Statistica Sinica Preprint No: SS-2024-0234	
Title	Estimation and Inference of Change Points in Functional Regression Time Series
Manuscript ID	SS-2024-0234
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202024.0234
Complete List of Authors	Shivam Kumar, Haotian Xu, Haeran Cho and Daren Wang
Corresponding Authors	Haotian Xu
E-mails	haotian.xu2@icloud.com
Notice: Accepted author version.	

Statistica Sinica

Estimation and Inference of Change Points in Functional Regression Time Series

Shivam Kumar, Haotian Xu, Haeran Cho and Daren Wang

University of Chicago, Auburn University, University of Bristol and University of California San Diego

Abstract: In this paper, we study the estimation and inference of change points under a functional linear regression model with changes in the slope function. We present a novel Functional Regression Binary Segmentation (FRBS) algorithm which is computationally efficient as well as achieving consistency in multiple change point detection. This algorithm utilizes the predictive power of piece-wise constant functional linear regression models in the reproducing kernel Hilbert space framework. We further propose a refinement step that improves the localization rate of the initial estimator output by FRBS, and derive asymptotic distributions of the refined estimators for two different regimes determined by the magnitude of a change. To facilitate the construction of confidence intervals for underlying change points based on the limiting distribution, we propose a consistent block-type long-run variance estimator. Our theoretical investigation accommodates temporal dependence and heavy-tails in both the functional covariates and the measurement errors. Empirical performance of our method is demonstrated through extensive simulation studies and applications to financial and economic datasets.

Key words and phrases: Change points, functional regression, time series, temporal dependence, heavy-tail.

1. Introduction

Functional Data Analysis (FDA) studies data that are represented as random functions. The infinite dimension of functional data poses a significant challenge to the development of statistical methods. We refer to Wang et al. (2016) for a comprehensive overview of the FDA. Extensive treatments of the subject can also be found in Ramsay and Silverman (2002), Kokoszka and Zhang (2012), Hsing and Eubank (2015), and Kokoszka and Reimherr (2017). Functional Principal Component Analysis (FPCA), a pivotal approach in FDA, focuses on characterizing the dominant modes of variation in random functions. Seminal contributions to the development and application of FPCA include, for example, Ramsay and Silverman (2005) and Yao et al. (2005). Another important approach in this area employs strategies based on Reproducing Kernel Hilbert Space (RKHS) for estimating the mean, covariance, and slope functions, as demonstrated in Cai and Yuan (2010). Unlike non-parametric methods such as FPCA, the RKHS-based approach selects the most representative functional features in an adaptive manner from an RKHS.

Functional time series analysis is an important area within FDA, focusing on functional data with temporal dependence. From the modeling perspective, Cai et al. (2000) focused on functional regression via local linear modeling; Kowal et al. (2017) investigated functional linear models; and Kowal et al. (2019) explored functional autoregression. To analysis functional time series, Panaretos and Tavakoli (2013) employed a Fourier analysis-based approach and Rubín and Panaretos (2020) considered the estimation of the dynamics of functional time series in a sparse sampling regime. We refer to Koner and Staicu (2023) for a comprehensive survey.

In this paper, we focus on a functional linear regression model with the slope function changing in a piece-wise constant manner. Given the data sequence $\{(y_j, X_j)\}_{j=1}^n$, we consider the model

$$y_j = \langle \beta_j^*, X_j \rangle_{\mathcal{L}^2} + \varepsilon_j, \qquad 1 \le j \le n, \tag{1.1}$$

where $\{y_j\}_{j=1}^n$ are the scalar responses, $\{X_j\}_{j=1}^n$ the functional covariates, $\{\varepsilon_j\}_{j=1}^n$ the centered noise sequence, and $\{\beta_j^*\}_{j=1}^n$ the true slope functions. Here, we denote $\langle \beta_j^*, X_j \rangle_{\mathcal{L}^2} = \int \beta_j^*(u) X_j(u) du$. We assume that there exists a collection of time points $\{\eta_k\}_{k=0}^{\mathcal{K}+1} \subset \{0, 1, \dots, n\}$ with $0 = \eta_0 < \eta_1 < \dots <$

$\eta_{\mathcal{K}} < \eta_{\mathcal{K}+1} = n$ such that

$$\beta^*_j \neq \beta^*_{j+1} \quad \text{if and only if} \quad j \in \{\eta_1, \ldots, \eta_{\mathcal{K}}\}. \tag{1.2}$$

We refer to the model specified in (1.1) and (1.2) as the functional linear regression model with change points. Our goals are twofold: to estimate the locations of the change points consistently, and to derive the limiting distributions of these estimators and consequently construct an asymptotically valid confidence interval around each change point.

The considered problems are part of the vast body of change point analysis. The primary interest of change point analysis is to detect the presence of change points and estimate their locations in various data types. Wang et al. (2020) and Sullivan and Woodall (2000) have addressed the detection of changes in the mean and covariance of a sequence of fixed-dimensional multivariate data, while Wang and Samworth (2018) and Kaul et al. (2023) have focused on high-dimensional situations. In functional settings, Dette and Kutta (2021) studied the detection of changes in the eigensystem, while Li et al. (2022), Harris et al. (2022), and Madrid Padilla et al. (2022) considered problems related to detecting changes in the mean and Jiao et al. (2023) considered changes in the covariance. Change point detection problems within this context have also been investigated in the Bayesian framework, e.g. Li and Ghosal (2021). Beyond estimation of change points, the limiting distributions of change point estimators have been studied in high-dimensional regression (Xu et al., 2024; Kaul and Michailidis, 2021), multivariate non-parametric (Madrid Padilla et al., 2023) as well as functional (Aue et al., 2009, 2018) settings.

Despite these contributions, the estimation and inference of change points in functional linear regression settings remain unaddressed, and this paper aims to fill this gap. To this end, we first propose a two-step procedure based on RKHS, to detect and locate the multiple change points. Then, we investigate limiting distributions of change point estimators and introduce a new method to construct a confidence interval for each change point. This requires the estimation of long-run variance in the presence of temporal dependence which is of independent interest on its own, as highlighted by studies such as Khismatullina and Vogt (2020) and Hörmann and Kokoszka (2010).

In studying the theoretical properties, we adopt a general framework that only requires the existence of sixth moments and a polynomial decay of α-mixing coefficients for both functional covariates and noise sequences, which greatly expands the applicability of our method. We also allow for local changes that tend to zero with the increasing sample size, and the number of change points may diverge.

1.1 List of contributions

We briefly summarize the main contributions made in this paper below.

- To the best of our knowledge, our work is the first attempt at estimating and inferring change points in functional linear regression settings. Our theory only requires weak moment assumptions and accommodates temporal dependence and a diverging number of change points. In addition to deriving the error bound on change point estimators, we establish the corresponding minimax lower bound, thereby demonstrating the optimality of the proposed change point estimator.

- To facilitate the practical feasibility of our inference procedure, we introduce a block-type long-run variance estimator and prove its consistency. This estimator is subsequently employed to construct an asymptotically valid confidence interval for each change point.

- We demonstrate the numerical performance of our proposed method through extensive numerical examples and applications to financial and economic datasets. Our approach numerically outperforms alternative change point estimation methods that rely on FPCA or high-dimensional regression methods.

1.2 Basics of RKHS

This section briefly reviews the basics of RKHS that are relevant to functional linear regression. We refer to Wainwright (2019) for a detailed introduction to RKHS.

