On Functional Processes with Multiple Discontinuities

Jialiang Li*

Department of Statistics and Applied Probability

National University of Singapore

Yaguang Li

School of Management, University of Science and Technology of China

Tailen Hsing

Department of Statistics, University of Michigan, USA

Abstract

We consider the problem of estimating multiple change points for a functional data process. There are numerous examples in science and finance in which the process of interest may be subject to some sudden changes in the mean. The process data that are not in a close vicinity of any change point can be analyzed by the usual nonparametric smoothing methods. However, the data close to change points and contain the most pertinent information of structural breaks need to be handled with special care. This paper considers a half-kernel approach that addresses the inference of the total number, locations, and jump sizes of the changes. Convergence rates and asymptotic

*Email:stalj@nus.edu.sg

distributional results for the proposed procedures are thoroughly investigated. Simulations are conducted to examine the performance of the approach, and a number of real data sets are analyzed to provide an illustration.

Keywords: change point, functional data, convergence rate, kernel smoothing, phase transition, structural break

1 Introduction

Nonparametric techniques for functional data analysis (FDA) are well established and widely applied in practice. See Ramsay and Silverman (2005), Horváth and Kokoszka (2012), Hsing and Eubank (2015), and Wang et al. (2015) for an overview. For identification of meaningful subgroups and distinct patterns in functional data, earlier works include Chiou and Li (2007), Chiou and Müller (2014) and Chiou (2012). In this paper we attempt to identify different sub-regions of the functional process delineated by discontinuities of the mean function. This focus is different from those earlier discussion on *change points* with functional data.

Many scientific and business inquiries require the modelling of functional data with multiple change points. For example, a phase of a thermodynamic system and the states of matter have uniform physical properties. During a phase transition of a given medium, certain properties of the medium change, often discontinuously, as a result of the change of some external condition such as temperature or pressure. Another example is the financial and economic forecasting problem where two consecutive business cycles may each be described by a smooth process but disconnected at the joining time caused by some unexpected drastic change in economic outlook. Identification of the change-point structure for a period with multiple business cycles may improve the forecasting for production, prices, interest rates and other financial measurements.

We first fix our scope for functional data in this paper. In a typical functional-data setting, a sample of n curves are observed at a set of discrete, cross-sectional points; denote by m_i the number of observations for curve i, $i = 1, \cdots, n$. Generally

speaking, there are two different sample settings (Cai and Yuan, 2011). Let T_{ij} denote the location of the j-th observation for the i-th curve. In a *common design*, the observations are sampled at the same locations across curves, that is, $T_{1j} = T_{2j} = \cdots = T_{nj} =: T_j$ for all $j = 1, \ldots, m$; mostly likely, the T_j are points on a fixed grid. A more general setting is when the T_{ij} are independently sampled from a bounded interval $[a, b]$, which we shall refer to as *independent design*. One may also distinguish between dense and sparse functional data (see, e.g., Li and Hsing, 2010; Zhang and Wang, 2016), which roughly correspond to whether each m_i is larger than some power of n or all m_i's are bounded by a finite positive number. Our theoretical results will focus on the independent design setting since doing so allows us to treat the sparse and dense settings in a unified manner. However, we note that when the data are dense, the differences between the common design and independent design settings are small both in theory and computations.

Numerous studies have focused on detecting change points in the mean for univariate series. Wu and Chu (1993) proposed a kernel-type estimators of jump points in the fixed-design nonparametric regression model. Csörgö and Horváth (1997) provided an account of early results in change-point detection mainly for data with independent and identically distributed (i.i.d.) error terms. Zhang (2016) discussed a nonparametric test for change-point detection under dependent and non-stationary errors. Górecki et al. (2018) proposed a test for shifts in the mean of a heteroscedastic time series which accommodates the possibility of changing variability. See Qiu et al. (1991), Spokoiny et al. (1998), Qiu (2005) and Gijbels et al. (2007) for additional results on nonparametric change-point detection. For multivariate high dimensional dependent series, Xu et al. (2016) considered detecting change points for the variance for blocked time series and dependent panel data. Gromenko et al. (2017) extended the cumulative sum paradigm test statistics for detecting whether the pattern of a spatio-temporal process has changed over a time period. Wang and Samworth (2018) studied the change point in high-dimensional time series of which the mean structure changes in a sparse subset of the coordinates. Recently, Li et al. (2019) proposed a new nonparametric procedure to detect

change points in the mean of high-dimensional time series data taking into account both spatial and temporal dependence.

The research on change-point detection for functional data only began recently and most of the existing works focus on the change-point detection in the mean of functional observations. For example, Berkes et al. (2009) considered a (single) break in the mean function assuming that the error terms are i.i.d. curves, and Aue et al. (2009) assumed that the change size can be both a constant or approaching zero as the sample size tends to infinity. The results in Berkes et al. (2009) and Aue et al. (2009) were generalized by Aston and Kirch (2012) to deal with an epidemic change model, where after the initial change the mean changes back to the original level after some time. Aue et al. (2014) studied the change of regression operator in a functional regression model, and Aue et al. (2018) developed a functional test statistics for detecting and dating structural breaks in functional data. Bardsley et al. (2017) introduced several asymptotic methods to test the null hypothesis that the mean structure of a sequence of curves does not change by taking into account their range and other aspects of their shapes. These aforementioned works on change points in the context of functional data all have quite different focuses from ours in this paper.

Most of the earlier works on change point detection focused on single change point estimation. The problem of multiple change points is more challenging as the estimation of the unknown number of change points adds an extra layer of difficulty. Recent research has paid more attention to computational and theoretical development for multiple change point. Fryzlewicz (2014) considered a wild binary segmentation method for consistent estimation of the number and locations of multiple change points for independent data. Li and Jin (2018) introduced another approach for multiple change point detection where a non-concave penalization approach was adopted. Their framework can be extended to covariate change point in regression analysis. However, these approaches are only suitable for independent data with parametric functionals. For discontinuous nonparametric curve estimation, Xia and Qiu (2015) proposed a jump information criterion to consistently detect the number

of jumps through kernel smoothing. Their work is only pertinent to cross-sectional data and cannot be directly applied to functional data.

In this paper, we introduce a nonparametric framework for change-point analysis based on functional data. We detect the multiple jump discontinuities in the mean function using a local linear smoother that employs a one-sided kernel. After the jump locations are identified, we can then estimate the jump sizes as well as the smooth component of the mean function and the covariance function. While the within-curve dependency in the functional-data setting poses extra challenges compared with nonparametric regression, it also leads to richer theoretical developments. We will prove a range of convergence rates for our estimates depending on the denseness of the observation points on each curve. In particular, our results for the jump location/size estimation show a "phase transition" of rates as observations per curve changes from sparse to dense. This is akin to the results on mean function estimation in, e.g., Li and Hsing (2010); Cai and Yuan (2011); Zhang and Wang (2016). Moreover, the asymptotic normality of the mean estimator is also discussed after the change points are consistently detected, where a pointwise confidence band can be constructed for the estimated mean function. Our results are both general and sharp. Different types of sampling weights are considered and our convergence rates reflect clearly the roles of the smoothing parameters. Furthermore, the functional process is not limited to the Gaussian process and can have heavy-tailed marginal distributions, as demonstrated in the simulation studies.

This paper is organized as follows. In Section 2, we introduce the model and data structure as well as all the estimation procedures. We establish the theoretical properties of the procedures in Section 3, where we also discuss the optimal rates of change point estimation and their connections to prominent results in the literature. Some simulation studies are provided in Section 4, and real data applications in Section 5. All proofs are included in Appendix.

2 Model Assumptions and Methodology

2.1 Model Assumptions

Let $\{X(t), t \in [a,b]\}$ be a stochastic process defined on a fixed bounded interval $[a,b]$. Without loss of generality, we assume that $[a,b] = [0,1]$. Denote the mean and covariance function of $X(t)$ by

$$\mu(t) = \mathbb{E}\{X(t)\}, \qquad R(s,t) = \mathrm{cov}\{X(s), X(t)\},$$

which are assumed to exist. Let

$$U(t) := X(t) - \mu(t)$$

denote the zero-mean stochastic component of $X(t)$.

The goal is to estimate the mean and covariance function based on noisy functional data $\{(Y_{ij}, T_{ij}), i = 1, \ldots, n, j = 1, \ldots, m_i\}$, where

$$Y_{ij} = X_i(T_{ij}) + e_{ij} = \mu(T_{ij}) + U_i(T_{ij}) + e_{ij}, \tag{1}$$

where the X_i's are i.i.d. realizations of X, the T_{ij}'s are i.i.d. random observational points with density function $f_T(\cdot)$, and the e_{ij}'s are i.i.d. random errors with mean zero and finite variance σ^2. The index T may be the temperature variable in a physics research study, the time variable in a financial time series, or the location variable in a spatial or image data. In our asymtotic theory, we allow m_i to vary with n and we will address scenarios where the m_i's are bounded or tend to ∞.

The mean function μ is a usually assumed to be smooth in the functional data literature (cf. Li and Hsing, 2010; Zhang and Wang, 2016). However, as explained in Section 1, there are scenarios where it would be more realistic to allow μ to have discontinuities. We consider the situation that the mean function $\mu(t)$ have jump discontinuities at M different locations, $\tau_1, \tau_2, \ldots, \tau_M \in (0,1)$, such that

$$\mu(t) = \nu(t) + \sum_{k=1}^{M} d_k I(t \geq \tau_k), \tag{2}$$

where $\nu(t)$ is a smooth function with a bounded second-order derivative, and $\{d_k : k = 1, \ldots, M\}$ are the jump sizes. For convenience, assume that $|d_1| > |d_2| > \cdots > |d_M| > 0$. Our goal is to conduct inference on the $(\tau_k, d_k), M, \nu$ and R.

2.2 Mean function estimation

Since the mean function $\mu(t)$ is smooth except at the change points $\tau_1, \ldots, \tau_M$, one may adopt the familiar local linear regression estimator (Fan and Gijbels, 1996) to deal with the smooth part of μ. Let $K(\cdot)$ be a bounded symmetric kernel function with support $[-1, 1]$ and $K_h(t) = (1/h)K(t/h)$ where h is bandwidth. Recall that the local linear estimator $\widehat{\mu}(t; h)$ of $\mu(t)$ in Li and Hsing (2010) can be obtain as $\widehat{a}_0$, where

$$(\widehat{a}_0, \widehat{a}_1) = \underset{a_0, a_1}{\operatorname{argmin}} \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \{Y_{ij} - a_0 - a_1(T_{ij} - t)\}^2 K_h(T_{ij} - t). \tag{3}$$

and the weight w_i is attached to each observation for the ith subject such that $\sum_{i=1}^{n} m_i w_i = 1$. It follows that

$$\widehat{\mu}(t; h) = \frac{R_0(t)S_2(t) - R_1 S_1(t)}{S_0(t)S_2(t) - S_1^2(t)}, \tag{4}$$

where

$$\begin{aligned} S_r(t) &= \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_h(T_{ij} - t)\{(T_{ij} - t)/h\}^r, \\ R_r(t) &= \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_h(T_{ij} - t)\{(T_{ij} - t)/h\}^r Y_{ij}, \ r = 0, 1, 2. \end{aligned} \tag{5}$$

This estimator was studied in detail in Li and Hsing (2010) and Zhang and Wang (2016) assuming that $\mu(t)$ has bounded second derivative. Since $\widehat{\mu}(t; h)$ is a local estimator, so long as t is not within a distance of h from any τ_k, the properties of $\widehat{\mu}(t; h)$ are identical to those found in Li and Hsing (2010) and Zhang and Wang (2016). Thus, one approach to estimating the τ_k might be to focus on the locations where $\widehat{\mu}$ changes rapidly. However, this is not an effective approach.

Let

$$Y_{ij}^* = Y_{ij} - \sum_{k=1}^{M} d_k I(T_{ij} \geq \tau_k) \tag{6}$$

for which the mean function is $\nu(t)$ by (2). Then the local linear estimator is

$$\widehat{\nu}^*(t; h) = \frac{R_0^*(t)S_2(t) - R_1^*(t)S_1(t)}{S_0(t)S_2(t) - S_1^2(t)},$$

where $R_r^*(t)$ is obtained by replacing Y_{ij} with Y_{ij}^* in $R_r(t)$. We can refer to $\widehat{\nu}^*(t;h)$ the "oracle" estimator of $\nu(t)$ as we assume the jump locations and sizes are known. Clearly,

$$R_r(t) - R_r^*(t) = \sum_{i=1}^{n}\sum_{j=1}^{m_i} K_h(T_{ij}-t)\left(\frac{T_{ij}-t}{h}\right)^r \sum_{k=1}^{M} d_k I(T_{ij}\geq \tau_k). \tag{7}$$

Observe that, for t not within a distance of h from any τ_k, we have $\widehat{\mu}(t;h)-\widehat{\nu}^*(t;h) = \sum_k d_k I(t\geq\tau_k)$ and therefore

$$\widehat{\mu}(t;h) - \mu(t) = \widehat{\nu}^*(t;h) - \nu(t). \tag{8}$$

For t within a distance of h from some τ_k, the relationship between $\widehat{\mu}(t;h)$ and $\widehat{\nu}^*(t;h)$ is more complicated and we will defer that to Section 3. In the rest of the paper, we focus on the case where the (τ_k, d_k) are unknown.

2.3 Estimating discontinuities

In this subsection, we consider estimating the unknown locations and sizes of the discontinuities of μ using a so-called half-kernel approach. Asymmetric, one-sided, kernels were sometimes used to address the boundary issues in the early literature of kernel smoothing (cf. Gasser and Muller, 1984; Rice, 1984). Qiu et al. (1991) and Xia and Qiu (2015) considered a difference kernel estimator based one-sided local linear regression to detect discontinuities in the mean regression curve. Wu and Zhao (2007) proposed a test statistic using the difference between the left and right local averages of observations for identifying the existence of structural breaks in trends. Recently Cao et al. (2016) used the half-kernel approach to handle asynchronous longitudinal data.

Let

$$K_-(t) = K(t)\mathbf{1}_{\{t\in[-1,0)\}} \quad \text{and} \quad K_+(t) = K(t)\mathbf{1}_{\{t\in[0,1]\}}$$

be the two one-sided kernel functions corresponding to the symmetric kernel function $K(\cdot)$. Denote

$$K_{h,\ell}(t) = (1/h)K_\ell(t/h), \ell = \pm.$$

for a bandwidth h. Then, for a given interior point $t \in [h, 1-h]$, by using the data points in the two halves $[t-h, t)$ and $[t, t+h]$ of the neighborhood $[t-h, t+h]$ we obtain the two one-sided local linear estimators $\widehat{\mu}_-(t;h)$ and $\widehat{\mu}_+(t;h)$, which are simply obtained in the same way as (3) and (4) using K_- and K_+ as the kernels. Now, define

$$\Delta_n(t;h) = \widehat{\mu}_+(t;h) - \widehat{\mu}_-(t;h). \tag{9}$$

Note that if $[t-h, t)$ does not contain any τ_k, then, as a special case of $\widehat{\mu}(t;h)$, $\widehat{\mu}_-(t;h)$ estimates $\mu(t)$ and its asymptotic properties are no different from the general local linear estimator for continuous $\mu(t)$; the same can be said for $\widehat{\mu}_+(t)$ if $[t, t+h]$ does not contain any τ_k. In particular, if $[t-h, t+h]$ does not contain any τ_k, then both $\widehat{\mu}_-(t;h), \widehat{\mu}_+(t;h)$ estimate $\mu(t)$ under suitable conditions and therefore $\Delta_n(t;h)$ is expected to be small. On the other hand, as we shall see below, if a point t is within a distance h of some τ_k, say $t = \tau_k + \lambda h$ for some $\lambda \in [-1, 1]$ and a suitable bandwidth h, then $\widehat{\mu}_-(t;h)$ and $\widehat{\mu}_+(t;h)$ would be estimating quantities that differ roughly by the jump size. These ideas motivate the approach to be introduced below.

For clarity, formally define the h-neighborhoods of the τ_k as follows. Let

$$D_h = [h, 1-h], \quad D_{h,k} = [\tau_k - h, \tau_k + h],\ k = 1, \ldots, M,$$

and

$$D_{h,M_1 \to M_2} = \bigcup_{k=M_1}^{M_2} D_{h,k}, \quad \bar{D}_{h,M_1 \to M_2} = D_h \backslash D_{h,M_1 \to M_2},\ 1 \le M_1 \le M_2 \le M.$$

As explained earlier, $\bar{D}_{h,1\to M}$ contains no change points and the detection criterion $|\Delta_n(t;h)|$ is expected to be small in $\bar{D}_{h,1\to M}$. On the other hand, $|\Delta_n(t;h)|$ is expected to reflect the jump sizes in any of the $D_{h,k}$.

We first consider the estimation of the τ_k. To do so we begin by assuming that M is known but will later return to the case of unknown M. Let

$$\widehat{\tau}_1 = \widehat{\tau}_1(h_\tau) := \underset{t \in D_{h_\tau}}{\operatorname{argmax}} |\Delta_n(t;h_\tau)| \tag{10}$$

be the first change point candidate, where h_τ is the bandwidth used for this step. Similarly, $\widehat{\tau}_2$ is the maximizer of $|\Delta_n(t;h_\tau)|$ in the set $D_{h_\tau}\backslash\widehat{D}_{\epsilon h_\tau,1}$, where $\widehat{D}_{\epsilon h_\tau,1} = [\widehat{\tau}_1 - \epsilon h_\tau, \widehat{\tau}_1 + \epsilon h_\tau]$, with $\epsilon > 1$, is the region excluding the ϵh_τ-neighborhood of the first identified change point. There is flexibility in choosing ϵ but for convenience we will let $\epsilon = 2$ in our procedure. The recursive formula of $\widehat{\tau}_k$ is

$$\widehat{\tau}_k = \widehat{\tau}_k(h_\tau) := \underset{t\in D_{h_\tau}\backslash\bigcup_{i=1}^{k-1}\widehat{D}_{\epsilon h_\tau,i}}{\operatorname{argmax}} |\Delta_n(t;h_\tau)|, \quad k = 1,\dots M, \tag{11}$$

where $\widehat{D}_{\epsilon h_\tau,i} = [\widehat{\tau}_i - \epsilon h_\tau, \widehat{\tau}_i + \epsilon h_\tau]$.

For estimating d_k, an obvious candidate is $\Delta_n(\widehat{\tau}_k;h_\tau)$ (cf. Xia and Qiu, 2015). However, this approach does not achieve the optimal rate in some cases for functional data. Thus, we propose the following modified approach. Theorem 3.2 shows that $|\widehat{\tau}_k - \tau_k| = O(h_\tau^2 + h_\tau\varsigma_n(h_\tau))$, where

$$\varsigma_n(h) = \left\{\log n \sum_{i=1}^{n} m_i\big((m_i-1)+1/h\big)w_i^2\right\}^{1/2}. \tag{12}$$

Thus, if we pick ρ_d to be of a bigger order of magitude than $h_\tau^2 + h_\tau\varsigma_n(h_\tau)$, then $\widehat{\tau}_k - \rho_d < \tau_k$ and $\widehat{\tau}_k + \rho_d > \tau_k$ for large n. This motivates the estimator

$$\widehat{d}_k(h_\tau, h_d, \rho_d) = \widehat{\mu}_+(\widehat{\tau}_k + \rho_d; h_d) - \widehat{\mu}_-(\widehat{\tau}_k - \rho_d; h_d). \tag{13}$$

The considerations for selecting h_τ and h_d are not entirely the same, as will be seen in Theorem 3.2 and the discussions in Section 3.3.

