Change-point Detection for Sparse and Dense Functional Data in General Dimensions

Carlos Misael Madrid Padilla
Department of Mathematics
University of Notre Dame
cmadridp@nd.edu

Daren Wang
Department of Statistics
University of Notre Dame
dwang24@nd.edu

Zifeng Zhao
Mendoza College of Business
University of Notre Dame
zzhao2@nd.edu

Yi Yu
Department of Statistics
University of Warwick
yi.yu.2@warwick.ac.uk

Abstract

We study the problem of change-point detection and localisation for functional data sequentially observed on a general d-dimensional space, where we allow the functional curves to be either sparsely or densely sampled. Data of this form naturally arise in a wide range of applications such as biology, neuroscience, climatology and finance. To achieve such a task, we propose a kernel-based algorithm namely functional seeded binary segmentation (FSBS). FSBS is computationally efficient, can handle discretely observed functional data, and is theoretically sound for heavy-tailed and temporally-dependent observations. Moreover, FSBS works for a general d-dimensional domain, which is the first in the literature of change-point estimation for functional data. We show the consistency of FSBS for multiple change-point estimation and further provide a sharp localisation error rate, which reveals an interesting phase transition phenomenon depending on the number of functional curves observed and the sampling frequency for each curve. Extensive numerical experiments illustrate the effectiveness of FSBS and its advantage over existing methods in the literature under various settings. A real data application is further conducted, where FSBS localises change-points of sea surface temperature patterns in the south Pacific attributed to El Niño. The code to replicate all of our experiments can be found at https://github.com/cmadridp/FSBS.

1 Introduction

Recent technological advancement has boosted the emergence of functional data in various application areas, including neuroscience [e.g. 11, 28], finance [e.g. 13], transportation [e.g. 10], climatology [e.g. 7, 14] and others. We refer the readers to [37] - a comprehensive review, for recent development of statistical research in functional data analysis.

In this paper, we study the problem of change-point detection and localisation for functional data, where the data are observed sequentially as a time series and the mean functions are piecewise stationary, with abrupt changes occurring at unknown time points. To be specific, denote $\mathcal{D}$ as a general d-dimensional space that is homeomorphic to $[0, 1]^d$, where $d \in \mathbb{N}^+$ is considered as arbitrary but fixed. We assume that the observations $\{(x_{t,i}, y_{t,i})\}_{t=1,i=1}^{T,n} \subseteq \mathcal{D} \times \mathbb{R}$ are generated based on

$$y_{t,i} = f_t^*(x_{t,i}) + \xi_t(x_{t,i}) + \delta_{t,i}, \text{ for } t = 1, \dots, T \text{ and } i = 1, \dots, n. \quad (1)$$

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

In this model, $\{x_{t,i}\}_{t=1,i=1}^{T,n} \subseteq \mathcal{D}$ denotes the discrete grids where the (noisy) functional data $\{y_{t,i}\}_{t=1,i=1}^{T,n} \subseteq \mathbb{R}$ are observed, $\{f_t^* : \mathcal{D} \to \mathbb{R}\}_{t=1}^T$ denotes the deterministic mean functions, $\{\xi_t : \mathcal{D} \to \mathbb{R}\}_{t=1}^T$ denotes the functional noise and $\{\delta_{t,i}\}_{t=1,i=1}^{T,n} \subseteq \mathbb{R}$ denotes the measurement error. We refer to Assumption 1 below for detailed technical conditions on the model.

To model the unstationarity of sequentially observed functional data which commonly exists in real world applications, we assume that there exist $K \in \mathbb{N}$ change-points, namely $0 = \eta_0 < \eta_1 < \cdots < \eta_K < \eta_{K+1} = T$, satisfying that $f_t^* \neq f_{t+1}^*$, if and only if $t \in \{\eta_k\}_{k=1}^K$. Our primary interest is to accurately estimate $\{\eta_k\}_{k=1}^K$.

Due to the importance of modelling unstationary functional data in various scientific fields, this problem has received extensive attention in the statistical change-point literature, see e.g. [3], [6], [18], [40], [4] and [12]. Despite the popularity, we identify a few limitations in the existing works. Firstly, both the methodological validity and theoretical guarantees of all these papers require fully observed functional data without measurement error, which may not be realistic in practice. Secondly, most existing works focus on the single change-point setting and to our best knowledge, there is no consistency result of multiple change-point estimation for functional data. Lastly but most importantly, existing algorithms only consider functional data with support on $[0, 1]$ and thus are not applicable to functional data with multi-dimensional domain, a type of data frequently encountered in neuroscience and climatology.

In view of the aforementioned three limitations, in this paper, we make several theoretical and methodological contributions, summarized below.

• In terms of methodology, our proposed kernel-based change-point detection algorithm, functional seeded binary segmentation (FSBS), is computationally efficient, can handle discretely observed functional data contaminated with measurement error, and allows for temporally-dependent and heavy-tailed data. FSBS, in particular, works for a general d-dimensional domain with arbitrary but fixed $d \in \mathbb{N}^+$. This level of generality is the first time seen in the literature.

• In terms of theory, we show that under standard regularity conditions, FSBS is consistent in detecting and localising multiple change-points. We also provide a sharp localisation error rate, which reveals an interesting phase transition phenomenon depending on the number of functional curves observed T and the sampling frequency for each curve n. To the best of our knowledge, the theoretical results we provide in this paper are the sharpest in the existing literature.

• A striking case we handle in this paper is that each curve is only sampled at one point, i.e. $n = 1$. To the best of our knowledge, all the existing functional data change-point analysis papers assume full curves are observed. We not only allow for discrete observation, but carefully study this most extreme sparse case $n = 1$ and provide consistent localisation of the change-points.

• We conduct extensive numerical experiments on simulated and real data. The result further supports our theoretical findings, showcases the advantages of FSBS over existing methods and illustrates the practicality of FSBS.

• A byproduct of our theoretical analysis is new theoretical results on kernel estimation for functional data under temporal dependence and heavy-tailedness. This set of new results *per se* are novel, enlarging the toolboxes of functional data analysis.