For any compact set $\mathcal{T}$, denote the space of square-integrable functions defined on $\mathcal{T}$ as $\mathcal{L}^2(\mathcal{T}) = \{f : \mathcal{T} \to \mathbb{R} : \|f\|_{\mathcal{L}^2}^2 = \int_{\mathcal{T}} f^2(u)du < \infty\}$. For any $f, g \in \mathcal{L}^2(\mathcal{T})$, let $\langle f, g\rangle_{\mathcal{L}^2} = \int_{\mathcal{T}} f(u)g(u)\, du$. For a linear map F from $\mathcal{L}^2(\mathcal{T})$ to $\mathcal{L}^2(\mathcal{T})$, define $\|F\|_{\mathrm{op}} = \sup_{\|h\|_{\mathcal{L}^2}=1} \|F(h)\|_{\mathcal{L}^2}$. A kernel function $R : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ is a symmetric and nonnegative definite function. The integral operator L_R of R is a linear map from $\mathcal{L}^2(\mathcal{T})$ to $\mathcal{L}^2(\mathcal{T})$ is defined as $L_R(f)(\cdot) = \int_{\mathcal{T}} R(\cdot, u)f(u)\, du$. Suppose in addition that R is bounded. Then, Mercer's theorem (e.g. Theorem 12.20 of Wainwright (2019)) implies that there exists a set of orthonormal eigenfunctions $\{\psi_l^R\}_{l=1}^\infty \subset \mathcal{L}^2(\mathcal{T})$ and a sequence of nonnegative eigenvalues $\{\theta_l^R\}_{l=1}^\infty$ sorted non-increasingly, such that $R(u_1, u_2) = \sum_{l=1}^\infty \theta_l^R \psi_l^R(u_1)\psi_l^R(u_2)$. Thus, we have that $L_R(\psi_l^R) = \theta_l^R \psi_l^R$. Define the RKHS generated by R as

$$\mathcal{H}(R) = \left\{ f \in \mathcal{L}^2(\mathcal{T}) : \|f\|_{\mathcal{H}(R)}^2 = \sum_{l=1}^\infty \frac{\langle f, \psi_l^R\rangle_{\mathcal{L}^2}^2}{\theta_l^R} < \infty \right\}.$$

For any $f, g \in \mathcal{H}(R)$, denote

$$\langle f, g\rangle_{\mathcal{H}(R)} = \sum_{l=1}^\infty \frac{\langle f, \psi_l^R\rangle_{\mathcal{L}^2} \langle g, \psi_l^R\rangle_{\mathcal{L}^2}}{\theta_l^R}. \tag{1.3}$$

Define $R^{1/2}(u_1,u_2)=\sum_{l=1}^{\infty}\sqrt{\theta_l^R}\psi_l^R(u_1)\psi_l^R(u_2)$. Thus, $L_{R^{1/2}}(\psi_l^R)=\sqrt{\theta_l^R}\psi_l^R$. It follows that $L_{R^{1/2}}:\mathcal{L}^2(\mathcal{T})\to\mathcal{H}(R)$ is bijective and distance-preserving. In addition, if $\{\Phi_l\}_{l=1}^{\infty}$ is a $\mathcal{L}^2(\mathcal{T})$ basis, then $\{L_{R^{1/2}}(\Phi_l)\}_{l=1}^{\infty}$ is a basis of $\mathcal{H}(R)$. For any $f,g\in\mathcal{L}^2(\mathcal{T})$, denote

$$R[f,g]=\iint_{\mathcal{T}\times\mathcal{T}} f(u_1)R(u_1,u_2)g(u_2)\,du_1du_2.$$

Let R_1 and R_2 be any generic kernel functions. We denote the composition of R_1 and R_2 as $R_1R_2(u_1,u_2)=\int_{\mathcal{T}}R_1(u_1,v)R_2(v,u_2)dv$.

1.3 Notation and organization

For two positive real number sequences $\{a_j\}_{j=1}^{\infty}$ and $\{b_j\}_{j=1}^{\infty}$, we write $a_j\lesssim b_j$ or $a_j=O(b_j)$ if there exists an absolute positive constant C such that $a_j\le Cb_j$. We denote $a_j\asymp b_j$, if $a_j\lesssim b_j$ and $b_j\lesssim a_j$. We write $a_j=o(b_j)$ if $\lim_{j\to\infty}b_j^{-1}a_j\to 0$. For a sequence of $\mathbb{R}$-valued random variables $\{X_j\}_{j=1}^{\infty}$, we denote $X_j=O_{\mathbb{P}}(a_j)$ if $\lim_{M\to\infty}\limsup_j\mathbb{P}(|X_j|\ge Ma_j)=0$. We denote $X_j=o_{\mathbb{P}}(a_j)$ if $\limsup_j\mathbb{P}(|X_j|\ge Ma_j)=0$ for all $M>0$. The convergences in distribution and probability are respectively denoted by $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{P}$. With slight abuse of notations, for any positive integers s and e where $0\le s<e<n$, we use $(s,e]$ to denote the set $(s,e]\cap\{1,\ldots,n\}$.

The rest of the paper is organized as follows. Section 2 introduces our new methodology for estimating multiple change points within functional linear regression settings. Section 3 studies the theoretical properties of the proposed estimators, establishing their minimax optimality and limiting distributions. In Section 4, we discuss the construction of confidence intervals around each change point and provide an asymptotically valid procedure for the long-run variance estimation. Finally, Section 5 performers a real data analysis on the Standard and Poor's 500 index dataset. The implementation of the proposed methodology can be found at https://github.com/civamkr/FRBS. Extensive simulation studies, an additional real data analysis and all proofs are collected in the supplementary material.

2. Change point estimation

In this section, we introduce our method for change point estimation under the functional linear regression model defined in (1.1). To motivate our approach, we first consider a closely related two-sample testing problem in the functional linear regression setting. Given data $\{(y_j, X_j)\}_{j=1}^n$ generated from (1.1), consider

$$H_0 : \beta^*_{s+1} = \ldots = \beta^*_e \quad \text{vs.} \quad H_a : \beta^*_{s+1} = \ldots = \beta^*_t \neq \beta^*_{t+1} = \ldots = \beta^*_e,$$

where $0 < s < t < e \leq n$. In other words, we are interested in testing whether there is a change in the slope function at time t within the interval

$(s,e]$. The corresponding likelihood ratio statistic is

$$\widehat{W}_t^{s,e} = \log\left(\frac{\max_{\beta\in\mathcal{H}(K)} \mathfrak{L}\left(\{y_j, X_j\}_{j=s+1}^e, \beta\right)}{\max_{\beta_1\in\mathcal{H}(K)} \mathfrak{L}\left(\{y_j, X_j\}_{j=s+1}^t, \beta_1\right) \max_{\beta_2\in\mathcal{H}(K)} \mathfrak{L}\left(\{y_j, X_j\}_{j=t+1}^e, \beta_2\right)}\right) \tag{2.4}$$

where, assuming for the moment that $\{\epsilon_j\}_{j=1}^n$ are i.i.d. standard normal, we have the likelihood function

$$\mathfrak{L}(\{y_j, X_j\}_{j=s+1}^e, \beta) = \prod_{j=s+1}^{e} (2\pi)^{-1/2} \exp\left(-(y_j - \langle X_j, \beta\rangle_{\mathcal{L}^2})^2/2\right),$$

and $\mathcal{H}(K)$ denotes RKHS corresponding to kernel K defined in Assumption 2 below. Note that (2.4) can be further simplified to

$$\begin{aligned}\widehat{W}_t^{s,e} \;&=\; \sum_{j=s+1}^{e}\left(y_j - \langle X_j, \widehat{\beta}_{(s,e]}\rangle_{\mathcal{L}^2}\right)^2 - \sum_{j=s+1}^{t}\left(y_j - \langle X_j, \widehat{\beta}_{(s,t]}\rangle_{\mathcal{L}^2}\right)^2 \\ &\quad - \sum_{j=t+1}^{e}\left(y_j - \langle X_j, \widehat{\beta}_{(t,e]}\rangle_{\mathcal{L}^2}\right)^2, \end{aligned} \tag{2.5}$$

where $\widehat{\beta}_{(s,e]}$ is the maximum likelihood estimator of the slope function based on $\{(y_j, X_j)\}_{j=s+1}^e$. Inspired by Cai and Yuan (2012), we consider the following penalized estimator

$$\widehat{\beta}_{(s,e]} = \underset{\beta\in\mathcal{H}(K)}{\arg\min}\left\{\frac{1}{(e-s)}\sum_{j\in(s,e]}(y_j - \langle X_j, \beta\rangle_{\mathcal{L}^2})^2 + \lambda_{e-s}\|\beta\|^2_{\mathcal{H}(K)}\right\}, \tag{2.6}$$

where λ_{e-s} is a tuning parameter to ensure the smoothness of the estimator. While (2.6) is an optimization problem in an infinite-dimensional space, the solution can be found in a finite-dimensional subspace via the representer theorem in RKHS (Yuan and Cai, 2010), which is shown to be statistically and computationally efficient.