After obtaining the estimates $\widehat{\tau}_k, \widehat{d}_k$, we can define the following estimate for the mean function

$$\widehat{\mu}(t;h_\tau,h_d,\rho_d) = \widehat{\nu}(t;h_\tau) + \sum_{k=1}^{M}\widehat{d}_k(h_\tau,h_d,\rho_d)I(t\ge\widehat{\tau}_k(h_\tau)) \tag{14}$$

where $\widehat{\nu}(t;h_\tau)$ is the local linear estimator using bandwidth h_τ, with Y_{ij} replaced by

$$\widetilde{Y}_{ij}(h_\tau,h_d,\rho_d) := Y_{ij} - \sum_{k=1}^{M}\widehat{d}_k(h_\tau,h_d,\rho_d)I(T_{ij}\ge\widehat{\tau}_k(h_\tau)).$$

Note that the same notation $\widehat{\mu}$ is used in (4) and (13) to streamline notation. However, we distinguish the two by the different tuning parameters.

In applications, M is generally unknown and has to be estimated. A simple and effective way to determine M is to leverage the idea that the uniform rate of $\Delta(t;h_\tau)$ at continuity points is $h_\tau^2+\varsigma_n(h_\tau)$ (cf. Theorem 3.1) under rather general conditions, with ς_n defined by (12). Thus, one can proceed in the same way as before in estimating τ_k but would stop whenever

$$\sup_{t\in D_{h_\tau}\setminus\bigcup_{i=1}^{k-1}\widehat{D}_{\epsilon h_\tau,i}}|\Delta_n(t;h_\tau)|<\text{ some threshold }\zeta_n \tag{15}$$

and declare $\widehat{M}=k-1$, where ζ_n is of a bigger order of magnitude than $h_\tau^2+\varsigma_n(h_\tau)$. We will establish that $\widehat{M}\to M$ with probability one under suitable conditions (cf. Theorem 3.4).

Remark 1. *(i) After obtaining the mean estimate $\widehat{\mu}(t;h_\tau,h_d,\rho_d)$ by (14), the corresponding covariance estimator $\widehat{R}(s,t;h_\tau,h_d,\rho_d,h_R)$ can be obtained following the approaches in Li and Hsing (2010) and Zhang and Wang (2016) by a local linear smoother applied to the raw residual covariances, $C_{ijl}=[Y_{ij}-\widehat{\mu}(T_{ij})][Y_{il}-\widehat{\mu}(T_{il})]$ with another bandwidth h_R. Briefly, choose some appropriate weight w_i for each C_{ijl} such that $\sum_{i=1}^n m_i^2\mathsf{w}_i=1$ and let $\widehat{R}(s,t)=\widehat{b}_0$, where*

$$(\widehat{b}_0,\widehat{b}_1,\widehat{b}_2)=\underset{b_0,b_1,b_2}{\operatorname{argmin}}\sum_{i=1}^n\Bigg[\mathsf{w}_i\sum_{l\neq j}\{C_{ijl}-b_0-b_1(T_{ij}-s)-b_2(T_{il}-t)\}^2 \times K_{h_R}(T_{ij}-s)K_{h_R}(T_{il}-t)\Bigg], \tag{16}$$

with $\sum_{l\neq j}$ denoting sum over all $l,j=1,\ldots,m_i$ such that $l\neq j$. We also note that for dense functional data, after obtaining $\widehat{\mu}(t)$, it is possible to estimate the individual U_i, the stochastic component of X_i, by smoothing the residuals $\widehat{\varepsilon}_{ij}=Y_{ij}-\widehat{\mu}(T_{ij}), i=1,\ldots,m_i$.

(ii) To implement the procedures described above, a fully data-driven method for the tuning parameters is appealing. In general, cross validation is a widely accepted approach for selecting bandwidths and can be adopted here. We will postpone the details of this until Section 4.1.

3 Theoretical Properties

In this section, we conduct a theoretical investigation of the properties of the procedures introduced in Section 2. Standard notations for asymptotic relations between two real sequences $\{a_n\}$ and $\{b_n\}$ will be used. For instance, $a_n \lesssim b_n$ means $a_n = O(b_n)$; $a_n \prec b_n$ and $a_n \succ b_n$ mean $a_n = o(b_n)$ and $b_n = o(a_n)$, respectively; $a_n \asymp b_n$ stands for $a_n \lesssim b_n$ and $b_n \lesssim a_n$. Also, $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$ for any two real numbers a, b.

Throughout the rest of the paper we always assume that $h = h_n \to 0$.

3.1 Technical assumptions

For clarity, we collect most of the technical conditions in this subsection.

Let $\psi(x)$ be a function on $(0, \infty)$ such that $\psi(x)/x$ is monotone increasing for large enough x. For a kernel function K, define

$$g(\lambda) = g_K(\lambda) := \int_{|\lambda|}^1 K_+^*\big(u\big)du, \ \lambda \in [-1, 1], \tag{17}$$

where

$$K_\ell^*\big(u\big) = K_\ell(u)\frac{\upsilon_{\ell,2} - \upsilon_{\ell,1}u}{\upsilon_{\ell,0}\upsilon_{\ell,2} - \upsilon_{\ell,1}^2} \quad \text{and} \quad \upsilon_{\ell,r} = \int_{-1}^1 t^r K_\ell(t)dt, \ \ell = +, -. \tag{18}$$

Also, define

$$\mathcal{G}_n(t; h) = \sum_{k=1}^M d_k g\left(\frac{t - \tau_k}{h}\right) I(t \in D_{h,k}).$$

The following conditions are needed to facilitate our theoretical results.

(C1) $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ which is is Lipschitz continuous and $\upsilon_2 = \int t^2 K(t)dt < \infty$, and $\|K\|^2 = \int K(t)^2 dt < \infty$.

(C2) The kernel K is such that $|g(\lambda)|$ has a unique maximum of 1 at 0.

(C3) For some constants $m_T > 0$ and $M_T < \infty$, the density function f_T of T_{ij} satisfies $m_T \le f_T(t) \le M_T$ for all $t \in [0, 1]$ and f_T is differentiable with a bounded derivative.

(C4) The smooth component $\nu(t)$ of the mean function is twice differentiable and the second derivative $\nu^{(2)}(t)$ is uniformly bounded.

(C5) $\mathbb{E}(\psi(|e_{ij}|)) < \infty$ and $\mathbb{E}\left(\sup_{t\in[0,1]} \psi(|U(t)|)\right) < \infty$.

(C6) $\varsigma_n(h) \to 0$ and

$$\psi\left(\frac{n}{\log n} h\varsigma_n(h)\right) \succ \frac{n}{\log n}. \tag{19}$$

where $\varsigma_n(h)$ is defined in (12).

(C7) $\sup_n\{n \max_i(m_i w_i)\} \leq B < \infty$

We briefly discuss the conditions. (C1) is a standard condition for the kernel function in smoothing problems. (C2) is less common but we conjecture it holds universally. We need this condition explicitly to ensure the desirable rates in the estimates. An intuitive explanation of this is given in (a) of Remark 2. We have examined a large number of kernels, either those commonly used or numerically generated, and have not encountered a single one that does not satisfy (C2). However, we do not yet know how to prove it for a general K. The condition is stated as a result in the supplement of Xia and Qiu (2015) but the proof appears to be incorrect. (C3)–(C4) are standard assumptions in the context of FDA and local polynomial smoothing. In (C4), if ν has bounded high-order derivatives, then the corresponding higher-order local polynomial regression procedure would lead to better rates of convergence. (C5) and (C6) are more general than the standard conditions in the literature where $\psi(x)$ is taken as x^α for some $\alpha \geq 2$ (cf. Li and Hsing, 2010; Zhang and Wang, 2016). While there is little incentive to entertain other choices of ψ for the smooth mean estimation problem, it is different for the change point estimation problem. For instance, for a process X for which the marginal distributions have exponentially decaying tail probabilities (e.g., Gaussian), the almost sure convergence rate of n^{-1} for the jump point location estimation problem can be achieved by using a ψ that is exponentially increasing. See the discussions in Section 3.3 for details.

3.2 Main results

We present the following results by first assuming that one has prior knowledge about the number of jump locations M. The jump sizes and locations $\{(d_k, \tau_k) : k = 1, \ldots, M\}$ can be consistently estimated with $\Delta_n(t; h)$, defined in (9), using the approach introduced in Section 2. The various convergence rates are described by the following results.

Theorem 3.1. *Assume that* (C1) and (C3)–(C7) *hold. Then, with probability one,*

(i) $\sup_{t \in \bar{D}_{h,1\to M}} |\Delta_n(t; h)| = O\left(h^2 + \varsigma_n(h)\right)$,

(ii) $\sup_{t \in D_{h,1\to M}} |\Delta_n(t; h) - \mathcal{G}_n(t; h)| = O\left(h + \varsigma_n(h)\right)$.

The proof of (i) of Theorem 3.1 is based on the idea that both $\widehat{\mu}_+(t)$ and $\widehat{\mu}_-(t)$ estimate $\mu(t)$ at rate $O(h^2 + \varsigma_n(h))$ for $t \in \bar{D}_{h,1\to M}$. This is in keeping with what was obtained in Li and Hsing (2010) for a smooth mean function. Theorem 3.1 shows that the $\Delta_n(t; h)$ behaves like $d_k g\left(\frac{t-\tau_k}{h}\right)$ for t in the h-neighborhood of some τ_k and is of vanishing order if this is not the case. This means that, for large n, our algorithm focuses on these neighborhoods to search for change points.

The following is a key result of this paper which gives the rates of convergence of $\widehat{\tau}_k, \widehat{d}_k$ defined in (10)-(14).

Theorem 3.2. *Assume that* (C1)–(C7) *hold for* $h = h_\tau$ *and* h_d. *Let* ρ_d *be another tuning parameter such that* $h_\tau(h_\tau + \varsigma_n(h_\tau)) \prec \rho_d \lesssim h_\tau + \varsigma(h_\tau)$. *Then we have, with probability one,*

(i) $\max_{k=1,\ldots,M} |\widehat{\tau}_k(h_\tau) - \tau_k| = O\left(h_\tau^2 + h_\tau \varsigma_n(h_\tau)\right)$,

(ii) $\max_{k=1,\ldots,M} |\widehat{d}_k(h_\tau, h_d, \rho_d) - d_k| = O\left(h_\tau + \varsigma_n(h_\tau) + h_d^2 + \varsigma_n(h_d)\right)$,

(iii) $\sup_{t \in \bar{D}_{\rho_d,1\to M}} |\widehat{\mu}(t; h_\tau, h_d, \rho_d) - \mu(t)| = O\left(h_\tau + \varsigma_n(h_\tau) + h_d^2 + \varsigma_n(h_d)\right)$.

A few remarks are in order.

Remark 2. *(a) We mentioned in the discussion following Theorem 3.1 that, asymptotically,* $\Delta_n(t; h)$ *behaves like* $d_k g\left(\frac{t-\tau_k}{h}\right)$ *for* t *in the* h*-neighborhood of some*

τ_k. *Condition (C2) then enables us to control where the maximum of* $|\Delta_n(t;h)|$ *is achieved in this interval.*

(b) *The proof of Theorem 3.2 shows that (iii) remains true if the set* $\bar{D}_{\rho_d,1\to M}$ *is replaced by any set* E_n *such that the minimum distance between points in* E_n *and the* τ_k *is of a larger order than* $h_\tau^2 + h_\tau\varsigma_n(h_\tau)$. *An example of such* E_n *that only depends on data is* $\left(\cup_{k=1}^M[\widehat{\tau}_k(h_\tau)-\rho_n,\widehat{\tau}_k(h_\tau)-\rho_n]\right)^c$.

(c) *As stated in Theorem 3.2, the choice of* h_d *depends on* h_τ. *Indeed, one possibility for* ρ_d *is* $h_\tau + \varsigma(h_\tau)$. *In order to evaluate the optimal rates in (i)-(iii), one would choose* h_τ, h_d *to minimize the stated error bounds there while taking into account (C6). We will defer the discussions until Section 3.3.*

(d) *Applying Theorem 3.2, the theoretical properties of the estimator defined for the covariance in Remark 1 can be established in the same way as in Li and Hsing (2010) and are omitted.*

Next we establish the asymptotic normality of the mean function estimator in the smooth part.

Theorem 3.3. *Assume that the conditions of Theorem 3.2 with* $\psi(x) = x^3$. *Also, assume that*

$$\frac{\sum_{i=1}^n m_i w_i^3(1/h_\tau^2 + (m_i-1)/h_\tau + (m_i-1)(m_i-2))}{\left(\sum_{i=1}^n m_i w_i^2(1/h_\tau + (m_i-1)\right)^{3/2}} \to 0, \tag{20}$$

$$\min\left(h_\tau/\sum_{i=1}^n m_i w_i^2, 1/\sum_{i=1}^n m_i(m_i-1)w_i^2\right)(h_\tau + \varsigma_n(h_\tau) + h_d^2 + \varsigma_n(h_d))^2 \to 0, \tag{21}$$

and

$$h_\tau\sum_{i=1}^n m_i(m_i-1)w_i^2/\sum_{i=1}^n m_i w_i^2 \to C_0 \in [0,\infty]. \tag{22}$$

Then for any interior point $t \in \bar{D}_{\rho_d,1\to M}$,

$$\Gamma^{-1/2}\left[\widehat{\mu}(t) - \mu(t) - (1/2)h_\tau^2\upsilon_2\nu^{(2)}(t)\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Gamma = \frac{\sum_{i=1}^n m_i w_i^2}{h_\tau}\|K\|^2\frac{R(t,t)+\sigma^2}{f_T(t)} + \left(\sum_{i=1}^n m_i(m_i-1)w_i^2\right)R(t,t).$$

Remark 3. *Theorem 3.3 is useful for providing the asymptotic pointwise confidence band for the estimated mean function. In the covariance function, the unknown quantities* Γ*,* $R(t,t)$*,* σ^2 *and* $f_T(t)$ *can be replaced by their consistent estimators. In practice,* $R(t,t)$ *and* σ^2 *can be estimated using a local linear smoother as described in Remark 1, and the density* $f_T(t)$ *can be estimated via a kernel density estimation (Sheather and Jones, 1991). It is possible to extend our theoretical results further to achieve functional weak convergence by verifying the process tightness conditions. See Degras (2011) and Zheng et al. (2014) for some recent examples. Such development may facilitate the construction of simultaneous confidence band. However, it will involve a high level of technicality and is beyond the scope of this paper.*

The following corollary provides the central limit theorem of $\Delta_n(t;h)$ which can be used for testing whether a given point is a change point.

Corollary 3.1. *Assume that the conditions of Theorem 3.2 hold with* $\psi(x) = x^3$*. Also, assume (20), (22) and* $\min\left(h_\tau / \sum_{i=1}^n m_i w_i^2, 1/\sum_{i=1}^n N_i w_i^2\right)(h_\tau + \varsigma_n(h_\tau))^2 \to 0$ *for bandwidth* h_τ*. Then for any point* $t \notin \{\tau_1, \ldots, \tau_M\}$*,*

$$\Omega^{-1/2}\Delta_n(t) \xrightarrow{d} \mathcal{N}(0,1)$$

where

$$\Omega = \frac{2\sum_{i=1}^n m_i w_i^2}{h_\tau}\|K_+^*\|^2 \frac{R(t,t)+\sigma^2}{f_T(t)} + 2\left(\sum_{i=1}^n m_i(m_i-1)w_i^2\right)R(t,t).$$

Corollary 3.1 provides the basis for a statistical test; in particular, one can reject the null hypothesis that t is not a change point at level α if

$$|\Delta_n(t)\Omega^{-1/2}| > z_{1-\alpha/2},$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of the standard normal distribution. If Ω contains unknown components, then they can be replaced by consistent estimates as discussed in Remark 3. Note that this test does not adjust for multiple comparisons and cannot be regarded as a global test. In particular, it cannot be used to determine M when it is unknown. However, the procedure described immediately before Remark 1 in Section 2.3 for estimating M is consistent, as the following result shows.

Theorem 3.4. *Assume that conditions in Theorem 3.2 hold and the threshold parameter ζ_n in (15) satisfies $\zeta_n \to 0$ and $\zeta_n/\{h_\tau^2 + \varsigma_n(h_\tau)\} \to \infty$. Then $\widehat{M} \to M$ with probability one.*

3.3 Discussions on optimal rates for $\widehat{\tau}_k$ and $\widehat{d}_k$

In this subsection we discuss optimal choices of the tuning parameters h_τ, h_d for estimating τ_k, d_k by $\widehat{\tau}_k$ and $\widehat{d}_k$, respectively, based on Theorem 3.2. We will focus on two examples of $\psi(x)$, namely, $\psi(x) = x^\alpha$ for some $\alpha > 2$ and $\psi(x) = e^x$.

First consider the optimal convergence rate of $\widehat{\tau}_k$. For simplicity, consider $w_i = 1/(nm_i)$ in Li and Hsing (2010) let $\bar{m}_n = \left(n^{-1}\sum_{i=1}^n m_i^{-1}\right)^{-1}$, the harmonic mean of the m_i. Then,

$$\varsigma_n(h) = O\left(\left\{\frac{\log(n)}{n}\left(1 + \frac{1}{h\bar{m}_n}\right)\right\}^{1/2}\right). \tag{23}$$

By (i) of Theorem 3.2, we have that

$$\max_{k=1,\ldots,M} |\widehat{\tau}_k(h_\tau) - \tau_k| = O\left(h_\tau^2 + \left\{\frac{\log n}{n}\left(h_\tau^2 + \frac{h_\tau}{\bar{m}_n}\right)\right\}^{1/2}\right). \tag{24}$$

For any h_τ that satisfies (C6), we consider the rate calculation for two specific functions $\psi(x)$.

(a) Consider $\psi(x) = x^\alpha$, with $\alpha > 2$. In view of (23) and (24), the optimal h_τ is obtained by selecting the smallest h_τ that satisfies (C6), i.e.,

$$\log n/(n\bar{m}_n) \prec h_\tau, \tag{25}$$

$$\left(\log n/n\right)^{1-2/\alpha} \prec h_\tau^2 + h_\tau/\bar{m}_n. \tag{26}$$

It suffices to focus on (26), which amounts to

$$h_\tau = B_n\left(\left(\frac{\log n}{n}\right)^{1/2-1/\alpha} \wedge \bar{m}_n\left(\frac{\log n}{n}\right)^{1-2/\alpha}\right) \tag{27}$$

where $B_n \to \infty$ slowly. After some calculations, the rate in (24) can be seen to be

$$\begin{aligned} h_\tau^2 + h_\tau\varsigma(h_\tau) &= B_n^2\bar{m}_n^2(\log n/n)^{2-4/\alpha} + B_n(\log n/n)^{1-1/\alpha} \\ &= B_n^2\left\{\bar{m}_n^2 \wedge (n/\log n)^{1-2/\alpha}\right\}(\log n/n)^{2-4/\alpha} + B_n(\log n/n)^{1-1/\alpha}. \end{aligned}$$

Thus, for $\bar{m}_n$ bounded, the rate dominated by $B_n(\log n/n)^{1-1/\alpha}$ for any $B_n \to \infty$. However, if $\bar{m}_n$ is not bounded, then the rate increases in $\bar{m}_n$, with the worst possible rate $B_n(\log n/n)^{1-2/\alpha}$ achieved when $\bar{m}_n > (n/\log n)^{1/2-1/\alpha}$. This is counter to the intuition that denser data lead to better rates.

(b) Next consider $\psi(x) = e^x$, where (C6) is essentially

$$\left(h_\tau^2 + \frac{h_\tau}{\bar{m}_n}\right)^{1/2} \left(\frac{n}{\log n}\right)^{1/2} - \log n + \log\log n \to \infty, \tag{28}$$

which can be achieved if

$$\left(h_\tau^2 + \frac{h_\tau}{\bar{m}_n}\right)^{1/2} \left(\frac{n}{\log n}\right)^{1/2} \geq \log n$$

Arguing as in the previous case, let

$$h_\tau = \left(\left(\frac{\log^3 n}{n}\right)^{1/2} \wedge \bar{m}_n \frac{\log^3 n}{n}\right).$$

The rate in (24) can be seen to be $\log^2 n/n$.