Notation and definition. For any function $f : [0, 1]^d \to \mathbb{R}$ and for $1 \le p < \infty$, define $\|f\|_p = (\int_{[0,1]^d} |f(x)|^p \, dx)^{1/p}$ and for $p = \infty$, define $\|f\|_\infty = \sup_{x \in [0,1]^d} |f(x)|$. Define $\mathcal{L}_p = \{f : [0, 1]^d \to \mathbb{R}, \|f\|_p < \infty\}$. For any vector $s = (s_1, \ldots, s_d)^\top \in \mathbb{N}^d$, define $|s| = \sum_{i=1}^d s_i$, $s! = s_1! \cdots s_d!$ and the associated partial differential operator $D^s = \frac{\partial^{|s|}}{\partial x_1^{s_1} \cdots \partial x_d^{s_d}}$. For $\alpha > 0$, denote $\lfloor \alpha \rfloor$ to be the largest integer smaller than α. For any function $f : [0, 1]^d \to \mathbb{R}$ that is $\lfloor \alpha \rfloor$-times continuously differentiable at point x_0, denote by $f_{x_0}^\alpha$ its Taylor polynomial of degree $\lfloor \alpha \rfloor$ at x_0, which is defined as $f_{x_0}^\alpha(x) = \sum_{|s| \le \lfloor \alpha \rfloor} \frac{(x - x_0)^s}{s!} D^s f(x_0)$. For a constant $L > 0$, let $\mathcal{H}^\alpha(L)$ be the set of functions $f : [0, 1]^d \to \mathbb{R}$ such that f is $\lfloor \alpha \rfloor$-times differentiable for all $x \in [0, 1]^d$ and satisfy $|f(x) - f_{x_0}^\alpha(x)| \le L|x - x_0|^\alpha$, for all $x, x_0 \in [0, 1]^d$. Here $|x - x_0|$ is the Euclidean distance between $x, x_0 \in \mathbb{R}^d$. In non-parametric statistical literature, $\mathcal{H}^\alpha(L)$ are often referred to as the class of Hölder smooth functions. We refer the interested readers to [30] for more detailed discussion on Hölder smooth functions.

For two positive sequences $\{a_n\}_{n\in\mathbb{N}^+}$ and $\{b_n\}_{n\in\mathbb{N}^+}$, we write $a_n = O(b_n)$ or $a_n \lesssim b_n$ if $a_n \leq Cb_n$ with some constant $C > 0$ that does not depend on n, and $a_n = \Theta(b_n)$ or $a_n \asymp b_n$ if $a_n = O(b_n)$ and $b_n = O(a_n)$.

2 Functional seeded binary segmentation

2.1 Problem formulation

Detailed model assumptions imposed on model (1) are collected in Assumption 1. For notational simplicity, without loss of generality, we set the general d-dimensional domain $\mathcal{D}$ to be $[0,1]^d$, as the results apply to any $\mathcal{D}$ that is homeomorphic to $[0,1]^d$.

Assumption 1. *The data* $\{(x_{t,i}, y_{t,i})\}_{t=1,i=1}^{T,n} \subseteq [0,1]^d \times \mathbb{R}$ *are generated based on model* (1).

a. *(Discrete grids) The grids* $\{x_{t,i}\}_{t=1,i=1}^{T,n} \subseteq [0,1]^d$ *are independently sampled from a common density function* $u : [0,1]^d \to \mathbb{R}$. *In addition, there exist constants* $r > 0$ *and* $L > 0$ *such that* $u \in \mathcal{H}^r(L)$ *and that* $\inf_{x\in[0,1]^d} u(x) \geq \tilde{c}$ *with an absolute constant* $\tilde{c} > 0$.

b. *(Mean functions) For* $r > 0$ *and* $L > 0$, *we have* $f_t^* \in \mathcal{H}^r(L)$. *The minimal spacing between two consecutive change-points* $\Delta = \min_{k=1}^{K+1}(\eta_k - \eta_{k-1})$ *satisfies that* $\Delta = \Theta(T)$.

c. *(Functional noise) Let* $\{\varepsilon_i, \varepsilon_0'\}_{i\in\mathbb{Z}}$ *be i.i.d. random elements taking values in a measurable space* S_ξ *and* g *be a measurable function* $g : S_\xi^\infty \to \mathcal{L}_2$. *The functional noise* $\{\xi_t\}_{t=1}^T \subseteq \mathcal{L}_2$ *takes the form*

$$\xi_t = g(\mathcal{G}_t), \quad \textit{with } \mathcal{G}_t = (\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t).$$

There exists an absolute constant $q \geq 3$, *such that* $\mathbb{E}(\|\xi_t\|_\infty^q) < C_{\xi,1}$ *for some absolute constant* $C_{\xi,1}$. *Define a coupled process*

$$\xi_t^* = g(\mathcal{G}_t^*), \quad \textit{with } \mathcal{G}_t^* = (\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t).$$

We have $\sum_{t=1}^\infty t^{1/2-1/q}\{\mathbb{E}\|\xi_t - \xi_t^*\|_\infty^q\}^{1/q} < C_{\xi,2}$ *for some absolute constant* $C_{\xi,2} > 0$.

d. *(Measurement error) Let* $\{\epsilon_i, \epsilon_0'\}_{i\in\mathbb{Z}}$ *be i.i.d. random elements taking values in a measurable space* S_δ *and* $\tilde{g}_n$ *be a measurable function* $\tilde{g}_n : S_\delta^\infty \to \mathbb{R}^n$. *The measurement error* $\{\delta_t\}_{t=1}^T \subseteq \mathbb{R}^n$ *takes the form*

$$\delta_t = \tilde{g}_n(\mathcal{F}_t), \quad \textit{with } \mathcal{F}_t = (\ldots, \epsilon_{-1}, \epsilon_0, \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t).$$

There exists an absolute constant $q \geq 3$, *such that* $\max_{i=1}^n \mathbb{E}(|\delta_{t,i}|^q) < C_{\delta,1}$ *for some absolute constant* $C_{\delta,1}$. *Define a coupled process*

$$\delta_t^* = \tilde{g}_n(\mathcal{F}_t^*), \quad \textit{with } \mathcal{F}_t^* = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_{t-1}, \epsilon_t).$$

We have $\max_{i=1}^n \sum_{t=1}^\infty t^{1/2-1/q}\{\mathbb{E}|\delta_{t,i} - \delta_{t,i}^*|^q\}^{1/q} < C_{\delta,2}$ *for some absolute constant* $C_{\delta,2} > 0$.

Assumption 1**a** allows the functional data to be observed on discrete grids and moreover, we allow for different grids at different time points. The sampling distribution μ is required to be lower bounded on the support $[0,1]^d$, which is a standard assumption widely used in the nonparametric literature [e.g. 33]. Here, different functional curves are assumed to have the same number of grid points n. We remark that this is made for presentation simplicity only. It can indeed be further relaxed and the main results below will then depend on both the minimum and maximum numbers of grid points.

Note that Assumption 1**a** does not impose any restriction between the sampling frequency n and the number of functional curves T, and indeed our method can handle both the dense case where $n \gg T$ and the sparse case where n can be upper bounded by a constant. Besides the random sampling scheme studied here, another commonly studied scenario is the fixed design, where it usually assumes that the sampling locations $\{x_i\}_{i=1}^n$ are common to all functional curves across time. We remark that while we focus on the random design here, our proposed algorithm can be directly applied to the fixed design case without any modification. Furthermore, its theoretical justification under the fixed design case can be established similarly with minor modifications, which is omitted.