To illustrate the effectiveness of the likelihood ratio statistics $\widehat{W}_t^{s,e}$ in revealing the location of a change point, we visualize in Figure 1 the statistic $\widehat{W}_t^{s,e}$ when the interval $(s,e]$ contains a single change point at η. Here, we also display the population counterpart

$$W_t^{s,e} = \frac{(t-s)(e-t)}{(e-s)} \Sigma[\beta^*_{(s,t]} - \beta^*_{(t,e]}, \beta^*_{(s,t]} - \beta^*_{(t,e]}], \tag{2.7}$$

where $\beta^*_{(s,e]} = (e-s)^{-1} \sum_{j=s+1}^{e} \beta^*_j$, and Σ is the covariance operator of $\{X_j\}_{j=1}^{\infty}$, the centered and stationary covariate sequence, i.e. $\Sigma(u_1, u_2) = \mathbb{E}(X_1(u_1)X_1(u_2))$. We observe that $\widehat{W}_t^{s,e}$ closely approximates $W_t^{s,e}$, which is a 'tent-shape' function in t and is maximized at the single change point η, and thus $\widehat{W}_t^{s,e}$ attains its maximum close to η (in fact, exactly at η in this example).

Thus motivated, we propose a two-step method for change point estimation in functional regression time series. In Step 1, we adopt a computationally efficient algorithm that scans $\widehat{W}_t^{s,e}$ at strategically selected intervals to generate preliminary estimators. Then, in Step 2, we utilize the preliminary estimators to develop the final estimators with enhanced accuracy.

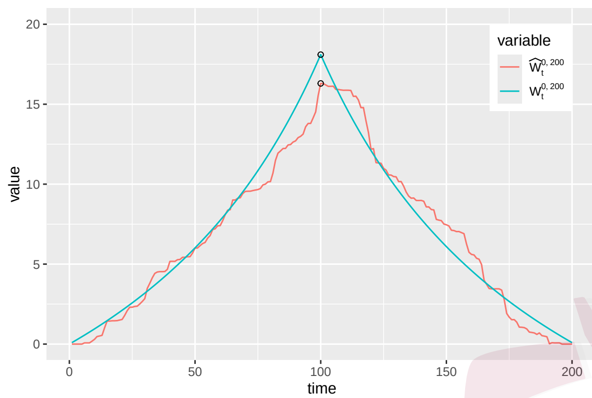

Figure 1: Plot of $\widehat{W}_t^{s,e}$ and its population version $W_t^{s,e}$ with $s = 0$ and $e = 200$ over $t = 1, \ldots, 199$. The data are simulated under Scenario I detailed in Section A.1 in the supplementary material with $n = 200$ and a change point occurs at $\eta = 100$. The estimator $\widehat{\beta}_{(s,e]}$ is obtained with $\lambda_{e-s} = 0.2$. Both $\widehat{W}_t^{s,e}$ and $W_t^{s,e}$ achieve their maximum at $t = 100$.

Step 1: preliminary estimator

In Step 1, our goal is to achieve consistency in multiple change point detection with computational efficiency. To this end, we employ the seeded binary segmentation algorithm proposed by Kovács et al. (2023), which leverages a deterministic, multi-resolution scaffold of "seeded intervals":

Definition 1 (Seeded intervals)**.** Let n be the length of a given time series and Ω a given integer satisfying $0 < \Omega < n$. Letting $M = \lceil \log_2 (n/\Omega) \rceil + 1$ be the total number of layers, denote $\mathfrak{l}_k = n/2^{k-1}$ and $\mathfrak{b}_k = \mathfrak{l}_k/2 = n/2^k$, for the layer index $k = 1, \ldots, M$. Then, the collection of seeded intervals is

$$\mathcal{J}^{(\Omega)} = \bigcup_{k=1}^{M} \mathcal{J}_k \quad \text{where} \quad \mathcal{J}_k = \bigcup_{i=1}^{2^k-1} \left\{ \Big[\lceil (i-1)\mathfrak{b}_k \rceil, \lfloor (i-1)\mathfrak{b}_k + \mathfrak{l}_k \rfloor \Big] \right\},$$

where $\mathcal{J}_k$ is the seeded intervals in the k-th layer.

In this construction, each $\mathcal{J}_k$ consists of short, equally spaced intervals whose centers coincide with the endpoints of the previous level. This carefully designed structure guarantees that every true change point is covered by exactly two of the intervals at the finest level, while keeping the total number of intervals modest, since

$$|\mathcal{J}^{(\Omega)}| = \sum_{k=1}^{\lceil \log_2(\frac{n}{\Omega}) \rceil + 1} (2^k - 1) = 2^{\lceil \log_2(\frac{n}{\Omega}) \rceil + 2} - 3 - \left\lceil \log_2 \left(\frac{n}{\Omega} \right) \right\rceil \leq 8 \left(\frac{n}{\Omega} \right). \tag{2.8}$$

Evaluating the likelihood-ratio statistic $\widehat{W}_t^{s,e}$ defined in (2.5) at these strategically selected intervals, yields a total computational cost that scales essentially linearly with the sample size n, as discussed later in Remark 1. Algorithm 1 outlines the procedure of computing the preliminary change point estimators, which is called with $(s, e] = (0, n]$. This algorithm recursively detects change points based on scanning the likelihood ratio statistics. Specifically, using the set of seeded intervals, the algorithm iteratively identifies the shortest interval associated with a strong signal for a change (in the sense that $\widehat{W}_t^{s_m,e_m}, (s_m, e_m] \in \mathcal{J}^{(\Omega)}$, exceeds a threshold τ). After computing the statistics we adopt the rule of Baranowski et al. (2019): among all seeds whose maximum statistic exceeds a universal threshold τ, we retain the shortest one. This ensures, with high probability, that the selected interval contains exactly one change point and its width is at least of the order $O(\Omega)$. The maximizer of the likelihood statistic within this interval is then recorded as a preliminary estimator for the change point. Upon detection of each change point, it stores the estimator and proceeds to search for further change points separately within the sections of the data determined by two consecutive estimators previously detected. In the absence of a change point within a data section $(s, e]$, we expect all $\widehat{W}_t^{s,e}$, $s < t < e$, to fall below the given threshold τ, in which case the algorithm excludes the interval $(s, e]$ from further consideration. In addition to the threshold τ, Algorithm 1 requires the choice of the regularization parameter λ_{e-s} for the local estimation of the slope function, which takes the form $\lambda_{e-s} = \omega(e-s)^{-2r/(2r+1)}$ with some $\omega > 0$ and r that controls the regularity of the regression coefficient (see Assumption 2). The choice of these tuning parameters are discussed in Section A in the supplementary material.

Algorithm 1: Functional Regression Binary Segmentation.

FRBS $\big((s,e], \mathcal{J}^{(\Omega)}, \omega, \tau\big)$

INPUT: Data $\{(y_j, X_j)\}_{j=1}^n$, seeded intervals $\mathcal{J}^{(\Omega)}$, tuning parameters $\omega, \tau > 0$.