We next briefly discuss the optimal convergence rate of $\widehat{d}_k$. First, ρ_d depends on h_τ, which we have already addressed above. Thus, it is sufficient to focus on h_d. In view of the error bound $h_d^2 + \varsigma_n(h_d)$, the considerations on that are similar to the bandwidth selection problem when the mean function is continuous; see Li and Hsing (2010). If $\bar{m}_n$ is bounded and $\psi(x) = x^\alpha$, $\alpha > 2.5$, then the optimal h_d is $(\log n/n)^{1/5}$ which leads to the rate of $(\log n/n)^{2/5}$. On the other hand, if $\bar{m}_n \gtrsim (n/\log n)^{1/4}$ and and $\psi(x) = x^\alpha$, $\alpha > 4$, then any bandwidth h_d satisfying $\bar{m}_n^{-1} \lesssim h_d \lesssim (\log n/n)^{1/4}$ leads to the rate $(\log n/n)^{1/2}$. Unlike for $\widehat{\tau}_k$, the rates will not improve if $\psi(x) = e^x$. The overall rate for $\widehat{d}_k$ can be obtained by combining ρ_d and $h_d^2 + \varsigma_n(h_d)$.

For convenience, a brief summary of the preceding discussion on rates is given Table 1.

Remark 4. *A number of papers have studied the problem of estimating jump points in the nonparametric regression problem. For example, Müller and Song (1997), Grégoire and Hamrouni (2002) and Huh and Park (2004) all established the rates for estimating the discontinuities to be $O_p(n^{-1})$ based on different approaches under*

Table 1: Summary of convergence rates for break locations $\widehat{\tau}_k$ and jump sizes $\widehat{d}_k$.

	$\widehat{\tau}_k$		$\widehat{d}_k$	
$\psi(x)$	$\bar{m}_n \lesssim 1$	$\bar{m}_n \gtrsim (n/\log n)^{1/2-1/\alpha}$	$\bar{m}_n \lesssim 1$	$\bar{m}_n \gtrsim (n/\log n)^{1/4}$
x^α	$B_n(\log n/n)^{1-1/\alpha}$	$B_n(\log n/n)^{1-2/\alpha}$	$(\log n/n)^{2/5}$	$(\log n/n)^{1/2}$
e^x		$\log^2 n/n$		

the finite variance assumption for the errors. Recall that $\psi(x) = e^x$ is needed in our approach in order to achieve the almost sure rate of $O(n^{-1})$. Xia and Qiu (2015), on the other hand, showed that the almost sure rate of estimating the jump points to be $B_n(\log n/n)^{2/3}$ for any $B_n \to \infty$. That result can be roughly compared to ours in the sparse functional setting with $\alpha = 3$.

4 Tuning parameter selection and simulation studies

In this section, we discuss the selection of tuning parameters in our procedures from data and evaluate the performance of the procedures based on two simulation examples.

4.1 Data-driven tuning parameter selection

The methods introduced in this paper depend crucially on the tuning parameter h_τ, h_d, and ζ_n. While the theory provides some guidance, in practice we need to determine these tuning parameters using a data-driven approach. In this section we provide some details on how to choose these tuning parameters.

First, we suggest using the cut-off value from the test constructed from Corollary 3.1 as the threshold ζ_n for determining M in (15). Specifically, given h_τ, one can choose ζ_n to be $\max_t z_{1-\alpha/2}\Omega^{1/2}(t)$, where the unknown variance terms $R(t,t)$ and σ^2 can be estimated by plugging in the local linear smoothers after we obtain the residuals from fitting the initial one-sided local linear regression, similar to the

calculation steps described in Remark 1. Then, we can adopt a cross validation approach to select the optimal bandwidths h_τ and h_d. In the simulation studies in the next two subsections, we consider the K-fold cross validation score

$$\mathrm{CV}(h_\tau, h_d) = \frac{1}{K}\sum_{k=1}^{K}\left[\sum_{i\in\mathcal{I}_k} w_i \sum_{j=1}^{m_i}\left(Y_{ij} - \widehat{\mu}_{-k}(T_{ij}; h_\tau, h_d)\right)^2\right],$$

where $\mathcal{I}_1, \ldots, \mathcal{I}_K$ form a partition of $\{1, \ldots, n\}$, and $\widehat{\mu}_{-k}(\cdot; h_\tau, h_d)$ is our estimator in (14), with bandwidths $h_\tau, \rho_d = h_\tau$, and h_d, based all the data Y_{ij} except the ones for which $i \in \mathcal{I}_k$. Then select h_τ and h_d as the minimizer of $\mathrm{CV}(h_\tau, h_d)$ by searching over a 2-dimensional grid. We have examined this approach using $K = 5$ and 10 and found satisfactory performance in numerical studies (cf. Sections 4.2 and 4.3)

4.2 Simulation 1

In this section, we evaluate the numerical performance of our proposed approach through simulations. We consider the following mean function settings:

- Setting 1: $\mu(t) = \sin(2\pi t) + \cos(2\pi t) + t^2 + \sum_{k=1}^{3} d_k I(t \geq \tau_k)$,

with $M = 3$ change points at $(\tau_1, \tau_2, \tau_3) = (0.25, 0.5, 0.75)$ and jump sizes $(d_1, d_2, d_3) = (0.5, -0.4, 0.4)$, and

- Setting 2: $\mu(t) = \sin(2\pi t) + \sum_{k=1}^{5} d_k I(t \geq \tau_k)$,

with $M = 5$ change points at $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5) = (0.167, 0.333, 0.500, 0.667, 0.833)$ and jump sizes $(d_1, d_2, d_3, d_4, d_5) = (0.45, -0.5, 0.45, -0.5, 0.45)$. The true covariance function is designed as

$$R(s,t) = \sum_{k=1}^{3} \xi_k \psi_k(s)\psi_k(t), \qquad s, t \in [0,1],$$

where $\xi_k = 1/(k+1)^2,\ k = 1, 2, 3$, and

$$\psi_1(t) = 1, \qquad \psi_2(t) = \sqrt{2}\sin(2\pi t), \qquad \psi_3(t) = \sqrt{2}\cos(2\pi t).$$

The data are then generated from model (1) with $T_{ij} \sim$ Uniform$[0,1]$, $U_{ij} = \sum_{k=1}^{3} a_{ik}\psi_k(T_{ij})$ where a_{ik} are independent from $N(0, \xi_k)$ and $e_{ij} \sim N(0, 0.2^2)$.

In order to investigate the phase transition of convergence rate of change points, we consider the following functional data structures: $n = 400$ and $m_i \sim \text{Poisson}(m)$ where $m = 10, 20$, or 50 for different settings. The cases of $m = 10$ and $m = 50$ may be viewed as representing sparse and dense functional data, respectively, whereas that of $m = 20$ represents an intermediate scenario between the two extremes.

In addition, we investigate the case of having a small number of curves n, while the observations for each curve are dense. Specifically we consider the combinations $n = 50$, $m = 300$, and $n = 100$, $m = 100$. These cases resemble the real data analysis in the following section, in which we have relatively small n and very large m_i for each curve. In each setting, we generated $L = 200$ simulation runs.

We follow similar simulation settings in Zhang and Wang (2016) and compare three weight schemes for the estimation. The OBS scheme assigns the same weight to each observation, that is, $w_i = 1/\sum_{i=1}^n m_i$ (Yao et al., 2005); the SUBJ scheme assigns the same weight to each subject, that is, $w_i = 1/(nm_i)$ (Li and Hsing, 2010); the mixture of the OBS and SUBJ schemes, that is, $w_i = \alpha/(\sum_{i=1}^n m_i) + (1-\alpha)/(nm_i(m_i-1))$ for some $0 \le \alpha \le 1$. Here, we choose $\alpha = c_2/(c_1 + c_2)$ where

$$c_1 = \frac{1}{h_\tau \sum_{i=1}^n m_i} + \frac{\sum_{i=1}^n m_i^2}{(\sum_{i=1}^n m_i)^2},$$

and

$$c_2 = \left(\frac{1}{h_\tau (n^{-1} \sum_{i=1}^n m_i^{-1})} + 1 \right) \frac{1}{n}.$$

We report the results with the bandwidths h_τ and h_d selected by the 5-fold cross validation procedure proposed in Section 4.1. The performance based on 10-fold cross validation is quite similar.

To assess the estimation performance of multiple change point detection, we report the centered mean, $\text{Mean}(\widehat{M}) - M$, and the centered median, $\text{Med}(\widehat{M}) - M$, of the number of change point detected by our method and the frequency of correct estimation of the number of change point, $\mathcal{E}(\widehat{M} = M)$. To investigate the accuracy of the estimated change point locations, we calculate the means of the two Hausdorff distances (cf. Boysen et al., 2009; Xia and Qiu, 2015), $\max_i \min_j |\tau_i - \widehat{\tau}_j|$ and $\max_j \min_i |\tau_i - \widehat{\tau}_j|$, where τ and $\widehat{\tau}$ denote a true change point location and its

estimate, respectively. The performance of $\widehat{\mu}$ is evaluated by the mean integrated squared error (MISE),

$$\text{MISE}(\widehat{\mu}) = \frac{1}{L}\sum_{l=1}^{L}\int_a^b [\widehat{\mu}_{[l]}(t) - \mu(t)]^2 dt,$$

where $\{\widehat{\mu}_{[l]}, l = 1, \ldots, L\}$ are estimates of the mean functions obtained from $L = 200$ replications. We denote by $\widehat{\mu}_{obs}, \widehat{\mu}_{subj}$, and $\widehat{\mu}_{\alpha}$ the corresponding results for the OBS, SUBJ, and mixture weight mean estimators, respectively.

Tables 2–3 present the various performance measures of the proposed change point detection procedure for $m = 10, 20$, and 50. Observe that the performance improves as the m_i increases while the size of improvement becomes smaller once the data becomes dense. Clearly, under the dense setting $m = 50$, the proposed change point detection performance is more satisfactory with smaller biases for $\widehat{M}$, $\widehat{\tau}_j$ and $\widehat{d}_k$, and higher frequency of achieving exactly $\widehat{M} = M$. It is relatively more difficult to identify the number of change points correctly based on sparse functional data , especially when the number of change points is large, as observed in setting 2.

The MISEs of Table 4 for $\widehat{\mu}$ suggest overall satisfactory performance of our proposed estimator. It supports our theory the estimation of $\widehat{\mu}$ can still be consistent with small m even though the change point detection performance in practice may be less satisfactory. The estimation results for all cases appear to be insensitive to the three mixing schemes. Moreover, we compare our method with the standard functional mean estimation without considering the break points in the mean function (Li and Hsing, 2010; Zhang and Wang, 2016). Let $\widehat{\mu}^{ZW}$ denotes the mean estimation in Zhang and Wang (2016), which can be implemented by the R package `fdapace`. Not surprisingly, the entries in Table 4 show that our method has much smaller MISE for the mean estimation than the existing approaches assuming smooth mean function. The optimal bandwidths in this simulation study were selected by a 5-fold CV procedure. We report the average of optimal bandwidths selected in simulations in Table 5. We also provide the bandwidths selected by a CV procedure for the estimate $\widehat{\mu}^{ZW}$, denoted by h_{ZW}. It seems that our data driven procedure tends to select a slightly larger bandwidth h_d for jump size estimation

Table 2: Performance of change point detection for setting 1 with $n = 400$ in simulation 1. The corresponding standard deviations are in the parentheses.

		$\boldsymbol{m = 10}$	$\boldsymbol{m = 20}$	$\boldsymbol{m = 50}$
SUBJ	$\text{Mean}(\widehat{M}) - M$	-0.125	0.035	0.020
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	61.5	76.5	95.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0494 (0.0632)	0.0237 (0.0557)	0.0069 (0.0341)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.1063 (0.1444)	0.0321 (0.0856)	0.0050 (0.0499)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.3023 (0.2286)	0.1539 (0.1356)	0.0876 (0.0344)
OBS	$\text{Mean}(\widehat{M}) - M$	-0.030	-0.095	0.050
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	67.0	81.4	96.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0556 (0.0685)	0.0102 (0.0364)	0.0054 (0.0269)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0923 (0.1294)	0.0379 (0.0978)	0.0001 (0.0001)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.3213 (0.2656)	0.1488 (0.1163)	0.0821 (0.0329)
MIX	$\text{Mean}(\widehat{M}) - M$	-0.133	-0.070	-0.015
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	70.9	84.0	97.5
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0395 (0.0593)	0.0147 (0.0452)	0.0006 (0.0084)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0846 (0.1254)	0.0377 (0.0966)	0.0050 (0.0351)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.3045 (0.2385)	0.1632 (0.1476)	0.0894 (0.0364)

than h_τ for change point location detection. This empirical observation affirms the theoretical results. Furthermore, we found that one tends to select smaller bandwidth when using $\widehat{\mu}^{ZW}$ in these settings with breaks.

Table 6 summarizes the coverage probabilities of the 95% pointwise confidence band after detecting the change points. The coverage rates in general improves as m_i increases. Furthermore, Figure 1 displays representative smoothed mean estimation after consistently detecting the change point in simulation 1 along with

Table 3: Performance of change point detection for setting 2 with $n = 400$ in simulation 1. The corresponding standard deviations are in the parentheses.

		$\boldsymbol{m = 10}$	$\boldsymbol{m = 20}$	$\boldsymbol{m = 50}$
SUBJ	$\text{Mean}(\widehat{M}) - M$	-0.62	-0.080	0.030
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	62.0	94.0	97.5
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0094 (0.0221)	0.0015 (0.0114)	0.0023 (0.0148)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0845 (0.1236)	0.0112 (0.0440)	0.0001 (0.0007)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2078 (0.1896)	0.1257 (0.0431)	0.0908 (0.0320)
OBS	$\text{Mean}(\widehat{M}) - M$	-0.500	-0.030	0.010
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	65.5	96.0	99.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0108 (0.0243)	0.0017 (0.0092)	0.0008 (0.0079)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0727 (0.1129)	0.0068 (0.0299)	0.0001 (0.0001)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2260 (0.1915)	0.1413 (0.1075)	0.0907 (0.0308)
MIX	$\text{Mean}(\widehat{M}) - M$	-0.650	0.015	-0.005
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	62.5	93.0	99.5
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0110 (0.0261)	0.0051 (0.0225)	0.0001 (0.0008)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0883 (0.1297)	0.0057 (0.0284)	0.0008 (0.0113)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2227 (0.1909)	0.1365 (0.0783)	0.0899 (0.0328)

the 95% pointwise confidence band constructed from asymptotic normal approximation. The selected curve estimation corresponds to the case with median MISE among the L simulations. In Figure 2, we also show the empirical distributions of $\widehat{\sigma}^2$. The estimated variance is close to the true variance value $0.2^2 = 0.04$ in all cases and is insensitive to the mixing scheme.

The computing times of our complete estimation procedure (including mean and covariance function estimation) along with the jump points detection are sum-

Table 4: MISE ($\times 10^{-2}$) of $\widehat{\mu}_{obs}$, $\widehat{\mu}_{subj}$ and $\widehat{\mu}_{\alpha}$ with $n = 400$ in simulation 1. The corresponding standard deviations are in the parentheses.

		$m = 10$	$m = 20$	$m = 50$
Setting 1	$\widehat{\mu}_{obs}$	0.3320 (0.1373)	0.1436 (0.0676)	0.0538 (0.0173)
	$\widehat{\mu}_{obs}^{ZW}$	0.4724 (0.0795)	0.3380 (0.0532)	0.2233 (0.0311)
	$\widehat{\mu}_{subj}$	0.3347 (0.1315)	0.1543 (0.0658)	0.0529 (0.0170)
	$\widehat{\mu}_{subj}^{ZW}$	0.4781 (0.1093)	0.3420 (0.0524)	0.2271 (0.0287)
	$\widehat{\mu}_{\alpha}$	0.3354 (0.1307)	0.1455 (0.0653)	0.0526 (0.0177)
	$\widehat{\mu}_{\alpha}^{ZW}$	0.4785 (0.0921)	0.3387 (0.0525)	0.2254 (0.0293)
Setting 2	$\widehat{\mu}_{obs}$	0.3429 (0.1640)	0.1342 (0.0543)	0.0558 (0.0166)
	$\widehat{\mu}_{obs}^{ZW}$	0.6669 (0.1087)	0.5091 (0.0473)	0.2698 (0.0260)
	$\widehat{\mu}_{subj}$	0.3781 (0.1885)	0.1354 (0.0719)	0.0557 (0.0196)
	$\widehat{\mu}_{subj}^{ZW}$	0.6889 (0.1089)	0.5191 (0.0564)	0.2760 (0.0246)
	$\widehat{\mu}_{\alpha}$	0.3652 (0.1784)	0.1303 (0.0542)	0.0530 (0.0185)
	$\widehat{\mu}_{\alpha}^{ZW}$	0.6806 (0.1002)	0.5168 (0.0461)	0.2732 (0.0259)

marized in Table 7, where jump locations are detected over 101 grid points equally spaced on the interval $[0, 1]$. The computations were done using a desktop computer with Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz and 14 GB of RAM; the operating system was 64-bit Windows, and R version 4.0.4 (2021-02-15) was used. To speed up computations involving loops, `C++` was implemented through `Rcpp`. While it is inevitable that more computing time is needed as the number of subjects n and/or the number of observations m_i increase, our procedure is still very efficient for large sample sizes.

4.3 Simulation 2

We next generate data assuming a heavy-tailed process to illustrate the applicability of our proposed methods. We consider the same true mean and covariance function settings given in the preceding subsection, except that the data are gener-

Table 5: Bandwidths selected in simulation 1.

		Setting 1				Setting 2			
(n, m)	Weight	ζ_n	h_τ	h_d	h_{ZW}	ζ_n	h_τ	h_d	h_{ZW}
$(400, 10)$	SUBJ	0.312	0.063	0.097	0.037	0.258	0.061	0.097	0.031
	OBS	0.317	0.062	0.095	0.037	0.248	0.061	0.096	0.030
	MIX	0.307	0.066	0.099	0.039	0.253	0.060	0.096	0.030
$(400, 20)$	SUBJ	0.193	0.059	0.093	0.029	0.225	0.063	0.095	0.023
	OBS	0.235	0.056	0.092	0.030	0.219	0.063	0.095	0.022
	MIX	0.234	0.057	0.091	0.030	0.185	0.062	0.094	0.023
$(400, 50)$	SUBJ	0.177	0.045	0.089	0.016	0.184	0.050	0.090	0.010
	OBS	0.173	0.047	0.088	0.017	0.170	0.048	0.092	0.010
	MIX	0.171	0.050	0.083	0.016	0.170	0.051	0.087	0.010
$(100, 100)$	SUBJ	0.220	0.056	0.088	0.028	0.216	0.061	0.095	0.017
	OBS	0.220	0.057	0.093	0.028	0.215	0.061	0.092	0.018
	MIX	0.217	0.058	0.089	0.027	0.217	0.060	0.092	0.019
$(50, 300)$	SUBJ	0.250	0.053	0.085	0.020	0.247	0.058	0.091	0.010
	OBS	0.250	0.049	0.085	0.021	0.247	0.059	0.092	0.010
	MIX	0.251	0.053	0.086	0.024	0.248	0.058	0.092	0.011

ated assuming that $T_{ij} \sim \text{Uniform}[0, 1]$, $U_{ij} = \sum_{k=1}^{3} a_{ik}\psi_k(T_{ij})$ with a_{ik} distributed as t with 3 degree of freedom, and e_{ij} is equal to $\sqrt{0.03}$ multiplied by a t random variable with 3 degree of freedom.

Eyeballing Tables 8–9, we see that the accuracy of change-point detection increases as the observations become denser. Similarly, examining Tables 10, the MISEs for $\widehat{\mu}$ suggest good overall performance of our proposed estimator. The average of optimal bandwidths selected in simulation 2 are reported in Table 11, and Table 12 summarizes the coverage probabilities of the 95% pointwise confidence band after detecting the change points.