The observed functional data have mean functions $\{f_t^*\}_{t=1}^T$, which are assumed to be Hölder continuous in Assumption 1**b**. Note that the Hölder parameters in Assumption 1**a** and **b** are both denoted by r. We remark that different smoothness are allowed and we use the same r here for notational simplicity. This sequence of mean functions is our primary interest and is assumed to possess a piecewise constant pattern, with the minimal spacing Δ being of the same order as T. This assumption essentially requires that the number of change-points is upper bounded. It can also be further relaxed and we will have more elaborated discussions on this matter in Section 5.

Our model allows for two sources of noise - functional noise and measurement error, which are detailed in Assumption 1**c** and **d**, respectively. Both the functional noise and the measurement error are allowed to possess temporal dependence and heavy-tailedness. For temporal dependence, we adopt the physical dependence framework by [39], which covers a wide range of time series models, such as ARMA and vector AR models. It further covers popular functional time series models such as functional AR and MA models [18]. We also remark that Assumption 1**c** and **d** impose a short range dependence, which is characterized by the absolute upper bounds $C_{\xi,2}$ and $C_{\delta,2}$. Further relaxation is possible by allowing the upper bounds $C_{\xi,2}$ and $C_{\delta,2}$ to vary with the sample size T.

The heavy-tail behavior is encoded in the parameter q. In Assumption 1**c** and **d**, we adopt the same quantity q for presentational simplicity and remark that different heavy-tailedness levels are allowed. An extreme example is that when $q = \infty$, the noise is essentially sub-Gaussian. Importantly, Assumption 1**d** does not impose any restriction on the cross-sectional dependence among measurement errors observed on the same time t, which can be even perfectly correlated.

2.2 Kernel-based change-point detection

To estimate the change-point $\{\eta_k\}_{k=1}^K$ in the mean functions $\{f_t^*\}_{t=1}^T$, we propose a kernel-based cumulative sum (CUSUM) statistic, which is simple, intuitive and computationally efficient. The key idea is to recover the unobserved $\{f_t^*\}_{t=1}^T$ from the observations $\{(x_{t,i}, y_{t,i})\}_{t=1,i=1}^{T,n}$ based on kernel estimation.

Given a kernel function $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ and a bandwidth parameter $h > 0$, we define $K_h(x) = h^{-d}K(x/h)$ for $x \in \mathbb{R}^d$. Given the random grids $\{x_{t,i}\}_{t=1,i=1}^{T,n}$ and a bandwidth parameter $\bar{h}$, we define the density estimator of the sampling distribution $u(x)$ as

$$\hat{p}(x) = \hat{p}_{\bar{h}}(x) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n} K_{\bar{h}}(x - x_{t,i}), \quad x \in [0,1]^d.$$

Given $\hat{p}(x)$ and a bandwidth parameter $h > 0$, for any time $t = 1, 2, \cdots, T$, we define the kernel-based estimation for $f_t^*(x)$ as

$$F_{t,h}(x) = \frac{\sum_{i=1}^n y_{t,i}K_h(x - x_{t,i})}{n\hat{p}(x)}, \quad x \in [0,1]^d. \tag{2}$$

Based on the kernel estimation $F_{t,h}(x)$, for any integer pair $0 \le s < e \le T$, we define the CUSUM statistic as

$$\widetilde{F}_{t,h}^{(s,e]}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{l=s+1}^{t} F_{l,h}(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{l=t+1}^{e} F_{l,h}(x), \quad x \in [0,1]^d. \tag{3}$$

The CUSUM statistic defined in (3) is the cornerstone of our algorithm and is based on two kernel estimators $\hat{p}(\cdot)$ and $F_{t,h}(\cdot)$. At a high level, the CUSUM statistic $\widetilde{F}_{t,h}^{(s,e]}(\cdot)$ estimates the difference in mean between the functional data in the time intervals $(s, t]$ and $(t, e]$. In the functional data analysis literature, other popular approaches for mean function estimation are reproducing kernel Hilbert space based methods and local polynomial regression. However, to our best knowledge, existing works based on the two approaches typically require that the functional data are temporally independent and it is not obvious how to extend their theoretical guarantees to the temporal dependence case. We therefore choose the kernel estimation method owing to its flexibility in terms of both methodology and theory and we derive new theoretical results on kernel estimation for functional data under temporal dependence and heavy-tailedness. We would like to point out that in the existing literature, kernel-based change-point estimation methods are used in detecting change-points in nonparametric models [e.g. 23, 1, 24, 27].

For multiple change-point estimation, a key ingredient is to isolate each single change-point with well-designed intervals in $[0, T]$. To achieve this, we combine the CUSUM statistic in (3) with a modified version of the seeded binary segmentation (SBS) proposed in [22]. SBS is based on a collection of deterministic intervals defined in Definition 1.

Definition 1 (Seeded intervals). *Let $\mathcal{K} = \lceil C_{\mathcal{K}} \log\log(T) \rceil$, with some sufficiently large absolute constant $C_{\mathcal{K}} > 0$. For $k \in \{1, \ldots, \mathcal{K}\}$, let $\mathcal{J}_k$ be the collection of $2^k - 1$ intervals of length $l_k = T2^{-k+1}$ that are evenly shifted by $l_k/2 = T2^{-k}$, i.e.*

$$\mathcal{J}_k = \{(\lfloor (i-1)T2^{-k} \rfloor, \lceil (i-1)T2^{-k} + T2^{-k+1} \rceil], \quad i = 1, \ldots, 2^k - 1\}.$$

The overall collection of seeded intervals is denoted as $\mathcal{J} = \cup_{k=1}^{\mathcal{K}} \mathcal{J}_k$.

The essential idea of the seeded intervals defined in Definition 1 is to provide a multi-scale system of searching regions for multiple change-points. SBS is computationally efficient with a computational cost of the order $O(T \log(T))$ [22].

Based on the CUSUM statistic and seeded intervals, Algorithm 1 summarises the proposed functional seeded binary segmentation algorithm (FSBS) for multiple change-point estimation in sequentially observed functional data. There are three main tuning parameters involved in Algorithm 1, the kernel bandwidth $\bar{h}$ in the estimation of the sampling distribution, the kernel bandwidth h in the estimation of the mean function and the threshold parameter τ for declaring change-points. Their theoretical and numerical guidance will be presented in Sections 3.1 and 4, respectively.

