the realization of this PPP-based (or PPPP-based) diagnostics, we notice from the simulations shown in Sect. 5.2 that all the PPP (and PPPP) departure increases with the sample size n (comparing plots in the panel for smaller n with those in the panel for larger n). Indeed, as a result of the large sample property, all the departure eventually converges to 0.5 with probability 1 (or the power of detecting absolute departure converges to 1) as the sample size exceeds certain large value, and therefore the missed change-points can hardly be discerned by the graphical presentation of PPP's (or PPPP's). However, since the rate of convergence varies across PPP's (or PPPP's) associated with different cumulative statistics, further assessment of this convergence rate would likely aid our ability to detect change-points.

In our application of (3.4) to CCRM where the diagnostic statistics $\mathbf{T}$ is multivariate, PPPP of each component of $\mathbf{T}$ is calculated separately. That is, our calculation draws posterior predictive samples of T_j from $\int p(T_j | \psi) \pi(\psi | \mathbf{h} \backslash T_j) d\psi$. As an alternative approach, one may treat $\mathbf{T}$ as an entity and calculate (3.4) based on $\mathbf{T}$ and $\mathbf{U} = \psi_{SS} \backslash \mathbf{T}$. This latter approach in our case will however limit the use of (3.4) to only proper priors, since $\pi(\psi | \psi_{SS} \backslash \mathbf{T}) = \pi(\psi)$, and hence (3.4) is reduced to the prior predictive p -value.

Finally, although this paper confines its scope to a special subset of competing risks models, the technique described in Sect. 4 should in principle be generalized to models that involve covariates or a different survival distribution. For the former, (4.1), (4.2), (4.6), and (4.7) are obtained by first modeling $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ in terms of $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{X}\boldsymbol{\beta})$ and $\boldsymbol{\rho} = \boldsymbol{\rho}(\mathbf{X}\boldsymbol{\gamma})$, and then deriving posteriors of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, where $\mathbf{X}$ denotes the covariate matrix. For the latter, it remains future research to formalize the general elicitation of suitable $\mathbf{T}$.

Acknowledgments

An earlier version of this paper was presented at the Seventh World Meeting of International Society for Bayesian Analysis. Work on this paper was supported by facilities within the Veterans Evidence-based Research, Dissemination, and Implementation Center, the National Institute of Diabetes & Digestive & Kidney Diseases under grants No. K25 DK075092-01. The authors also wish to acknowledge helpful comments from the referees and the Editor.

References

- Barry D, Hartigan JA. A Bayesian analysis for change point problems. *J Am Stat Assoc.* 1993; 88:309–319.
- Bayarri MJ, Berger J. P-values for composite null models (with discussion). *J Am Stat Assoc.* 2000; 95:1127–1142.
- Berger, JO. Statistical decision theory and Bayesian analysis. New York: Springer; 1985.
- Chen, M-H.; Shao, Q-M.; Ibrahim, JG. Monte Carlo methods in Bayesian computation. New York: Springer; 2000.
- Chiang, CL. Introduction to stochastic processes in biostatistics. New York: Wiley; 1968.