Figure 3 also displays representative smoothed mean estimation after consis-

Table 6: Coverage probabilities of the 95% pointwise confidence band in simulation 1.

(n, m)		$\widehat{\mu}_{subj}$	$\widehat{\mu}_{obs}$	$\widehat{\mu}_{\alpha}$
$(400, 10)$	Setting 1	91.0	91.4	91.8
	Setting 2	90.0	90.5	90.3
$(400, 20)$	Setting 1	96.7	97.1	96.9
	Setting 2	97.2	97.3	97.4
$(400, 50)$	Setting 1	98.9	99.5	97.9
	Setting 2	99.5	98.9	96.5
$(100, 100)$	Setting 1	98.8	98.9	99.5
	Setting 2	99.6	99.7	99.6
$(50, 300)$	Setting 1	99.7	99.4	99.6
	Setting 2	99.6	99.7	99.8

Table 7: Computing times (seconds) with 5-folds CV over a prefixed grid of bandwidths in simulation 1. Total is the entire computing time while CPD is the time spent on change point detection after all tuning parameters are fixed at optimal values.

		Setting 1		Setting 2	
n	m	Total	CPD	Total	CPD
400	10	10.71	0.08	10.34	0.10
400	20	33.69	0.35	33.12	0.38
400	50	259.91	1.39	255.02	1.28
100	100	164.72	0.39	158.51	0.52
50	300	770.14	1.28	804.19	1.25

tently detecting the change points in simulation 2 along with the 95% pointwise confidence band constructed from the asymptotic normal approximation.

Additional simulation results with different data generating settings are included

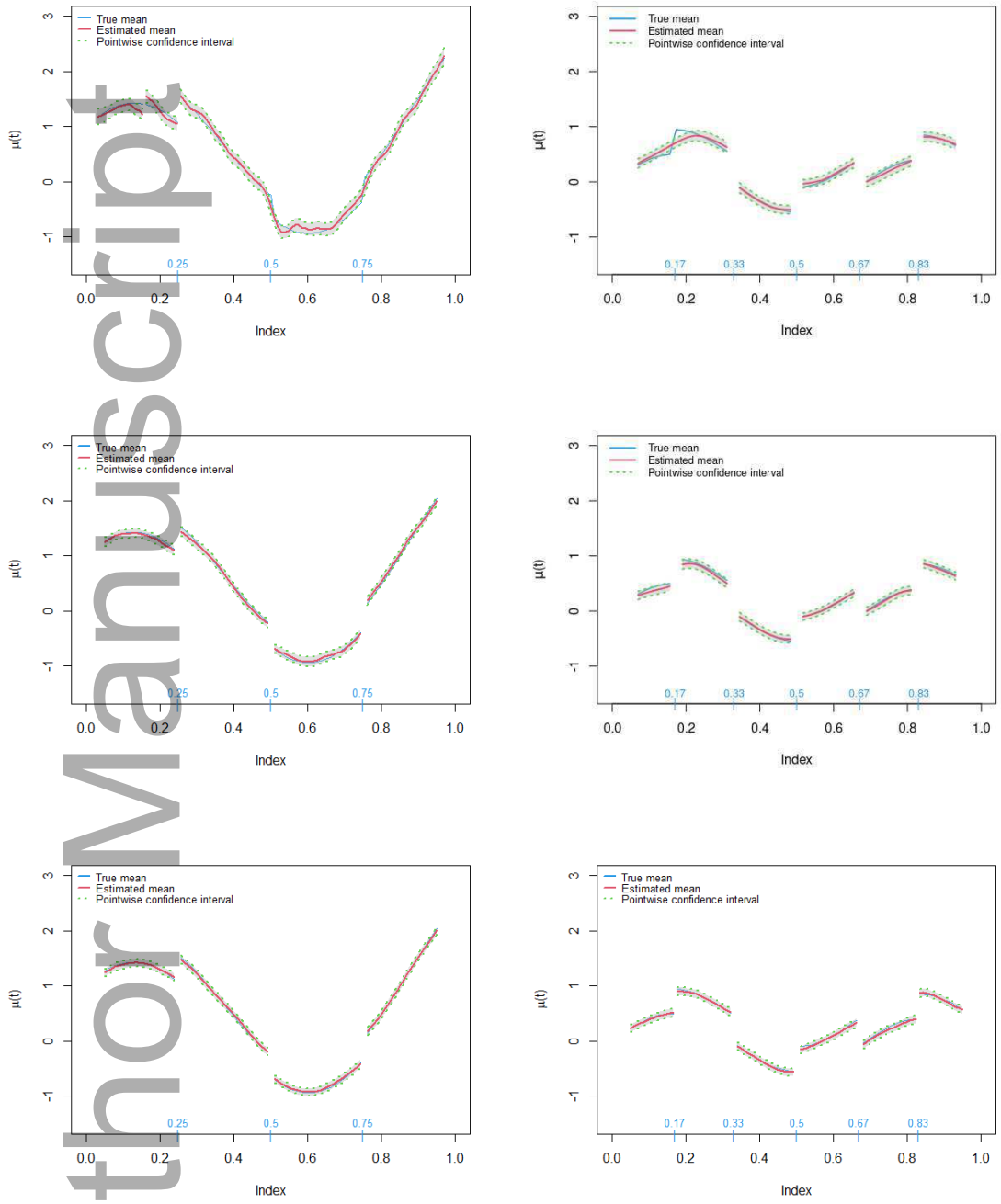


Figure 1: Smoothed mean function estimation and 95% pointwise confidence band from asymptotic normal approximation with the mixture weight in simulation 1. The two columns correspond to setting 1 and setting 2, and the three rows correspond to $m = 10, 20, 50$, respectively.

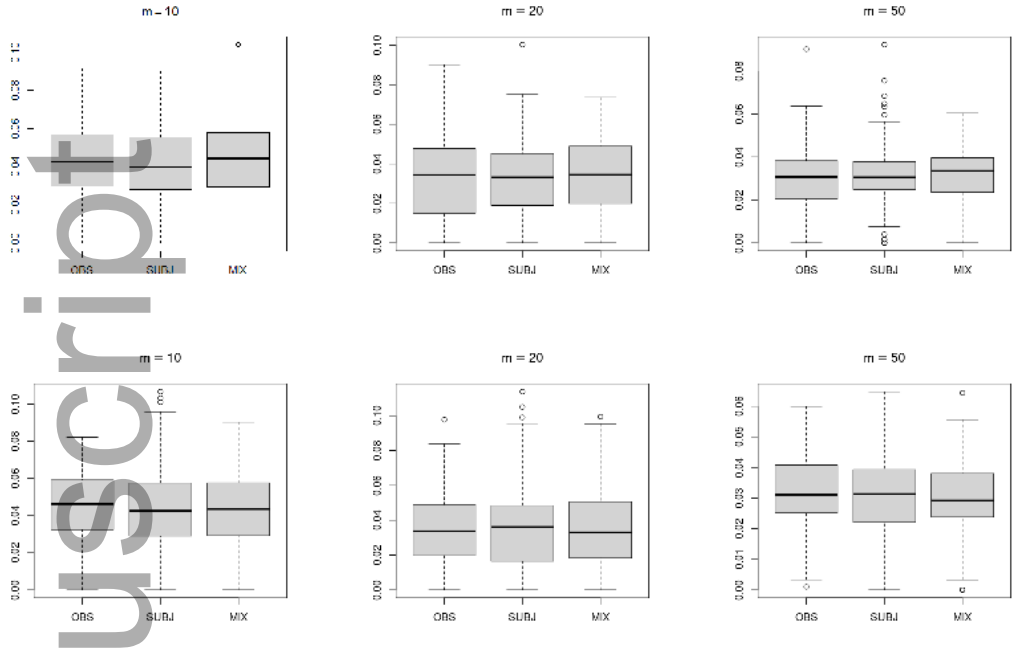


Figure 2: Box plots for the estimated $\widehat{\sigma}^2$ in simulation 1. The two rows correspond to setting 1 and setting 2, respectively.

in the supplementary file of this paper.

5 Case studies

5.1 Average Value Weighted Returns

We first consider an average value weighted returns data. Value weighted return is a type of weighting methodology to calculate portfolio return that gives a weight to an asset in the portfolio based on the asset's market value. The data set can be freely downloaded from Kenneth French's website: `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. The original data set consists of the daily simple returns of $n = 49$ industry portfolios from 1927 to 2020. Let Y_{ij} be the daily return of the ith portfolio at the jth time point, $i = 1, \ldots, 49$, $j = 1, \ldots, m_i$. These data have already motivated some well-known financial factor models (Li et al., 2018). In this section, we choose the most recent $m_i = 354$ observations from December 2018 through April 2020 for an illustration of our

Table 8: Performance of change point detection for setting 1 with $n = 400$ in simulation 2. The corresponding standard deviations are in the parentheses.

		$m = 10$	$m = 20$	$m = 50$
SUBJ	$\text{Mean}(\widehat{M}) - M$	-0.429	0.085	0.040
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	56.4	76.9	95.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0383 (0.0595)	0.0284 (0.0568)	0.0069 (0.0308)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.1455 (0.1681)	0.0309 (0.0871)	0.0025 (0.0249)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.3256 (0.2471)	0.1609 (0.1213)	0.0897 (0.0359)
OBS	$\text{Mean}(\widehat{M}) - M$	-0.314	0.111	0.035
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	60.8	77.3	96.5
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0363 (0.0574)	0.0300 (0.0582)	0.0043 (0.0251)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.1233 (0.1729)	0.0255 (0.0793)	0.0013 (0.0177)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2792 (0.1931)	0.1609 (0.1373)	0.0768 (0.0335)
MIX	$\text{Mean}(\widehat{M}) - M$	-0.390	0.120	-0.020
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	59.1	77.5	96.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0355 (0.0559)	0.0312 (0.0599)	0.0018 (0.0181)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.1412 (0.1640)	0.0288 (0.0786)	0.0075 (0.0427)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2883 (0.2053)	0.1695 (0.1426)	0.0845 (0.0348)

proposed methods, and investigate the mean return curves across different industries $\mu(t)$ over time.

Figures 4 and 5 present the results for the estimated mean function $\widehat{\mu}(t)$. Thirteen change points $\widehat{\tau}_k$ are detected by our proposed method at 2018-12-24, 2019-01-22, 2019-02-11, 2019-03-04, 2019-05-13, 2019-06-04, 2019-06-24, 2019-07-15, 2019-08-22, 2019-10-02, 2019-10-31, 2019-12-06, 2020-02-10, 2020-03-16, 2020-04-03. The estimated jump sizes $\widehat{d}_k$ at the 15 time points are 5.659, -1.265, 1.620, -1.346,

Table 9: Performance of change point detection for setting 2 with $n = 400$ in simulation 2. The corresponding standard deviations are in the parentheses.

		$m = 10$	$m = 20$	$m = 50$
SUBJ	$\text{Mean}(\widehat{M}) - M$	0.075	-0.080	0.010
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	81.0	92.0	99.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0236 (0.0391)	0.0033 (0.0143)	0.0007 (0.0069)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0260 (0.0457)	0.0153 (0.0447)	0.0001 (0.0001)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2633 (0.2054)	0.1530 (0.1257)	0.0861 (0.0309)
OBS	$\text{Mean}(\widehat{M}) - M$	0.125	0.005	0.020
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	81.0	94.5	98.0
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0227 (0.0367)	0.0043 (0.0187)	0.0022 (0.0154)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0267 (0.0471)	0.0060 (0.0284)	0.0001 (0.0001)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2532 (0.2085)	0.1417 (0.0953)	0.0894 (0.0339)
MIX	$\text{Mean}(\widehat{M}) - M$	0.175	-0.025	-0.005
	$\text{Med}(\widehat{M}) - M$	0	0	0
	$\mathcal{E}(\widehat{M} = M)[\%]$	80.5	94.5	99.5
	$\max_i \min_j \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0262 (0.0408)	0.0040 (0.0165)	0.0001 (0.0001)
	$\max_j \min_i \lvert\tau_i - \widehat{\tau}_j\rvert$	0.0244 (0.0432)	0.0104 (0.0401)	0.0008 (0.0113)
	$\max_k \lvert\widehat{d}_k - d_k\rvert$	0.2539 (0.2027)	0.1509 (0.1204)	0.0891 (0.0318)

2.052, 1.774, -1.032, -1.119, -2.347, 2.184, 1.426, -1.029, 1.157, -0.979 and 6.518, respectively. Other than these 15 breaks, the return curve follows a smooth-changing pattern. We also treat the data as independent samples and apply the nonparametric regression method by Xia and Qiu (2015) to estimate the mean function. The resulting curve (green curve in Figure 4) contains eight change points that are all detected by our method except the one at 2020-02-28. This approach ignores the data correlation and may not be appropriate for the daily asset data. Figure

Table 10: MISE ($\times 10^{-2}$) of $\widehat{\mu}_{obs}$, $\widehat{\mu}_{subj}$ and $\widehat{\mu}_{\alpha}$ with $n = 400$ in simulation 2. The corresponding standard deviations are in the parentheses.

		$m = 10$	$m = 20$	$m = 50$
Setting 1	$\widehat{\mu}_{obs}$	0.3375 (0.1419)	0.1400 (0.0605)	0.0494 (0.0176)
	$\widehat{\mu}_{obs}^{ZW}$	0.4820 (0.1069)	0.3356 (0.0595)	0.2195 (0.0301)
	$\widehat{\mu}_{subj}$	0.3741 (0.1386)	0.1495 (0.0648)	0.0558 (0.0198)
	$\widehat{\mu}_{subj}^{ZW}$	0.4996 (0.0874)	0.3345 (0.0498)	0.2224 (0.0333)
	$\widehat{\mu}_{\alpha}$	0.3387 (0.1337)	0.1460 (0.0689)	0.0542 (0.0211)
	$\widehat{\mu}_{\alpha}^{ZW}$	0.4725 (0.0851)	0.3360 (0.0506)	0.2235 (0.0311)
Setting 2	$\widehat{\mu}_{obs}$	0.3171 (0.1685)	0.1298 (0.0519)	0.0540 (0.0002)
	$\widehat{\mu}_{obs}^{ZW}$	0.6585 (0.1182)	0.5035 (0.0508)	0.2679 (0.0250)
	$\widehat{\mu}_{subj}$	0.3421 (0.1691)	0.1415 (0.0678)	0.0525 (0.0177)
	$\widehat{\mu}_{subj}^{ZW}$	0.6628 (0.1027)	0.5062 (0.0475)	0.2742 (0.0229)
	$\widehat{\mu}_{\alpha}$	0.3185 (0.1627)	0.1331 (0.0655)	0.0535 (0.0175)
	$\widehat{\mu}_{\alpha}^{ZW}$	0.6485 (0.0954)	0.5078 (0.0499)	0.2721 (0.0242)

4 also displays change points detected by a double CUSUM method for panel data (Cho, 2016), which is implemented by R package `hdbinseg`. Three change points are detected at 2018-12-24, 2020-02-21, and 2020-03-23. We note that the first change point was also detected by our method. This approach, however, only gives piecewise constant description of the data. Our results may provide a better separation of different financial and economical cycles over the time period for all asset types across different industries. The 95% pointwise confidence band is shown in Figure 5, where we have a wider confidence band in the period of March 2020 compared to other periods.

To compare with the single series analysis, we also present the estimated individual mean functions for 4 selected industry portfolios in Figure 6. We note that the estimated change points from these individual industry are distinct but closely related to the change points detected by our approach for the functional series.

Table 11: Bandwidths selected in simulation 2.

		Setting 1				Setting 2			
(n, m)	Weight	ζ_n	h_τ	h_d	h_{ZW}	ζ_n	h_τ	h_d	h_{ZW}
$(400, 10)$	SUBJ	0.355	0.066	0.099	0.037	0.237	0.063	0.096	0.030
	OBS	0.333	0.066	0.099	0.038	0.230	0.063	0.096	0.031
	MIX	0.340	0.065	0.097	0.038	0.230	0.063	0.097	0.029
$(400, 20)$	SUBJ	0.249	0.059	0.095	0.030	0.244	0.061	0.094	0.030
	OBS	0.239	0.061	0.093	0.029	0.231	0.064	0.095	0.022
	MIX	0.242	0.059	0.095	0.029	0.241	0.062	0.094	0.022
$(400, 50)$	SUBJ	0.177	0.047	0.089	0.017	0.177	0.048	0.092	0.010
	OBS	0.178	0.047	0.088	0.016	0.173	0.048	0.088	0.010
	MIX	0.258	0.050	0.089	0.017	0.254	0.052	0.090	0.010
$(100, 100)$	SUBJ	0.216	0.053	0.095	0.027	0.209	0.059	0.092	0.017
	OBS	0.214	0.054	0.091	0.028	0.212	0.059	0.093	0.018
	MIX	0.216	0.053	0.090	0.027	0.212	0.058	0.094	0.016
$(50, 300)$	SUBJ	0.247	0.053	0.077	0.024	0.259	0.057	0.081	0.011
	OBS	0.243	0.055	0.086	0.025	0.259	0.058	0.081	0.011
	MIX	0.248	0.052	0.086	0.023	0.246	0.058	0.092	0.011

In practice investigators could combine information from both sources to provide useful macroeconomic analysis and produce sensible predictions.

We next consider monthly maximum average value weighted returns for 49 industry portfolios with a longer time frame from January 1990 to April 2020. This leads to a functional series $\{Y_{ij} : i = 1, \cdots, 49; j = 1 \cdots, 364\}$. Figure 7 indicates that 10 jumps $\widehat{\tau}_k$ are detected by our proposed method, at Feb 1992, Oct 1998, Apr 2000, Nov 2001, May 2006, Jun 2008, Apr 2010, Jun 2012, Feb 2015, Jan 2019. These breaks correspond to the recoveries after the financial crises during the periods 1997–1998 and 2008–2012. The estimated jump sizes $\widehat{d}_k$ at the 10 time points are -0.673, -1.009, -2.146, -1.028, 0.996, 2.747, 2.186, 1.135, -1.114 and

Table 12: Coverage probabilities of the 95% pointwise confidence band in simulation 2.

(n, m)		$\widehat{\mu}_{subj}$	$\widehat{\mu}_{obs}$	$\widehat{\mu}_{\alpha}$
$(400, 10)$	Setting 1	91.1	91.8	91.2
	Setting 2	89.8	91.7	90.8
$(400, 20)$	Setting 1	96.9	97.1	96.7
	Setting 2	97.1	97.6	97.8
$(400, 50)$	Setting 1	98.0	98.7	98.4
	Setting 2	98.7	98.9	98.5
$(100, 100)$	Setting 1	99.3	99.4	99.5
	Setting 2	99.5	99.7	99.8
$(50, 300)$	Setting 1	99.5	99.7	99.8
	Setting 2	99.7	99.8	99.7

-2.538, respectively. In contrast, the green curve estimate using the approach of Xia and Qiu (2015) only identifies four change points at Oct 1998, Nov 2001, Apr 2010, Jan 2019 (also all detected by our method) while the change point detection method by Cho (2016) identifies two change points at Mar 2000 (also detected by our method) and Jul 2003. Similarly, we have a wide confidence band where the data has a high variation. To compare with the individual estimate, we further present the estimated individual function for 4 selected industry types in Figure 8. One can observe that the traditional single curve change point detection method may be inadequate.