Algorithm 1 Functional Seeded Binary Segmentation. FSBS $((s, e], \bar{h}, h, \tau)$

INPUT: Data $\{x_{t,i}, y_{t,i}\}_{t=1,i=1}^{T,n}$, seeded intervals $\mathcal{J}$, tuning parameters $\bar{h}, h, \tau > 0$.

Initialization: If $(s, e] = (0, n]$, set $\mathbf{S} \leftarrow \varnothing$ and set $\rho \leftarrow \log(T) n^{-1} h^{-d}$. Furthermore, sample $\lceil \log(T) \rceil$ points from $\{x_{t,i}\}_{t=1,i=1}^{T,n}$ uniformly at random without replacement and denote them as $\{u_m\}_{m=1}^{\lceil \log(T) \rceil}$. Estimate the sampling distribution evaluated at $\{\hat{p}_{\bar{h}}(u_m)\}_{m=1}^{\lceil \log(T) \rceil}$.

for $\mathcal{I} = (\alpha, \beta] \in \mathcal{J}$ and $m \in \{1, \ldots, \lceil \log(T) \rceil\}$ **do**
  if $\mathcal{I} = (\alpha, \beta] \subseteq (s, e]$ and $\beta - \alpha > 2\rho$ **then**
    $A_m^{\mathcal{I}} \leftarrow \max_{\alpha+\rho \le t \le \beta-\rho} |\widetilde{F}_{t,h}^{(\alpha,\beta]}(u_m)|$, $D_m^{\mathcal{I}} \leftarrow \arg\max_{\alpha+\rho \le t \le \beta-\rho} |\widetilde{F}_{t,h}^{(\alpha,\beta]}(u_m)|$
  else
    $(A_m^{\mathcal{I}}, D_m^{\mathcal{I}}) \leftarrow (-1, 0)$
  end if
end for
$(m^*, \mathcal{I}^*) \leftarrow \arg\max_{m=1,\ldots,\lceil \log(T) \rceil, \mathcal{I} \in \mathcal{J}} A_m^{\mathcal{I}}$.
if $A_{m^*}^{\mathcal{I}^*} > \tau$ **then**
  $\mathbf{S} \leftarrow \mathbf{S} \cup D_{m^*}^{\mathcal{I}^*}$
  FSBS $((s, D_{m^*}^{\mathcal{I}^*}], \bar{h}, h, \tau)$
  FSBS $((D_{m^*}^{\mathcal{I}^*}, e], \bar{h}, h, \tau)$
end if

OUTPUT: The set of estimated change-points $\mathbf{S}$.

Algorithm 1 is conducted in an iterative way, starting with the whole time course, using the multi-scale seeded intervals to search for the point according to the largest CUSUM value. A change-point is declared if the corresponding maximum CUSUM value exceeds a pre-specified threshold τ and the whole sequence is then split into two with the procedure being carried on in the sub-intervals.

Algorithm 1 utilizes a collection of random grid points $\{u_m\}_{m=1}^{\lceil \log(T) \rceil} \subseteq \{x_{t,i}\}_{t=1,i=1}^{T,n}$ to detect changes in the functional data. For a change of mean functions at the time point η with $\|f_{\eta+1}^* - f_\eta^*\|_\infty > 0$, we show in the appendix that, as long as $\lceil \log(T) \rceil$ grid points are sampled, with high probability, there is at least one point $u_{m'} \in \{u_m\}_{m=1}^{\lceil \log(T) \rceil}$ such that $|f_{\eta+1}^*(u_{m'}) - f_\eta^*(u_{m'})| \asymp \|f_{\eta+1}^* - f_\eta^*\|_\infty$. Thus, this procedure allows FSBS to detect changes in the mean functions without evaluating functions on a dense lattice grid and thus improves computational efficiency.

3 Main Results

3.1 Assumptions and theory

We begin by imposing assumptions on the kernel function $K(\cdot)$ used in FSBS.

Assumption 2 (Kernel function)**.** *Let $K(\cdot) : \mathbb{R}^d \to \mathbb{R}^+$ be compactly supported and satisfy the following conditions.*

a. *The kernel function $K(\cdot)$ is adaptive to the Hölder class $\mathcal{H}^r(L)$, i.e. for any $f \in \mathcal{H}^r(L)$, it holds that $\sup_{x\in[0,1]^d} \left| \int_{[0,1]^d} K_h(x-z) f(z)\,\mathrm{d}z - f(x)\right| \le \tilde{C}h^r$, where $\tilde{C} > 0$ is a constant that only depends on L.*

b. *The class of functions $\mathcal{F}_K = \{K(x - \cdot)/h : \mathbb{R}^d \to \mathbb{R}^+, h > 0\}$ is separable in $\mathcal{L}_\infty(\mathbb{R}^d)$ and is a uniformly bounded VC-class. This means that there exist constants $A, \nu > 0$ such that for every probability measure Q on $\mathbb{R}^d$ and every $u \in (0, \|K\|_\infty)$, it holds that $\mathcal{N}(\mathcal{F}_K, \mathcal{L}_2(Q), u) \le (A\|K\|_\infty/u)^v$, where $\mathcal{N}(\mathcal{F}_K, \mathcal{L}_2(Q), u)$ denotes the u-covering number of the metric space $(\mathcal{F}_K, \mathcal{L}_2(Q))$.*

Assumption 2 is a standard assumption in the nonparametric literature, see [15, 16], [20], [35] among many others. These assumptions hold for most commonly used kernels, including uniform, polynomial and Gaussian kernels.

Recall the minimal spacing $\Delta = \min_{k=1}^{K+1}(\eta_k - \eta_{k-1})$ defined in Assumption 1**b**. We further define the jump size at the kth change-point as $\kappa_k = \|f^*_{\eta_k+1} - f^*_{\eta_k}\|_\infty$ and define $\kappa = \min_{k=1}^{K}$ as the minimal jump size. Assumption 3 below details how strong the signal needs to be in terms of κ and Δ, given the grid size n, the number of functional curves T, smoothness parameter r, dimensionality d and moment condition q.

Assumption 3 (Signal-to-noise ratio, SNR)**.** *There exists an arbitrarily-slow diverging sequence $C_{\mathrm{SNR}} = C_{\mathrm{SNR}}(T)$ such that*

$$\kappa\sqrt{\Delta} > C_{\mathrm{SNR}} \log^{\max\{1/2,5/q\}}(T)\left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)^{1/2}.$$

We are now ready to present the main theorem, showing the consistency of FSBS.

