Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Bias-Correction and Test for Mark-Point Dependence with Replicated Marked Point Processes

Ganggang Xu, Jingfei Zhang, Yehua Li & Yongtao Guan

To cite this article: Ganggang Xu, Jingfei Zhang, Yehua Li & Yongtao Guan (2024) Bias-Correction and Test for Mark-Point Dependence with Replicated Marked Point Processes, Journal of the American Statistical Association, 119:545, 217-231, DOI: 10.1080/01621459.2022.2106234

To link to this article: https://doi.org/10.1080/01621459.2022.2106234

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2024, VOL. 119, NO. 545, 217–231: Theory and Methods
https://doi.org/10.1080/01621459.2022.2106234

Bias-Correction and Test for Mark-Point Dependence with Replicated Marked Point Processes

Ganggang Xu[a], Jingfei Zhang[a], Yehua Li[b], and Yongtao Guan[a]

[a]Department of Management Science, University of Miami, Coral Gables, FL; [b]Department of Statistics, University of California, Riverside, CA

ABSTRACT

Mark-point dependence plays a critical role in research problems that can be fitted into the general framework of marked point processes. In this work, we focus on adjusting for mark-point dependence when estimating the mean and covariance functions of the mark process, given independent replicates of the marked point process. We assume that the mark process is a Gaussian process and the point process is a log-Gaussian Cox process, where the mark-point dependence is generated through the dependence between two latent Gaussian processes. Under this framework, naive local linear estimators ignoring the mark-point dependence can be severely biased. We show that this bias can be corrected using a local linear estimator of the cross-covariance function and establish uniform convergence rates of the bias-corrected estimators. Furthermore, we propose a test statistic based on local linear estimators for mark-point independence, which is shown to converge to an asymptotic normal distribution in a parametric $\sqrt{n}$-convergence rate. Model diagnostics tools are developed for key model assumptions and a robust functional permutation test is proposed for a more general class of mark-point processes. The effectiveness of the proposed methods is demonstrated using extensive simulations and applications to two real data examples. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received July 2022
Accepted July 2022

KEYWORDS
Mark-point dependence;
Marked point processes

1. Introduction

In many scientific fields, numerical variables of interest are commonly observed at some random event times for a collection of subjects. For example, Fok, Ramsay, Abrahamowicz, and Fortin (2012) considered systemic lupus erythematosus disease activity index (SLEDAI) scores of patients at times of flare episodes; Gervini and Baur (2020) studied the bid prices of Palm M515 personal digital assistants on week-long eBay auctions. In these examples, the events in turns are flare episodes and bids, while the associated numerical variables are SLEDAI score and bid price. The event times in each example are random and can be viewed as a realization from a point process, whereas the numerical variables are often referred to as marks. The random event times and the marks together form a so-called marked point process (Illian et al. 2008). The marks are often well defined over the entire study domain, but not just at event times. For example, SLEDAI scores could potentially be obtained at any time (Fok, Ramsay, Abrahamowicz, and Fortin 2012), and it is therefore reasonable to assume a separate mark process that generated the SLEDAI scores for each patient.

Marked point process data commonly arise in the analysis of longitudinal data with irregularly scattered observation times, where independent observations from different subjects are typically available. Tools from functional data analysis have been used to model such data by treating the mark processes as random functions and the observed mark values as discrete observations on the functions (Hsing and Eubank 2015). While there has been extensive recent literature on this topic (e.g., Yao, Müller, and Wang 2005; Chen and Müller 2012; Zhang and Wang 2016; Wang, Wong, and Zhang 2021), most of existing work rely on a convenient but restrictive assumption stipulating that the marks and points are independent; we will refer to this assumption as *mark-point independence* in this article. Potential mark-point dependence is not considered except in a few papers (e.g., Fok, Ramsay, Abrahamowicz, and Fortin 2012; Gervini and Baur 2020). In real applications, however, the mark-point independence assumption may be invalid. For example, the SLEDAI scores are expected to be high at times of flare episodes (Fok, Ramsay, Abrahamowicz, and Fortin 2012) and hence the SLEDAI scores and the flare episode times may be correlated. Ignoring such mark-point dependence can lead to biased estimation results for the mark process. Hence, testing mark-point independence and correcting any biases caused by such dependence can have a major impact on statistical practice in this area.

In this article, we consider the problem of estimating the mean and covariance functions of the mark process nonparametrically and testing mark-point independence, when independent replicates of the marked point process are available. To that end, we assume that the point process in each replicate is a log-Gaussian Cox process (LGCP; Møller, Syversveen, and Waagepetersen 1998). Mark-point dependence can be modeled

CONTACT Yongtao Guan yguan@bus.miami.edu Department of Management Science, University of Miami, Coral Gables, FL 33124.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA.

through the correlation between the latent Gaussian process defining the LGCP and the mark process which is also assumed to be Gaussian. Similar assumptions have also been made in Diggle, Menezes, and Su (2010) and Gervini and Baur (2020). Under the proposed modeling framework, we show that the naive local linear estimators of the mean and covariance functions that ignore the mark-point dependence are biased, where the biases can be corrected by using a local linear estimator of the cross-covariance function between the mark process and the latent Gaussian process for the point process; see Theorems 1–2 for more details. The resulting bias-corrected estimators are shown to be uniformly consistent for their respective population counterparts.

Our proposed approach estimates the mean and covariance functions via nonparametric smoothing. Unlike the likelihood-based approaches (e.g., Fok, Ramsay, Abrahamowicz, and Fortin 2012; Gervini and Baur 2020), the proposed estimators do not require fully specifying the data generating mechanism of the marked point process. In particular, there is no need to model the point process other than assuming it to be an LGCP and requires no explicit assumption on how the mark-point dependence is generated. In this sense, these estimators are therefore more robust to model misspecifications than the existing likelihood-based methods. A second advantage of the proposed method is computational since the local linear estimators can be efficiently computed given the selected bandwidths. In contrast, Gervini and Baur (2020) used the Karhunen-Loève expansion to approximate the mark process and the Gaussian process for the point process separately with some basis functions. The resulting model parameters are estimated by a penalized maximum likelihood approach, which can be computationally intensive with several tuning parameters to be selected.

Finally, another important contribution of this work is the introduction of a new testing procedure for mark-point independence. To the best of our knowledge, the proposed test is the first formal test designed for marked point processes with replicates. The limiting distribution of the proposed test statistic under mark-point independence is shown to be normal with mean 0 and a variance that can be estimated using observed data. Surprisingly, the proposed test statistic converges to its limiting distribution at the classical parametric rate of $n^{1/2}$, even though it is constructed based on some nonparametric estimators; see Section 4.2 and Theorem 3 for details. Our test relies on the assumptions that the mark process is Gaussian and the point process is an LGCP. We describe a set of diagnostic tools to check these assumptions, and propose a functional permutation test for mark-point independence that does not rely on distributional assumptions of the underlying marked-point process. Our simulation studies demonstrate the validity and power of the proposed functional permutation test for a variety of mark-point process models.

We note that some methods are developed to test the mark-point dependence (see, e.g., Schlather, Ribeiro, and Diggle 2004; Guan and Afshartous 2007; Zhang 2014, 2017) in a single spatial mark-point process, though the majority require the marked point process to be stationary, which can be implausible in many applications. For example, both bid intensity and bid price may increase with time during an auction (Gervini and Baur 2020); as a result, neither the mark (i.e., bid price) process nor the point (i.e., bid time) process is stationary.

The rest of the article is organized as follows. In Section 2, we describe our model. In Section 3, we discuss the naive mean and covariance estimators for the mark process and their biases in the presence of mark-point dependence. In Section 4, we propose an estimator for the cross-covariance function between the mark process and point processes, based on which we propose bias-corrected estimators for the mean and covariance functions of the mark process; we also propose a testing procedure for the mark-point independence. Asymptotic properties of the proposed estimators and the test statistic are studied in Section 5. In Section 6, we describe a functional permutation test and some diagnostic tools to assess model assumptions. Numerical performances of the proposed methods are illustrated by simulation studies in Section 7 and two real datasets in Section 8. Finally, some concluding remarks are provided in Section 9, and implementation details, additional numerical results, together with technical proofs, are collected in the supplementary materials.

2. Model Specification

Consider a marked point process defined over a time window $\mathcal{T} \subset \mathbb{R}$. Let $\{[s, Z_i(s)] : s \in N_i, i = 1, \ldots, n\}$ denote n independent realizations of the process, where $N_i = \{s_{ij} : s_{ij} \in \mathcal{T}, j = 1, \ldots, n_i\}$ is the set of n_i events from the ith point process and $Z_i(s)$ is the associated mark for an event at $s \in \mathcal{T}$. We assume that the random event times in N_i are generated by an LGCP, with the latent intensity function

$$\lambda_i(s) = \lambda_0(s) \exp\left[X_i(s)\right], \tag{1}$$

for $s \in \mathcal{T}$. In the above, $\lambda_0(\cdot)$ is a baseline intensity function and $X_i(\cdot)$ is a latent zero-mean Gaussian process with a variance function $\sigma_X^2(\cdot)$, $i = 1, \ldots, n$. Conditional on the latent intensity function, an LGCP is simply an inhomogeneous Poisson process. The first- and second-order marginal intensity functions of the point process are therefore

$$\rho(s) \equiv \mathbb{E}\left[\lambda_i(s)\right] = \lambda_0(s) \exp\left[\sigma_X^2(s)/2\right], \tag{2}$$

$$\rho_2(s,t) \equiv \mathbb{E}\left[\lambda_i(s)\lambda_i(t)\right] = \rho(s)\rho(t) \exp\left[C_X(s,t)\right], \tag{3}$$

where $C_X(s,t) = \mathrm{cov}[X_1(s), X_1(t)]$, $s, t \in \mathcal{T}$.

We assume that the mark process is well defined for all $s \in \mathcal{T}$. More specifically,

$$Z_i(s) = \mu(s) + Y_i(s) + e_i(s), \tag{4}$$

where $\mu(\cdot)$ is some deterministic function, $Y_i(\cdot)$ is a zero-mean Gaussian process with a variance function $\sigma_Y^2(\cdot)$, and $e_i(\cdot)$ is a zero-mean Gaussian white noise process with a variance function $\sigma_e^2(\cdot)$, $i = 1, \ldots, n$.

Denote $C_Y(s,t) = \mathrm{cov}[Y_1(s), Y_1(t)]$ and $C_{XY}(s,t) = \mathrm{cov}[X_1(s), Y_1(t)]$, for any $s, t \in \mathcal{T}$. When $C_{XY}(\cdot,\cdot) \not\equiv 0$, the mark process and the point process are not independent. We are interested in estimating the mean function $\mu(\cdot)$ and the covariance function $C_Y(\cdot,\cdot)$ based on the observed data. We remark that if the point process reduces to Poisson, that is, $C_X(\cdot,\cdot) \equiv 0$, it always holds that $C_{XY}(\cdot,\cdot) \equiv 0$.

3. Naive Estimation of the Mean and Covariance Functions

When the mark process and the point process are independent, local linear estimators for $\mu(\cdot)$ and $C_Y(\cdot,\cdot)$ are well studied (e.g., Yao, Müller, and Wang 2005). We refer to these estimators as the naive estimators. Define $K_{1,h_\mu}(s) = h_\mu^{-1}K_1(s/h_\mu)$, where $K_1(\cdot)$ is a kernel function and h_μ is a bandwidth. Then, the naive local linear estimator for $\mu(s), s \in \mathcal{T}$, is defined as $\widetilde{\mu}(s) = \widetilde{\beta}_{0,\mu}$, where $\widetilde{\beta}_{0,\mu}$ is obtained by minimizing

$$L_{n,\mu}(\beta_{0,\mu},\beta_{1,\mu}) = \sum_{i=1}^{n}\sum_{u\in N_i}[Z_i(u) - \beta_{0,\mu} - \beta_{1,\mu}(u-s)]^2 K_{1,h_\mu}(u-s)$$

with respect to $\beta_{0,\mu}$ and $\beta_{1,\mu}$. Let $\boldsymbol{\phi}_{h_\mu}(s) = (1, s/h_\mu)^\top$, and denote by $\mathbf{e}_p$ a p-dimensional vector whose first entry equals 1 and all other entries equal 0. Then,

$$\widetilde{\mu}(s) = \mathbf{e}_2^\top\left[\widehat{\mathbf{A}}_{n,h_\mu,1}(s)\right]^{-1}\widehat{\mathbf{A}}_{n,h_\mu,2}(s), \quad s\in\mathcal{T}, \text{ where (5)}$$

$$\widehat{\mathbf{A}}_{n,h_\mu,1}(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u\in N_i}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)^\top \quad (6)$$

$$\widehat{\mathbf{A}}_{n,h_\mu,2}(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u\in N_i}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)Z_i(u). \quad (7)$$

Define $K_{2,h_y}(s,t) = h_y^{-2}K_2(s/h_y, t/h_y)$, where $K_2(\cdot,\cdot)$ is a bivariate kernel function and h_y is a bandwidth. Let $\widetilde{W}_i(u,v) = [Z_i(u) - \widetilde{\mu}(u)][Z_i(v) - \widetilde{\mu}(v)]$. The naive local linear estimator for $C_Y(s,t)$, $s,t\in\mathcal{T}$, is defined as $\widetilde{C}_Y(s,t) = \widetilde{\beta}_{0,y}$, where $\widetilde{\beta}_{0,y}$ is obtained by minimizing

$$L_{n,C_Y}(\beta_{0,y},\beta_{1,y},\beta_{2,y}) = \sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}\left[\widetilde{W}_i(u,v) - \beta_{0,y} - \beta_{1,y}(u-s) - \beta_{2,y}(v-t)\right]^2 K_{2,h_y}(u-s,v-t)$$

with respect to $\beta_{0,y}$, $\beta_{1,y}$, and $\beta_{2,y}$. In the above, the $\neq$ sign indicates $u\neq v$. Let $\boldsymbol{\psi}_{h_y}(s,t) = (1, s/h_y, t/h_y)^\top$. The local linear estimator $\widetilde{C}_Y(s,t)$ can be written as

$$\widetilde{C}_Y(s,t) = \mathbf{e}_3^\top\left[\widehat{\mathbf{B}}_{n,h_y,1}(s,t)\right]^{-1}\widehat{\mathbf{B}}_{n,h_y,2}(s,t), \quad s,t\in\mathcal{T}, \text{ where} \quad (8)$$

$$\widehat{\mathbf{B}}_{n,h_y,1}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}K_{2,h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t)^\top, \quad (9)$$

$$\widehat{\mathbf{B}}_{n,h_y,2}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}K_{2,h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t)\widetilde{W}_i(u,v). \quad (10)$$

When the mark process and the point process are independent, $\widetilde{\mu}(\cdot)$ and $\widetilde{C}_Y(\cdot,\cdot)$ are consistent estimators for $\mu(\cdot)$ and $C_Y(\cdot,\cdot)$ under mild conditions (Yao, Müller, and Wang 2005; Li and Hsing 2010). However, in the presence of mark-point dependence, both estimators will have non-negligible biases for their target parameters as we will show in the following two sections. To facilitate the derivations, we first present a technical lemma.