5.2 Monthly U.S. Treasuries

We consider the monthly U.S. treasuries from January 1983 to September 2010. In total, there are $n = 15$ interest rates with maturities of 3, 6, 9, 12, 18, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. Each maturity has a monthly yield curve of interest rates with $m_i = 333$ observations spanning the observation period. These

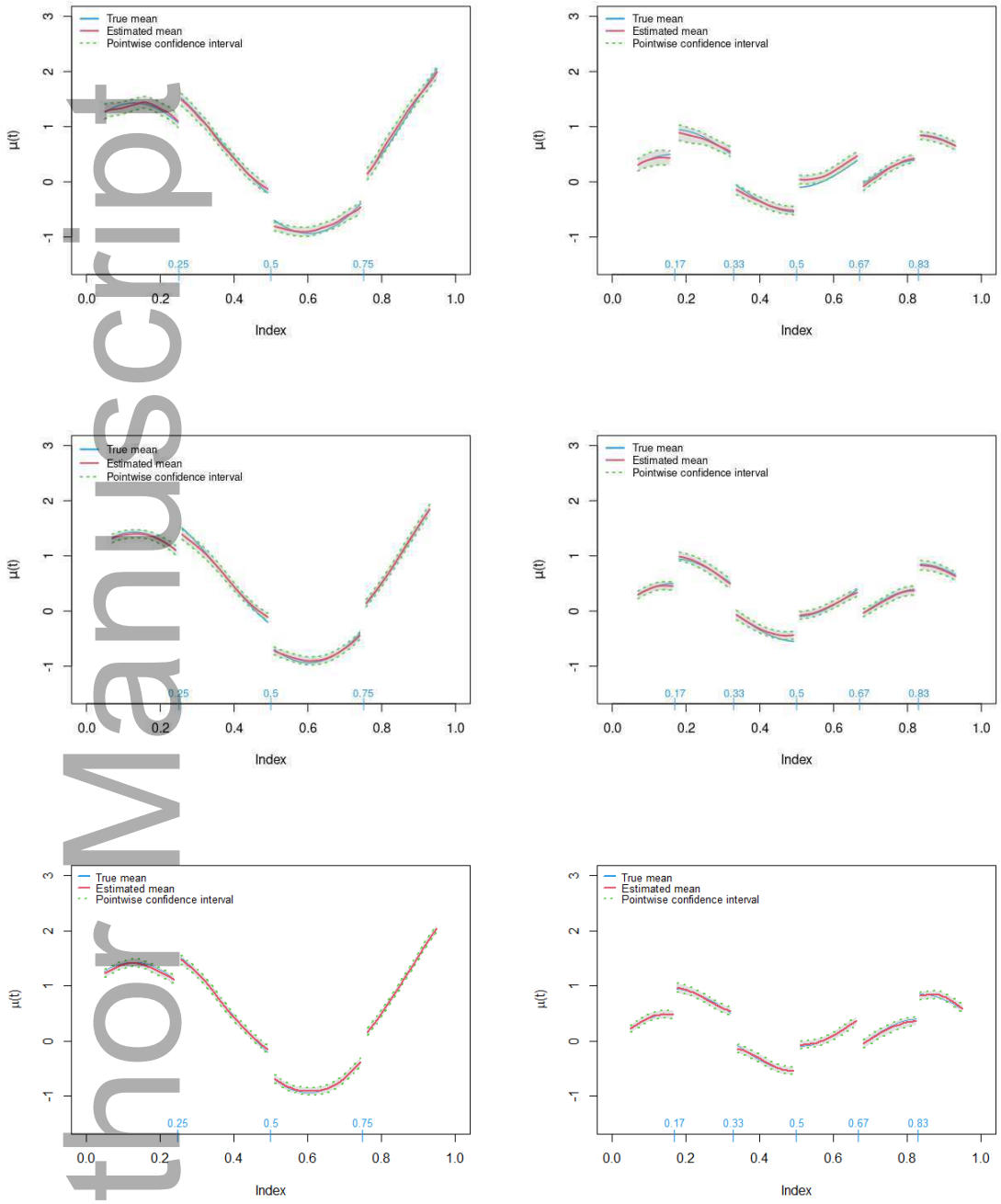


Figure 3: Smoothed mean function estimation and 95% pointwise confidence band from asymptotic normal approximation with the mixture weight in simulation 2. The two columns correspond to setting 1 and setting 2, and the three rows correspond to $m = 10, 20, 50$, respectively.

data have already been studied with different statistical models in Diebold and Li (2006); Chen and Li (2011); Chen and Niu (2014). The time period not only covers two financial crises in 1987 and 2008, but also the tenure of the Fed Chairmen Paul Volcker (1979 - 1987), Alan Greenspan (1987 - 2006) and Ben Bernanke (2006 - 2014). Some interesting questions are whether the monetary policies changed with the chairman, a statement which most would probably agree with, and whether there were lasting effects from the two crises on the movement of interest rates. We apply our proposed method to address these questions.

Figure 9 presents the mean curve estimation results. Four break points $\widehat{\tau}_k$ are detected at Mar 1987, Feb 1994, Aug 1999, Dec 2007, which includes two aforementioned historical financial crises. The change in Feb 1994 corresponds to the event when the Federal Reserve unexpectedly raised interest rates and caused the bond market crisis or the so-called *Great Bond Massacre*. The change in Aug 1999 reflects a transition to an economic recovery phase. The estimated jump sizes $\widehat{d}_k$ at the 4 change points are 1.554, 2.595, 2.036, and -1.828. The results using Xia and Qiu (2015) and Cho (2016) were similarly presented as in the preceding example. We observe again that the method by Xia and Qiu (2015) detected fewer change points. The results from applying Cho (2016) also divide the whole period into five segments with 4 change points, located at slightly different time points.

The supplementary file contains more data analysis results. The code for simulation studies and real data analysis are provided in Github and directly downloadable. The `R` package `FPMD` to implement our methods can be installed from Github at `https://github.com/liygCR/FPMD`.

6 Discussion

The change point detection approach introduced in this paper has a broad range of applications. In our real-data examples, we considered two economic data sets for which the functional series are indexed by calendar time. More applications exist in medical and other scientific investigations where the index t could be any continuous variable. For example, in genomic studies researchers may be interested

in identifying short intervals of copy number variations in multiple DNA sequences; in environmental monitoring, it may be important to detect anomalies such as the presence of hazardous materials or intruders typically existing only in some small areas. Our proposed change point detection method produces meaningful segmentation of such functional data and suggests important subgroups or sub-regions delineated by discontinuous boundaries.

An immediate extension of our approach is change point detections for multivariate functional data. In such problems, one might be also interested in identifying structural changes of the covariance function as well as the mean function. Indeed, many economic patterns are impacted by the underlying volatility or correlation process that is subject to sudden breaks triggered by external events. In such situations, both the theoretical development and the computational burden present new challenges. We will address these in future research.

Acknowledgments

We are grateful to the Co-Editor, Associate Editor and two referees for their constructive comments. The work was partly supported by Academic Research Funds R-155-000-205-114, R-155-000-195-114 and Tier 2 MOE funds in Singapore MOE2017-T2-2-082: R-155-000-197-112 (Direct cost) and R-155-000-197-113 (IRC), as well as Grant DMS-1916226 in U.S.A.

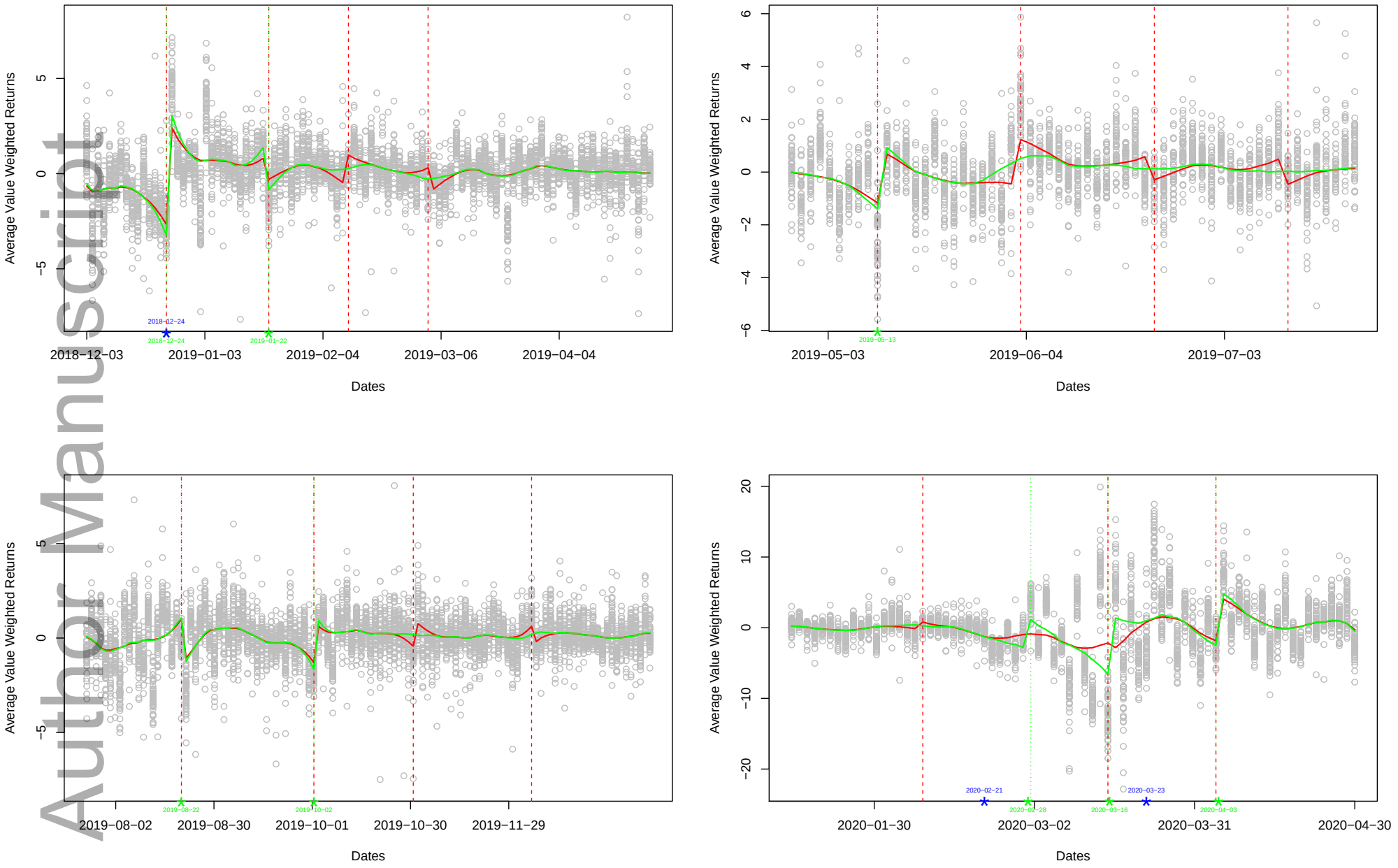


Figure 4: Average value weighted returns for 49 industry portfolios. The solid red curve is the mean function estimate using our method, and the vertical dash lines indicate the change point locations. The green curve is the estimate by using Xia and Qiu (2015). The blue points are change points detected by Cho (2016).

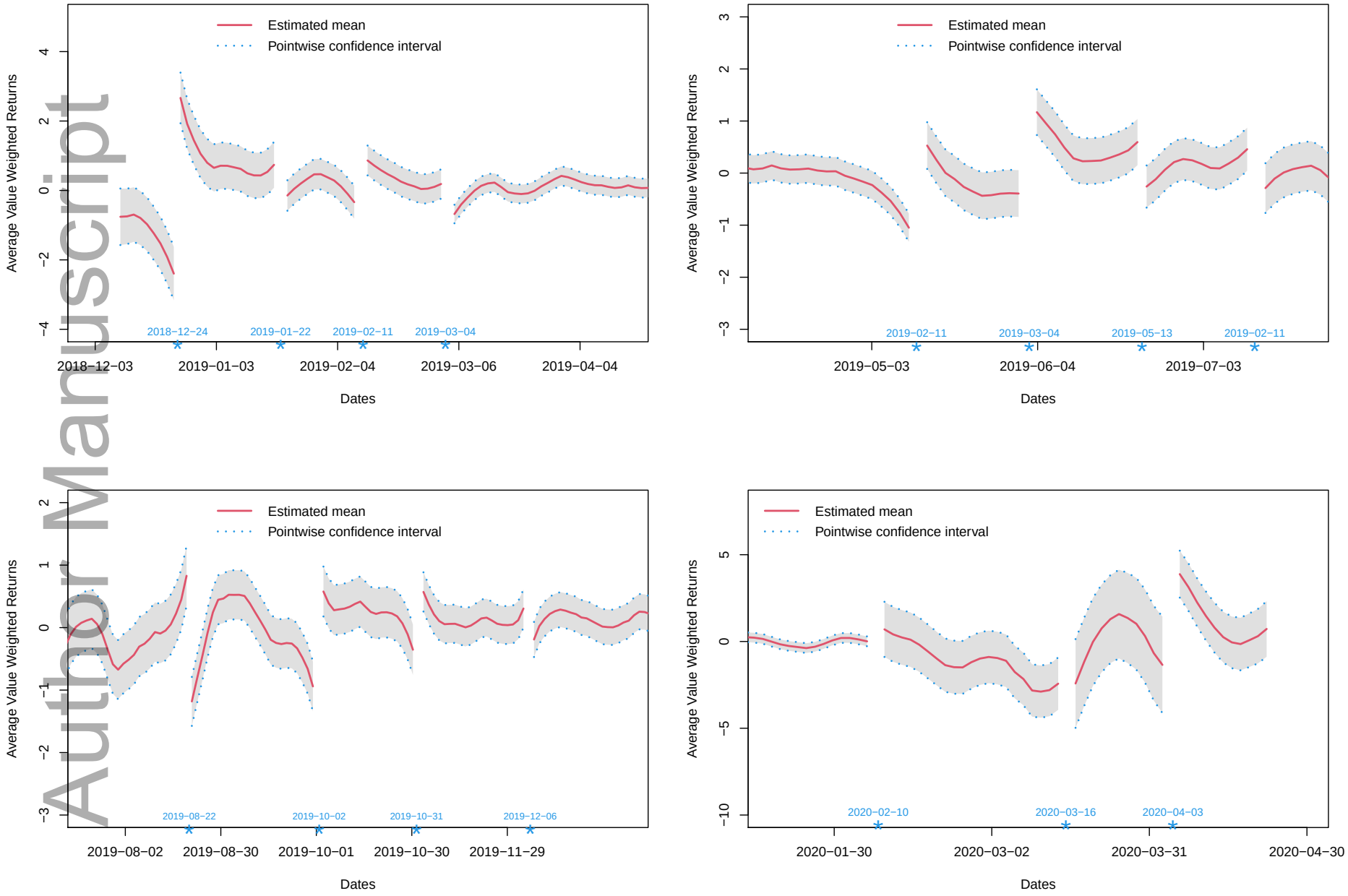


Figure 5: Smoothed mean function estimation and 95% pointwise confidence band from asymptotic normal.

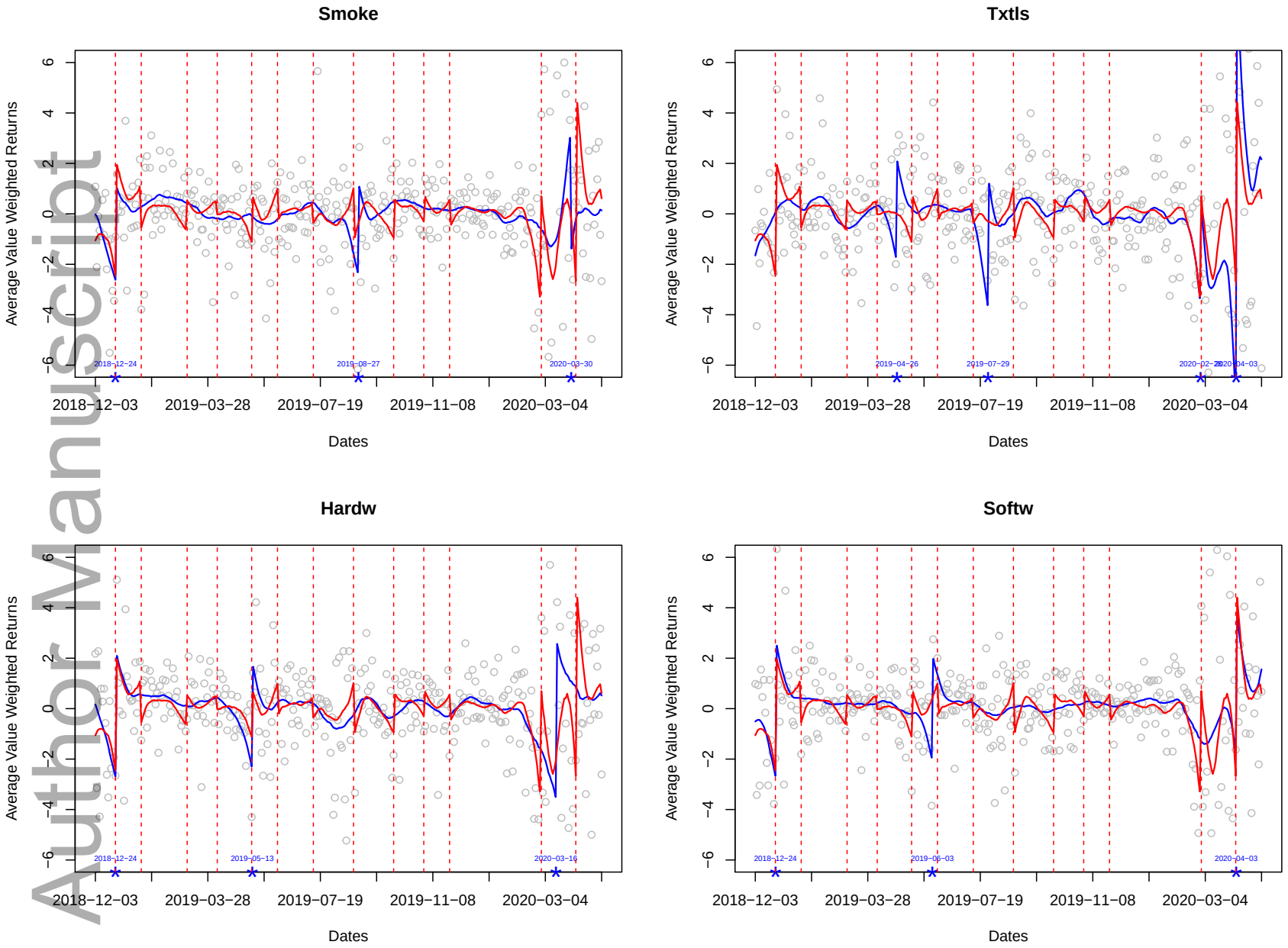


Figure 6: Individual curve estimate of average value weighted returns for 4 industry portfolios. The blue solid curve is the individual estimate and the red curve is $\widehat{\mu}(t)$ using our approach.