Initialize: If $(s,e] = (0,n]$, the estimated change point set $\widehat{\mathcal{B}} \leftarrow \emptyset$. Compute and store $\left\{\widehat{W}_t^{s,e}\right\}_{\substack{t\in(s,e]\\(s,e]\in\mathcal{J}^{(\Omega)}}}$ (see (2.5)) with $\lambda_{e-s} = \omega\,(e-s)^{-2r/(2r+1)}$.

for $(s_m, e_m] \in \mathcal{J}^{(\Omega)}$ **do**

  if $(s_m, e_m] \subset (s,e]$ **then**

    $b_m \leftarrow \arg\max_{s_m<t<e_m} \widehat{W}_t^{s_m,e_m}$

    $a_m \leftarrow \widehat{W}_{b_m}^{s_m,e_m}$

  else $a_m \leftarrow 0$

$\mathcal{M}^{s,e} := \{m : a_m > \tau\}$

if $\mathcal{M}^{s,e} \neq \emptyset$ **then**

  $m^* \leftarrow \arg\min_{m\in\mathcal{M}^{s,e}} |e_m - s_m|$

  $\widehat{\mathcal{B}} \leftarrow \widehat{\mathcal{B}} \cup \{b_{m^*}\}$

  FRBS$\big((s, b_{m^*}], \mathcal{J}^{(\Omega)}, \omega, \tau\big)$

  FRBS$\big((b_{m^*}, e], \mathcal{J}^{(\Omega)}, \omega, \tau\big)$

OUTPUT: $\widehat{\mathcal{B}}$.

Step 2: refined estimator

Let $\widehat{\mathcal{B}} = \{\widehat{\eta}_k, 1 \le k \le \widehat{\mathcal{K}} : \widehat{\eta}_1 < \ldots < \widehat{\eta}_{\widehat{\mathcal{K}}}\}$ denote the set of preliminary change point estimators returned by Algorithm 1. In this step, we produce the refined estimators $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ with enhanced accuracy. As we shall establish in Theorem 1, each estimator obtained in Step 1 is consistent, but further refinement is needed to achieve optimal localization, which further enables the derivation of the asymptotic distribution of the resultant estimator. To this end, we construct a smaller window that, with high probability, contains exactly one change point and moreover, the change point is contained well-within the interval which ensures sufficient balance of data on both sides to maintain statistical power. Within this window, we minimize a cost function to identify the best split of the interval. Specifically, for each $k = 1, \ldots, \widehat{\mathcal{K}}$, the final estimator is obtained as

$$\widetilde{\eta}_k = \underset{s_k<t<e_k}{\arg\min}\, \mathcal{Q}_k(t), \quad \text{where} \tag{2.9}$$

$$\mathcal{Q}_k(t) = \sum_{j=s_k+1}^{t} \Big(Y_j - \langle X_j, \widehat{\beta}_{(s_k,\widehat{\eta}_k]}\rangle_{\mathcal{L}^2}\Big)^2 + \sum_{j=t+1}^{e_k} \Big(Y_j - \langle X_j, \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\rangle_{\mathcal{L}^2}\Big)^2,$$

$$s_k = 9\widehat{\eta}_{k-1}/10 + \widehat{\eta}_k/10 \quad \text{and} \quad e_k = \widehat{\eta}_k/10 + 9\widehat{\eta}_{k+1}/10. \tag{2.10}$$

The specific constants (9/10 and 1/10) in (2.10) are arbitrary; any convex combination would work equally well in theory. We select these values for convenience and their good practical performance as observed in Xu et al. (2024). As shown later in Section 3.2, the refined estimator $\widetilde{\eta}_k$ attains the rate of localization matching the minimax lower bound, and thus is minimax optimal.

3. Theoretical properties

In this section, we establish theoretical properties of the change point estimators proposed in Section 2. We first introduce the required assumptions for the model (1.1)–(1.2), which permit temporal dependence and heavy-tailedness in the data.

To quantify the degree of temporal dependence, we adopt the α-mixing coefficient which is a standard tool in the time series literature. Recall that a stochastic process $\{Z_t\}_{t\in\mathbb{Z}}$ is said to be α-mixing (strong mixing) if

$$\alpha(k) = \sup_{t\in\mathbb{Z}} \alpha\left(\sigma(Z_s, s \le t), \sigma(Z_s, s \ge t+k)\right) \to 0, \quad \text{as} \ \ k \to \infty,$$

where we write $\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A}, B\in\mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|$ for any two σ-fields $\mathcal{A}$ and $\mathcal{B}$. Assumption 1 concerns the distributions of the functional covariates and the noise sequence.

Assumption 1. (i) The functional covariate sequence $\{X_j\}_{j=1}^n \subset \mathcal{L}^2(\mathcal{T})$ satisfies $\mathbb{E}[X_j] = 0$, $\mathbb{E}[\|X_j\|_{\mathcal{L}^2}^2] < \infty$, and for any $f \in \mathcal{L}^2(\mathcal{T})$, there exists some constant $c > 0$ such that $(\mathbb{E}[\langle X_j, f\rangle_{\mathcal{L}^2}^6])^{1/6} \le c(\mathbb{E}[\langle X_j, f\rangle_{\mathcal{L}^2}^2])^{1/2}$.

(ii) The noise sequence $\{\varepsilon_j\}_{j=1}^n \subset \mathbb{R}$ satisfies $\mathbb{E}[\varepsilon_j|X_j] = 0$ and $\mathbb{E}[\varepsilon_j^6|X_j] < \infty$.

(iii) The sequence $\{(X_j, \varepsilon_j)\}_{j=1}^n$ is stationary and α-mixing with the mixing coefficients satisfying $\sum_{k=1}^{\infty} k^{1/3}\alpha^{1/3}(k) < \infty$.

Under Assumption 1, the functional covariate and the noise sequences are allowed to possess heavy tails. In particular, Assumption 1 (i) assumes that the 6-th moment of the random variable $\langle X_j, f\rangle_{\mathcal{L}^2}$ is bounded by its second moment, which holds if, for example, each X_j is a Gaussian process. Similar assumptions on the moments of the functional covariate are made in Cai and Yuan (2012) for the investigation into the penalized slope estimator in (2.6) in the stationary setting. The α-mixing condition essentially requires that $\alpha(k) = o(1/k^4)$, allowing the mixing coefficient to decay at a polynomial rate.

Assumption 2. (i) The slope function satisfies $\beta_j^* \in \mathcal{H}(K)$, for all $j = 1, \ldots, n$, where $\mathcal{H}(K)$ is the RKHS generated by the kernel function K.
(ii) The covariance operator Σ of $\{X_j\}_{j=1}^{\infty}$ and the kernel function K satisfy

$$K^{1/2}\Sigma K^{1/2}(t_1, t_2) = \sum_{l\geq 1} \mathfrak{s}_l \phi_l(t_1)\phi_l(t_2),$$

where $\{\phi_l\}_{l=1}^{\infty}$ are the eigenfunctions and $\{\mathfrak{s}_l\}_{l=1}^{\infty}$ the corresponding eigenvalues satisfying $\mathfrak{s}_l \asymp l^{-2r}$ for some constant $r > 1$.

Assumption 2 (i) requires that the slope functions are in the RKHS generated by the kernel function K, which regularizes the smoothness of the slope function. Assumption 2 (ii) requires that the function $K^{1/2}\Sigma K^{1/2}$ admits an eigen-decomposition with polynomially decaying eigenvalues, which controls the regularity of regression prediction. Both Assumption 2 (i) and

(ii) are also found in Cai and Yuan (2012).

Under the model (1.1), we define the change size of two consecutive slope functions as $\kappa_k^2 = \Sigma[\beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}]$. The form of κ_k^2 is closely related to the population counterpart (defined in (2.6)) of the likelihood ratio statistics $\widehat{W}_t^{s,e}$ in (2.5). In fact, if the time interval $(s, e]$ contains only one change point η_k, the statistic $\widehat{W}_{\eta_k}^{s,e}$ converges asymptotically to $W_{\eta_k}^{s,e}$, which in turn satisfies

$$\begin{aligned} W_{\eta_k}^{s,e} &= \frac{(\eta_k - s)(e-\eta_k)}{(e-s)} \Sigma[\beta^*_{(s,\eta_k]} - \beta^*_{(\eta_k,t]}, \beta^*_{(s,\eta_k]} - \beta^*_{(\eta_k,t]}] \\ &= \frac{(\eta_k - s)(e-\eta_k)}{(e-s)} \kappa_k^2. \end{aligned} \tag{3.11}$$

The detectability of each change point η_k depends on both the change size κ_k and how far it is from the adjacent change points. Let us define the minimal change size and the minimal spacing of change points as

$$\kappa = \min_{1\le k\le \mathcal{K}} \kappa_k \quad \text{and} \quad \Delta = \min_{1\le k\le \mathcal{K}+1} (\eta_k - \eta_{k-1}),$$

respectively. Assumption 3 specifies the signal-to-noise condition for the consistency of our method in terms of κ and Δ.