Lemma 1. Let X, Y_1, and Y_2 be three normal random variables with means 0, μ_1 and μ_2 and variances σ_X^2, σ_1^2, and σ_2^2, respectively. Then it holds that,

$$\mathbb{E}[Y_1\exp(X)] = [\mu_1 + \operatorname{cov}(X,Y_1)]\exp(\sigma_X^2/2), \quad (11)$$

$$\mathbb{E}[Y_1^2\exp(X)] = \{\sigma_1^2 + [\mu_1 + \operatorname{cov}(X,Y_1)]^2\}\exp(\sigma_X^2/2), \quad (12)$$

$$\mathbb{E}[Y_1Y_2\exp(X)] = \{\operatorname{cov}(Y_1,Y_2) + [\operatorname{cov}(X,Y_1) + \mu_1][\operatorname{cov}(X,Y_2) + \mu_2]\}\exp(\sigma_X^2/2). \quad (13)$$

The proof is given in the supplementary materials.

3.1. Bias of $\widetilde{\mu}(\cdot)$

To derive the bias of $\widetilde{\mu}(s)$, $s\in\mathcal{T}$, we first note that

$$\sum_{u\in N_i}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)Z_i(u) = \int_{\mathcal{T}}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)Z_i(u)N_i(du),$$

where $N_i(du)$ denotes the random number of events from the ith point process in an infinitesimal time interval du at $u\in\mathcal{T}$. Then, we have that for any $s\in\mathcal{T}$,

$$\mathbf{A}_{h_\mu,2}(s) \equiv \mathbb{E}\left[\widehat{\mathbf{A}}_{n,h_\mu,2}(s)\right] = \int_{\mathcal{T}}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)\mathbb{E}[Z_i(u)N_i(du)],$$

where $\widehat{\mathbf{A}}_{n,h_\mu,2}(s)$ is as defined in (7). It follows from (2) and (4) that

$$\mathbf{A}_{h_\mu,2}(s) = \int_{\mathcal{T}}\mu(u)\rho(u)K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)du + \int_{\mathcal{T}}\lambda_0(u)\mathbb{E}\{Y_i(u)\exp[X_i(u)]\}K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)du,$$

where $\rho(\cdot)$ is as defined in (2). Since $X_i(u)$ and $Y_i(u)$, $u\in\mathcal{T}$, are both normal random variables, it immediately follows from (11) that $\mathbb{E}\{Y_i(u)\exp[X_i(u)]\} = C_{XY}(u,u)\exp\left[\sigma_X^2(u)/2\right]$. If $\mu(\cdot)$ and $C_{XY}(\cdot,\cdot)$ are smooth functions such that $\mu(u)\approx\mu(s)$ and $C_{XY}(u,u)\approx C_{XY}(s,s)$ for any u in a small neighborhood around s, then $\mathbf{A}_{h_\mu,2}(s)\approx[\mu(s) + C_{XY}(s,s)]\,\mathbf{a}_{h_\mu}(s)$, where

$$\mathbf{a}_{h_\mu}(s) = \int_{\mathcal{T}}\rho(u)K_{1,h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}(u-s)du. \quad (14)$$

Note that, due to the definition of $\boldsymbol{\phi}_{h_\mu}(\cdot)$, $\mathbf{a}_{h_\mu}(s)$ is the first column of the following matrix

$$\mathbf{A}_{h_\mu,1}(s) \equiv \mathbb{E}\left[\widehat{\mathbf{A}}_{n,h_\mu,1}(s)\right] = \int_{\mathcal{T}} \rho(u) K_{1,h_\mu}(u-s) \boldsymbol{\phi}_{h_\mu}(u-s)\boldsymbol{\phi}_{h_\mu}^{\top}(u-s)du, \quad (15)$$

where $\widehat{\mathbf{A}}_{n,h_\mu,1}(s)$ is as defined in (6). By the law of large numbers, it holds under mild conditions that $|\widehat{\mathbf{A}}_{n,h_\mu,1}(s) - \mathbf{A}_{h_\mu,1}(s)| \xrightarrow{p} 0$ and $|\widehat{\mathbf{A}}_{n,h_\mu,2}(s) - \mathbf{A}_{h_\mu,2}(s)| \xrightarrow{p} 0$, as $n \to \infty$, where $\xrightarrow{p}$ stands for convergence in probability. It then follows from the definition of $\widetilde{\mu}(\cdot)$ in (5) and the continuous mapping theorem that for any $s \in \mathcal{T}$,

$$\widetilde{\mu}(s) = \mathbf{e}_2^{\top}\left[\mathbf{A}_{h_\mu,1}(s)\right]^{-1}\mathbf{a}_{h_\mu}(s)\left[\mu(s) + C_{XY}(s,s)\right] \times[1+o_p(1)] \xrightarrow{p} \mu(s) + C_{XY}(s,s) \equiv \mu^*(s). \quad (16)$$

In other words, $\widetilde{\mu}(\cdot)$ is asymptotically unbiased for $\mu^*(\cdot)$, but not for $\mu(\cdot)$ unless $C_{XY}(\cdot,\cdot) \equiv 0$.

3.2. Bias of $\widetilde{C}_Y(\cdot,\cdot)$

For ease of presentation, we replace $\widetilde{W}_i(u,v)$ in (10) by $W_i(u,v) = [Z_i(u) - \mu^*(u)][Z_i(v) - \mu^*(v)]$. This is not of particular concern since $\widetilde{\mu}(\cdot)$ converges to $\mu^*(\cdot)$ uniformly in probability; see Theorem 1 in Section 5. Now define $N_i^{(2)}(du,dv) = N_i(du)N_i(dv)I(u \neq v)$, where $I(\cdot)$ is an indicator function. Note that for any real function $f(\cdot,\cdot)$, it holds that

$$\sum_{u,v\in N_i}^{\neq} f(u,v) = \int_{\mathcal{T}^2} f(u,v) N_i^{(2)}(du,dv).$$

Then, for any $s,t \in \mathcal{T}$, it holds that

$$\mathbf{B}_{h_y,2}(s,t) \equiv \mathbb{E}\left[\widehat{\mathbf{B}}_{n,h_y,2}(s,t)\right] = \int_{\mathcal{T}^2} K_{2,h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t) \mathbb{E}\left[W_i(u,v)N_i^{(2)}(du,dv)\right],$$

where $\widehat{\mathbf{B}}_{n,h_y,2}(s)$ is as defined in (10). Since for any $u,v \in \mathcal{T}$,

$$\mathbb{E}\left[W_i(u,v)N_i^{(2)}(du,dv)\right] = \lambda_0(u)\lambda_0(v)\mathbb{E}\left\{W_i(u,v)\exp\left[X_i(u)+X_i(v)\right]\right\}dudv,$$

it follows from (3) and (13) that for any $s,t \in \mathcal{T}$,

$$\mathbf{B}_{h_y,2}(s,t) = \int_{\mathcal{T}^2} \rho_2(u,v) C_Y^*(u,v) K_{2,h_y}(u-s,v-t) \boldsymbol{\psi}_{h_y}(u-s,v-t)dudv,$$

where $\rho_2(u,v)$ is as defined in (3) and

$$C_Y^*(u,v) = C_Y(u,v) + C_{XY}(u,v)C_{XY}(v,u). \quad (17)$$

If $C_Y(\cdot,\cdot)$ and $C_{XY}(\cdot,\cdot)$ are both smooth functions such that $C_Y(u,v) \approx C_Y(s,t)$ and $C_{XY}(u,v) \approx C_{XY}(s,t)$ for (u,v) in a small neighborhood around (s,t), then $\mathbf{B}_{h_y,2}(s,t) \approx C_Y^*(s,t)\mathbf{b}_{h_y}(s,t)$, where

$$\mathbf{b}_{h_y}(s,t) = \int_{\mathcal{T}^2} \rho_2(u,v) K_{2,h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t)dudv. \quad (18)$$

Note that, due to the definition of $\boldsymbol{\psi}_{h_y}(\cdot,\cdot)$, $\mathbf{b}_{h_y}(s,t)$ is the first column of the matrix

$$\mathbf{B}_{h_y,1}(s,t) \equiv \mathbb{E}\left[\widehat{\mathbf{B}}_{n,h_y,1}(s,t)\right] = \int_{\mathcal{T}^2} \rho_2(u,v) K_{2,h_y}(u-s,v-t) \boldsymbol{\psi}_{h_y}(u-s,v-t)\boldsymbol{\psi}_{h_y}^{\top}(u-s,v-t)dudv, \quad (19)$$

where $\widehat{\mathbf{B}}_{n,h_y,1}(s)$ is as defined in (9). By the law of large numbers, under suitable conditions, $|\widehat{\mathbf{B}}_{n,h_y,1}(s,t) - \mathbf{B}_{h_y,1}(s,t)| \xrightarrow{p} 0$ and $|\widehat{\mathbf{B}}_{n,h_y,2}(s,t) - \mathbf{B}_{h_y,2}(s,t)| \xrightarrow{p} 0$ as $n \to \infty$. It then follows from the definition of $\widetilde{C}_Y(\cdot,\cdot)$ in (8) and the continuous mapping theorem that

$$\widetilde{C}_Y(s,t) = \mathbf{e}_3^{\top}\left[\mathbf{B}_{h_y,1}(s,t)\right]^{-1}\mathbf{b}_{h_y}(s,t)C_Y^*(s,t) \times[1+o_p(1)] \xrightarrow{p} C_Y^*(s,t), \quad s,t \in \mathcal{T}. \quad (20)$$

This shows that the naive covariance function estimator $\widetilde{C}_Y(\cdot,\cdot)$ given in (8) is an asymptotically unbiased estimator for $C_Y^*(\cdot,\cdot)$ defined in (17), but not for $C_Y(\cdot,\cdot)$ unless $C_{XY}(\cdot,\cdot) \equiv 0$.

4. Bias-Correction and Test for Mark-Point Independence

In this section, we propose a local linear estimator for the cross-covariance function $C_{XY}(\cdot,\cdot)$, based on which bias-corrected estimators for $\mu(\cdot)$ and $C_Y(\cdot,\cdot)$ are constructed. We also propose a formal testing procedure for mark-point independence. The bandwidth selection procedure is detailed in Section S.1.1, supplementary materials.