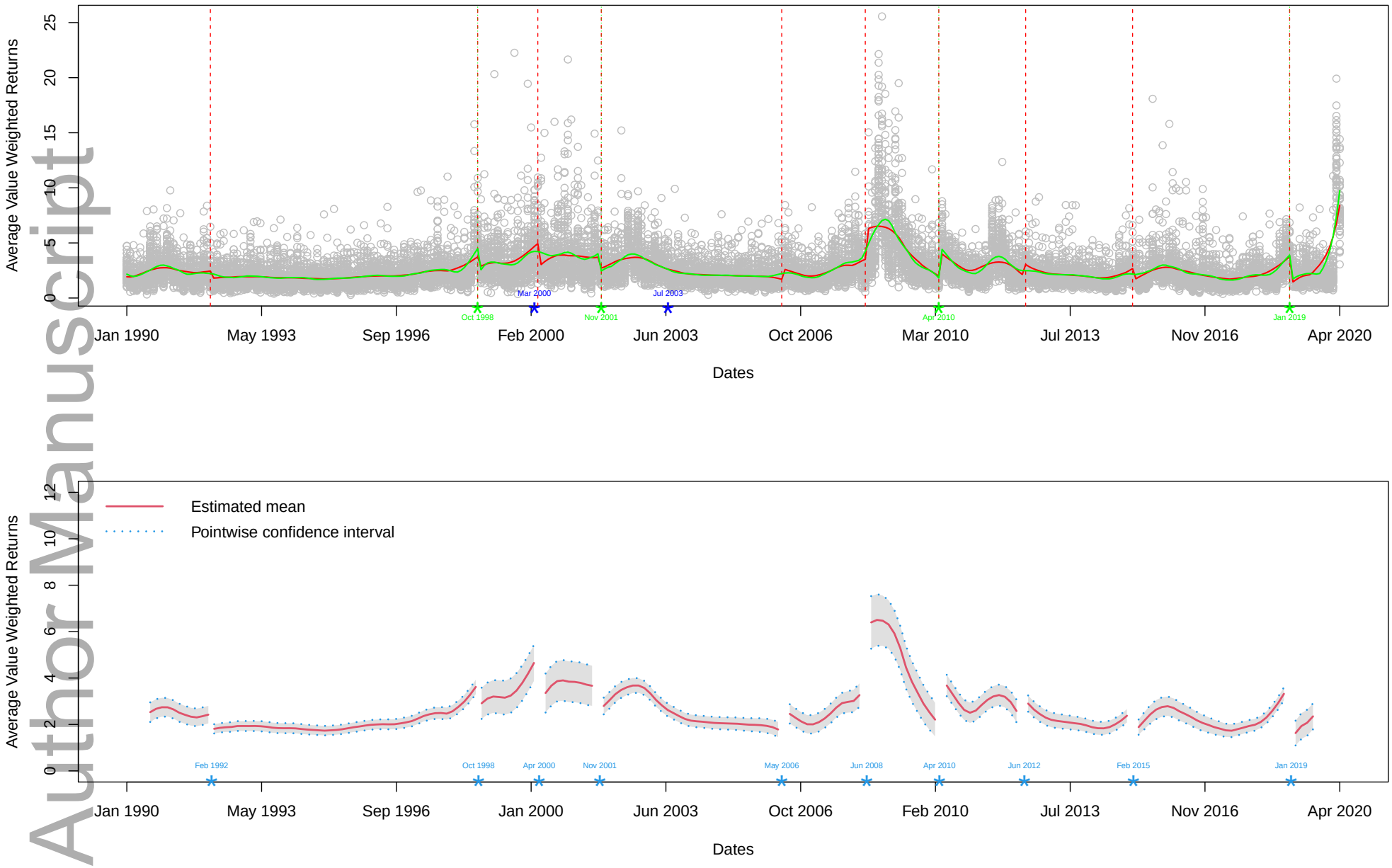


Figure 7: Monthly maximum average value weighted returns for 49 industry portfolios. Top panel: The red solid line is the mean function estimate $\widehat{\mu}(t)$, the vertical dash lines show the change point locations, the green curve is the estimate using Xia and Qiu (2015), and the blue points are change detected by Cho (2016). Bottom panel: $\widehat{\mu}(t)$ and 95% pointwise confidence band using our method.

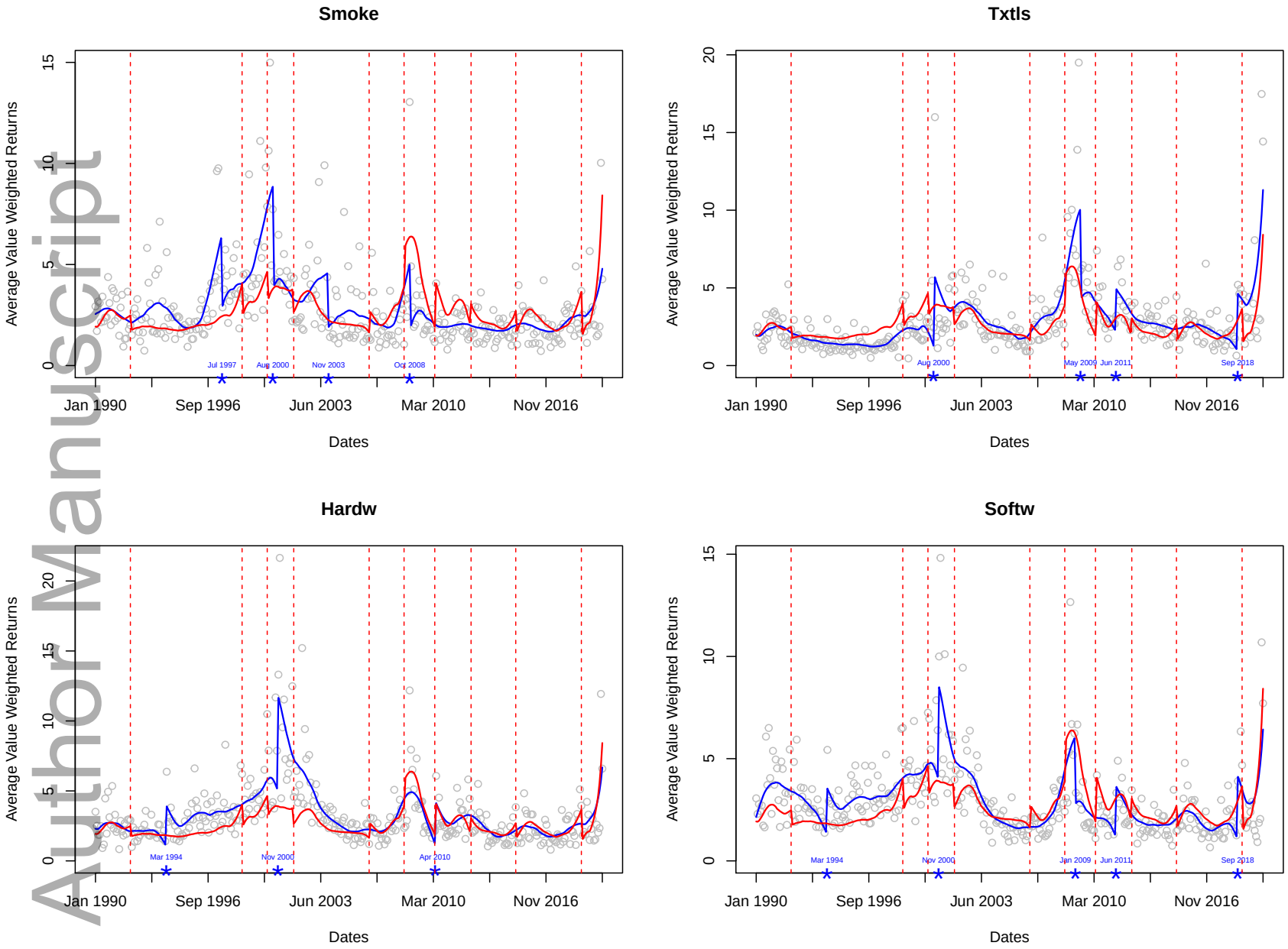


Figure 8: Individual curve estimates of monthly maximum average value weighted returns for 4 industry portfolios.The blue curve is the individual estimate and the red curve is $\widehat{\mu}(t)$ using our approach.

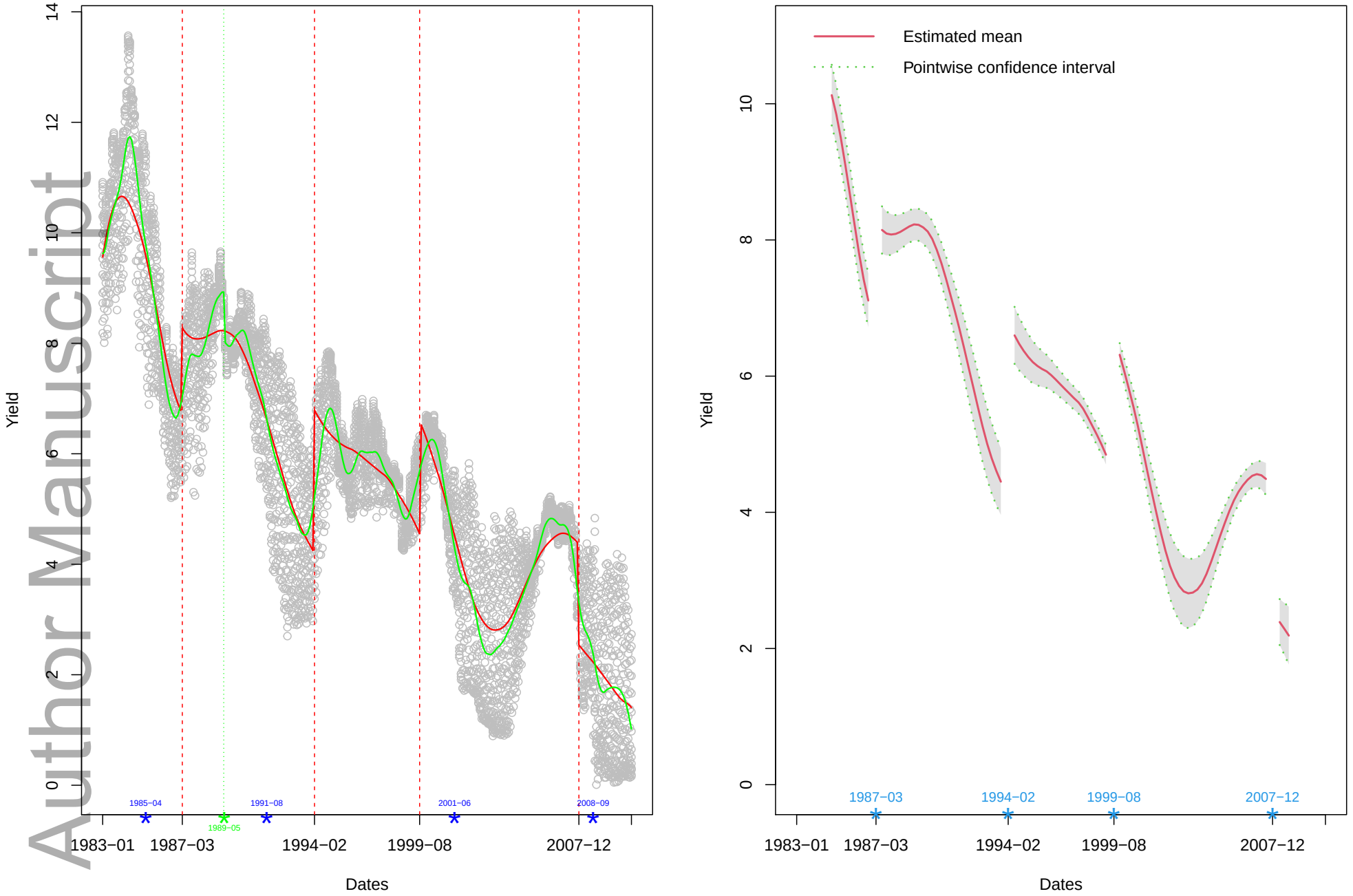


Figure 9: End-of-month U.S. Treasury bill and bond yields at maturities ranging from three months to ten years. Left panel: The red solid line is the mean function estimate $\widehat{\mu}(t)$, the vertical dash lines show the change points $\widehat{\tau}_k$, the green curve is obtained by using Xia and Qiu (2015), and the blue points are breaks detected by Cho (2016). Right panel: Mean function estimation and

Appendix: Technical Details

A. Notations and Basic Lemmas

For easy reference, we first collect all the notations needed in the proof here.

Let $K_h(u) = (1/h)K(u/h)$ and $K_{h,(r)}(u) = (1/h)K(u/h)(u/h)^r, r = 0, 1$. For $r = 0, 1$ and $\tau \in (0, 1)$, define

$$S_r(t, \tau) = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_{h,(r)}(T_{ij} - t) I(T_{ij} \geq \tau), \tag{A.1}$$

and

$$H(t, \tau) = \frac{S_0(t, \tau)S_2(t) - S_1(t, \tau)S_1(t)}{S_0(t)S_2(t) - S_1^2(t)} - I(t \geq \tau). \tag{A.2}$$

Let

$$K^*(u) = K(u)\frac{\upsilon_2 - \upsilon_1 u}{\upsilon_0\upsilon_2 - \upsilon_1^2} \text{ where } \upsilon_j = \int_{-1}^{1} K(u)u^j du, \tag{A.3}$$

and define

$$\widetilde{H}(\lambda) = \int_{-1}^{1} K^*(u) I(u \geq -\lambda) du - I(\lambda \geq 0), \ \lambda \in [-1, 1]. \tag{A.4}$$

Note that $\widetilde{H}(\lambda)$ can also be expressed as

$$\widetilde{H}(\lambda) = I(-1 \leq \lambda < 0) \int_{-\lambda}^{1} K^*(u) du - I(0 \leq \lambda \leq 1) \int_{-1}^{-\lambda} K^*(u) du. \tag{A.5}$$

Furthermore, for $\ell = \pm$, let $S_{\ell,r}(t)$, $R_{\ell,r}(t)$, $K_{\ell,h}(u)$, $K_{\ell,h,(r)}(u)$, $K_\ell^*(u)$, $S_{\ell,r}(t, \tau)$, $H_\ell(t, \tau)$ and $\widetilde{H}_\ell(\lambda)$ be defined in the same way as $S_r(t)$, $R_r(t)$ (in (5)), $K_h(u)$, $K_{h,(r)}(u)$, $K^*(u)$, $S_r(t, \tau)$, $H(t, \tau)$ and $\widetilde{H}(\lambda)$, respectively, except with K_ℓ replacing K.

The following lemma slightly generalizes Lemmas 1 - 3 in Li and Hsing (2010). As usual, let $t_1 \wedge t_2 = \min(t_1, t_2)$ and $t_1 \vee t_2 = \max(t_1, t_2)$.

Lemma 1. *Assume that (C5), (C6) and (C7) hold for some ψ and some bandwidth sequence $h = h_n$ tending to 0. Let $\mathscr{Z}_{ij} = U_i(T_{ij})$ or e_{ij} for $1 \leq i \leq n, 1 \leq j \leq m_i$,*

and

$$\begin{aligned} G_n(t_1,t_2) &= \sum_{i=1}^{n}\left\{ w_i \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(T_{ij} \in [t_1 \wedge t_2, t_1 \vee t_2]) \right\}, \\ G(t_1,t_2) &= \mathbb{E}\{G_n(t_1,t_2)\} \\ V_n(t,h) &= \sup_{|u|\le h} |G_n(t,t+u) - G(t,t+u)|, \ h>0. \end{aligned} \tag{A.6}$$

Then

(i) $\sup_{t\in[0,1]} V_n(t,h) = O\left(h\varsigma_n(h)\right)$ *a.s.*

(ii) $\sup_{t\in[0,1]} \varsigma_n^{-1}(h)|D_{r,n}(t) - \mathbb{E}\{D_{r,n}(t)\}| = O(1)$ *a.s., where, for any nonnegative integer* r, $\ell = +,-$,

$$D_{r,n}(t) := \sum_{i=1}^{n}\left[w_i \sum_{j=1}^{m_i} K_{h,(r)}(T_{ij}-t)\mathscr{Z}_{ij} \right].$$

Proof. The details of this proof will be omitted since they mirror those for Lemmas 1 and 2 of Li and Hsing (2010). The only major difference is that we bound $G_n(t,t+u) - G_n^*(t,t+u)$ by the following argument using Markov's inequality:

$$\begin{aligned} &\big(G_n(t,t+u) - G_n^*(t,t+u)\big) \\ &\quad\le \sum_{i=1}^{n}\left\{ w_i \sum_{j=1}^{m_i} \mathscr{Z}_{ij} I(\mathscr{Z}_{ij} > Q_n) \right\} \\ &\quad\le (Q_n/\psi(Q_n)) \sum_{i=1}^{n}\left\{ w_i \sum_{j=1}^{m_i} \psi(\mathscr{Z}_{ij}) I(\mathscr{Z}_{ij} > Q_n) \right\} \\ &\quad\le (Q_n/\psi(Q_n)) \sum_{i=1}^{n}\left\{ w_i \sum_{j=1}^{m_i} \psi(\mathscr{Z}_{ij}) \right\} \end{aligned}$$

since $\psi(x)/x$ is increasing. In Li and Hsing (2010), $\psi(x)$ was assumed to be x^α for some $\alpha > 2$. □

The following lemmas combine ideas from Li and Hsing (2010) and Xia and Qiu (2015).

Lemma 2. *Assume that (C1)–(C3) and (C6)–(C7) hold, and the kernel function* K *is uniformly Lipschitz continuous. Then, for any interior point* $\tau \in (0,1)$, *with*

probability one

$$\sup_{\lambda\in[-1,1]}\left|H(\tau+\lambda h,\tau)-\widetilde{H}(\lambda)\right|=O\left(h+\varsigma_n(h)\right)\ \ as\ h\to 0.$$

Proof. For convenience, let $t=\tau+\lambda h$. By assumptions (C2) and (C8),

$$\begin{aligned}\mathbb{E}[S_r] &= \frac{1}{h}\sum_{i=1}^{n} w_i \sum_{j=1}^{m_i}\mathbb{E}\left[K\left(\frac{T_{ij}-t}{h}\right)\left(\frac{T_{ij}-t}{h}\right)^r\right]\\ &= \sum_{i=1}^{n} w_i m_i \int_{-1}^{1} K(u)u^r f_T(uh+t)du\\ &= \sum_{i=1}^{n} w_i m_i f_T(t)\upsilon_r + O(h)\\ &= f_T(t)\upsilon_r + O(h),\end{aligned}$$

where $\sum_{i=1}^{n} w_i m_i = 1$ and $\upsilon_r = \int_{-1}^{1} t^r K(t)dt$. By (ii) of Lemma 1, we conclude that, uniformly for $t\in[a+h,b-h]$,

$$S_r(t)=\sum_{i=1}^{n} w_i m_i f_T(t)\upsilon_r + O\left(h+\varsigma_n(h)\right) \tag{A.7}$$

where $r=0,1,2$ and w_i satisfies (C6)–(C7). The same rate can also be similarly seen to hold for boundary points. Then, by (A.7), we have

$$\begin{aligned}&H(t,\tau)\\ &= \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_h(T_{ij}-t)\frac{S_2\left(\frac{T_{ij}-t}{h}\right)-S_1\left(\frac{T_{ij}-t}{h}\right)}{S_0\left(\frac{T_{ij}-t}{h}\right)S_2\left(\frac{T_{ij}-t}{h}\right)-S_1\left(\frac{T_{ij}-t}{h}\right)^2} I(T_{ij}\ge\tau)\\ &\quad - I(t\ge\tau) \qquad\qquad (A.8)\\ &= \frac{1}{\sum_{i=1}^{n} w_i m_i f_T(t)}\sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_h\left(\frac{T_{ij}-t}{h}\right) I(T_{ij}\ge\tau)\frac{\upsilon_2-\upsilon_1\left(\frac{T_{ij}-t}{h}\right)}{\upsilon_0\upsilon_2-\upsilon_1^2}\\ &\quad - I(t\ge\tau)+O\left(h+\varsigma_n(h)\right).\end{aligned}$$

The expectation of the first term is

$$\begin{aligned}&\frac{1}{f_T(t)}\mathbb{E}\left[K_h\left(\frac{T-t}{h}\right)I(T\ge\tau)\frac{\upsilon_2-\upsilon_1\left(\frac{T-t}{h}\right)}{\upsilon_0\upsilon_2-\upsilon_1^2}\right]\\ &= \frac{1}{f_T(t)}\int_{-1}^{1} K^*(u)I(u\ge-\lambda)f_T(uh+t)du\end{aligned}$$

$$= \int_{-1}^{1} K^*(u) I(u \geq -\lambda) du + O(h)$$

where $K^*(u)$ is as in (A.3). Thus, by (A.5) and applying Lemma 1 on the first term in (A.8), we obtain $H(t,\tau) = \widetilde{H}(\lambda) + O\left(h + \varsigma_n(h)\right)$ a.s. □

Define

$$D_{-,h,k} = [\tau_k - h, \tau_k],\ D_{+,h,k} = [\tau_k, \tau_k + h],\ k = 1, \ldots, M,$$

and, for $\ell = \pm$,

$$D_{\ell,h,M_1 \to M_2} = \bigcup_{k=M_1}^{M_2} D_{\ell,h,k}, \quad \bar{D}_{\ell,h,M_1 \to M_2} = D \backslash D_{\ell,h,M_1 \to M_2},\ 1 \leq M_1 \leq M_2 \leq M.$$

Lemma 3. *Assume the assumptions for Theorem 3.1. Then the following hold with probability one:*

(i) $\sup_{t \in \bar{D}_{+,h,1\to M}} |\widehat{\mu}_-(t;h) - \mu(t)|$ *and* $\sup_{t \in \bar{D}_{-,h,1\to M}} |\widehat{\mu}_+(t;h) - \mu(t)|$ *are both* $O(h^2 + \varsigma_n(h))$*, and*

(ii) $\sup_{\lambda \in [-1,1]} \left| \widehat{\mu}(\tau_k + \lambda h; h) - \mu(\tau_k + \lambda h) - d_k \widetilde{H}(\lambda) \right| = O\left(h + \varsigma_n(h)\right)$ *for* $k = 1, \ldots, M$.