Theorem 1. *Under Assumptions 1, 2 and 3, let $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ be the estimated change-points by FSBS detailed in Algorithm 1 with data $\{x_{t,i}, y_{t,i}\}_{t=1,i=1}^{T,n}$, bandwidth parameters $\bar{h} = C_{\bar{h}}(Tn)^{-\frac{1}{2r+d}}$, $h = C_h(Tn)^{\frac{-1}{2r+d}}$ and threshold parameter $\tau = C_\tau \log^{\max\{1/2,5/q\}}(T)\left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)^{1/2}$, for some absolute constants $C_{\bar{h}}, C_h, C_\tau > 0$. It holds that*

$$\mathbb{P}\left\{\widehat{K} = K;\ |\widehat{\eta}_k - \eta_k| \le C_{\mathrm{FSBS}} \log^{\max\{1,10/q\}}(T)\left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)\kappa_k^{-2},\ \forall k = 1,\ldots,K\right\} \ge 1 - 3\log^{-1}(T),$$

where $C_{\mathrm{FSBS}} > 0$ is an absolute constant.

In view of Assumption 3 and Theorem 1, we see that with properly chosen tuning parameters and with probability tending to one as the sample size T grows, the output of FSBS estimates the correct number of change-points and

$$\max_{k=1}^{K} |\widehat{\eta}_k - \eta_k|/\Delta \lesssim \left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)\log^{\max\{1,10/q\}}(T)/(\kappa^2\Delta) = o(1),$$

where the last inequality follows from Assumption 3. The above inequality shows that there exists a one-to-one mapping from $\{\widehat{\eta}_k\}_{k=1}^{K}$ to $\{\eta_k\}_{k=1}^{K}$, assigning by the smallest distance.

3.2 Discussions on functional seeded binary segmentation (FSBS)

From sparse to dense regimes. In our setup, each curve is only observed at n discrete points and we allow the full range of choices of n, representing from sparse to dense scenarios, all accompanied with consistency results. In the most sparse case $n = 1$, Assumption 3 reads as $\kappa\sqrt{\Delta} \gtrsim T^{d/(4r+2d)} \times$ a logarithmic factor, under which the localisation error is upper bounded by $T^{d/(2r+d)}\kappa^{-2}$, up to a logarithmic factor. To the best of our knowledge, this challenging case has not been dealt in the existing change-point detection literature for functional data. In the most dense case, we can heuristically let $n = \infty$ and for simplicity let $q = \infty$ representing the sub-Gaussian noise case. Assumption 3 reads as $\kappa\sqrt{\Delta} \asymp \log^{1/2}(T)$ and the localisation error is upper bounded by $\kappa^{-2}\log(T)$. Both the SNR ratio and localisation error are the optimal rate in the univariate mean change-point localisation problem [36], implying the optimality of FSBS in the dense situation.

Tuning parameters. There are three tuning parameters involved. In the CUSUM statistic (3), we specify that the density estimator of the sampling distribution is a kernel estimator with bandwidth $\bar{h} \asymp (Tn)^{-1/(2r+d)}$. Due to the independence of the observation grids, such a choice of the bandwidth follows from the classical nonparametric literature [e.g. 33] and is minimax-rate optimal in terms of the estimation error. For completeness, we include the study of $\hat{p}(\cdot)$'s theoretical properties in Appendix B. In practice, there exist different default methods for the selection of $\bar{h}$, see for example the function `Hpi` from the R package `ks` ([9]).

The other bandwidth tuning parameter h is also required to be $h \asymp (Tn)^{-1/(2r+d)}$. Despite that we allow for physical dependence in both functional noise and measurement error, we show that the same order of bandwidth (as $\bar{h}$) is required under Assumption 1. This is an interesting finding, if not surprising. This particular choice of h is due to the fact that the physical dependence put forward by [39] is a short range dependence condition and does not change the rate of the sample size.

The threshold tuning parameter τ is set to be a high-probability upper bound on the CUSUM statistics when there is no change-point and is in fact of the form $\tau = C_\tau \log^{\max\{1/2,5/q\}}(T)\sqrt{n^{-1}h^{-d}+1}$. This also reflects the requirement on the SNR detailed in Assumption 3, that $\kappa\sqrt{\Delta} \gtrsim \tau$.

Phase transition. Recall that the number of curves is T and the number of observations on each curve is n. The asymptotic regime we discuss is to let T diverge, while allowing all other parameters, including n, to be functions of T. In Theorem 1, we allow a full range of cases in terms of the relationship between n and T. As a concrete example, when the smooth parameter $r = 2$, the jump size $\kappa \asymp 1$ and in the one-dimensional case $d = 1$, with high probability (ignoring logarithmic factors for simplicity),

$$\max_{k=1}^{K} |\widehat{\eta}_k - \eta_k| = O_p(T^{\frac{1}{5}} n^{-\frac{4}{5}} + 1) = \begin{cases} O_p(1), & n \geq T^{1/4}; \\ O_p(T^{\frac{1}{5}} n^{-\frac{4}{5}}), & n \leq T^{1/4}. \end{cases}$$

This relationship between n and T was previously demonstrated in the mean function estimation literature [e.g. 8, 41], where the observations are discretely sampled from independently and identically distributed functional data. It is shown that the minimax estimation error rate also possesses the same phase transition between n and T, i.e. with the transition boundary $n \asymp T^{1/4}$, which agrees with our finding under the change-point setting.

Physical dependence and heavy-tailedness In Assumption 1**c** and **d**, we allow for physical dependence type temporal dependence and heavy-tailed additive noise. As we have discussed, since the physical dependence is in fact a short range dependence, all the rates involved are the same as those in the independence cases, up to logarithmic factors. Having said this, the technical details required in dealing with this short range dependence are fundamentally different from those in the independence cases. From the result, it might be more interesting to discuss the effect of the heavy-tail behaviours, which are characterised by the parameter q. It can be seen from the rates in Assumption 3 and Theorem 1 that the effect of q disappears and it behaves the same as if the noise is sub-Gaussian when $q \geq 10$.

4 Numerical Experiments

4.1 Simulated data analysis

We compare the proposed FSBS with state-of-the-art methods for change-point detection in functional data across a wide range of simulation settings. Specifically, we compare with three competitors: BGHK in [6], HK in [18] and SN in [40]. All three methods estimate change-points via examining mean change in the leading functional principal components of the observed functional data. BGHK is designed for temporally independent data while HK and SN can handle temporal dependence via the estimation of long-run variance and the use of self-normalization principle, respectively. All three methods require fully observed functional data. In practice, they convert discrete data to functional observations by using B-splines with 20 basis functions.

For the implementation of FSBS, we adopt the Gaussian kernel. Following the standard practice in kernel density estimation, the bandwidth h is selected by the function `Hpi` in the R package `ks` ([9]). The tuning parameter τ and the bandwidth h are chosen by cross-validation, with evenly-indexed data being the training set and oddly-indexed data being the validation set. For each pair of candidate (h, τ), we obtain change-point estimators $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ on the training set and compute the validation loss $\sum_{k=1}^{\widehat{K}} \sum_{t\in[\widehat{\eta}_k, \widehat{\eta}_{k+1})} \sum_{i=1}^{n} \{(\widehat{\eta}_{k+1} - \widehat{\eta}_k)^{-1} \sum_{t=\widehat{\eta}_k+1}^{\widehat{\eta}_{k+1}} F_{t,h}(x_{t,i}) - y_{t,i}\}^2$. The pair (h, τ) is then chosen to be the one corresponding to the lowest validation loss.