4.1. Bias-Corrected Estimation of the Mean and Covariance Functions

Following similar steps to obtain (8), the local linear estimator for $C_{XY}(s,t)$, $s,t \in \mathcal{T}$, can be defined as $\widehat{C}_{XY}(s,t) = \widehat{\beta}_{0,xy}$, where $\widehat{\beta}_{0,xy}$ is obtained by minimizing

$$L_{n,C_{XY}}(\beta_{0,xy},\beta_{1,xy},\beta_{2,xy}) = \sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq} \left[Z_i(v) - \widetilde{\mu}(v) - \beta_{0,xy} - \beta_{1,xy}(u-s) - \beta_{2,xy}(v-t)\right]^2 K_{2,h_{xy}}(u-s,v-t)$$

with respect to $\beta_{0,xy}$, $\beta_{1,xy}$, and $\beta_{2,xy}$, and h_{xy} is a bandwidth. It can be shown that

$$\widehat{C}_{XY}(s,t) = \mathbf{e}_3^{\top}\left[\widehat{\mathbf{B}}_{n,h_{xy},1}(s,t)\right]^{-1}\widehat{\mathbf{B}}_{n,h_{xy},3}(s,t), \quad s,t \in \mathcal{T}, \quad (21)$$

where $\widehat{\mathbf{B}}_{n,h_{xy},1}(s,t)$ is as defined in (9) with bandwidth h_{xy}, and

$$\widehat{\mathbf{B}}_{n,h_{xy},3}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}[Z_i(v)-\widetilde{\mu}(v)] K_{2,h_{xy}}(u-s,v-t)\boldsymbol{\psi}_{h_{xy}}(u-s,v-t). \quad (22)$$

We will show below that $\widehat{C}_{XY}(s,t)$ is asymptotically unbiased for $C_{XY}(s,t)$, despite the fact that $\widetilde{\mu}(\cdot)$ is a biased estimator for $\mu(\cdot)$. For ease of presentation, we replace $\widetilde{\mu}(\cdot)$ in (22) by $\mu^*(\cdot)$ as we did previously when studying the bias of $\widetilde{C}_Y(\cdot,\cdot)$. Then,

$$\begin{aligned}\mathbf{B}_{h_{xy},3}(s,t) &\equiv \mathbb{E}\left[\widehat{\mathbf{B}}_{n,h_{xy},3}(s,t)\right] \\ &= \int_{\mathcal{T}^2} K_{2,h_{xy}}(u-s,v-t)\boldsymbol{\psi}_{h_y}(u-s,v-t) \\ &\quad \mathbb{E}\left\{[Z_i(v)-\mu^*(v)]N_i^{(2)}(du,dv)\right\}.\end{aligned}$$

Note that for any $u,v\in\mathcal{T}$,

$$\begin{aligned}&\mathbb{E}\left\{[Z_i(v)-\mu^*(v)]N_i^{(2)}(du,dv)\right\} \\ &= \lambda_0(s)\lambda_0(s)\mathbb{E}\left\{[Z_i(v)-\mu^*(v)]\exp[X_i(u)+X_i(v)]\right\}dudv,\end{aligned}$$

which is equal to $\rho_2(u,v)C_{XY}(u,v)dudv$ due to (3) and (11). If $C_{XY}(\cdot,\cdot)$ is a smooth function in a small neighborhood around (s,t), then $\mathbf{B}_{h_{xy},3}(s,t)\approx C_{XY}(s,t)\mathbf{b}_{h_{xy}}(s,t)$, where $\mathbf{b}_{h_{xy}}(s,t)$ is as defined in (18) with bandwidth h_{xy}. Recall that $\mathbf{b}_{h_{xy}}(s,t)$ is the first column of the matrix $\mathbf{B}_{h_{xy},1}(s,t)$. It then follows from the same steps used to obtain (20) that $\mathbb{E}\left[\widehat{C}_{XY}(s,t)\right]\approx C_{XY}(s,t)$, that is, $\widehat{C}_{XY}(s,t)$ is an asymptotically unbiased estimator for $C_{XY}(s,t)$ for any $s,t\in\mathcal{T}$.

In light of (16) and (20) and given $\widehat{C}_{XY}(\cdot,\cdot)$, it is natural to consider the following bias corrected estimators for $\mu(s)$ and $C_Y(s,t)$ for any $s,t\in\mathcal{T}$,

$$\widehat{\mu}(s) = \widetilde{\mu}(s)-\widehat{C}_{XY}(s,s), \quad (23)$$

$$\widehat{C}_Y(s,t) = \widetilde{C}_Y(s,t)-\widehat{C}_{XY}(s,t)\widehat{C}_{XY}(t,s), \quad (24)$$

where $\widetilde{\mu}(\cdot)$ and $\widetilde{C}_Y(\cdot,\cdot)$ are as defined in (5) and (8), respectively. As we shall show in Section 5, $\widehat{\mu}(\cdot)$ and $\widehat{C}_Y(\cdot,\cdot)$ are uniformly consistent for $\mu(\cdot)$ and $C_Y(\cdot,\cdot)$, respectively.

4.2. Test for Mark-Point Independence

In practice, it may be difficult to know in advance whether there exists mark-point dependence. In this section, we propose a formal procedure to test the hypothesis of mark-point independence, that is, $H_0: C_{XY}(\cdot,\cdot)\equiv 0$. To motivate our test statistic, we first temporarily assume that $\mu^*(s)$ and $\sigma_Z^2(s)$, where $\sigma_Z^2(s)=\mathrm{var}[Z(s)]$, are both known for any $s\in\mathcal{T}$. Consider the following two random sums:

$$S_{n,1} = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}[Z_i(v)-\mu^*(v)]^2 \quad \text{and}$$

$$S_{n,2} = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}\sigma_Z^2(v).$$

Under H_0, it is straightforward to see that $\mathbb{E}(S_{n,1})=\mathbb{E}(S_{n,2})$. More generally, note that

$$\mathbb{E}(S_{n,1}) = \int_{\mathcal{T}^2}\lambda_0(u)\lambda_0(v)\mathbb{E}\left\{[Z_i(v)-\mu^*(v)]^2\exp[X_i(u)+X_i(v)]\right\}dudv.$$

Recall that $\mu^*(v)=\mu(v)+C_{XY}(v,v)$ for any $v\in\mathcal{T}$. By the definition of $\rho_2(\cdot,\cdot)$ in (3) and using (12) in Lemma 1, we can then verify that

$$\mathbb{E}(S_{n,1}) = \int_{\mathcal{T}^2}\rho_2(u,v)\left[\sigma_Z^2(v)+C_{XY}(u,v)^2\right]dudv.$$

Combined with the fact that $\mathbb{E}(S_{n,2})=\int_{\mathcal{T}^2}\rho_2(u,v)\sigma_Z^2(v)dudv$, it holds that

$$\mathbb{E}(S_{n,1}) = \mathbb{E}(S_{n,2}) + \int_{\mathcal{T}^2}\rho_2(u,v)C_{XY}(u,v)^2dudv. \quad (25)$$

Since the integral term in (25) is strictly nonnegative and equals zero under the null, a test statistic can be formed based on the difference between $S_{n,1}$ and $S_{n,2}$.

In practice, both $\mu^*(\cdot)$ and $\sigma_Z^2(\cdot)$ are unknown, and thus $S_{n,1}$ and $S_{n,2}$ cannot be calculated exactly. For $\mu^*(\cdot)$, we estimate it with $\widetilde{\mu}(\cdot)$ as defined in (5). For $\sigma_Z^2(\cdot)$, we consider its naive local linear estimator $\widetilde{\sigma}_Z^2(s)=\widetilde{\beta}_{0,\sigma}$, where $\widetilde{\beta}_{0,\sigma}$ is obtained by minimizing

$$\begin{aligned}&L_{n,\sigma}(\beta_{0,\sigma},\beta_{1,\sigma}) = \sum_{i=1}^{n}\sum_{u\in N_i} \\ &\left\{[Z_i(u)-\widetilde{\mu}(u)]^2-\beta_{0,\sigma}-\beta_{1,\sigma}(u-s)\right\}^2 \\ &K_{1,h_\sigma}(u-s)\end{aligned}$$

with respect to $\beta_{0,\sigma}$ and $\beta_{1,\sigma}$, and h_σ is a bandwidth. It immediately follows from the same arguments used to obtain (5) that for any $s\in\mathcal{T}$,

$$\widetilde{\sigma}_Z^2(s) = \mathbf{e}_2^\top\left[\widehat{\mathbf{A}}_{n,h_\sigma,1}(s)\right]^{-1}\widehat{\mathbf{A}}_{n,h_\sigma,3}(s), \text{ where}$$

$$\widehat{\mathbf{A}}_{n,h_\sigma,3}(s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{u\in N_i}K_{1,h_\sigma}(u-s)\boldsymbol{\phi}_{h_\sigma}(u-s)[Z_i(u)-\widetilde{\mu}(u)]^2. \quad (26)$$

As we shall show in Section 5, $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}_Z^2(\cdot)$ are uniformly consistent for $\mu^*(\cdot)$ and $\sigma_Z^2(\cdot)$, respectively. The uniform convergence of $\widetilde{\sigma}_Z^2(\cdot)$ to $\sigma_Z^2(\cdot)$ is surprising considering that $\widetilde{C}_Y(s,s)$ is biased for $C_Y(s,s)$, that is, the variance of $Y(s)$, for $s\in\mathcal{T}$, under mark-point dependence.

Given $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}_Z^2(\cdot)$, we propose a test statistic of the following form

$$T_n = \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}[Z_i(v)-\widetilde{\mu}(v)]^2 - \frac{1}{n}\sum_{i=1}^{n}\sum_{u,v\in N_i}^{\neq}\widetilde{\sigma}_Z^2(v). \quad (27)$$

Theorem 3 in Section 5 shows that, under some conditions, when $h_\mu=o(n^{-1/4})$ and $h_\sigma=o(n^{-1/4})$, $\sqrt{n}T_n$ converges to $N(0,\Omega_T)$ in distribution under H_0. The specific form of Ω_T is given in Theorem 3, based on which an empirical estimator can be derived, as given in Section S.1.2 of the supplementary materials. See also Section S.1.1, supplementary materials for the bandwidth selection procedure.

5. Theoretical Properties

In this section, we investigate theoretical properties of the proposed estimators and test statistic. We consider the asymptotic framework where the time domain $\mathcal{T}$ is bounded while the number of replicates n goes to infinity. For any matrix (or vector) $\mathbf{A}$, denote $\|\mathbf{A}\|_{\max} = \max_{i,j}(|a_{ij}|)$ as the max norm, and $\lambda_{\min}(\mathbf{A})$ as the smallest eigenvalues of $\mathbf{A}$. We use $\text{diag}\{d_1, \ldots, d_n\}$ to denote an $n \times n$ diagonal matrix with diagonal elements $d_1, \ldots, d_n$. Finally, for a function $f(\cdot)$, we denote $f^{(j)}(\cdot)$ as its jth derivative for some $j \geqslant 1$.

The following assumptions are sufficient for our theoretical investigation.

[C1] Assume that $Z(s)$ is well defined for any $s \in \mathcal{T}$, and $\rho(\cdot)$, $\mu(\cdot)$, $\mu^*(\cdot)$ and $\sigma_Z^2(\cdot)$ are third-order continuously differentiable on $\mathcal{T}$ with bounded first-, second-, and third-order derivatives, and there exists a constant $\rho_0 > 0$ such that $\rho(s) > \rho_0$ for all $s \in \mathcal{T}$.

[C2] Assume that $K_1(\cdot)$ and $K_2(\cdot,\cdot)$ have supports $[-1,1]$ and $[-1,1]^2$, respectively.

(a) Define $\boldsymbol{\sigma}_{K_1,\boldsymbol{\phi}} = \int_{-1}^{1} s^2 K_1(s)\boldsymbol{\phi}_1(s)ds$ and $\mathbf{Q}_{K_1,\boldsymbol{\phi}}^{\{j\}} = \int_{-1}^{1} K_1^j(s)\boldsymbol{\phi}_1(s)\boldsymbol{\phi}_1^\top(s)ds$ for $j = 1, 2$. Assume that $\lambda_{\min}\left(\mathbf{Q}_{K_1,\boldsymbol{\phi}}^{\{1\}}\right) > 0$.

(b) Define $\boldsymbol{\sigma}_{K_2,\boldsymbol{\psi}}(w) = \int_{-1}^{1}\int_{-1}^{1}\left[(1-w)s^2 + wt^2\right] K_2(s,t)\boldsymbol{\psi}_1(s,t)dsdt$ and

$$\mathbf{Q}_{K_2,\boldsymbol{\psi}}^{\{j\}} = \int_{-1}^{1}\int_{-1}^{1} K_2^j(s,t)\boldsymbol{\psi}_1(s,t)\boldsymbol{\psi}_1^\top(s,t)dsdt,$$

for $j = 1, 2$.

Assume that $\lambda_{\min}\left(\mathbf{Q}_{K_2,\boldsymbol{\psi}}^{\{1\}}\right) > 0$.

[C3] Assume that (a) $h_\mu \to 0$ as $n \to \infty$ such that $nh_\mu^2/[\log(n)]^2 \to \infty$; (b) $h_\sigma/h_\mu \to c$ as $n \to \infty$ for some constant $c > 0$.

[C4] Assume that $\rho_2(\cdot,\cdot)$, $C_Y(\cdot,\cdot)$ and $C_{XY}(\cdot,\cdot)$ are third-order continuously differentiable on $\mathcal{T}^2$ with bounded first-, second-, and third-order partial derivatives.

[C5] Assume that (a) $h_y \to 0$ as $n \to \infty$ such that $nh_y^2/[\log(n)]^2 \to \infty$; (b) $h_{xy} \to 0$ as $n \to \infty$ such that $nh_{xy}^2/[\log(n)]^2 \to \infty$; (c) $h_\mu^{3/2}/h_y \to 0$ and $h_\mu^{3/2}/h_{xy} \to 0$ as $n \to \infty$.

Assumptions C1 and C4 impose some conditions on the smoothness of the intensity functions and the mean, variance and covariance functions. Assumption C2 specifies some desirable properties of the kernel functions used for the proposed local linear estimators and is satisfied by many popular kernel functions such as the Epanechnikov kernel. Assumption C3 warrants that the bandwidths h_μ and h_σ are of the same order and neither can approach 0 at a rate faster than $\log(n)/\sqrt{n}$ as $n \to \infty$. In addition, Assumption C5 requires that h_y and h_{xy} decay at a rate slower than $\log(n)/\sqrt{n}$ as $n \to \infty$ and h_μ cannot be too large compared to h_y or h_{xy}. Similar conditions have been commonly used in the existing literature, see, for example, Yao, Müller, and Wang (2005) and Li and Hsing (2010). The first theorem investigates the uniform convergence of the three naive estimators when there is possible mark-point dependence.