Proof. (i) It follows from the proof of Theorem 3.1 of Li and Hsing (2010), applying Lemma 1, that

$$\sup_{t \in D_h} |\widehat{\nu}^*(t;h) - \nu(t)| = O\left(h^2 + \varsigma_n(h)\right). \tag{A.9}$$

Note that the required smoothness assumption on ν for this to work is that ν has a bounded second-order derivative at every point in $[t-h, t+h] = t + [-h, h]$, where $[-h, h]$ is the support of K_h. Now, for $t \in \bar{D}_{-,h,1\to M}$, μ has a bounded second-order derivative at every point in $[t, t+h] = t + [0, h]$, where $[0, h]$ is the support of $K_{+,h}$. Thus, the exact same approach applies to give the rate of $\widehat{\mu}_+$. The rate for $\widehat{\mu}_-$ on $\bar{D}_{+,h,1\to M}$ is similar.

(ii) Consider $t \in D_{h,k}$. By (7) we have

$$\widehat{\mu}(t;h) - \widehat{\nu}^*(t) = \sum_{s:\tau_s < t} d_s + d_k H(t, \tau_k).$$

As a result, for $t = \tau_k + \lambda h, \lambda \in [-1, 1]$,

$$\widehat{\mu}(t;h) - \mu(t) - \widetilde{H}(\lambda) = \widehat{\nu}^*(t) - \nu(t) + d_k\big(H(\tau_k + \lambda h, \tau_k) - \widetilde{H}(\lambda)\big).$$

Thus, (ii) follows from (A.9) and Lemma 2. □

B. Proof of Theorem 3.1 and Theorem 3.4

Proof of Theorem 3.1 . By (i) of Lemma 3, we have

$$\begin{aligned}\sup_{t\in\bar{D}_{h,1\to M}} |\Delta_n(t;h)| &= \sup_{t\in\bar{D}_{h,1\to M}} |\widehat{\mu}_+(t;h) - \widehat{\mu}_-(t;h)| \\ &\le \sup_{t\in\bar{D}_{h,1\to M}} |\widehat{\mu}_+(t;h) - \mu(t)| + \sup_{t\in\bar{D}_{h,1\to M}} |\widehat{\mu}_-(t;h) - \mu(t)| \\ &= O\left(h^2 + \varsigma_n(h)\right) \quad a.s.\end{aligned}$$

which is (i). To provie (ii), first observe that $g(\lambda) = \widetilde{H}_+(\lambda) - \widetilde{H}_-(\lambda)$. Note that (ii) of Lemma 3 is true for any kernel. Hence, applying that with one-sided kernels K_ℓ, we have for all k

$$\begin{aligned}&\sup_{\lambda\in[-1,1]} |\Delta_n(\tau_k + \lambda h; h) - d_k g(\lambda)| \\ &\le \sup_{\lambda\in[-1,1]} |\widehat{\mu}_+(\tau_k + \lambda h; h) - \mu(\tau_k + \lambda h) - d_k\widetilde{H}_+(\lambda)| \\ &\quad + \sup_{\lambda\in[-1,1]} |\widehat{\mu}_-(\tau_k + \lambda h; h) - \mu(\tau_k + \lambda h) - d_k\widetilde{H}_-(\lambda)| \\ &= O\left(h + \varsigma_n(h)\right) \quad a.s.\end{aligned}$$

□

Proof of Theorem 3.4. By (i) of Theorem 3.1,

$$\sup_{t\in\bar{D}_{h_\tau,1\to M}} |\Delta_n(t)| = O\left(h_\tau^2 + \varsigma_n(h_\tau)\right) \qquad a.s.$$

Thus, there is a set A satisfying $P(A) = 1$ such that for all $\omega \in A$, there exits $N = N(\omega)$ such that when $n > N$, we have

$$\sup_{t\in\bar{D}_{h_\tau,1\to M}} |\Delta_n(t)| \le \zeta_n.$$

On the other hand, by (ii) of Theorem 3.1, we have

$$\min_{k\in\{1,\ldots,M\}} |\Delta_n(\tau_k)| \succ \zeta_n \qquad a.s.$$

The choice $\epsilon = 2$ ensures that, asymptotically, the same jump point will not be identified more than once. Therefore, with probability one, for all large n there will be exactly M separate regions where $|\Delta_n(t)|$ exceeds the threshold ζ_n. Thus, the conclusion of the result follows. □

C. Proof of Theorem 3.2

Proof. By (i) and (ii) of Theorem 3.1, we conclude that there is a set A with $\mathbb{P}(A) = 1$, where, for each $\omega \in A$, there exist a constant $C = C(\omega)$ and a positive integer $N = N(\omega)$ such that

$$\sup_{t\in D} |\Delta_n(t) - \mathcal{G}_n(t)| \le \frac{C}{2}(h_\tau + \varsigma_n(h_\tau)) \quad \text{for all } n \ge N. \tag{A.10}$$

We first focus on $\widehat{\tau}_1$, which is the global maximizer of $|\Delta_n(t)|$. Observe that, by (A.10), $\widehat{\tau}_1$ must be in $D_{h_\tau,1}$ for all large n. Focusing on the subset $D_{h_\tau,1}$ of the global set D in (A.10), we have

$$\sup_{\lambda\in[-1,1]} |\widehat{\mathcal{G}}^*(\lambda) - \mathcal{G}^*(\lambda)| \le \frac{C}{2}(h_\tau + \varsigma_n(h_\tau)) \quad \text{for all } n \ge N, \tag{A.11}$$

where $\widehat{\mathcal{G}}^*(\lambda) := |\Delta_n(\tau_1+\lambda h_\tau)|, \mathcal{G}^*(\lambda) := |\mathcal{G}_n(\tau_1+\lambda h_\tau)|, \lambda \in [-1,1]$, and $\widehat{\tau}_1 = \tau_1 + \widehat{\lambda} h_\tau$ with $\widehat{\lambda}$ the maximizer of $\widehat{\mathcal{G}}^*(\lambda)$. Denote $a_n = h_\tau + \varsigma_n(h_\tau)$ and pick any $\eta \in (0, K_+^*(0))$. Let $W_n = (-C_1 a_n, C_1 a_n)$ where $C_1 := \frac{C}{|d_1|(K_+^*(0)-\eta)}$. It follows that

$$\sup_{\lambda\in[-1,1]\cap W_n^c} \mathcal{G}^*(\lambda) = |d_1| \sup_{\lambda\in[-1,1]\cap W_n^c} |g(\lambda)| = |d_1| g(C_1 a_n)$$

since $|g|$ is continuous and by (C2) achieves its maximum at 0. Also, since K_+^* is continuous at 0, for all large n

$$\inf_{\lambda\in(0,C_1 a_n)} -g'(\lambda) = \inf_{\lambda\in(0,C_1 a_n)} K_+^*(\lambda) \ge K_+^*(0) - \eta.$$

Thus, by the mean value theorem, for all large n

$$\begin{aligned}\mathcal{G}^*(0) - \sup_{\lambda\in[-1,1]\cap W_n^c} \mathcal{G}^*(\lambda) &= |d_1| g(0) - |d_1| g(-C_1 a_n) \\ &\ge |d_1| C_1 a_n \inf_{\lambda\in(0,C_1 a_n)} -g'(\lambda) \ge C a_n,\end{aligned}$$

or, equivalently,

$$\mathcal{G}^*(0) - Ca_n \geq \sup_{\lambda \in [-1,1] \cap W_n^c} \mathcal{G}^*(\lambda). \tag{A.12}$$

By (A.11) and the fact that $\widehat{\lambda}$ is the maximizer of $\widehat{\mathcal{G}}^*(\lambda)$, we have, for all $n \geq N$,

$$\begin{aligned}\mathcal{G}^*(0) - \mathcal{G}^*(\widehat{\lambda}) &\leq |\mathcal{G}^*(0) - \widehat{\mathcal{G}}^*(0)| + \{\widehat{\mathcal{G}}^*(0) - \widehat{\mathcal{G}}^*(\widehat{\lambda})\} + |\widehat{\mathcal{G}}^*(\widehat{\lambda}) - \mathcal{G}^*(\widehat{\lambda})| \\ &< \frac{C}{2}a_n + 0 + \frac{C}{2}a_n = Ca_n,\end{aligned}$$

and hence

$$\mathcal{G}^*(0) - Ca_n < \mathcal{G}^*(\widehat{\lambda}). \tag{A.13}$$

Combining (A.12) and (A.13) leads to the conclusion that $\widehat{\lambda} \in W_n$. Therefore,

$$|\widehat{\lambda} - \lambda| < C_1 a_n \quad \text{or} \quad |\widehat{\tau}_1 - \tau_1| < h_\tau C_1 a_n \quad \text{for all } n \geq N. \tag{A.14}$$

In accordance with the procedure described in Section 2.3, we now delete the interval $[\widehat{\tau}_1 - \epsilon h_\tau, \widehat{\tau}_1 + \epsilon h_\tau]$ with $\epsilon > 1$. By (A.14), for all $n \geq N$, we have $[\tau_1 - h_\tau, \tau_1 + h_\tau] \subset [\widehat{\tau}_1 - \epsilon h_\tau, \widehat{\tau}_1 + \epsilon h_\tau]$ and $\tau_2, \ldots, \tau_M \in [\widehat{\tau}_1 - \epsilon h_\tau, \widehat{\tau}_1 + \epsilon h_\tau]^c$. Thus, the above argument can be applied iteratively for $\tau_2, \ldots, \tau_M$.

(ii) By part (i), Since $\widehat{\tau}_k - \rho_d \in \bar{D}_{+,h,1\to M}$ and $\widehat{\tau}_k + \rho_d \in \bar{D}_{-,h,1\to M}$ a.s. for large n, (i) of Lemma 3 implies that

$$\begin{aligned}\widehat{\mu}_-(\widehat{\tau}_k - \rho_d; h_d) &= \mu(\widehat{\tau}_k - \rho_d) + O\left(h_d^2 + \varsigma_n(h_d)\right), \\ \widehat{\mu}_+(\widehat{\tau}_k + \rho_d; h_d) &= \mu(\widehat{\tau}_k + \rho_d) + O\left(h_d^2 + \varsigma_n(h_d)\right).\end{aligned}$$

It follows that

$$\begin{aligned}\widehat{d}_k - d_k &= \widehat{\mu}_+(\widehat{\tau}_k + \rho_d; h_d) - \widehat{\mu}_-(\widehat{\tau}_k - \rho_d; h_d) - d_k \\ &= \mu(\widehat{\tau}_k + \rho_d) - \mu(\widehat{\tau}_k - \rho_d) - d_k + O\left(h_d^2 + \varsigma_n(h_d)\right) \\ &= \nu(\widehat{\tau}_k + \rho_d) - \nu(\widehat{\tau}_k - \rho_d) + O\left(h_d^2 + \varsigma_n(h_d)\right) \\ &= O\left(\rho_d + h_d^2 + \varsigma_n(h_d)\right),\end{aligned}$$

where we applied the assumption that ν has a uniformly bounded derivative in the last step.

(iii) In the proof below, we will hide the tuning parameters to simplify notation where $\widehat{\mu}(t) = \widehat{\mu}(t; h_\tau, h_d, \rho_d)$. Note that the bandwidth used in $\widehat{\nu}^*(t)$ and $\widehat{\nu}(t)$ is h_τ, but the latter also depends on the bandwidth h_d, which is needed for $\widehat{d}_k$. However, for ease of presentation we leave the bandwidths out of the notations of these estimators in the proof below. It follows that

$$\begin{aligned}
&\widehat{\mu}(t) - \mu(t) \\
&= \widehat{\nu}(t) + \sum_{k=1}^{M} \widehat{d}_k I(t \geq \widehat{\tau}_k) - \left(\nu(t) + \sum_{k=1}^{M} d_k I(t \geq \tau_k) \right) \\
&= (\widehat{\nu}^*(t) - \nu(t)) + \widehat{\nu}(t) - \widehat{\nu}^*(t) + \sum_{k=1}^{M} \widehat{d}_k I(t \geq \widehat{\tau}_k) - \sum_{k=1}^{M} d_k I(t \geq \tau_k).
\end{aligned} \tag{A.15}$$

With the notation of $S_r(t, \tau)$ and $H(t, \tau)$ in (A.1) and (A.2), we can write

$$\begin{aligned}
&\widehat{\nu}(t) - \widehat{\nu}^*(t) + \sum_{k=1}^{M} \widehat{d}_k I(t \geq \widehat{\tau}_k) - \sum_{k=1}^{M} d_k I(t \geq \tau_k) \\
&= \sum_{k=1}^{M} d_k \left(\frac{S_0(t, \tau_k) S_2(t) - S_1(t, \tau_k) S_1(t)}{S_0(t) S_2(t) - S_1^2(t)} - I(t \geq \tau_k) \right) \\
&\quad - \sum_{k=1}^{M} \widehat{d}_k \left(\frac{S_0(t, \widehat{\tau}_k) S_2(t) - S_1(t, \widehat{\tau}_k) S_1(t)}{S_0(t) S_2(t) - S_1^2(t)} - I(t \geq \widehat{\tau}_k) \right) \\
&= \sum_{k=1}^{M} d_k H(t, \tau_k) - \sum_{k=1}^{M} \widehat{d}_k H(t, \widehat{\tau}_k) \\
&= \sum_{k=1}^{M} d_k H(t, \tau_k) - \sum_{k=1}^{M} d_k H(t, \widehat{\tau}_k) + O\left(\rho_d + h_d^2 + \varsigma_n(h_d) \right).
\end{aligned} \tag{A.16}$$

Recall that $D_{h_\tau} = [h_\tau, 1 - h_\tau]$ and $D_{h_\tau, k} = [\tau_k - h_\tau, \tau_k + h_\tau]$. Now, define $\widehat{D}_{h_\tau, k} = [\widehat{\tau}_k - h_\tau, \widehat{\tau}_k + h_\tau]$. By the definition of $H(t, \tau)$, one can verify that

$$\begin{aligned}
&\sum_{k=1}^{M} d_k H(t, \tau_k) = 0 \text{ for } t \notin \cup_{k=1}^{M} D_{h_\tau, k}, \\
&\sum_{k=1}^{M} d_k H(t, \widehat{\tau}_k) = 0 \text{ for } t \notin \cup_{k=1}^{M} \widehat{D}_{h_\tau, k}.
\end{aligned} \tag{A.17}$$

Since the sets $D_{h_\tau, k} \cup \widehat{D}_{h_\tau, k}, k = 1, \ldots, M$, are disjoint, it suffices to focus on the rate of $H(t, \tau_k) - H(t, \widehat{\tau}_k)$ on $D_{h_\tau, k} \cup \widehat{D}_{h_\tau, k}$ for each k. Write

$$D_{h_\tau, k} \cup \widehat{D}_{h_\tau, k} = (D_{h_\tau, k} \cap \widehat{D}_{h_\tau, k}) \cup (D_{h_\tau, k} \backslash \widehat{D}_{h_\tau, k}) \cup (\widehat{D}_{h_\tau, k} \backslash D_{h_\tau, k}).$$

As in the proof of Lemma 2, we can express each point t in $D_{h_\tau,k}$ as $t = \tau_k + \lambda h_\tau$ for some $\lambda \in [-1, 1]$. Likewise, each point t in $\widehat{D}_{h_\tau,k}$ can be expressed as $t = \widehat{\tau}_k + \widehat{\lambda} h_\tau$ for some $\widehat{\lambda} \in [-1, 1]$. Thus, by Lemma 2, uniformly for each $t \in D_{h_\tau,k} \cap \widehat{D}_{h_\tau,k}$,

$$H(t, \tau_k) - H(t, \widehat{\tau}_k) = \widetilde{H}(\lambda) - \widetilde{H}(\widehat{\lambda}) + O(h_\tau + \varsigma_n(h_\tau)). \tag{A.18}$$

It follows that $\widehat{\lambda} - \lambda = (\widehat{\tau}_k - \tau_k)/h_\tau = O\left(h_\tau + \varsigma_n(h_\tau)\right)$. Also, note that λ and $\widehat{\lambda}$ have different signs only for t between τ_k and $\widehat{\tau}_k$, which cannot be the case for $t \notin D_{\rho_d,k}$ by (i). Thus, by (A.5) and (A.18),

$$\sup_{t \in (D_{h_\tau,k} \cap \widehat{D}_{h_\tau,k}) \setminus D_{\rho_d,k}} |H(t, \tau_k) - H(t, \widehat{\tau}_k)| = O(h_\tau + \varsigma_n(h_\tau)). \tag{A.19}$$

Next, consider $H(t, \tau_k) - H(t, \widehat{\tau}_k)$ for $t \in (D_{h_\tau,k} \setminus \widehat{D}_{h_\tau,k}) \cup (\widehat{D}_{h_\tau,k} \setminus D_{h_\tau,k})$. Observe that in this case, one of $H(t, \tau_k)$ and $H(t, \widehat{\tau}_k)$ is exactly equal to zero and the other term is $\widetilde{H}(\lambda) + O(h_\tau + \varsigma_n(h_\tau))$ where $1 - |\lambda| = O(h_\tau + \varsigma_n(h_\tau))$. Hence, again we have

$$\sup_{t \in (D_{h_\tau,k} \setminus \widehat{D}_{h_\tau,k}) \cup (\widehat{D}_{h_\tau,k} \setminus D_{h_\tau,k})} |H(t, \tau_k) - H(t, \widehat{\tau}_k)| = O(h_\tau + \varsigma_n(h_\tau)). \tag{A.20}$$

Thus, (iii) follows from combining (A.15), (A.9), (A.16), (A.19) and (A.20). □

D. Proof of Theorem 3.3

Proof. By (A.15) and the proofs there, with probability one,

$$\widehat{\mu}(t) - \mu(t) = \widehat{\nu}^*(t) - \nu(t) + O(h_\tau + \varsigma_n(h_\tau) + h_d^2 + \varsigma_n(h_d))),$$

uniformly for $t \in \bar{D}_{\rho_d,1\to M}$. Thus, by (21), $\Gamma^{-1/2}(\widehat{\mu}(t) - \mu(t)) = \Gamma^{-1/2}(\widehat{\nu}^*(t) - \nu(t)) + o(1)$ for such t and the result is then a direct consequence of Theorem 3.1 of Zhang and Wang (2016). □