Assumption 3. Suppose that

$$\min\left\{ \frac{\kappa^2\Delta^2}{n \cdot n^{1/(2r+1)} \log^{1+2\xi}(n)}, \frac{\kappa^2 \Delta^{r/(2r+1)}}{\log^{1+2\xi}(\Delta)} \right\} \to \infty,$$

where $\xi > 0$ is some constant and r is defined in Assumption 2.

To establish the consistency of the preliminary estimators in Theorem 1, it is sufficient to have $\kappa^2\Delta^2/(n \cdot n^{1/(2r+1)} \log^{1+2\xi}(n)) \to \infty$. The additional requirement that $\kappa^2\Delta^{r/(2r+1)}/\log^{1+2\xi}(\Delta) \to \infty$, is made to derive the limiting distribution of the refined estimator in Theorem 2. When Δ is of the same order as n, since $r > 1$, the second condition in Assumption 3 is crucial and it is simplified to $\kappa^2\Delta^{r/(2r+1)} \to \infty$. Similar assumptions have been employed in Madrid Padilla et al. (2021, 2024) for nonparametric change point analysis, where the smoothness of the density function plays a similar role as r.

3.1 Consistency of the preliminary estimator

We first present the main theorem establishing the consistency of Algorithm 1 and the associated rate of localization.

Theorem 1. *Suppose that Assumptions 1, 2, and 3 hold. Let* $c_{\tau,1} > 32$ *and* $c_{\tau,2} \in (0, 1/20)$ *denote absolute constants. Suppose that* τ *satisfies*

$$c_{\tau,1}\left(\frac{n}{\Delta}\right) n^{1/(2r+1)} \log^{1+2\xi}(n) < \tau < c_{\tau,2}\kappa^2\Delta, \tag{3.12}$$

where r *and* ξ *are defined in Assumptions 2 and 3, respectively, and that* $\omega > 1/2$ *is any finite constant. Also, let* $\mathcal{J}^{(\Delta)}$ *be seeded intervals constructed according to Definition 1 with* Δ *defined in Assumption 3. Then,* *FRBS*$((0,n],\mathcal{J}^{(\Delta)},\omega,\tau)$ *outputs* $\widehat{\mathcal{B}}=\{\widehat{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ *which satisfies*

$$\mathbb{P}\left(\widehat{\mathcal{K}}=\mathcal{K};\ \max_{1\le k\le \mathcal{K}}\kappa_k^2\,|\widehat{\eta}_k-\eta_k|\le C_1\left(\frac{n}{\Delta}\right)\Delta^{1/(2r+1)}\log^{1+2\xi}(n)\right)\to 1$$

as $n\to\infty$*, where* $2<C_1<c_{\tau,1}/16$.

Theorem 1 shows that uniformly for all $k=1,\dots\mathcal{K}$,

$$|\widehat{\eta}_k-\eta_k|=O_{\mathbb{P}}\left(\kappa_k^{-2}\left(\frac{n}{\Delta}\right)\Delta^{1/(2r+1)}\log^{1+2\xi}(n)\right)=o_{\mathbb{P}}(\Delta),$$

where the last equality follows from Assumption 3.

Remark 1 (Computational complexity)**.** Let n be the number of functional observations, p the number of evaluation points on each curve after discretization (i.e. the dimension of the vector representing X_j), Ω the length of the finest intervals imposed when constructing seeded intervals, Δ the minimal spacing between consecutive change points, and $\mathcal{K}$ the true number of change points. Suppose $\Omega=\Delta$. The statistic $\widehat{W}_t^{s_m,e_m}$ requires computational cost $O\big((e_m-s_m)p^2\big)$. To optimize computation, we precompute and store all partial sums (of vectorized X_j) of the form $\sum_{t\in(s_m,e_m]}X_tX_t^\top\in\mathbb{R}^{p\times p}$ and $\sum_{t\in(s_m,e_m]}y_tX_t\in\mathbb{R}^p$, with total preprocessing cost $O(np^2)$. According to (2.8), the overall cost to compute and store the sorted statistics $\widehat{W}_t^{s_m,e_m}$ for all $(s_m,e_m)\in\mathcal{J}^{(\Delta)}$ is $O\left(\frac{n^2}{\Delta}p^2\log(n/\Delta)\right)$, where Δ denotes the length of the finest-layer intervals. Since there are at most $\mathcal{K}$ change points, this adds $O(\mathcal{K}n/\Delta)$ additional work from computations in Algorithm 1 other than the initialization step. The second (refinement) step further improves each preliminary estimator within two windows of width at most Δ; as this stage only requires recomputing slope functions, its cost is $O(\mathcal{K}\Delta p^2)$ for $\mathcal{K}$ change points. In summary, the overall computational complexity of our method is

$$O\left(\mathcal{K}\frac{n}{\Delta} + \mathcal{K}\Delta p^2 + \frac{n}{\Delta} np^2 \log(n/\Delta)\right),$$

which reduces to $O\left(np^2\right)$ when $\Delta = O(n)$.

3.2 Consistency and the limiting distributions of the refined estimators

In this subsection, we analyze the consistency and the limiting distributions of the refined change point estimators. We demonstrate that the limiting distribution of the refined change point estimator $\widetilde{\eta}_k$ is divided into two regimes determined by the change size κ_k, namely the non-vanishing regime where $\kappa_k \to \varrho_k$ for some positive constant $\varrho_k > 0$, and the vanishing regime where $\kappa_k \to 0$.

Theorem 2. *Suppose that Assumptions 1, 2, and 3 hold. Let* $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ *denote the refined change point estimators obtained as in* (2.9) *and assume that* $\widehat{\mathcal{K}} = \mathcal{K}$*.*

A *(Non-vanishing regime) For any given* $k \in \{1, \ldots, \widehat{\mathcal{K}}\}$*, suppose* $\kappa_k \to \varrho_k$ *as* $n \to \infty$*, with* $\varrho_k > 0$ *being an absolute constant. Then* $|\widetilde{\eta}_k - \eta_k| = O_{\mathbb{P}}(1)$*. In addition, as* $n \to \infty$*,* $\widetilde{\eta}_k - \eta_k \xrightarrow{\mathcal{D}} \arg\min_{\gamma \in \mathbb{Z}} S_k(\gamma)$ *where, for* $\gamma \in \mathbb{Z}$*,* $S_k(\gamma)$ *is a two-sided random walk defined as*

$$S_k(\gamma) = \begin{cases} \sum_{j=\gamma}^{-1} \left\{-2\varrho_k \left\langle X_j, \Psi_k \right\rangle_{\mathcal{L}^2} \varepsilon_j + \varrho_k^2 \left\langle X_j, \Psi_k \right\rangle_{\mathcal{L}^2}^2\right\} & \text{for} \quad \gamma < 0, \\ 0 & \text{for} \quad \gamma = 0, \\ \sum_{j=1}^{\gamma} \left\{2\varrho_k \left\langle X_j, \Psi_k \right\rangle_{\mathcal{L}^2} \varepsilon_j + \varrho_k^2 \left\langle X_j, \Psi_k \right\rangle_{\mathcal{L}^2}^2\right\} & \text{for} \quad \gamma > 0, \end{cases}$$

$$\text{with} \quad \Psi_k = \lim_{n \to \infty} \frac{\beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}}{\sqrt{\Sigma\left[\beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}, \beta^*_{\eta_{k+1}} - \beta^*_{\eta_k}\right]}}.$$

B *(Vanishing regime) For any given* $k \in \{1, \ldots, \widehat{\mathcal{K}}\}$*, suppose* $\kappa_k \to 0$ *as* $n \to \infty$*. Then* $|\widetilde{\eta}_k - \eta_k| = O_{\mathbb{P}}(\kappa_k^{-2})$*. In addition, as* $n \to \infty$*,*

$$\kappa_k^2(\widetilde{\eta}_k - \eta_k) \xrightarrow{\mathcal{D}} \arg\min_{\gamma \in \mathbb{R}} \{|\gamma| + \sigma_\infty(k)\mathbb{W}(\gamma)\},$$

$$\text{where} \quad \sigma_\infty^2(k) = 4 \lim_{n \to \infty} \operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\langle X_j, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}} \rangle_{\mathcal{L}^2} \varepsilon_j}{\kappa_k}\right), \tag{3.13}$$

and $\mathbb{W}(\gamma)$ *is a two-sided standard Brownian motion defined as*

$$\mathbb{W}(\gamma) = \begin{cases} \mathbb{B}_1(-\gamma) & \text{for} \quad \gamma < 0, \\ 0 & \text{for} \quad \gamma = 0, \\ \mathbb{B}_2(\gamma) & \text{for} \quad \gamma > 0, \end{cases}$$

with $\mathbb{B}_1(r)$ and $\mathbb{B}_2(r)$ denoting two independent standard Brownian motions.