Theorem 1 (Naive estimators). Under Assumptions C1–C5, we have that, as $n \to \infty$,

(a) $\sup_{s\in\mathcal{T}} |\widetilde{\mu}(s) - \mu(s) + C_{XY}(s,s)| = O_p\left\{h_\mu^2 + [\log(n)/(nh_\mu)]^{1/2}\right\}$;

(b) $\sup_{s\in\mathcal{T}} \left|\widetilde{\sigma}_Z^2(s) - \sigma_Z^2(s)\right| = O_p\left\{h_\sigma^2 + [\log(n)/(nh_\sigma)]^{1/2}\right\}$;

(c) $\sup_{s,t\in\mathcal{T}} \left|\widetilde{C}_Y(s,t) - C_Y(s,t) - C_{XY}(s,t)C_{XY}(t,s)\right| = O_p\left\{h_y^2 + \left[\log(n)/(nh_y^2)\right]^{1/2}\right\}$.

The proof is given in the supplementary materials.

Theorem 1 suggests that when there exists mark-point dependence (i.e., $C_{XY}(s,t) \not\equiv 0$), both the naive mean function estimator $\widetilde{\mu}(\cdot)$ and the naive covariance function estimator $\widetilde{C}_Y(\cdot,\cdot)$ are biased for their respective targets. The uniform convergence rates in Theorem 1 are comparable to the optimal rates in the literature (Li and Hsing 2010), where the non-diminishing biases caused by mark-point dependence were not considered. A somewhat surprising observation is that the local linear estimator $\widetilde{\sigma}_Z^2(\cdot)$ is still uniformly consistent for $\sigma_Z^2(\cdot)$, despite the existence of mark-point dependence. The next theorem studies the asymptotic properties of the proposed bias-corrected estimators.

Theorem 2 (Bias corrected estimators). Under Assumptions C1–C5, it holds as $n \to \infty$,

(a) $\sup_{s,t\in\mathcal{T}} \left|\widehat{C}_{XY}(s,t) - C_{XY}(s,t)\right| = O_p\left\{h_{xy}^2 + \left[\log(n)/(nh_{xy}^2)\right]^{1/2}\right\}$;

(b) $\sup_{s\in\mathcal{T}} |\widehat{\mu}(s) - \mu(s)| = O_p\left\{h_\mu^2 + h_{xy}^2 + [\log(n)/(nh_\mu)]^{1/2} + \left[\log(n)/(nh_{xy}^2)\right]^{1/2}\right\}$;

(c) $\sup_{s,t\in\mathcal{T}} \left|\widehat{C}_Y(s,t) - C_Y(s,t)\right| = O_p\left\{h_y^2 + h_{xy}^2 + \left[\log(n)/(nh_y^2)\right]^{1/2} + \left[\log(n)/(nh_{xy}^2)\right]^{1/2}\right\}$.

The proof is given in the supplementary materials.

The key result in Theorem 2 is part (a), which establishes uniform consistency of the local linear estimator $\widehat{C}_{XY}(\cdot,\cdot)$ to the cross-covariance function $C_{XY}(\cdot,\cdot)$ between the mark process and the point process. The proposed estimator achieves the same uniform convergence rate as the classical local linear estimator of the covariance function based on sparse functional data (Li and Hsing 2010). As a result of part (a), uniform consistency of the bias-corrected estimators $\widehat{\mu}(\cdot)$ and $\widehat{C}_Y(\cdot,\cdot)$ are subsequently established in (b) and (c). Note that the bandwidth h_{xy} appears in the uniform convergence rates in parts (b)–(c) because $\widehat{C}_{XY}(\cdot,\cdot)$ is used for bias corrections in $\widehat{\mu}(\cdot)$ and $\widehat{C}_Y(\cdot,\cdot)$.

Finally, we give the asymptotic distribution of the test statistic proposed in Section 4.2. Define $\rho_3(s,u,v) = \mathbb{E}[\lambda_i(s)\lambda_i(u)\lambda_i(v)]$, $\rho_4(s,t,u,v) = \mathbb{E}[\lambda_i(s)\lambda_i(t)\lambda_i(u)\lambda_i(v)]$, for

$s,t,u,v \in \mathcal{T}$. Expressions of $\rho_3(\cdot,\cdot,\cdot)$ and $\rho_4(\cdot,\cdot,\cdot,\cdot)$ can be derived with some algebra and be expressed in terms of $\sigma_X^2(\cdot)$ and $C_X(\cdot,\cdot)$. We omit their detailed expressions here.

Theorem 3 (Test statistic). Under Assumptions C1–C3 and assuming $nh_\mu^6 \to 0$ and $nh_\sigma^6 \to 0$ as $n \to \infty$, under $H_0 : C_{XY}(\cdot,\cdot) \equiv 0$, we have that,

$$\sqrt{n}\left[T_n + h_\sigma^2 \int_{\mathcal{T}} \mathcal{B}_{\sigma^2}(v)\rho(v)\tau(v)dv\right] \xrightarrow{d} N(0,\Omega_T),$$

where $\tau(s) = \frac{\int_{\mathcal{T}} \rho_2(u,s)du}{\rho(s)}$, $\mathcal{B}_{\sigma^2}(s) = \frac{(\sigma_Z^2)^{(2)}(s)}{2}\mathbf{e}_2^\top \left[\mathbf{Q}_{K_1,\boldsymbol{\phi}}^{\{1\}}\right]^{-1}$ $\boldsymbol{\sigma}_{K_1,\boldsymbol{\phi}}, s \in \mathcal{T}$, and

$$\begin{aligned}\Omega_T = {} & 2\int_{\mathcal{T}^4} C_Y^2(t,v)\rho_4(s,t,u,v)dsdtdudv \\ & +2\int_{\mathcal{T}^3} \left\{\sigma_Z^4(v) + [3-2\tau(s)]C_Y^2(s,v)\right\}\rho_3(s,u,v)dsdudv \\ & +2\int_{\mathcal{T}^2} [1-\tau(u)]\,[1-\tau(v)]\,C_Y^2(u,v)\rho_2(u,v)dudv \\ & +2\int_{\mathcal{T}} [1-\tau(v)]\,\tau(v)\rho(v)\sigma_Z^4(v)dv.\end{aligned}$$

Here $\xrightarrow{d}$ denotes convergence in distribution.

The proof is given in the supplementary materials.

We remark that the bias term in Theorem 3 is negligible if $\sqrt{n}h_\sigma^2 \to 0$. Theorem 3 indicates that although T_n is calculated using local linear estimators $\widetilde{\mu}(\cdot)$ and $\widetilde{\sigma}_Z(\cdot)$, it still achieves the parametric convergence rate $n^{-1/2}$ if $n^{1/4}h_\sigma \to 0$ as $n \to \infty$, which is a surprising result. Such a result enables us to conduct a valid hypothesis test for mark-point independence without specifying the complete data generating mechanism, provided that a consistent estimator of the variance Ω_T can be obtained. In practice, we estimate Ω_T by $\widehat{\Omega}_{T_n}$ given in Section S.1.2 of the supplementary materials. Validity of the resulting testing procedure and its empirical power are examined through simulations in Section 7.

6. Functional Permutation Test and Model Diagnostics

Our methods thus far are developed under two main assumptions, that is, the mark process is Gaussian and the point process is an LGCP. In this section, we propose a functional permutation test for testing the mark-point independence that does not reply on these two assumptions; we also devise diagnostic tools to check these two assumptions.

6.1. A Functional Permutation Test for Mark-Point Independence

Using the Karhunen-Loève expansion, the mark process $Z_i(\cdot)$ in (4) can be approximated by

$$Z_i(s) \approx \mu(s) + \sum_{k=1}^{p_Y} \xi_{ik}^Y \phi_k^Y(s) + e_i(s), \quad s \in \mathcal{T}, i = 1,\ldots,n, \quad (28)$$

where ξ_{ik}^Y's are uncorrelated random variables with mean 0 and variance η_k, with η_k being the kth largest eigenvalue of $C_Y(\cdot,\cdot)$, and $\phi_k^Y(\cdot)$ is the associated eigenfunction, $1 \leqslant k \leqslant p_Y$. Following similar proofs as in Section 5, we can show that under mark-point independence, the bias-corrected estimator $\widehat{C}_Y(\cdot,\cdot)$ in (24) is uniformly consistent for $C_Y(\cdot,\cdot)$, regardless of whether the mark process is Gaussian or the point process is an LGCP. Denoting the kth eigenvalue and eigenfunction of $\widehat{C}_Y(\cdot,\cdot)$ by $\widehat{\eta}_k$ and $\widehat{\phi}_k^Y(\cdot)$, respectively, the functional principal component (FPC) scores $\widehat{\boldsymbol{\xi}}_i^Y = (\widehat{\xi}_{i1}^Y,\ldots,\widehat{\xi}_{ip_Y}^Y)^\top$ for the ith process can be obtained by

$$\widehat{\boldsymbol{\xi}}_i^Y = \underset{\boldsymbol{\xi}}{\operatorname{argmin}} \sum_{s\in N_i} \left\{Z_i(s) - \widehat{\mu}(s) - \sum_{k=1}^{p_Y} \xi_{ik}^Y \widehat{\phi}_k^Y(s)\right\}^2, \quad (29)$$

where $\widehat{\mu}(\cdot)$ is as defined in (23). In practice, p_Y can be chosen as the smallest integer such that the percentage of cumulative variation explained by the first k components is greater than or equal to 95%, i.e., $\sum_{k=1}^{p_Y} \widehat{\eta}_k / \sum_{k=1}^{\infty} \widehat{\eta}_k \geqslant 0.95$.

Let $c(1),\ldots,c(n)$ be a random permutation of the sequence $1,\ldots,n$, and define $\widehat{e}_i(s) = Z_i(s) - \widehat{\mu}(s) - \sum_{k=1}^{p_Y} \widehat{\xi}_{ik}^Y \widehat{\phi}_k^Y(s), s \in N_i$ for all i. We obtain a set of permuted mark-point processes $\{[s, Z_i^c], s \in N_i\}_{i=1,\ldots,n}$ calculated as $Z_i^c(s) = \widehat{\mu}(s) + \sum_{k=1}^{p_Y} \widehat{\xi}_{c(i)k}^Y \widehat{\phi}_k^Y(s) + \widehat{e}_i^{\text{per}}(s)$, where $\widehat{\xi}_{c(i)k}^Y$'s are the permuted scores and $\widehat{e}_i^{\text{per}}(s)$'s are the permuted residuals. To permute the residuals, we start from a concatenation of the residual vectors from all subjects, denoted as $(\widehat{e}_i(s), s \in N_i, i = 1,\ldots,n)^\top$, and randomly permute all elements in this concatenated vector to get permuted residuals, denoted as $(\widehat{e}_i^{\text{per}}(s), s \in N_i, i = 1,\ldots,n)^\top$. We repeat the permutation for B times and obtain the p-value as

$$p\text{-value} = \frac{1}{B}\sum_{b=1}^{B} I\left(\left|T_n^{c_b}\right| > |T_n|\right), \quad (30)$$

where T_n and $T_n^{c_b}$ are the test statistics computed based on the observed data and the bth permuted data, respectively.

In the test described above, the permutations are designed to remove the mark-point dependence, if there is any, while preserving the marginal distributional properties of the mark process and the point process. We shall empirically demonstrate that the proposed functional permutation test is valid for various classes of marked point processes, and can be implemented together with the proposed test in Section 4.2. Theoretical justifications of the functional permutation test shall be deferred to future research. More implementation details of the functional permutation test are given in Section S.1.3 of the supplementary materials.

6.2. Model Diagnostics

If the mark process is Gaussian, the estimated FPC scores $\widehat{\boldsymbol{\xi}}_i^Y$'s in (29) should approximately follow the normal distribution. Consequently, we can check the normality of $\widehat{\boldsymbol{\xi}}_i^Y$'s using, for example, the QQ plot, to indirectly validate that the mark process is a Gaussian process.

To evaluate the validity of the LGCP assumption on the point process, we first derive nonparametric estimates of $\lambda_0(\cdot)$

and $C_X(\cdot,\cdot)$ under the LGCP assumption; see Section S.1.4 of the supplementary materials for details. Given these estimates, we simulate independent realizations from the LGCP defined by them and calculate some summary statistics from the observed and simulated point processes. A large difference between these summary statistics indicates that the LGCP assumption may be invalid. In this article, we consider the empirical nearest-neighbor distance distribution function (e.g., Møller and Waagepetersen 2003). For the observed data, this function can be calculated as

$$\widehat{G}(d) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{|N_i|}\sum_{u\in N_i} I[d_i(u) \leqslant d],$$

where $d_i(u)$ is the distance from $u \in N_i$ to its nearest neighbor in N_i. The same summary statistic can be calculated from each simulated point process, and we plot $\widehat{G}(d)$ against the average of those obtained from the simulated realizations over a range of d values, along with the upper and lower simulation envelopes. If the LGCP assumption holds, the plot should be roughly linear and contained in the simulation envelopes. See Section 8 for an illustration.