E. Proof of Corollary 3.1

Proof. Recall the definition of half-kernel estimator with bandwidth $h = h_\tau$,

$$\widehat{\mu}_\ell(t) = \frac{R_{\ell,0} S_{\ell,2} - R_{\ell,1} S_{\ell,1}}{S_{\ell,0} S_{\ell,2} - S_{\ell,1}^2},$$

where for $r = 0, 1, 2$,

$$S_{\ell,r}(t) = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_{h,\ell}(T_{ij} - t)\{(T_{ij} - t)/h\}^r,$$

$$R_{\ell,r}(t) = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_{h,\ell}(T_{ij} - t)\{(T_{ij} - t)/h\}^r Y_{ij}.$$

Then, we have

$$\widehat{\mu}_\ell(t) = \frac{R_{\ell,0}}{\widetilde{S}_{\ell,0}} - \frac{\widetilde{S}_{\ell,1}}{\widetilde{S}_{\ell,0}}\widehat{\mu}_\ell^{(1)}(t) + t\widehat{\mu}_\ell^{(1)}(t),$$

where $\widehat{\mu}_\ell^{(1)}(t) = \frac{1}{h}\frac{R_{\ell,1} - R_{\ell,0}S_{\ell,1}/S_{\ell,0}}{S_{\ell,2} - S_{\ell,1}^2/S_{\ell,0}}$ and

$$\widetilde{S}_{\ell,r} = \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} K_{h,\ell}(T_{ij} - t)T_{ij}^r, \qquad r = 0, 1.$$

If $t \in (0,1)\backslash\{\tau_1, \ldots, \tau_M\}$ is a differentiable point, following the proof of Theorem 3.1 of Zhang and Wang (2016), by the Cramér-Wold device and the Lyapunov central limit theorem due to the condition (20), we can achieve the asymptotic joint normality of $(R_{\ell,0} - \mathbb{E}[R_{\ell,0}], \widetilde{S}_{\ell,r} - \mathbb{E}[\widetilde{S}_{\ell,0}], \widetilde{S}_{\ell,1} - \mathbb{E}[\widetilde{S}_{\ell,1}])$ with convergence rate $[\min\{h/\sum_{i=1}^n m_i w_i^2, 1/\sum_{i=1}^n m_i(m_i-1)w_i^2\}]^{1/2}$. Explicitly, for $r, r' = 0, 1$, by Taylor expansion,

$$\mathbb{E}\tilde{S}_{\ell,r} = t^r f(t) + \frac{h^2}{2}\frac{\upsilon_{\ell,2}^2 - \upsilon_{\ell,1}\upsilon_{\ell,3}}{\upsilon_{\ell,0}\upsilon_{\ell,2} - \upsilon_{\ell,1}^2}(2rf^{(1)}(t) + t^r f^{(2)}(t)) + o(h^2);$$

$$\mathbb{E}R_{\ell,0} = \mu(t)f(t) + \frac{h^2}{2}\frac{\upsilon_{\ell,2}^2 - \upsilon_{\ell,1}\upsilon_{\ell,3}}{\upsilon_{\ell,0}\upsilon_{\ell,2} - \upsilon_{\ell,1}^2}(\mu(t)f^{(2)}(t) + 2\mu^{(1)}(t)f^{(1)}(t) + \mu^{(2)}(t)f(t)) + o(h^2);$$

$$\mathrm{Cov}(\tilde{S}_{\ell,r}, \tilde{S}_{\ell,r'}) = \frac{\sum_{i=1}^n m_i w_i^2}{h}\|K_\ell^*\|^2 t^{r+r'} f(t) + o\left(\frac{\sum_{i=1}^n m_i w_i^2}{h}\right);$$

$$\mathrm{Var}(R_{\ell,0}) = \frac{\sum_{i=1}^n m_i w_i^2}{h}\|K_\ell^*\|^2(\mu(t)^2 + R(t,t) + \sigma^2)f(t) + (\sum_{i=1}^n m_i(m_i-1)w_i^2)R(t,t)f(t)^2 + o\left(\frac{\sum_{i=1}^n m_i w_i^2}{h} + \sum_{i=1}^n m_i(m_i-1)w_i^2\right);$$

$$\mathrm{Cov}(R_{\ell,0}, \tilde{S}_{\ell,r}) = \frac{\sum_{i=1}^n m_i w_i^2}{h}\|K_\ell^*\|^2 t^r \mu(t)f(t) + o\left(\frac{\sum_{i=1}^n m_i w_i^2}{h}\right).$$

Then, the asymptotic normality of $\widehat{\mu}_\ell(t)$ follows from the delta method,

$$\widetilde{\Omega}^{-1/2}\left[\widehat{\mu}_\ell(t) - \frac{1}{2}h^2\frac{\upsilon_{\ell,2}^2 - \upsilon_{\ell,1}\upsilon_{\ell,3}}{\upsilon_{\ell,0}\upsilon_{\ell,2} - \upsilon_{\ell,1}^2}\mu^{(2)}(t) + o_p(h^2)\right] \xrightarrow{d} \mathcal{N}(0,1), \tag{A.21}$$

where

$$\widetilde{\Omega} = \frac{\sum_{i=1}^{n} m_i w_i^2}{h} \|K_\ell^*\|^2 \frac{R(t,t)+\sigma^2}{f_T(t)} + \left(\sum_{i=1}^{n} m_i(m_i-1) w_i^2 \right) R(t,t).$$

From (A.7) in Lemma 2 and $\sum_{i=1}^{n} w_i m_i = 1$, we have

$$S_{\ell,r}(t) = f_T(t)\upsilon_{\ell,r} + O(h + \varsigma_n(h)).$$

Then,

$$\begin{aligned}\widehat{\mu}_\ell(t) &= \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \frac{S_{\ell,2} - S_{\ell,1}\{(T_{ij}-t)/h\}}{S_{\ell,0}S_{\ell,2} - S_{\ell,1}^2} K_{h,\ell}(T_{ij}-t) Y_{ij} \\ &= \frac{1}{f_T(t)} \sum_{i=1}^{n} w_i \sum_{j=1}^{m_i} \frac{\upsilon_{\ell,2} - \upsilon_{\ell,1}\left(\frac{T_{ij}-t}{h}\right)}{\upsilon_{\ell,0}\upsilon_{\ell,2} - \upsilon_{\ell,1}^2} K_{h,\ell}\left(T_{ij}-t\right) Y_{ij} + O(h+\varsigma_n(h)).\end{aligned}$$

By the definition $\Delta_n(t) = \widehat{\mu}_+(t) - \widehat{\mu}_-(t)$, we have

$$\mathbb{E}[\Delta_n(t)] = O(h + \varsigma_n(h)). \tag{A.22}$$

Note that $\min\left(h/\sum_{i=1}^{n} m_i w_i^2, 1/\sum_{i=1}^{n} N_i w_i^2\right)(h+\varsigma_n(h))^2 \to 0$, from (A.21) and (A.22), we have

$$\Omega^{-1/2}\left[\Delta_n(t) - O(h+\varsigma_n(h))\right] \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\Omega = \frac{2\sum_{i=1}^{n} m_i w_i^2}{h_\tau} \|K_+^*\|^2 \frac{R(t,t)+\sigma^2}{f_T(t)} + 2\left(\sum_{i=1}^{n} m_i(m_i-1) w_i^2 \right) R(t,t),$$

which completes the proof.

□

References

Aston, J. A. and C. Kirch (2012). Detecting and estimating changes in dependent functional data. *Journal of Multivariate Analysis 109*, 204–220.

Aue, A., R. Gabrys, L. Horváth, and P. Kokoszka (2009). Estimation of a change-point in the mean function of functional data. *Journal of Multivariate Analysis 100*(10), 2254–2269.

Aue, A., S. Hörmann, L. Horváth, and M. Hušková (2014). Dependent functional linear models with applications to monitoring structural change. *Statistica Sinica*, 1043–1073.

Aue, A., G. Rice, and O. Sönmez (2018). Detecting and dating structural breaks in functional data without dimension reduction. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(3), 509–529.

Bardsley, P., L. Horváth, P. Kokoszka, and G. Young (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal 20*(1), 86–117.

Berkes, I., R. Gabrys, L. Horváth, and P. Kokoszka (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(5), 927–946.

Boysen, L., A. Kempe, V. Liebscher, A. Munk, O. Wittich, et al. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *The Annals of Statistics 37*(1), 157–183.

Cai, T. T. and M. Yuan (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *The Annals of Statistics 39*(5), 2330–2355.

Cao, H., J. Li, and J. Fine (2016). On last observation carried forward and asynchronous longitudinal regression analysis. *Electronic Journal of Statistics 10*, 1155–1180.

Chen, Y. and B. Li (2011). Forecasting yield curves in an adaptive framework. *Central European Journal of Economic Modelling and Econometrics 3*(4), 237–259.

Chen, Y. and L. Niu (2014). Adaptive dynamic nelson–siegel term structure model with applications. *Journal of Econometrics 180*(1), 98–115.

Chiou, J.-M. (2012). Dynamical functional prediction and classification, with application to traffic flow prediction. *The Annals of Applied Statistics 6*(4), 1588–1614.

Chiou, J.-M. and P.-L. Li (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(4), 679–699.

Chiou, J.-M. and H.-G. Müller (2014). Linear manifold modelling of multivariate functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(3), 605–626.

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics 10*(2), 2000–2038.

Csörgö, M. and L. Horváth (1997). *Limit theorems in change-point analysis*, Volume 18. John Wiley & Sons Inc.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 1735–1765.

Diebold, F. X. and C. Li (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics 130*(2), 337–364.

Fan, J. and I. Gijbels (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Volume 66. CRC Press.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

Gasser, T. and H. Muller (1984). Estimating regression functions and their derivatives by the kernel method. *Scandinavian Journal of Statistics 11*, 171–185.

Gijbels, I., A. Lambert, and P. Qiu (2007). Jump-preserving regression and smoothing using local linear fitting: a compromise. *Annals of the Institute of Statistical Mathematics 59*(2), 235–272.

Górecki, T., L. Horváth, and P. Kokoszka (2018). Change point detection in heteroscedastic time series. *Econometrics and Statistics 7*, 63–88.

Grégoire, G. and Z. Hamrouni (2002). Change point estimation by local linear smoothing. *Journal of Multivariate Analysis 83*(1), 56–83.

Gromenko, O., P. Kokoszka, and M. Reimherr (2017). Detection of change in the spatiotemporal mean function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(1), 29–50.

Horváth, L. and P. Kokoszka (2012). *Inference for functional data with applications*, Volume 200. Springer Science & Business Media.

Hsing, T. and R. Eubank (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*, Volume 997. John Wiley & Sons.

Huh, J. and B. Park (2004). Detection of a change point with local polynomial fits for the random design case. *Australian & New Zealand Journal of Statistics 46*(3), 425–441.

Li, J. and B. Jin (2018). Multi-threshold accelerated failure time model. *The Annals of Statistics 46*(6A), 2657–2682.

Li, J., M. Xu, P.-S. Zhong, and L. Li (2019). Change point detection in the mean of high-dimensional time series data under dependence. *arXiv preprint arXiv:1903.07006*.

Li, J., W. Zhang, and E. Kong (2018). Factor models for asset returns based on transformed factors. *Journal of Econometrics 207*(2), 432–448.

Li, Y. and T. Hsing (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics 38*(6), 3321–3351.

Müller, H.-G. and K.-S. Song (1997). Two-stage change-point estimators in smooth regression models. *Statistics & Probability Letters 34*(4), 323–335.

Qiu, P. (2005). *Image processing and jump regression analysis*, Volume 599. John Wiley & Sons.

Qiu, P., C. Asano, and X. Li (1991). Estimation of jump regression functions. *Bulletin of Informatics and Cybernetics 24*(3-4), 197–212.

Ramsay, J. and B. Silverman (2005). *Functional Data Analysis*. Springer Science & Business Media.

Rice, J. (1984). Boundary modification for kernel regression. *Communications in Statistics, Theory and Methods 13*, 893–900.

Sheather, S. J. and M. C. Jones (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological) 53*(3), 683–690.

Spokoiny, V. G. et al. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *The Annals of Statistics 26*(4), 1356–1378.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2015). Review of functional data analysis. *Annu. Rev. Statist 1*, 41.

Wang, T. and R. J. Samworth (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 80*(1), 57–83.

Wu, J. and C. Chu (1993). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics 21*(3), 1545–1566.

Wu, W. B. and Z. Zhao (2007). Inference of trends in time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69*(3), 391–410.

Xia, Z. and P. Qiu (2015). Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika 102*(2), 397–408.

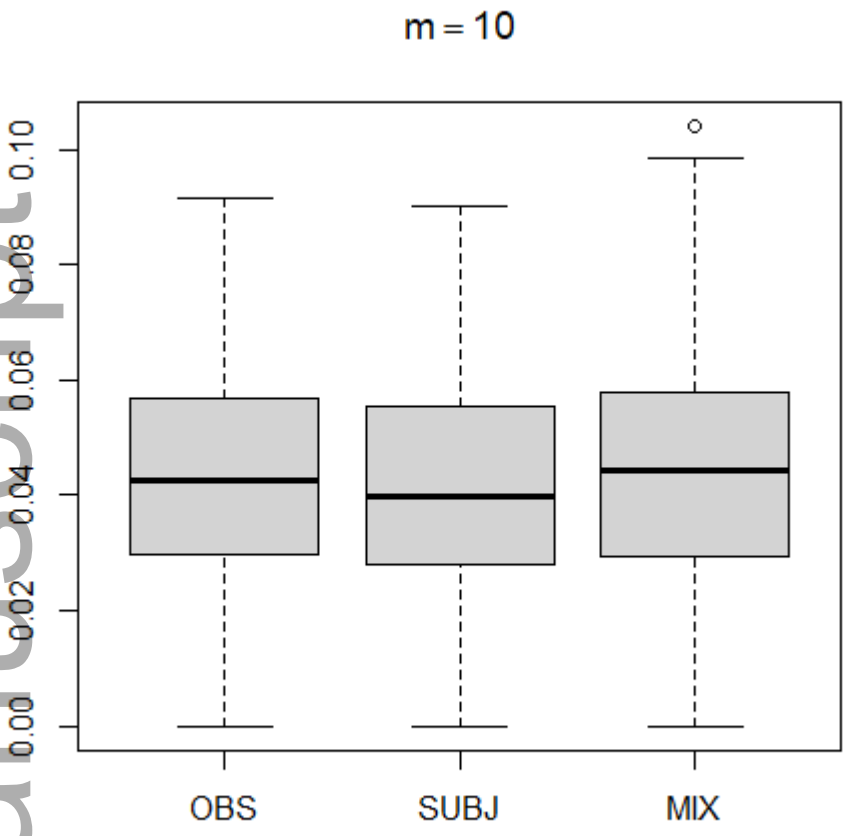
Xu, M., P.-S. Zhong, and W. Wang (2016). Detecting variance change-points for blocked time series and dependent panel data. *Journal of Business & Economic Statistics 34*(2), 213–226.

Yao, F., H.-G. Müller, and J.-L. Wang (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*(470), 577–590.

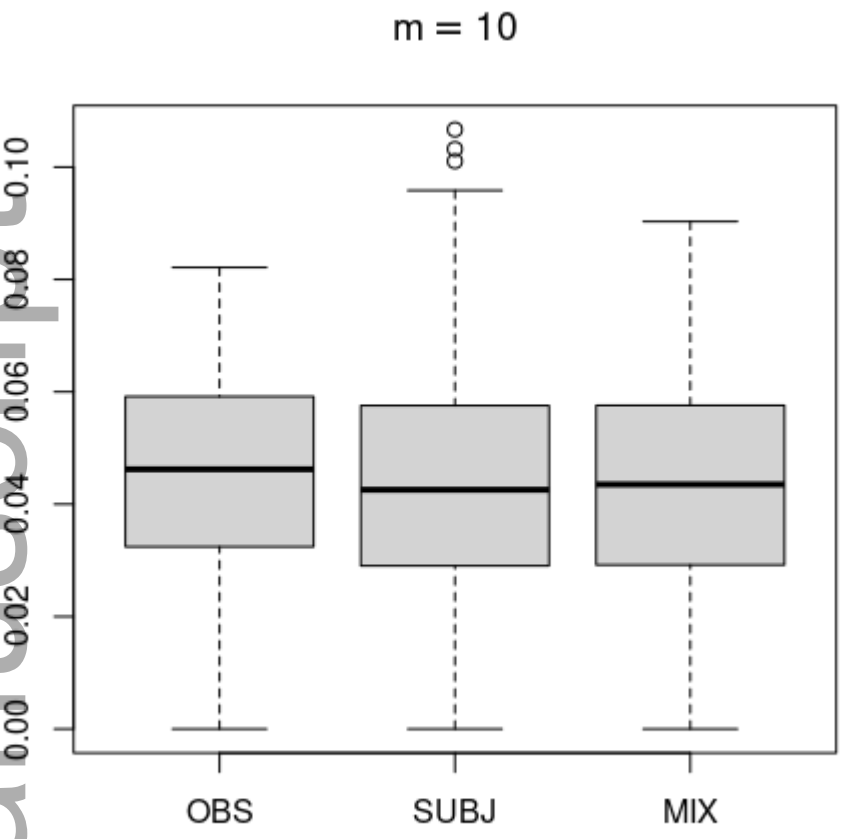
Zhang, T. (2016). Testing for jumps in the presence of smooth changes in trends of nonstationary time series. *Electronic Journal of Statistics 10*(1), 706–735.

Zhang, X. and J.-L. Wang (2016). From sparse to dense functional data and beyond. *The Annals of Statistics 44*(5), 2281–2321.

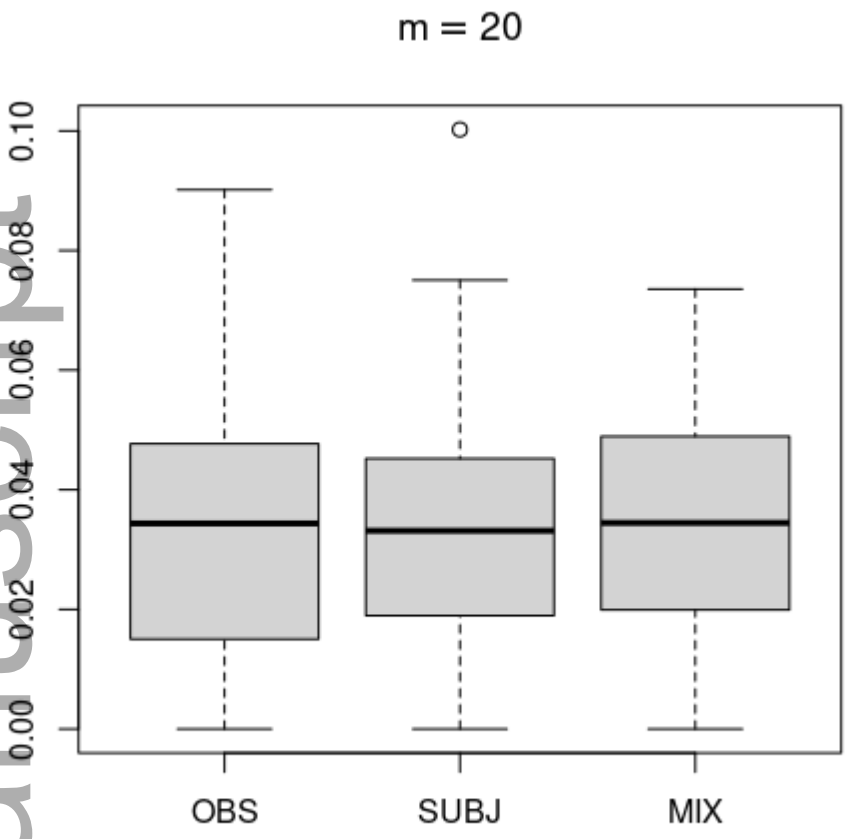
Zheng, S., L. Yang, and W. K. Härdle (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association 109*(506), 661–673.



rssb_12493_f2_err-n400-m10-setting1.png



rssb_12493_f2_err-n400-m10-setting2.png



rssb_12493_f2_err-n400-m20-setting1.png