We consider five different scenarios for the observations $\{x_{ti}, y_{ti}\}_{t=1,i=1}^{T,n}$. For all scenarios 1-5, we set $T = 200$. Given the dimensionality d, denote a generic grid point as $x = (x^{(1)}, \cdots, x^{(d)})$. Scenarios 1 to 4 are generated based on model (1). The basic setting is as follows.

• **Scenario 1 (S1)** Let $(n, d) = (1, 1)$, the unevenly-spaced change-points be $(\eta_1, \eta_2) = (30, 130)$ and the three distinct mean functions be $6\cos(\cdot)$, $6\sin(\cdot)$ and $6\cos(\cdot)$.

• **Scenario 2 (S2)** Let $(n, d) = (10, 1)$, the unevenly-spaced change-points be $(\eta_1, \eta_2) = (30, 130)$ and the three distinct mean functions be $2\cos(\cdot)$, $2\sin(\cdot)$ and $2\cos(\cdot)$.

• **Scenario 3 (S3)** Let $(n, d) = (50, 1)$, the unevenly-spaced change-points be $(\eta_1, \eta_2) = (30, 130)$ and the three distinct mean functions be $\cos(\cdot)$, $\sin(\cdot)$ and $\cos(\cdot)$.

• **Scenario 4 (S4)** Let $(n, d) = (10, 2)$, the unevenly-spaced change-points be $(\eta_1, \eta_2) = (100, 150)$ and the three distinct mean functions be 0, $3x^{(1)}x^{(2)}$ and 0.

For **S1**-**S4**, the functional noise is generated as $\xi_t(x) = 0.5\xi_{t-1}(x) + \sum_{i=1}^{50} i^{-1} b_{t,i} h_i(x)$, where $\{h_i(x) = \prod_{j=1}^{d} (1/\sqrt{2})\pi \sin(ix^{(j)})\}_{i=1}^{50}$ are basis functions and $\{b_{t,i}\}_{t=1,i=1}^{T,50}$ are i.i.d. standard normal random variables. The measurement error is generated as $\delta_t = 0.3\delta_{t-1} + \epsilon_t$, where $\{\epsilon_t\}_{t=1}^{T}$ are i.i.d. $\mathcal{N}(0, 0.5I_n)$. We observe the noisy functional data $\{y_{ti}\}_{t=1,i=1}^{T,n}$ at grid points $\{x_{ti}\}_{t=1,i=1}^{T,n}$ independently sampled from $\mathrm{Unif}([0, 1]^d)$.

Scenario 5 is adopted from [40] for densely-sampled functional data without measurement error.

• **Scenario 5 (S5)** Let $(n, d) = (50, 1)$, the evenly-spaced change-points be $(\eta_1, \eta_2) = (68, 134)$ and the three distinct mean functions be 0, $\sin(\cdot)$ and $2\sin(\cdot)$.

The grid points $\{x_{ti}\}_{i=1}^{50}$ are 50 evenly-spaced points in $[0, 1]$ for all $t = 1, \cdots, T$. The functional noise is generated as $\xi_t(\cdot) = \int_{[0,1]} \psi(\cdot, u)\xi_{t-1}(u)\,\mathrm{d}u + \epsilon_t(\cdot)$, where $\{\epsilon_t(\cdot)\}_{t=1}^{T}$ are independent standard Brownian motions and $\psi(v, u) = 1/3 \exp((v^2 + u^2)/2)$ is a bivariate Gaussian kernel.

S1-S5 represent a wide range of simulation settings including the extreme sparse case **S1**, sparse case **S2**, the two-dimensional domain case **S4**, and the densely sampled cases **S3** and **S5**. Note that **S1** and **S4** can only be handled by FSBS as for **S1** it is impossible to estimate a function via B-spline based on one point and for **S4**, the domain is of dimension 2.

Evaluation result: For a given set of true change-points $\mathcal{C} = \{\eta_k\}_{k=1}^{K}$, we evaluate the accuracy of the estimator $\{\widehat{\eta}_k\}_{k=1}^{\widehat{K}}$ by the difference $|\widehat{K} - K|$ and the Hausdorff distance $d(\hat{\mathcal{C}}, \mathcal{C})$, defined by $d(\hat{\mathcal{C}}, \mathcal{C}) = \max\{\max_{x\in\hat{\mathcal{C}}} \min_{y\in\mathcal{C}}\{|x - y|\}, \max_{y\in\hat{\mathcal{C}}} \min_{x\in\mathcal{C}}\{|x - y|\}\}$. For $\hat{\mathcal{C}} = \varnothing$, we use the convention that $|\widehat{K} - K| = K$ and $d(\hat{\mathcal{C}}, \mathcal{C}) = T$.

For each scenario, we repeat the experiments 100 times and Figure 1 summarizes the performance of FSBS, BGHK, HK and SN. Tabulated results can be found in Appendix A.1. As can be seen, FSBS consistently outperforms the competing methods by a wide margin and demonstrates robust behaviour across the board for both sparsely and densely sampled functional data.

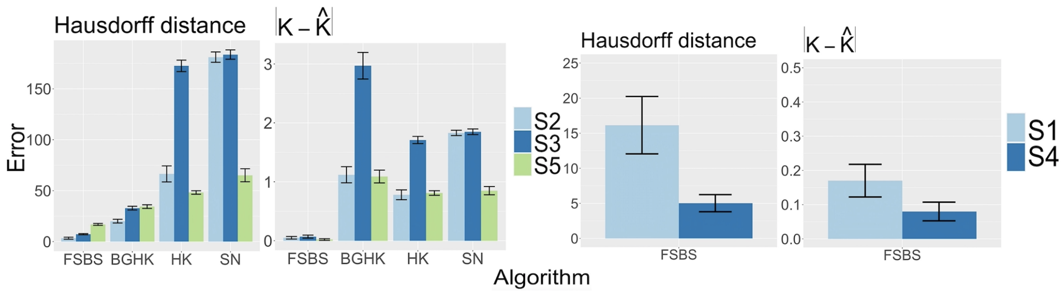

Figure 1: Bar plots for simulation results of **S1-S5**. Each bar reports the mean and standard deviation computed based on 100 experiments. From left to right, the first two plots correspond to the Hausdorff distance and $|K - \hat{K}|$ in **S2**, **S3** and **S5**. The last two plots correspond to **S1** and **S4**.