7. Simulation Studies

In this section, we evaluate the finite sample performance of the proposed methods. The mark processes are generated from $Z_i(s) = \mu(s) + Y_i(s) + e_i(s)$, $s \in [0,1]$, $i = 1,\ldots,n$, with

$$\begin{aligned}\mu(s) &= \frac{1}{2}\text{Beta}_{3,7}(s) + \frac{1}{2}\text{Beta}_{7,3}(s),\\ Y_i(s) &= \sum_{k=1}^{3}\xi_{ik}^Y\phi_k^Y(s), \quad e_i \sim N(0,1^2),\end{aligned} \tag{31}$$

where $\text{Beta}_{a,b}(\cdot)$ is the density function of the Beta(a,b) distribution, $\boldsymbol{\xi}_i^Y = (\xi_{i1}^Y, \xi_{i2}^Y, \xi_{i3}^Y)^\top$'s are iid random vectors with mean $\mathbf{0}$ and covariance matrix $\Sigma_Y = \text{diag}\{1, 0.6, 0.4\}$, and

$$\begin{aligned}\phi_1^Y(s) &= \sqrt{2}\sin(2\pi s), \quad \phi_2^Y(s) = \sqrt{2}\cos(2\pi s),\\ \phi_3^Y(s) &= \sqrt{2}\sin(4\pi s).\end{aligned} \tag{32}$$

The point processes are generated from an inhomogeneous Cox process according to (1), where $X_i(\cdot)$ therein has an isotropic covariance function of the form

$$\begin{aligned}C_X(s,t) &= \sigma_x^2\exp\left[-|s-t|^2/(2R^2)\right],\\ &\sigma_x > 0, R > 0, \text{ for any } s,t \in [0,1].\end{aligned} \tag{33}$$

For a given σ_x, we set $\lambda_0(s) = \frac{15}{8}(15+2s)\exp\left(-\frac{1}{2}\sigma_x^2\right)$, $s \in [0,1]$, such that there are on average 30 observed events for each subject if the point process is an LGCP. The parameters σ_x and R jointly control the clustering strength of the point process. We fix $R = 0.1$ and vary σ_x in different scenarios. In particular, when $\sigma_x = 0$, the resulting point process is an inhomogeneous Poisson process.

By standard stochastic process theory, the Karhunen-Loève decomposition of $X_i(\cdot)$ gives

$$X_i(s) = \sum_{k=1}^{\infty}\xi_{ik}^X\phi_k^X(s), \quad \text{with } \mathbb{E}\xi_{ik}^X = 0,$$

$$\text{var}(\xi_{ik}^X) = \eta_k^X, \text{ for any } s \in [0,1],$$

where $\{\eta_k^X, \phi_k^X(\cdot)\}$, $k = 1,2,\ldots$, are eigenvalue-eigenfunction pairs of the covariance function $C_X(\cdot,\cdot)$ defined in (33). Denote $\boldsymbol{\xi}_i^X = (\xi_{i1}^X, \xi_{i2}^X, \xi_{i3}^X)^\top$ such that $\mathbb{E}\boldsymbol{\xi}_i^X = \mathbf{0}$ and $\Sigma_X = \text{var}(\boldsymbol{\xi}_i^X) = \text{diag}\{\eta_1^X, \eta_2^X, \eta_3^X\}$. We let the correlation matrix between $\boldsymbol{\xi}_i^X$ and $\boldsymbol{\xi}_i^Y$ be of the form

$$\begin{aligned}\Sigma_{XY} &= \text{corr}\left(\boldsymbol{\xi}_i^X, \boldsymbol{\xi}_i^Y\right)\\ &= q\times\begin{pmatrix}1 & -0.5 & -0.25\\ 0.375 & 1 & -0.125\\ 0.125 & 0 & 1\end{pmatrix}, \quad \text{for } 0 \leqslant q \leqslant 0.8,\end{aligned} \tag{34}$$

and assume that the remaining ξ_{ik}^X's with $k \geqslant 4$ are independent of $\boldsymbol{\xi}_i^Y$. Under such a design, $X_i(\cdot)$ and $Y_i(\cdot)$ are independent when $q = 0$, and the strength of correlation between $X_i(\cdot)$ and $Y_i(\cdot)$ increases as q increases. The range of q is selected such that the variance-covariance matrix of the joint distribution of $(\boldsymbol{\xi}_i^X, \boldsymbol{\xi}_i^Y)$ is positive definite. When $q = 0.8$, the correlation between $X(s)$ and $Z(s)$ ranges from -0.16 to 0.53 for $s \in [0,1]$.

Finally, $\boldsymbol{\xi}_i^X$'s and $\boldsymbol{\xi}_i^Y$'s are jointly simulated as follows

$$\begin{pmatrix}\boldsymbol{\xi}_i^X\\ \boldsymbol{\xi}_i^Y\end{pmatrix} = \mathbf{R}\begin{pmatrix}\boldsymbol{\varepsilon}_i^X\\ \boldsymbol{\varepsilon}_i^Y\end{pmatrix}, \text{ with}$$

$$\begin{pmatrix}\Sigma_X & \Sigma_{XY}\\ \Sigma_{XY}^\top & \Sigma_Y\end{pmatrix} = \mathbf{R}^\top\mathbf{R}, \tag{35}$$

where $\mathbf{R}$ is an upper triangular matrix obtained through the Cholesky decomposition, $\boldsymbol{\varepsilon}_i^X$ and $\boldsymbol{\varepsilon}_i^Y$ are random vectors consisting of iid random variables with mean 0, variance 1, and marginal distributions $\mathbb{P}_X$ and $\mathbb{P}_Y$, respectively. Three types of marginal distributions are considered for $\mathbb{P}_X$ and $\mathbb{P}_Y$, namely, $N(0,1)$, referred to as `Gaussian`, centered exponential distribution with a rate 1, referred to as `Exp`, and scaled t-distribution with a degrees of freedom 4, referred to as `T4`. When both $\mathbb{P}_X$ and $\mathbb{P}_Y$ are Gaussian, (35) generates data under models (1) and (4). For ease of presentation, from now on, we shall denote, for example, (`Gaussian`, `Exp`) for the setting when $\boldsymbol{\varepsilon}_i^X$'s are Gaussian and $\boldsymbol{\varepsilon}_i^Y$'s are exponential.

7.1. Estimation Accuracy

We first compare the estimation accuracy of the proposed estimators in the presence of mark-point dependence. Specifically, there are four functions of interest as discussed below.

1. $\mu(\cdot)$: the mean function of the mark process. We consider both the naive estimator $\widetilde{\mu}(\cdot)$ defined in (5) and the bias-corrected estimator $\widehat{\mu}(\cdot)$ defined in (23). The estimation accuracy of each estimator is evaluated through the mean absolute deviation (MAD) defined as $\text{MAD}(\widetilde{\mu}) = \int_0^1|\widetilde{\mu}(s)-\mu(s)|\,ds$, and $\text{MAD}(\widehat{\mu}) = \int_0^1|\widehat{\mu}(s)-\mu(s)|\,ds$.
2. $\sigma_Z^2(\cdot)$: the variance function of the mark process. Given the naive estimator $\widetilde{\sigma}_Z^2(\cdot)$ defined in (26), we report $\text{MAD}(\widetilde{\sigma}_Z^2) = \int_0^1|\widetilde{\sigma}_Z^2(s)-\sigma_Z^2(s)|\,ds$.
3. $C_Y(\cdot,\cdot)$: the covariance function of the mark process. The estimation accuracy of the naive estimator $\widetilde{C}_Y(\cdot,\cdot)$ in (8) and

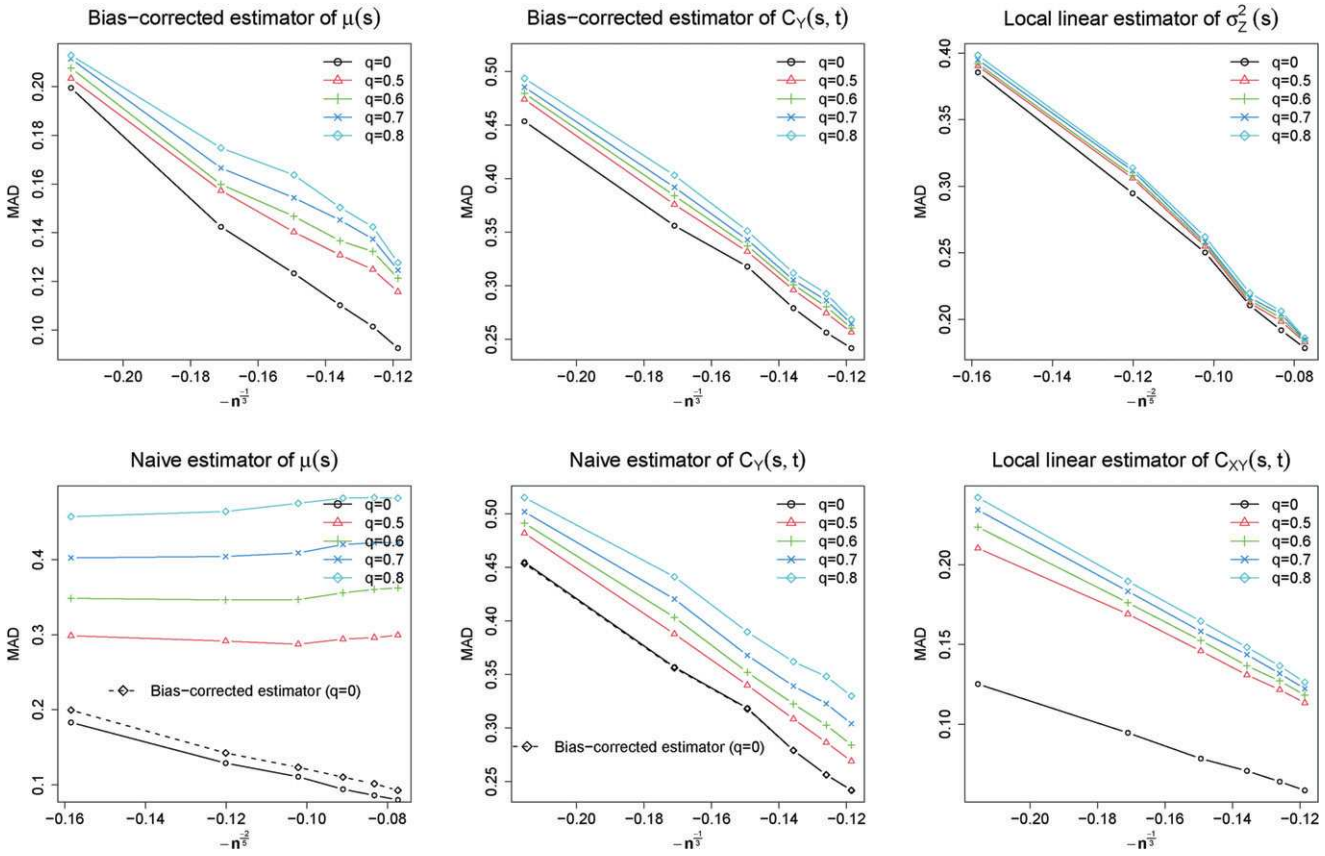

Figure 1. Estimation accuracy of local linear estimators under the (`Gaussian`, `Gaussian`) setting.

the bias-corrected estimator $\widehat{C}_Y(\cdot,\cdot)$ in (24) are compared through $\mathrm{MAD}(\widetilde{C}_Y) = \int_0^1\int_0^1 \left|\widetilde{C}_Y(s,t) - C_Y(s,t)\right| dsdt$, and $\mathrm{MAD}(\widehat{C}_Y) = \int_0^1\int_0^1 \left|\widehat{C}_Y(s,t) - C_Y(s,t)\right| dsdt$.

4. $C_{XY}(\cdot,\cdot)$: the cross-covariance function between the mark process and the point process. The estimation accuracy of the estimator $\widehat{C}_{XY}(\cdot,\cdot)$ defined in (21) is evaluated through $\mathrm{MAD}(\widehat{C}_{XY}) = \int_0^1\int_0^1 \left|\widehat{C}_{XY}(s,t) - C_{XY}(s,t)\right| dsdt$.

The Epanechnikov kernel is used with respective bandwidths selected following Section S.1.1, supplementary materials for each simulated dataset. We fix $\sigma_x = 1$ for the covariance function (33), and consider varying strength of correlation with $q = 0.5, 0.6, 0.7, 0.8$, as the sample size n varies from 100 to 600. Individuals with more than 200 time points are removed to enhance numerical stability. Summary statistics based on 500 simulation runs are illustrated in Figure 1 under the (`Gaussian`, `Gaussian`) setting.

It is seen from Figure 1 that $\mathrm{MAD}(\widehat{\mu})$ and $\mathrm{MAD}(\widetilde{\sigma}_Z^2)$ are approximately of orders $O(n^{-1/3})$ and $O(n^{-2/5})$, respectively. This agrees with our theoretical results in Theorems 1–2. Specifically, by Theorem 1, the optimal bandwidth for $\widetilde{\sigma}_Z^2(\cdot)$ is roughly of the order $O(n^{-1/5})$, giving the optimal convergence rate $O(n^{-2/5})$. Moreover, Theorem 2 indicates that the optimal convergence rate of $\widehat{\mu}(\cdot)$ is approximately of order $O(n^{-1/3})$, due to the bandwidth h_{xy} used in the bias correction term. Hence, the observed rates for $\mathrm{MAD}(\widehat{\mu})$ and $\mathrm{MAD}(\widetilde{\sigma}_Z^2)$ support our theoretical findings and validate the effectiveness of the bandwidth selection criteria proposed in Section S.1.1, supplementary materials.

Similarly, Theorem 2 suggests that $\mathrm{MAD}(\widehat{C}_{XY})$ and $\mathrm{MAD}(\widehat{C}_Y)$ are roughly of order $O(n^{-1/3})$ using the optimal bandwidths of order $O(n^{-1/6})$, and the graphic summaries in Figure 1 strongly corroborate these conclusions and the effectiveness of the bandwidth selection criteria proposed in Section S.1.1, supplementary materials. When there is no mark-point dependence ($q = 0$), it appears that the estimation efficiency loss caused by the unnecessary bias correction is minimal. We, therefore, recommend always using proposed bias corrections in practice.