Theorem 2 establishes the localization error bound with rate κ_k^{-2} for the refined change point estimator as well as the corresponding limiting distributions. This rate significantly improves upon that attained by the preliminary estimator derived in Theorem 1. Note that Theorem 2 assumes $\widehat{\mathcal{K}} = \mathcal{K}$, which holds asymptotically with probability tending to one by Theorem 1. We make a similar condition in Theorems 3–4 below. In the following Lemma 1, we further provide a matching lower bound:

Lemma 1. *Let $\{(y_j, X_j)\}_{j=1}^n$ be a functional regression time series following the models in (1.1)–(1.2) with $\mathcal{K} = 1$, and suppose that Assumptions 1 and 2 hold. Let $\mathbb{P}^n_{\kappa,\Delta}$ be the corresponding joint distribution. For any diverging sequence $\rho_n \to \infty$, consider the class of distributions*

$$\mathfrak{P} = \left\{ \mathbb{P}^n_{\kappa,\Delta} \ : \ \min\left\{ \frac{\kappa^2 \Delta^2}{n \cdot n^{1/(2r+1)} \log^{1+2\xi}(n)}, \frac{\kappa^2 \Delta^{r/(2r+1)}}{\log^{1+2\xi}(\Delta)} \right\} > \rho_n \right\}.$$

Then for sufficiently large n, it holds that

$$\inf_{\widehat{\eta}} \sup_{\mathbb{P} \in \mathfrak{P}} \mathbb{E}\left[|\widehat{\eta} - \eta(\mathbb{P})| \right] \geq \frac{1}{\kappa^2 e^2}.$$

The class of distributions $\mathfrak{P}$ encompasses all possible scenarios where Assumption 3 is satisfied. Lemma 1 complements the upper bound established in Theorem 2 in both the vanishing and the non-vanishing regimes.

The matching bounds in Theorem 2 and Lemma 1 indicate that our refined estimator is minimax optimal.

4. Confidence interval for the change points

In this section, we provide a practical way to construct confidence intervals for the true change points under the vanishing regime based on the limiting distribution derived in Theorem 2B. Since the limiting distribution in the vanishing regime contains an unknown long-run variance, we first discuss its consistent estimation.

4.1 Long run variance estimation

To utilize the limiting distribution in the vanishing regime derived in Theorem 2B, we first need to consistently estimate the long-run variance $\sigma^2_\infty(k)$ defined in (3.13). The long-run variance depends on the size of change κ_k at the change point η_k. To this end, we propose the plug-in estimator

$$\widehat{\kappa}_k = \sqrt{\widehat{\Sigma}_{(s_k,e_k]}\left[\widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]}\right]}, \tag{4.14}$$

where s_k and e_k are defined in (2.10), $\widehat{\beta}_{(s_k,\widehat{\eta}_k]}$ and $\widehat{\beta}_{(\widehat{\eta}_k,e_k]}$ are obtained as in (2.6), and

$$\widehat{\Sigma}_{(s_k,e_k)}(u_1,u_2) = \frac{1}{e_k - s_k}\sum_{j=s_k+1}^{e_k} X_j(u_1)X_j(u_2)$$

is the sample covariance operator for the functional data $\{X_j\}_{j=s_k+1}^{e_k}$. We show the consistency of $\widehat{\kappa}_k$ in Lemma C.1 in supplementary material. To estimate $\widetilde{\eta}_k$, we computed $\widehat{\beta}_{(s_k,\widehat{\eta}_k]}$ and $\widehat{\beta}_{(\widehat{\eta}_k,e_k]}$, where s_k and e_k are defined in (2.10). These quantities are also reused for the estimation of κ_k and subsequently in Algorithm 2. This step requires at most $O\left(\mathcal{K}np^2\right)$ operations, which is significantly less than the $O\left(\frac{n^2}{\Delta}p^2\right)$ complexity of the preliminary stage (see Remark 1).

For the estimation of $\sigma_\infty^2(k)$, we make use of a block-type strategy which has previously been adopted by Casini and Perron (2021) for the estimation of the long-run variance in a fixed-dimensional time series setting. In Algorithm 2, we outline our proposal for the estimation of $\sigma_\infty^2(k)$. The proposed method first partitions the data into mutually disjoint blocks of size $2q$ for some positive integer q, and filters out the intervals that contain change point estimators and ones that are adjacent to them. This filtering ensures that with high probability, the remaining intervals do not contain any change point. Let us denote the set of remaining intervals by $\mathcal{P}$. For each given $\mathcal{I} = (m, m+2q] \in \mathcal{P}$, we first compute the statistic

$$Z_j = \widehat{\kappa}_k^{-1} \cdot \left\langle X_j, \widehat{\beta}_{(s_k,\widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k,e_k]} \right\rangle_{\mathcal{L}^2} \cdot \left(y_j - \left\langle X_j, \widehat{\beta}_{(m,m+2q]} \right\rangle_{\mathcal{L}^2} \right)$$

at each $j \in \mathcal{I}$, which approximates the sequence $\kappa_k^{-1}\langle X_j, \beta^*_{\eta_k} - \beta^*_{\eta_{k+1}}\rangle_{\mathcal{L}^2}\varepsilon_j$. Then, we compute the scaled sample average of the centered sequence $Z_j -$

Z_{j+q}, $j = m+1, \ldots, m+q$, and denote it by $F_{\mathcal{I}}$. The estimator $\widehat{\sigma}^2_\infty(k)$ is obtained as the average of the square of $F_{\mathcal{I}}$ over $\mathcal{I} \in \mathcal{P}$.

Algorithm 2: Long-run Variance Estimation (LRV)

INPUT: Estimators $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{K}}$ and $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$, tuning parameter $q \in \mathbb{Z}^+$.

$\mathcal{N} \leftarrow \{1, \ldots, \lfloor n/2q \rfloor\} \setminus \bigcup_{k=1}^{\widehat{\mathcal{K}}} \{\lfloor \widehat{\eta}_k/2q \rfloor - 1, \lfloor \widehat{\eta}_k/2q \rfloor, \lfloor \widehat{\eta}_k/2q \rfloor + 1\}$

$\mathcal{P} \leftarrow \bigcup_{i \in \mathcal{N}} \{(2q(i-1), 2q\, i]\}$ $\triangleright$ $\mathcal{P}$ `is a collection of disjoint intervals in` $(0, n]$

for $k = 1, \ldots, \widehat{\mathcal{K}}$ **do**

  Compute s_k and e_k as in (2.10);

  for $\mathcal{I} = (m, m+2q] \in \mathcal{P}$ **do**

    for $j \in (m, m+2q]$ **do**

      $Z_j \leftarrow \widehat{\kappa}_k^{-1} \left\langle X_j, \widehat{\beta}_{(s_k, \widehat{\eta}_k]} - \widehat{\beta}_{(\widehat{\eta}_k, e_k]} \right\rangle_{\mathcal{L}^2} \left(y_j - \left\langle X_j, \widehat{\beta}_{(m, m+2q]} \right\rangle_{\mathcal{L}^2} \right)$

    $F_{\mathcal{I}} \leftarrow \sqrt{2}\, q^{-1/2} \left\{ \sum_{j=m+1}^{m+q} (Z_j - Z_{j+q}) \right\}$

  $\widehat{\sigma}^2_\infty(k) \leftarrow |\mathcal{P}|^{-1} \sum_{\mathcal{I} \in \mathcal{P}} (F_{\mathcal{I}})^2$

OUTPUT: $\{\widehat{\sigma}^2_\infty(k)\}_{k=1}^{\widehat{\mathcal{K}}}$.