- Denison, DGT.; Holmes, CC.; Mallick, BK.; Smith, AFM. Bayesian methods for nonlinear classification and regression. New York: Wiley; 2002.
- Gelman, A.; Carlin, J.; Stern, HS.; Rubin, D. Bayesian data analysis. London: Chapman & Hall; 1996a. Applied statistical science.
- Gelman A, Meng X-L, Stern HS. Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Stat Sin.* 1996b; 6:733–807.
- Glickman ME, Stern HS. A state-space model for national football league (NFL) scores. *J Am Stat Assoc.* 1998; 93:25–35.
- Guttman I. The use of the concept of a future observation in goodness-of-fit problems. *J Roy Stat Soc Ser B.* 1967; 29:83–100.
- Hinkley DV. Inference about the change-point from cumulative sum tests. *Biometrika.* 1971; 58:509–523.
- Hinkley DV, Hinkley EA. Inference about the change-point in a sequence of binomial variables. *Biometrika.* 1970; 57:477–488.
- James B, James KL, Siegmund D. Tests for a change-point. *Biometrika.* 1987; 74:71–83.
- Meng X-L. Posterior predictive p -values (in Bayesian inference and design). *Ann Stat.* 1994; 3:1142–1160.
- Page ES. A test for a change in a parameter occurring at an unknown point. *Biometrika.* 1955; 42:523–527.
- Pettitt AN. A simple cumulative sum type statistic for the change-point problem with zero-one observations. *Biometrika.* 1980; 67:79–84.
- Raftery AE, Akman VE. Bayesian analysis of a Poisson process with a change-point. *Biometrika.* 1986; 73:85–89.
- Robert, CP. The Bayesian choice: a decision-theoretic motivation. New York: Springer; 1994.
- Robins JM, van der Vaart A, Ventura V. The asymptotic distribution of p -values in composite null models. *J Am Stat Assoc.* 2000; 62:1143–1163.
- Rubin DB. Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann Stat.* 1984; 12:1151–1172.
- Smith AFM. A Bayesian approach to inference about a change-point in a sequence of random variables. *Biometrika.* 1975; 62:407–416.
- Stephens DA. Bayesian retrospective multiple-changepoint identification. *Appl Stat.* 1994; 43:159–178.
- Tsiatis A. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci.* 1975; 72:20–22. [PubMed: 1054494]

Fig. 1. Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with T under Model 1 (*left*) and Model 2 (*right*)

Fig. 2.
Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with C under Model 1 (row 1) and Model 2 (row 2)

Fig. 3. Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with T , where the data are generated from Model 1 ($x_v = 4$), but are fitted to two models that assume incorrect changepoints (column 1: $x_v = 7$; column 2: $x_v = 3$)

Fig. 4.

Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with C , where the data are generated from Model 1 ($x_v = 4$), but are fitted to an incorrect model that assumes $x_v = 7$

Fig. 5.

Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with T , where the data are generated from Model 1 ($x_v = 3.5$), but are fitted to two models that assume incorrect change-points (column 1: $x_v = 7$; column 2: $x_v = 4$)

Fig. 6.

Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with C , where the data are generated from Model 2 ($x_v = 3, 5$), but are fitted to an incorrect model that assumes $x_v = 7$

Fig. 7.

Trend plots for quantiles (.025, .10, .25, .50, .75, .90, .975) of PPP associated with C , where the data are generated from Model 2 ($x_v = 3, 5$), but are fitted to two incorrect models that assume $x_v = 4$ (row 1: $n = 500$; row 2: $n = 1000$) $x_v = 7$

Fig. 8.
Trend plots for means of PPP and PPPP associated with C , where the data are generated from Model 1 ($x_v = 4$), but are fitted to an incorrect model that assumes $x_v = 7$

Fig. 9. Trend plots for means of PPP and PPPP associated with C , where the data are generated from Model 2 ($x_v = 3, 5$), but are fitted to an incorrect model that assumes $x_v = 7$

Fig. 10.

Trend plots of means of PPP and PPPP associated with C, where the data are generated from Model 2 ($x_v = 3, 5$), but are fitted to an incorrect model that assumes $x_v = 4$

Table 1

The single change-point CCRM considered in Sect. 5.2

ψ	$u = 1$	$u = 2$
$[x_{v-1}, x_v)$	$[x_0, x_4)$	$[x_4, x_7)$
$\theta^{(u)}$	.10	.20
$\rho^{(u)}$	(.10,.30,.60)	(.20,.40,.40)

Table 2

The two change-points CCRM considered in Sect. 5.2

ψ	$u = 1$	$u = 2$	$u = 3$
$[x_{v-1}, x_v)$	$[x_0, x_3)$	$[x_3, x_5)$	$[x_5, x_7)$
$\theta^{(u)}$	.10	.15	.20
$\rho^{(u)}$	(.10,.30,.60)	(.20,.40,.40)	(.33,.33,.33)