Next, as expected, the naive estimator of $\mu(\cdot)$, that is, $\widetilde{\mu}(\cdot)$, suffers from considerably larger estimation errors than the bias-corrected estimators in all scenarios. To demonstrate this, in Figure 2, we give the mean and 95% quantile bands of 500 estimated functions under the (`Gaussian`, `Gaussian`) setting in the case of $q = 0.8$ and $n = 600$, where there is a larger bias for the naive estimator $\widetilde{\mu}(\cdot)$ than the bias-corrected estimator $\widehat{\mu}(\cdot)$.

Figure 1 suggests that the estimation errors of the bias-corrected estimator $\widehat{C}_Y(\cdot,\cdot)$ are generally smaller than the naive estimator $\widetilde{C}_Y(\cdot,\cdot)$. The improvement is moderate when the mark-point dependence is relatively small (e.g., when $q = 0.5$), but becomes more pronounced when q is larger. To further demonstrate this point, we summarize in Figure 2 the last two eigenfunctions of $\widetilde{C}_Y(\cdot,\cdot)$ and $\widehat{C}_Y(\cdot,\cdot)$ when $n = 600$ and $q = 0.8$, and compare them with those of $C_Y(\cdot,\cdot)$. The first

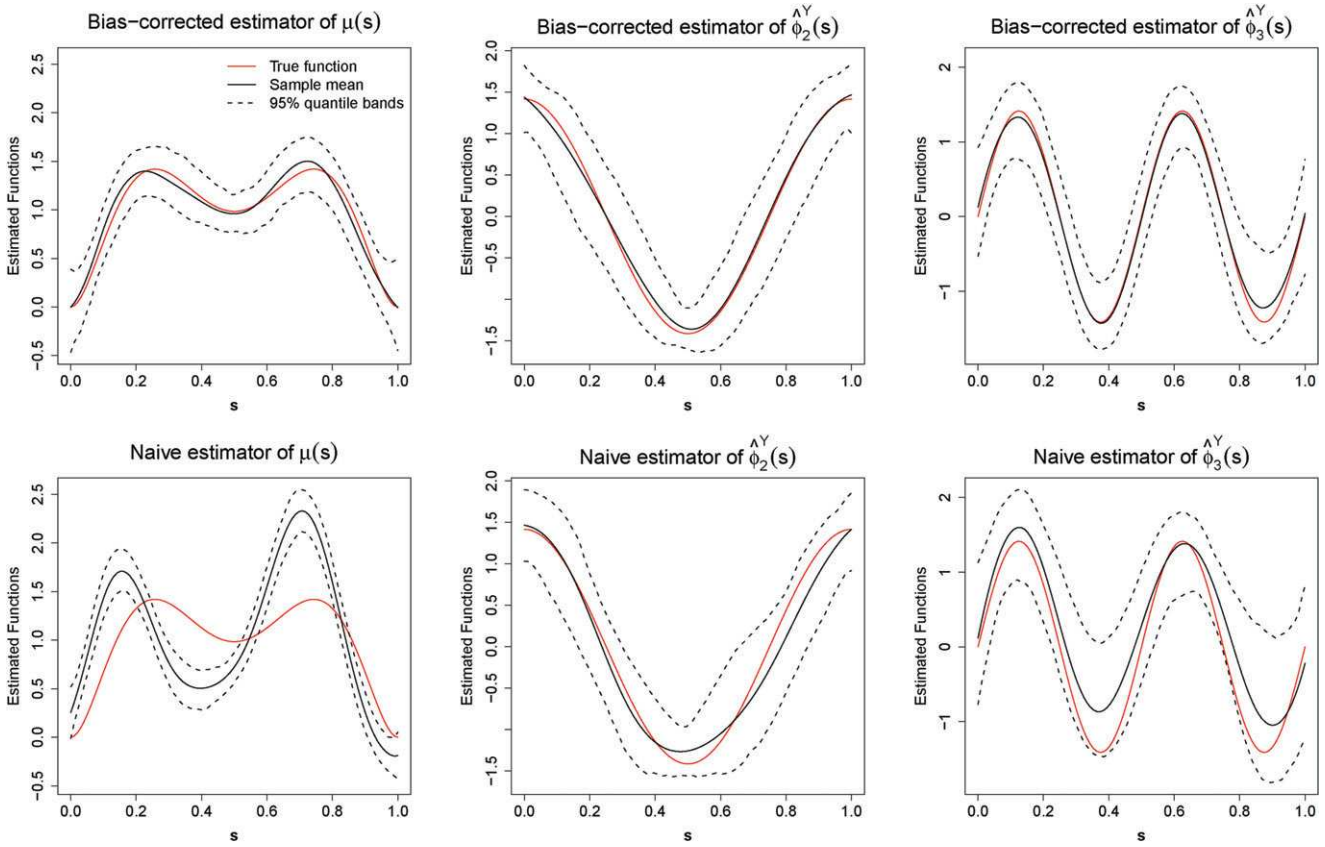

Figure 2. Estimation biases and 95% quantile bands for mean functions and last two eigenfunctions of $C_Y(\cdot,\cdot)$ when $n = 600$ and $q = 0.8$ under the (`Gaussian`, `Gaussian`) setting.

eigenfunctions of $\widetilde{C}_Y(\cdot,\cdot)$ and $\widehat{C}_Y(\cdot,\cdot)$ are similar and thus are not reported. It is clearly seen that $\widetilde{C}_Y(\cdot,\cdot)$ results in biased estimates of the eigenfunctions, but such biases are reduced with $\widehat{C}_Y(\cdot,\cdot)$. Compared to the biases observed in the estimation of $\mu(\cdot)$, the biases in the estimated eigenfunctions from the naive covariance function estimator are less pronounced, even in the case with $q = 0.8$. This suggests that the mark-point dependence may affect the estimation of the mean function $\mu(\cdot)$ more than the covariance function of the mark process.

Finally, additional simulation results for non-Gaussian $X_i(\cdot)$'s and $Y_i(\cdot)$'s are reported in Section S.3.1 of the supplementary materials. In such settings, the estimation bias resulted from the mark-point dependence may not be eliminated, but it can still be effectively reduced using the proposed bias correction procedure.

7.2. Validity of the Proposed Testing Procedures

In this section, we demonstrate the validity of the proposed testing procedures in Sections 4.2 and 6.1. For each simulated dataset, the permutation test is conducted using 500 random permutations. In our simulation settings, the null hypothesis of mark-point independence corresponds to the case with $q = 0$. We set $\sigma_x = 0, 0.5, 0.7, 1$ and vary the sample size n from 100 to 600. For each setting, we summarize test results based on 3000 simulated datasets. For computational efficiency, we first perform bandwidth selection using the cross-validation criteria in Section S.1.1, supplementary materials over 50 simulated datasets, apply the under smoothing corrections described in (S.1) to the average of the selected bandwidths, and hold the bandwidths fixed for all 3000 simulated datasets.

Figures 3 and 4 give empirical rejection rates of the proposed tests under H_0 at significance levels 0.05 and 0.10, respectively. Figures 3 suggest that under the (`Gaussian`, `Gaussian`) setting, the empirical rejection rates of both tests are close to the nominal levels in all scenarios when n is greater than 200. Under H_0, p-values from a valid test should follow Uniform$[0, 1]$. In the middle panels of Figure 3, we provide the empirical distribution of the p-values under the scenario $n = 500$ and $\sigma_x = 1$, which indeed approximately resembles the desired uniform distribution. To quantify the closeness between the empirical distribution $\widehat{F}_p(\cdot)$ of the p-values and the uniform distribution, we also consider the Mallow's distance $d_M(\widehat{F}_p) = \int_0^1 |\widehat{F}_p^{-1}(u) - u| du$, where a smaller value of $d_M(\widehat{F}_p)$ indicates that $\widehat{F}_p$ is closer to the uniform distribution. In the right panels of Figure 3, we show the values of $d_M(\widehat{F}_p)$ as a function of n under various case scenarios. As we can see, when n is sufficiently large, $d_M(\widehat{F}_p)$ falls into the envelope computed by the empirical quantiles of $d_M(\widehat{F}_B^{U,k})$, where $\widehat{F}_B^{U,k}$, $k = 1, \ldots, 1000$, are the empirical CDFs based on $B = 3000$ random numbers drawn from the Uniform$[0, 1]$. This result further supports the validity of the proposed testing procedures.

Figure 4 illustrates the rejection rates of the two tests when the mark process is not Gaussian and/or the point process is not an LGCP. While the asymptotic test still appears to be

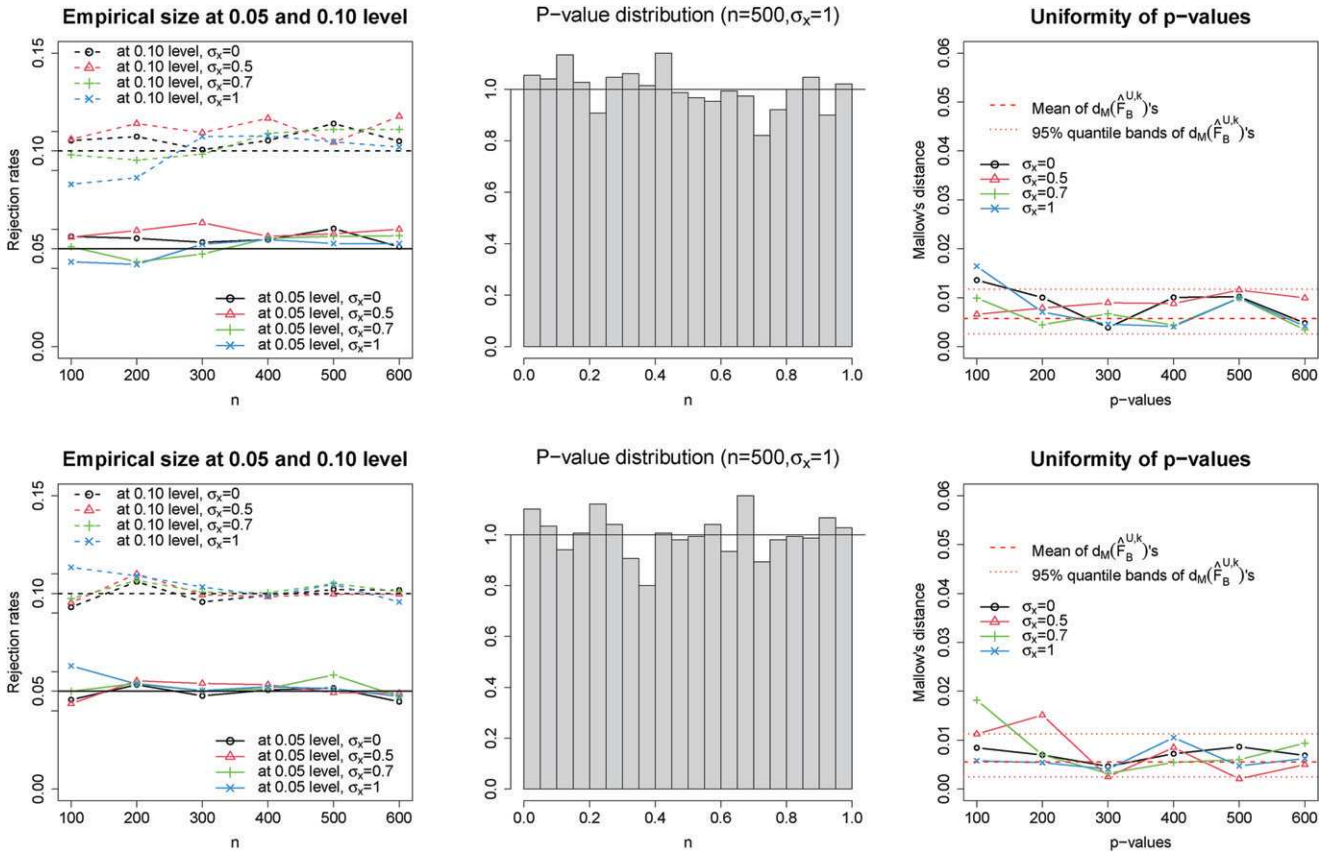

Figure 3. Validity of the proposed tests under the (`Gaussian`, `Gaussian`) setting. Top row: the asymptotic test; Bottom row: the permutation test.