Theorem 3. *Suppose that all the assumptions of Theorem 2 hold, $\kappa < \infty$, and that $\widehat{\mathcal{K}} = \mathcal{K}$. In Algorithm 2, let $\{(s_k, e_k)\}_{k=1}^{\mathcal{K}}$ be defined as in (2.10), $\{\widetilde{\eta}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ be the refined estimators as in (2.9), $\{\widehat{\kappa}_k\}_{k=1}^{\widehat{\mathcal{K}}}$ be defined as in (4.14), and q be an integer satisfying*

$$\left(\frac{\log^{2+2\xi}(\Delta)}{\kappa^2} \right)^{\frac{2r+1}{2r-1}} \ll q \ll \Delta. \tag{4.15}$$

Denote by $\{\widehat{\sigma}^2_\infty(k)\}_{k=1}^{\widehat{\mathcal{K}}}$ *the output of Algorithm 2. Then, for any given* $k \in \{1, \ldots, \widehat{\mathcal{K}}\}$, $\widehat{\sigma}^2_\infty(k) \xrightarrow{\mathbb{P}} \sigma^2_\infty(k)$ *as* $n \to \infty$.

The choice of the tuning parameter q needs to balance the bias within each interval and the variance across all intervals in $\mathcal{P}$. The practical choice of q is outlined in Section A.1 in the supplementary material.

4.2 Confidence interval construction

In this subsection, we outline the practical procedure for constructing an asymptotically valid confidence interval in the vanishing regime for each change point. For any given $k \in \{1, \ldots, \widehat{\mathcal{K}}\}$ and confidence level $\alpha \in (0, 1)$, a $100(1-\alpha)\%$ confidence interval for η_k is constructed in two steps:

Step I. Let $B \in \mathbb{N}$. For $b \in \{1, \ldots, B\}$, define

$$\widehat{u}_k^{(b)} = \underset{r\in(-\infty,\infty)}{\arg\min} \left(|r| + \widehat{\sigma}_\infty(k)\mathbb{W}^{(b)}(r)\right), \tag{4.16}$$

where $\widehat{\sigma}^2_\infty(k)$ is the long-run variance estimator from Algorithm 2, and

$$\mathbb{W}^{(b)}(r) = \begin{cases} \frac{1}{\sqrt{n}} \sum_{j=\lfloor nr \rfloor}^{-1} z_j^{(b)} & \text{for} \quad r < 0, \\ 0 & \text{for} \quad r = 0, \\ \frac{1}{\sqrt{n}} \sum_{j=1}^{\lceil nr \rceil} z_j^{(b)} & \text{for} \quad r > 0, \end{cases}$$

with $\{z_j^{(b)}\}_{j=-\infty}^{\infty}$ being i.i.d. standard normal random variables.

Step II. Let $\widehat{q}_{k,\alpha/2}$ and $\widehat{q}_{k,1-\alpha/2}$ be the $\alpha/2$-quantile and $(1-\alpha/2)$-quantile of the empirical distribution of $\{\widehat{u}_k^{(b)}\}_{b=1}^{B}$. Then, the confidence interval for η_k is constructed as

$$\left[\widetilde{\eta}_k + \frac{\widehat{q}_{k,\alpha/2}}{\widehat{\kappa}_k^2},\ \ \widetilde{\eta}_k + \frac{\widehat{q}_{k,1-\alpha/2}}{\widehat{\kappa}_k^2}\right], \tag{4.17}$$

where $\widehat{\kappa}_k^2$ is defined in (4.14).

Theorem 4. *Suppose that all the assumptions of Theorem 3 hold, and that* $\widehat{\mathcal{K}} = \mathcal{K}$. *For any given* $k \in \{1, \ldots, \widehat{\mathcal{K}}\}$ *and* $b = 1, \ldots, B$, *let* $\widehat{u}_k^{(b)}$ *be defined as in* (4.16). *Then, it holds that*

$$\frac{\kappa_k^2}{\widehat{\kappa}_k^2}\widehat{u}_k^{(b)} \xrightarrow{\mathcal{D}} \operatorname*{arg\,min}_{r\in\mathbb{R}} \left\{|r| + \sigma_\infty(k)\mathbb{W}(r)\right\} \quad \text{as} \quad n \to \infty.$$

Theorem 4 implies that the confidence intervals proposed in (4.17) is asymptotic valid in the *vanishing regime* considered in Theorem 2B. Confidence interval construction under the non-vanishing regime remains a challenging problem as the limiting distribution involves random quantities of typically unknown distributions. There are some recent attempts on this problem (e.g. Kaul and Michailidis, 2021; Ng et al., 2022; Cho and Kirch, 2022). However, to the best of our knowledge, there are few theoretical studies for confidence interval construction under the non-vanishing regime in the presence of temporal dependence.

5. Real data analysis

We consider the daily closing price of the S&P 500 index, from Jan-02-2019 to Jan-19-2023 (the data set is available at https://fred.stlouisfed.org/series/SP500). Inspired by a series of papers (e.g. Kokoszka and Zhang, 2012; Kokoszka and Reimherr, 2013), which study the predictability of stock prices using the intraday cumulative returns curves, we regress the daily returns (y_j) on the intratime cumulative return curves (X_j) of the previous one-month (i.e. 21 working days), and use our proposed FRBS as a tool to explore the potential changes in the relationship under the model (1.1). Specifically, we transform the closing price data (P_j) into the log-ratio of close price between two consecutive days in percent, i.e. $y_j = 100 \cdot \log(P_j/P_{j-1})$, and the discretized $X_j = (X_j(1), \ldots, X_j(20))^\top$, in percent,

$$X_j(k) = 100 \cdot \log(P_{j-k}/P_{j-21}), \quad k = 1, 2, \ldots, 20.$$

With j ranging as $j = 22, \ldots, 1271$, the sample size is $n = 1250$. Figure 2 plots y_j and X_j.

We refer to our combined procedure as 'FRBS'. With the tuning parameters selected as discussed in Section A.1 in the supplementary material, the proposed FRBS returns three change points at Jan-07-2020, Mar-11-2020,

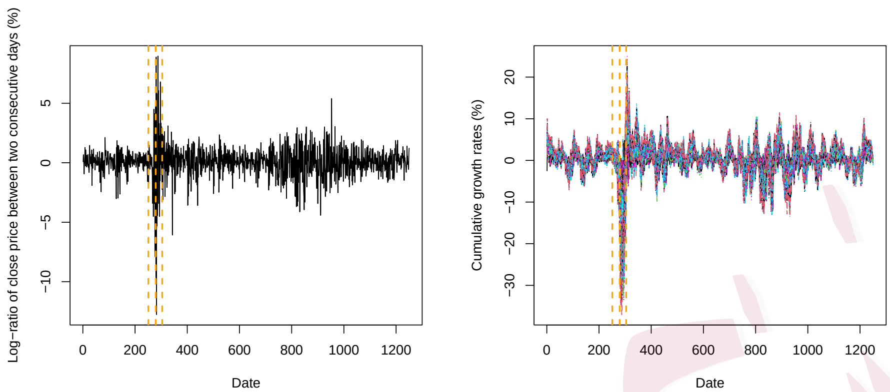

Figure 2: The log-ratio of close price between two consecutive days (y_j, left); the cumulative growth rate ($X_j(k)$, $1 \leq k \leq 20$, right). The refined change point estimators are marked by dashed orange lines.

and May-07-2020 as the preliminary estimators and Jan-30-2020, Mar-11-2020, and Apr-16-2020 as the refined ones. The first estimated change point, with a narrow 95% confidence interval [Jan-28-2020, Feb-03-2020], coincides with the date when WHO officially declared a Public Health Emergency of International Concern. This period reflects investor's concerns about the pandemic's impact on the global economy which led to increased market volatility and a significant sell-off. The second estimated change point, with a 95% confidence interval [Feb-20-2020, Mar-30-2020], matches the date when COVID-19 was characterized as a pandemic by WHO. This declaration confirmed the severity and global scale of the outbreak. During this period, many countries implemented lockdown measures, which lad to huge volatility in financial markets and a sharp drop in the S&P 500 index. The third estimated change point reflects that the initial impact of COVID-19 gradually settled. A series of economic and financial policies were introduced by the governments globally, and the market started to react to these policy changes. Our method produces a wide 95% confidence interval as [Mar-05-2020, May-18-2021].