4.2 Real data application

We consider the COBE-SSTE dataset [29], which consists of monthly average sea surface temperature (SST) from 1940 to 2019, on a 1 degree latitude by 1 degree longitude grid (48×30) covering Australia. The specific coordinates are latitude 10S-39S and longitude 110E-157E.

We apply FSBS to detect potential change-points in the two-dimensional SST. The implementation of FSBS is the same as the one described in Section 4.1. To avoid seasonality, we apply FSBS to the SST for the month of June from 1940 to 2019. We further conduct the same analysis separately for the month of July for robustness check.

For both the June and July data, two change-points are identified by FSBS, Year 1981 and 1996, suggesting the robustness of the finding. The two change-points might be associated with years when both the Indian Ocean Dipole and Oceanic Niño Index had extreme events [2]. The El Niño/Southern Oscillation has been recognized as an important manifestation of the tropical ocean-atmosphere-land coupled system. It is an irregular periodic variation in winds and sea surface temperatures over the tropical eastern Pacific Ocean. Much of the variability in the climate of Australia is connected with this phenomenon [5].

To visualize the estimated change, Figure 2 depicts the average SST before the first change-point Year 1981, between the two change-points, and after the second change-point Year 1996. The two rows correspond to the June and July data, respectively. As we can see, the top left corners exhibit different patterns in the three periods, suggesting the existence of change-points.

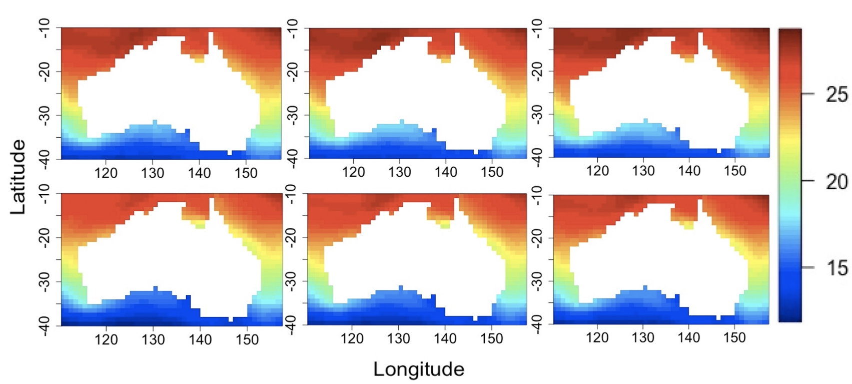

Figure 2: Average SST. From left to right: average SST from 1940 to 1981, average SST from 1982 to 1996, and average SST from 1997 to 2019. The top and bottom rows correspond to the June and July data respectively.

5 Conclusion

In this paper, we study change-point detection for sparse and dense functional data in general dimensions. We show that our algorithm FSBS can consistently estimate the change-points even in the extreme sparse setting with $n = 1$. Our theoretical analysis reveals an interesting phase transition between n and T, which has not been discovered in the existing literature for functional change-point detection. The consistency of FSBS relies on the assumption that the minimal spacing $\Delta \asymp T$. To relax this assumption, we may consider increasing $\mathcal{K}$ in Definition 1 to enlarge the coverage of the seeded intervals in FSBS and apply the narrowest over threshold selection method [Theorem 3 in 22]. With minor modifications of the current theoretical analysis, the consistency of FSBS can be established for the case of $\Delta \ll T$. Since such a relaxation does not add much more methodological insights to our paper, we omit this additional technical discussion for conciseness.

Acknowledgments and Disclosure of Funding

YY's research is partially funded by EPSRC EP/V013432/1.

References

[1] Arlot, S., Celisse, A. and Harchaoui, Z. [2019], 'A kernel multiple change-point algorithm via model selection', *Journal of machine learning research* **20**(162).

[2] Ashok, K., Guan, Z. and Yamagat, T. [2003], 'A look at the relationship between the enso and the indian ocean dipole', *Journal of the Meteorological Society of Japan. Ser. II* **81**(1), 41–56.

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

[4] Aue, A., Rice, G. and Sönmez, O. [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.

[5] Australian Government [2014], 'What is El Niño and how does it impact Australia?', `http://www.bom.gov.au/climate/updates/articles/a008-el-nino-and-australia.shtml`.

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

[7] Bonner, S. J., Newlands, N. K. and Heckman, N. E. [2014], 'Modeling regional impacts of climate teleconnections using functional data analysis', *Environmental and Ecological Statistics* **21**(1), 1–26.

[8] Cai, T. T. and Yuan, M. [2011], 'Optimal estimation of the mean function based on discretely sampled functional data: Phase transition', *The Annals of Statistics* **39**(5), 2330–2355.

[9] Chacón, J. E. and Duong, T. [2018], *Multivariate kernel smoothing and its applications*, Chapman and Hall/CRC.

[10] Chiou, J.-M., Zhang, Y.-C., Chen, W.-H. and Chang, C.-W. [2014], 'A functional data approach to missing value imputation and outlier detection for traffic flow data', *Transportmetrica B: Transport Dynamics* **2**(2), 106–129.

[11] Dai, X., Müller, H.-G., Wang, J.-L. and Deoni, S. C. [2019], 'Age-dynamic networks and functional correlation for early white matter myelination', *Brain Structure and Function* **224**(2), 535–551.

[12] Dette, H., Kokot, K. and Volgushev, S. [2020], 'Testing relevant hypotheses in functional time series via self-normalization', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82**(3), 629–660.

[13] Fan, Y., Foutz, N., James, G. M. and Jank, W. [2014], 'Functional response additive model estimation with online virtual stock markets', *The Annals of Applied Statistics* **8**(4), 2435–2460.

[14] Fraiman, R., Justel, A., Liu, R. and Llop, P. [2014], 'Detecting trends in time series of functional data: A study of antarctic climate change', *Canadian Journal of Statistics* **42**(4), 597–609.

[15] Giné, E. and Guillou, A. [1999], 'Laws of the iterated logarithm for censored data', *The Annals of Probability* **27**(4), 2042–2067.

[16] Giné, E. and Guillou, A. [2001], 'On consistency of kernel density estimators for randomly censored data: rates holding uniformly over adaptive intervals', *Annales de l'IHP Probabilités et statistiques* **37**(4), 503–522.

[17] Giné, E. and Guillou, A. [2002], 'Rates of strong uniform consistency for multivariate kernel density estimators', *Annales de l'Institut Henri Poincare (B) Probability and Statistics* **38**(6), 907–921.

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

[19] Jiang, H. [2017], Uniform convergence rates for kernel density estimation, *in* 'International Conference on Machine Learning', PMLR, pp. 1694–1703.