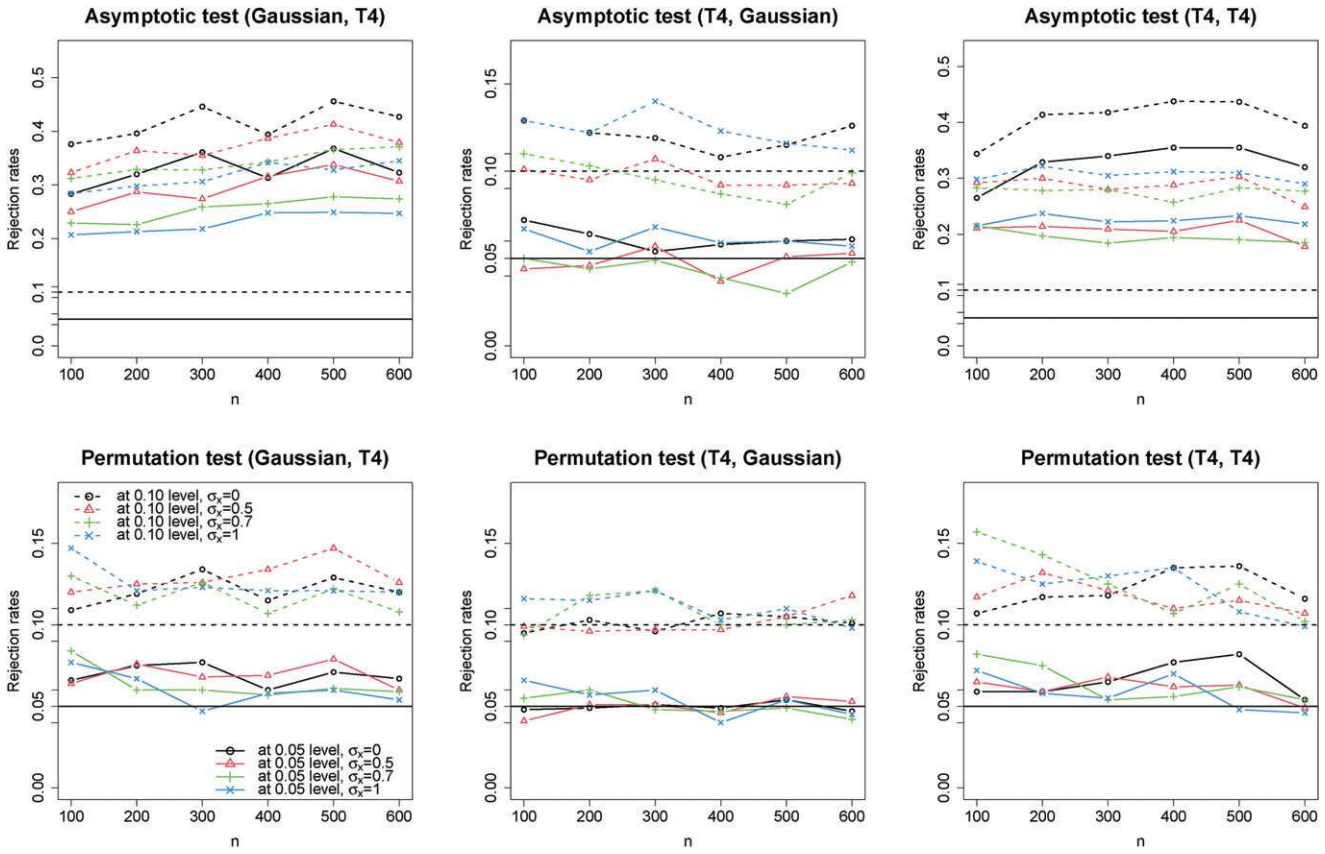

Figure 4. Rejection rates of two tests under mark-point independence in various distribution settings. Top row: the asymptotic test; Bottom row: the permutation test.

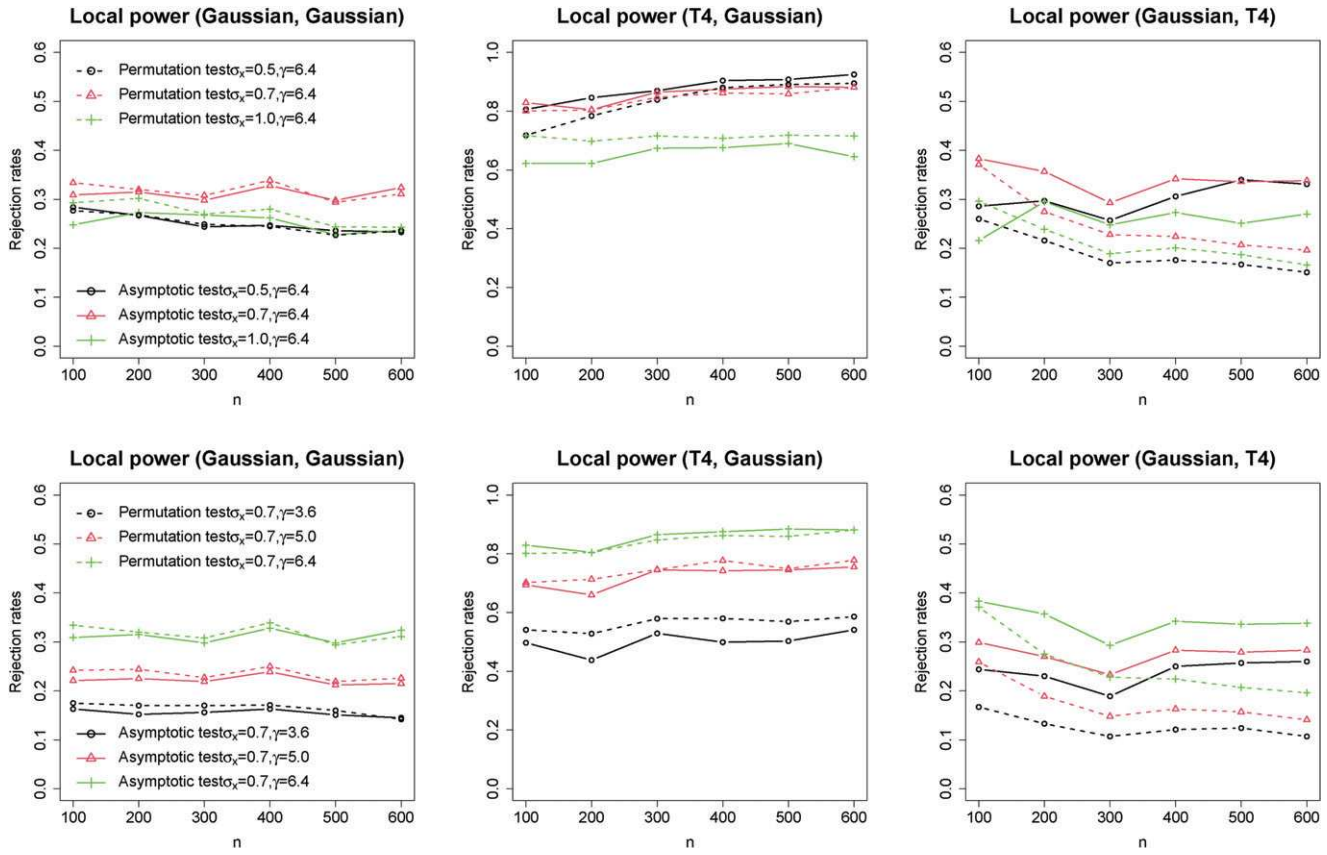

Figure 5. Rejection rates of two tests at the 0.05 level under the local alternatives with varying σ_X (top row) and γ (bottom row).

valid so long as the mark process is Gaussian, its rejection rates significantly exceed the nominal level when the mark process is non-Gaussian. In comparison, the permutation test seems to be close to the nominal sizes provided n is sufficiently large in all settings. Additional and similar simulation results can be found in Section S.3.2 of the supplementary materials.

7.3. Powers Against Local Alternatives

Next, we investigate the power of the proposed tests against local alternatives, which is a popular approach to study the test power, see, for example, Xu and Wang (2011). The asymptotic mean of $\sqrt{n}T_n$ in (27) when $C_{XY}(\cdot,\cdot) \not\equiv 0$ is $\sqrt{n}\int_{\mathcal{T}}\int_{\mathcal{T}} \rho_2(u,v)C_{XY}(u,v)^2 dudv$. Using the simulation setup as in (34) and define the local alternatives by letting $q^2 = n^{-1/2}\gamma$ for some constant $\gamma > 0$. If Theorem 3 holds, the asymptotic mean of $\sqrt{n}T_n$ should be proportional to γ and is independent of n. We set $\gamma = 3.6, 5.0, 6.4$, and refer to the resulting rejection probabilities as the local power. If the limiting distribution of the test statistic is correctly specified, the local power should stay constant as n increases. Data are simulated with varying values of n, σ_x and γ. Figure 5 illustrates the empirical rejection rates at the 0.05 significance level in various settings based on 1000 data replicates.

The left panel of Figure 5 suggests that under the (`Gaussian`, `Gaussian`) setting, the rejection rates of both tests stay roughly constant as n and γ increase. The local powers of both tests first increase as σ_x increases. This is because the magnitude of $C_{XY}(\cdot,\cdot)$ increases with σ_x by our simulation design. Correspondingly, the mean of the test statistic increases as suggested in (25), leading to a greater power. However, if σ_x continues to grow, the power of the proposed test will decrease. This can be explained by the fact that a larger σ_x also leads to a more clustered point process and consequently inflate the variance of the test statistic, which offsets the increase in the mean of the test statistic and thus leads to a reduced power. All these observations support our theoretical findings in Theorem 3 under the (`Gaussian`, `Gaussian`) setting. In other settings shown in Figure 5, the local powers of the permutation test are roughly constant as n increases, provided that n is sufficiently large. This further supports the validity of the proposed functional permutation test in more general settings. Similar results are observed when the `T4` distribution is replaced by the `Exp` distribution; see Section S.3.2 of the supplementary materials.

8. Real Data Analysis

8.1. EMA Smoking Data

In this section, we analyze the situational associations of smoking using the ecological momentary assessment (EMA) data. Smoking is found to be cued or suppressed by immediate situational factors, such as craving, mood, and social settings, and influences of these factors are modulated by gender (Todd 2004; Shiffman and Rathbun 2011). The data we analyze were collected in real-time from 302 smokers over 16 days (Shiffman et al. 2002). The participants were 43% male and 57% female.

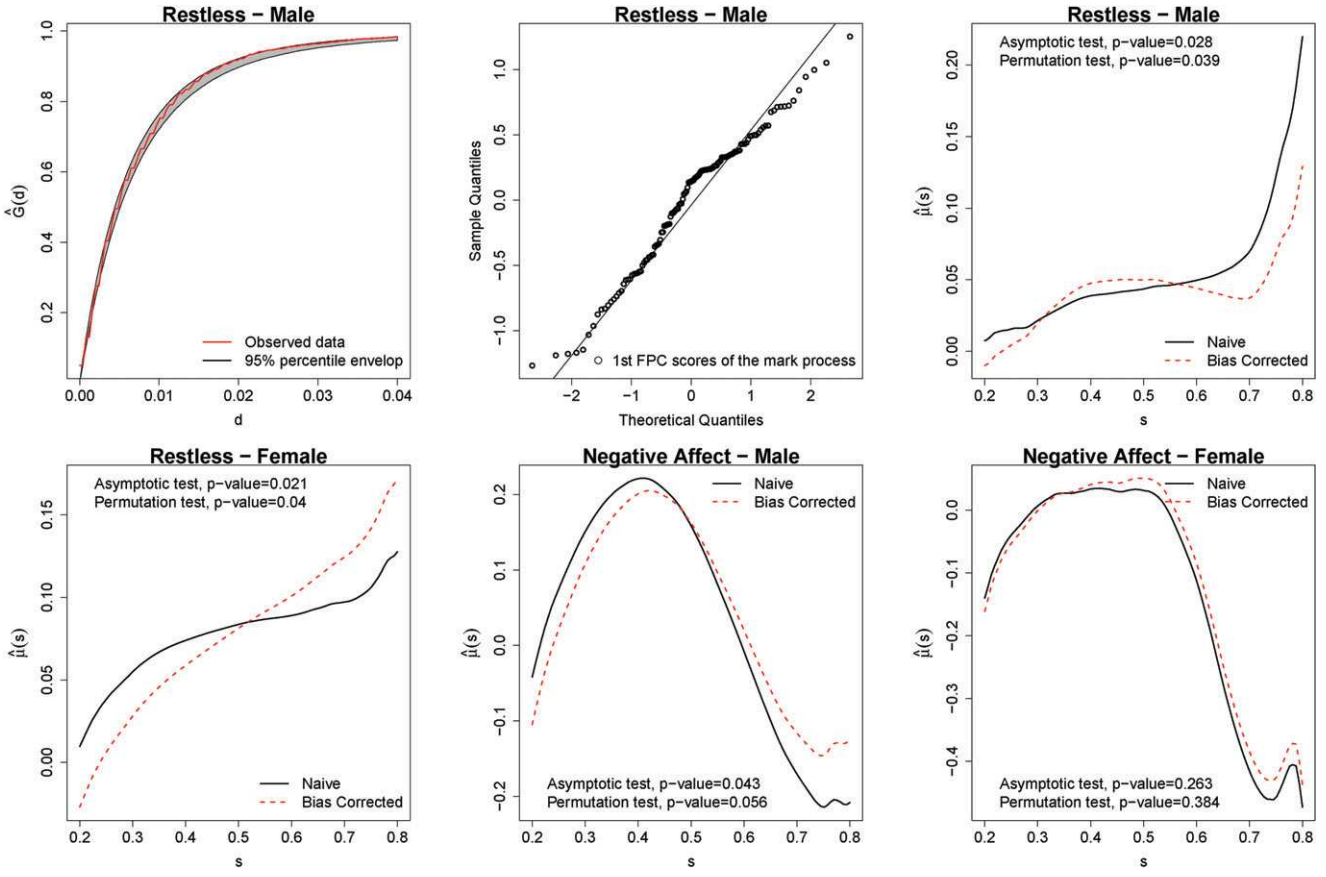

Figure 6. Top left: diagnostic plot for LGCP following Section 6.2; Top middle: QQ plot of the first-FPC scores following Section 6.1; Top right and bottom panel: Estimated daily mean functions of "restless" and "negative affect for male and female participants, respectively.