In comparison, we consider the same transformed y_j and X_j but regard X_j as a covariate vector of dimensional 20 and use high-dimensional linear regression (HDLR) with change points (Xu et al., 2024) to study the relationship between y_j and X_j. The HDLR algorithm outputs two change point estimators at Feb-18-2020 and Apr-14-2020.

In supplementary material, We preform an additional real data analysis on the U.S. Treasury yield to maturity curve versus inflation, which shows that FRBS is effective in identifying the structural breaks well documented in the economics literature.

6. Discussion

In this paper, we study the change point problem within the context of functional linear regression, with minimal assumptions accommodating temporal dependence and heavy-tailed distributions. Our contribution includes deriving the consistency and the limiting distribution of the change point estimators, a novel advancement in this functional framework. Additionally, we propose a theoretically sound and numerically robust long-run variance estimator to enhance the practicality of our findings. We offer the numerical implementation of our proposed approach which is shown to perform well on synthetic and real datasets.

The theoretical analysis has illuminated several challenging and intriguing directions for future research. One direction could involve devising asymptotically valid confidence intervals in the non-vanishing regime with respect to the size of the change. Another direction could focus on developing methods to simultaneously distinguish between different regimes of the size of change, motivated by their difference in the limiting distribution in Theorem 2.

Supplementary Materials

We collect extensive simulation studies, additional real data analysis and all the technical details in the online supplementary material.

References

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., 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.

Baranowski, R., Y. Chen, and P. Fryzlewicz (2019). Narrowest-over-threshold detection of multiple change points and change-point-like features. *Journal of the Royal Statistical Society Series B: Statistical Methodology 81*(3), 649–672.

Cai, T. and M. Yuan (2010). Nonparametric covariance function estimation for functional and longitudinal data. *University of Pennsylvania and Georgia inistitute of technology*.

Cai, T. T. and M. Yuan (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association 107*(499), 1201–1216.

Cai, Z., J. Fan, and Q. Yao (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association 95*(451), 941–956.

Casini, A. and P. Perron (2021). Minimax mse bounds and nonlinear var prewhitening for long-run variance estimation under nonstationarity. *arXiv preprint arXiv:2103.02235*.

Cho, H. and C. Kirch (2022). Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis 175*, 107552.

Dette, H. and T. Kutta (2021). Detecting structural breaks in eigensystems of functional time series. *Electronic Journal of Statistics 15*(1), 944 – 983.

Harris, T., B. Li, and J. D. Tucker (2022). Scalable multiple changepoint detection for functional data sequences. *Environmetrics 33*(2), e2710.

Hörmann, S. and P. Kokoszka (2010). Weakly dependent functional data. *The Annals of Statistics 38*(3), 1845 – 1884.

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

Jiao, S., R. D. Frostig, and H. Ombao (2023). Break point detection for functional covariance. *Scandinavian Journal of Statistics 50*(2), 477–512.

Kaul, A. and G. Michailidis (2021). Inference for change points in high dimensional mean shift models. *arXiv preprint arXiv:2107.09150*.

Kaul, A., H. Zhang, K. Tsampourakis, and G. Michailidis (2023). Inference on the change point under a high dimensional covariance shift. *Journal of Machine Learning Research 24*(168), 1–68.

Khismatullina, M. and M. Vogt (2020). Multiscale inference and long-run variance estimation in non-parametric regression with time series errors. *Journal of the Royal Statistical Society Series B: Statistical Methodology 82*(1), 5–37.

Kokoszka, P. and M. Reimherr (2013). Predictability of shapes of intraday price curves. *The*

Econometrics Journal 16(3), 285–308.

Kokoszka, P. and M. Reimherr (2017). *Introduction to functional data analysis*. CRC press.

Kokoszka, P. and X. Zhang (2012). Functional prediction of intraday cumulative returns. *Statistical Modelling 12*(4), 377–398.

Koner, S. and A.-M. Staicu (2023). Second-generation functional data. *Annual Review of Statistics and Its Application 10*, 547–572.

Kovács, S., P. Bühlmann, H. Li, and A. Munk (2023). Seeded binary segmentation: a general methodology for fast and optimal changepoint detection. *Biometrika 110*(1), 249–256.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2017). A bayesian multivariate functional dynamic linear model. *Journal of the American Statistical Association 112*(518), 733–744.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Functional autoregression for sparsely sampled data. *Journal of Business & Economic Statistics 37*(1), 97–109.

Li, J., Y. Li, and T. Hsing (2022). On functional processes with multiple discontinuities. *Journal of the Royal Statistical Society Series B: Statistical Methodology 84*(3), 933–972.

Li, X. and S. Ghosal (2021). Bayesian change point detection for functional data. *Journal of Statistical Planning and Inference 213*, 193–205.

Madrid Padilla, C. M., D. Wang, Z. Zhao, and Y. Yu (2022). Change-point detection for sparse and dense functional data in general dimensions. *Advances in Neural Information Processing Systems 35*, 37121–37133.

Madrid Padilla, C. M., H. Xu, D. Wang, O. H. Madrid Padilla, and Y. Yu (2023). Change point detection and inference in multivariable nonparametric models under mixing conditions. *arXiv preprint arXiv:2301.11491*.

Madrid Padilla, C. M., H. Xu, D. Wang, O. H. Madrid Padilla, and Y. Yu (2024). Change point detection and inference in multivariate non-parametric models under mixing conditions. *Advances in Neural Information Processing Systems 36*.

Madrid Padilla, O. H., Y. Yu, D. Wang, and A. Rinaldo (2021). Optimal nonparametric multivariate change point detection and localization. *IEEE Transactions on Information Theory 68*(3), 1922–1944.

Ng, W. L., S. Pan, and C. Y. Yau (2022). Bootstrap inference for multiple change-points in time series. *Econometric Theory 38*(4), 752–792.

Panaretos, V. M. and S. Tavakoli (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 568–603.

Ramsay, J. and B. Silverman (2005). Principal components analysis for functional data. *Functional data analysis*, 147–172.

Ramsay, J. O. and B. W. Silverman (2002). *Applied functional data analysis: methods and case studies*. Springer.

Rubín, T. and V. M. Panaretos (2020). Sparsely observed functional time series: estimation and prediction. *Electronic Journal of Statistics 14*(1), 1137 – 1210.

Sullivan, J. H. and W. H. Woodall (2000). Change-point detection of mean vector or covariance matrix shifts using multivariate individual observations. *IIE transactions 32*(6), 537–549.

Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge university press.

Wang, D., Y. Yu, and A. Rinaldo (2020). Univariate mean change point detection: Penalization, CUSUM and optimality. *Electronic Journal of Statistics 14*(1), 1917 – 1961.

Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). Functional data analysis. *Annual Review of Statistics and its application 3*, 257–295.

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.

Xu, H., D. Wang, Z. Zhao, and Y. Yu (2024). Change-point inference in high-dimensional regression models under temporal dependence. *The Annals of Statistics 52*(3), 999–1026.

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.

Yuan, M. and T. T. Cai (2010). A reproducing kernel Hilbert space approach to functional linear regression. *The Annals of Statistics 38*(6), 3412 – 3444.

University of Chicago

E-mail: Shivam.Kumar@chicagobooth.edu

Auburn University

E-mail: haotian.xu@auburn.edu

University of Bristol

E-mail: haeran.cho@bristol.ac.uk

University of California San Diego

E-mail: daw040@ucsd.edu