To be included in the study, a participant had to smoke at least 10 cigarettes per day and had been smoking for at least 2 years. Before data collection, participants were trained to use an Electronic Diary (ED) device designed to collect data in real-time. During the study period, the participants were instructed to record each cigarette in the ED, immediately before smoking. On about 4–5 randomly selected smoking occasions per day, the device administered an assessment. Each assessment provided several continuously measured mood-related ratings such as "negative affect" and "restless," with higher scores indicating more affective distress and stronger feelings of restlessness, respectively. See Shiffman et al. (2002) for more details about the data.

The first few days of monitoring were designed to allow the participants to become familiar with the ED. We therefore focus on days 4–16 in our analysis as suggested in Shiffman et al. (2002). In each day, the observation window for events, that is, $[0, 1]$, corresponded to all waking hours and was not subject-specific, that is, participants were measured on a common time interval; see Shiffman et al. (2002) for more details. For each participant, we aggregate all time points observed on different days into a single day for our analysis. As only 1.56% of total smoking events across all participants were within $[0, 0.2] \bigcup [0.8, 1]$, we focus on the time interval $[0.2, 0.8]$. See Figure S.7 in the supplementary materials for some sample trajectories. We estimate the naive and the bias-corrected estimators of the daily mean functions $\mu(\cdot)$ and covariance functions $C_Y(\cdot, \cdot)$ for two different marks (i.e., "negative affect" and "restless") of male and female participants, respectively; see Figure 6 and Section S.3.3 of the supplementary materials. The bandwidths are selected following the proposed procedures in Section S.1.1, supplementary materials.

Following Section 6.2, for the mark "restless" of the male group, Figure 6 shows that summary statistics from the observed data fell within the 95% percentile envelope based on 1000 simulated point processes, suggesting that the LGCP assumption for the point process is reasonable. The QQ plot of the first FPC scores of the mark process, which account for more than 95% of the total variations, is reasonably close to a straight line, indicating the mark process can be considered as Gaussian. Additional diagnostic plots for other marked point processes are given in Section S.3.3 of the supplementary materials, supporting a similar conclusion.

It is seen from the reported p-values in Figure 6 that at the significance level of 0.10, dependence between "negative affect" and smoking times is significant for males, but is insignificant for females. This is consistent with the findings in Shiffman and Rathbun (2011). On the other hand, "restless" and smoking times are found to have significant associations for both males and females at 0.05 significance level. This is an interesting finding that, to our best knowledge, has not been discussed before in the literature. The estimated mean functions $\mu(\cdot)$'s are illustrated in Figure 6. For "negative affect," it is seen that females have a much lower level of "negative affect" compared to

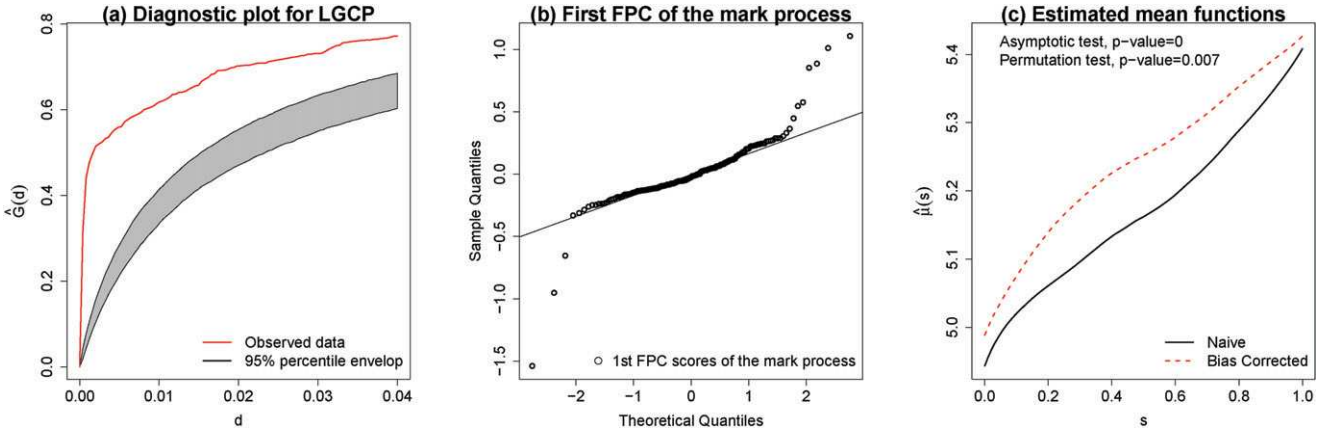

Figure 7. Model diagnostics and estimation for the auction data.

males. For males, there is a notable difference between the naive estimator and the bias-corrected estimator; this is expected as the small p-value indicates significant mark-point dependence.

For "restless," there are notable differences between the naive and bias-corrected estimators for both males and females. In the bias-corrected estimates, male participants are seen to exhibit a steady level of restlessness and then an increasing trend starting around 0.7, while female participants have an increasing level of restlessness throughout the day.

From the naive and bias-corrected estimates of the covariance function $C_Y(\cdot,\cdot)$, we see that the first eigenfunction is flat for both marks and both genders (not plotted). Interestingly, the first eigenfunction accounts for more than 93% of the variability for both marks and both genders. This suggests that the dominant mode of variation for "negative effect" and "restless" is the overall magnitude, and this holds for both male and female participants.

8.2. Ebay Online Auction Data

In this section, we analyze the bid prices of Palm M515 Personal Digital Assistants (PDA) on week-long eBay auctions that took place between March and May of 2003. This dataset contains the bid times and prices of 194 PDA auctions, which has been analyzed by many authors. See Figure S.10 of the supplementary materials for sample trajectories. Earlier analyses (Jank and Shmueli 2006) focused on dynamics of the bid price curves; Wu, Müller, and Zhang (2013) studied the bid times as point processes; Gervini and Baur (2020) modeled the bid times and bid prices jointly, since it was expected that items with lower prices tended to experience "bid sniping" (i.e., concentrated bidding activity close to the end of the auction), but they did not formally test for the mark-point independence.

Our analysis focuses on the last day of the seven-day auction period, which is rescaled to $[0,1]$. We remove subjects with fewer than 2 bids in the last day, resulting in a total sample size of $n = 174$. The bid prices are log-transformed as the mark process. We plot the diagnostic plots described in Section 6.2, and apply the testing procedures for mark-point independence in Sections 4.2 and 6.1. The diagnostic plots in Figure 7(a) and (b) suggest that the point process may not be an LGCP, as the observed nearest-neighbor distance distribution is not contained in the simulation envelope, and the mark process may not be Gaussian, as some departure from normality is seen in the QQ plot. From Figure 7(c), both the asymptotic and permutation tests reject the mark-point independence at the 0.05 level. The validity of the asymptotic test is sensitive to model misspecification, but the permutation test is more robust, as numerically demonstrated in our simulation. On one hand, the agreement between the asymptotic test and the permutation test suggests that one can probably reject the mark-point independence at the 0.05 level, supporting the well-recognized "bid sniping" effects toward the end of an auction. On the other hand, as demonstrated in Section S.3.1 of the supplementary materials, when the point process is not an LGCP and/or the mark process is not Gaussian, the proposed bias correction technique may not be able to eliminate the bias caused by the mark-point dependence and thus one needs to be cautious when interpreting the bias-corrected mean function.

9. Concluding Remarks

In this article, we propose a computationally efficient moment-based bias-correction procedure for estimating the mean and covariance functions of the mark process when there is mark-point dependence. We also propose two inferential procedures, including an asymptotic test and a functional permutation test, for testing the mark-point independence assumption.

While we focus on detecting mark-point dependence and correcting estimation bias caused by potential mark-point dependence, our work opens doors to a series of meaningful research directions such as bias-corrected statistical inference of functional principal component scores and bias-corrected testing of the mean functions of two or more groups (e.g., treatment vs. control, male vs. female). It is also of interest to extend the current methods to mark-point process data with more complicated structures, see, for example, Xu et al. (2020) and Yin et al. (2021).

Moreover, although our work has focused on temporal mark-point processes, it can be readily extended to model spatial mark-point processes if independent replicates are available, which requires using multi-dimensional kernel or product kernel functions. For future research, it is also of interest to consider the case where the marks are non-Gaussian (e.g., Poisson, Bino-

mial). In the case of Poisson marks, we can assume that

$$Z(s) \sim \text{Poisson}\left(\exp\left[\mu(s) + Y(s)\right]\right), \quad s \in \mathcal{T}, \tag{36}$$

where $\mu(\cdot)$ and $Y(\cdot)$ are defined similarly as those in (4). Our proposed framework can be extended to derive unbiased estimators for $\mu(\cdot)$ and $C_Y(\cdot,\cdot)$, that is, the covariance function of $Y(\cdot)$ in (36). See Section S.2 of the supplementary materials for more details. It is also of great interest to theoretically justify the functional permutation test for mark-point independence proposed in Section 6.1. We plan to investigate such directions in our future research.

Supplementary Materials

The supplementary material contains method implementation details, additional numerical results, and technical proofs.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

Zhang's research is supported by NSF grant DMS-2015190 and Guan's research is supported by NSF grant DMS-1810591.

References

Chen, K., and Müller, H.-G. (2012), "Modeling Repeated Functional Observations," *Journal of the American Statistical Association*, 107, 1599–1609. [217]

Diggle, P. J., Menezes, R., and Su, T.-l. (2010), "Geostatistical Inference under Preferential Sampling," *Journal of the Royal Statistical Society*, Series C, 59, 191–232. [218]

Fok, C. C. T., Ramsay, J. O., Abrahamowicz, M., and Fortin, P. (2012), "A Functional Marked Point Process Model for lupus Data," *Canadian Journal of Statistics*, 40, 517–529. [217,218]

Gervini, D., and Baur, T. J. (2020), "Joint Models for Grid Point and Response Processes in Longitudinal and Functional Data," *Statistica Sinica*, 30, 1905–1924. [217,218,230]

Guan, Y., and Afshartous, D. R. (2007), "Test for Independence between Marks and Points of Marked Point Processes: A Subsampling Approach," *Environmental and Ecological Statistics*, 14, 101–111. [218]

Hsing, T., and Eubank, R. (2015), *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*, Chichester: Wiley. [217]

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns*, Chichester: Wiley. [217]

Jank, W., and Shmueli, G. (2006), "Functional Data Analysis in Electronic Commerce Research," *Statistical Science*, 21, 155–166. [230]

Li, Y., and Hsing, T. (2010), "Uniform Convergence Rates for Nonparametric Regression and Principal Component Analysis in Functional/Longitudinal Data," *The Annals of Statistics*, 38, 3321–3351. [219,222]

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998), "Log Gaussian Cox Processes," *Scandinavian Journal of Statistics*, 25, 451–482. [217]

Møller, J., and Waagepetersen, R. P. (2003), *Statistical Inference and Simulation for Spatial Point Processes*, Boca Raton, FL: CRC Press. [224]

Schlather, M., Ribeiro, Paulo J., J., and Diggle, P. J. (2004), "Detecting Dependence between Marks and Locations of Marked Point Processes," *Journal of the Royal Statistical Society*, Series B, 66, 79–93. [218]

Shiffman, S., Gwaltney, C., Balabanis, M., Liu, K., Paty, J., Kassel, J., Hickcox, M., and Gnys, M. (2002), "Immediate Antecedents of Cigarette Smoking," *Journal of Abnormal Psychology*, 111, 531–545. [228,229]

Shiffman, S., and Rathbun, S. L. (2011), "Point Process Analyses of Variations in Smoking Rate by Setting, Mood, Gender, and Dependence," *Psychology of Addictive Behaviors*, 25, 501–510. [228,229]

Todd, M. (2004), "Daily Processes in Stress and Smoking: Effects of Negative Events, Nicotine Dependence, and Gender," *Psychology of Addictive Behaviors*, 18, 31–39. [228]

Wang, J., Wong, R. K. W., and Zhang, X. (2021), "Low-Rank Covariance Function Estimation for Multidimensional Functional Data," *Journal of the American Statistical Association* (to appear). [217]

Wu, S., Müller, H.-G., and Zhang, Z. (2013), "Functional Data Analysis for Point Processes with Rare Events," *Statistica Sinica*, 23, 1–23. [230]

Xu, G., Wang, M., Bian, J., Huang, H., Burch, T. R., Andrade, S. C., Zhang, J., and Guan, Y. (2020), "Semi-parametric Learning of Structured Temporal Point Processes," *The Journal of Machine Learning Research*, 21, 7851–7889. [230]

Xu, G., and Wang, S. (2011), "A Goodness-of-Fit Test of Logistic Regression Models for Case–Control Data with Measurement Error," *Biometrika*, 98, 877–886. [228]

Yao, F., Müller, H.-G., and Wang, J.-L. (2005), "Functional Data Analysis for Sparse Longitudinal Data," *Journal of the American Statistical Association*, 100, 577–590. [217,219,222]

Yin, L., Xu, G., Sang, H., and Guan, Y. (2021), "Row-clustering of a Point Process-valued Matrix," *Advances in Neural Information Processing Systems*, 34, 20028–20039. [230]

Zhang, T. (2014), "A Kolmogorov-Smirnov Type Test for Independence between Marks and Points of Marked Point Processes," *Electronic Journal of Statistics*, 8, 2557–2584. [218]

——— (2017), "On Independence and Separability between Points and Marks of Marked Point processes," *Statistica Sinica*, 27, 207–227. [218]

Zhang, X., and Wang, J. L. (2016), "From Sparse to Dense Functional Data and Beyond," *The Annals of Statistics*, 44, 2281–2321. [217]