[20] Kim, J., Shin, J., Rinaldo, A. and Wasserman, L. [2019], Uniform convergence rate of the kernel density estimator adaptive to intrinsic volume dimension, *in* 'International Conference on Machine Learning', PMLR, pp. 3398–3407.

[21] Kirch, C. [2006], Resampling methods for the change analysis of dependent data, PhD thesis, Universität zu Köln.

[22] Kovács, S., Li, H., Bühlmann, P. and Munk, A. [2020], 'Seeded binary segmentation: A general methodology for fast and optimal change point detection', *arXiv preprint arXiv:2002.06633* .

[23] Li, S., Xie, Y., Dai, H. and Song, L. [2015], 'M-statistic for kernel change-point detection', *Advances in Neural Information Processing Systems* **28**.

[24] Li, S., Xie, Y., Dai, H. and Song, L. [2019], 'Scan b-statistic for kernel change-point detection', *Sequential Analysis* **38**(4), 503–544.

[25] Liu, W., Xiao, H. and Wu, W. B. [2013], 'Probability and moment inequalities under dependence', *Statistica Sinica* pp. 1257–1272.

[26] National Oceanic and Atmospheric Administration [n.d.], 'ENSO effects across the northeastern Caribbean', `https://www.weather.gov/sju/climo_enso`.

[27] Padilla, O. H. M., Yu, Y., Wang, D. and Rinaldo, A. [2021], 'Optimal nonparametric multivariate change point detection and localization', *IEEE Transactions on Information Theory* .

[28] Petersen, A., Deoni, S. and Müller, H.-G. [2019], 'Fréchet estimation of time-varying covariance matrices from sparse data, with application to the regional co-evolution of myelination in the developing brain', *The Annals of Applied Statistics* **13**(1), 393–419.

[29] Physical Sciences Laboratory [2020], 'COBE SST2 and Sea-Ice', `https://psl.noaa.gov/data/gridded/data.cobe2.html`.

[30] Rigollet, P. and Vert, R. [2009], 'Optimal rates for plug-in estimators of density level sets', *Bernoulli* **15**(4), 1154–1178.

[31] Rinaldo, A. and Wasserman, L. [2010], 'Generalized density clustering', *The Annals of Statistics* **38**(5), 2678–2722.

[32] Sriperumbudur, B. and Steinwart, I. [2012], Consistency and rates for clustering with dbscan, *in* 'Artificial Intelligence and Statistics', PMLR, pp. 1090–1098.

[33] Tsybakov, A. B. [2009], *Introduction to Nonparametric Estimation*, Springer series in statistics, Springer, Dordrecht.

[34] Vershynin, R. [2018], *High-dimensional probability: An introduction with applications in data science*, Vol. 47, Cambridge university press.

[35] Wang, D., Lu, X. and Rinaldo, A. [2019], 'Dbscan: Optimal rates for density-based cluster estimation', *Journal of Machine Learning Research* .

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

[37] Wang, J.-L., Chiou, J.-M. and Müller, H.-G. [2016], 'Functional data analysis', *Annual Review of Statistics and Its Application* **3**, 257–295.

[38] Wikipedia, 2004 Atlantic hurricane season [2022], '2004 Atlantic hurricane season', `https://en.wikipedia.org/wiki/2004_Atlantic_hurricane_season#:~:text=The`.

[39] Wu, W. B. [2005], 'Nonlinear system theory: Another look at dependence', *Proceedings of the National Academy of Sciences* **102**(40), 14150–14154.

[40] Zhang, X., Shao, X., Hayhoe, K. and Wuebbles, D. J. [2011], 'Testing the structural stability of temporally dependent functional observations and application to climate projections', *Electronic Journal of Statistics* **5**, 1765–1796.

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

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] See Sections 3 and 4.
   (b) Did you describe the limitations of your work? [Yes] See Section 5.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Section 3.1.
   (b) Did you include complete proofs of all theoretical results? [Yes] See Appendix B.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] We have included detailed instructions needed to reproduce the numerical results in Section 4. We also submit our code to the supplementary materials.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 4.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] See Figure 1.
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] The computation is sufficiently fast and computational cost is not a concern.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Supplementary Materials

Additional numerical results and all technical details are included in the supplementary materials.

A Additional numerical results

A.1 Detailed simulation results

We present the tables containing the results of the simulation study in Section 4.1 of the main text. On each table, the mean over 100 repetitions is reported, and the numbers in parenthesis denote the standard errors. For the purpose of identifying underestimation and overestimation, we also include the proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive.

Table 1: **Scenario 1** ($n = 1, d = 1$ changes from $6\cos$-$6\sin$-$6\cos$)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.05	0.86	0.09	0.17 (0.05)	16.15 (4.09)

Changes occur at the times 30 and 130.

Table 2: **Scenario 2** ($n = 10, d = 1$, changes from $2\cos$-$2\sin$-$2\cos$)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.05	0.95	0	0.05 (0.02)	3.32 (1)
BGHK	0.58	0.42	0	1.12 (0.14)	20.11 (1.82)
HK	0.16	0.47	0.37	0.78 (0.08)	66.45 (7.87)
SN	0.04	0.03	0.93	1.83 (0.04)	181.11 (5.05)

Changes occur at the times 30 and 130.

Table 3: **Scenario 3** ($n = 50, d = 1$, changes from cos-sin-cos)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0	0.93	0.07	0.07 (0.03)	7.35 (0.54)
BGHK	0.85	0.15	0	2.97 (0.22)	32.88 (1.82)
HK	0	0.08	0.92	1.71 (0.06)	172.52 (5.61)
SN	0.02	0.04	0.94	1.85 (0.05)	183.63 (4.57)

Changes occur at the times 30 and 130.

Table 4: **Scenario 4** ($n = 10, d = 2$, changes from 0-$3x^{(1)}x^{(2)}$-0)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0	0.92	0.08	0.08 (0.021)	5.02 (1.25)

Changes occur at the times 100 and 150.

Table 5: **Scenario 5** ($n = 50, d = 1$, changes from 0-sin-2 sin)

Model	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
FSBS	0.02	0.98	0	0.02 (0.01)	16.9 (0.93)
BGHK	0.48	0.30	0.22	1.09 (0.11)	34.36 (1.78)
HK	0	0.19	0.81	0.81 (0.04)	48.24 (1.71)
SN	0.08	0.33	0.59	0.85 (0.07)	65.15 (6.38)

Changes occur at the times 68 and 134.

A.2 Details of Figure 2

We zoom in the top-left and top-right corners of each panel in Figure 2 and present in Figures A.2 and A.2, respectively. The top-left and top-right corners correspond to the northwest and northeast coasts of Australia, where the changes occur.

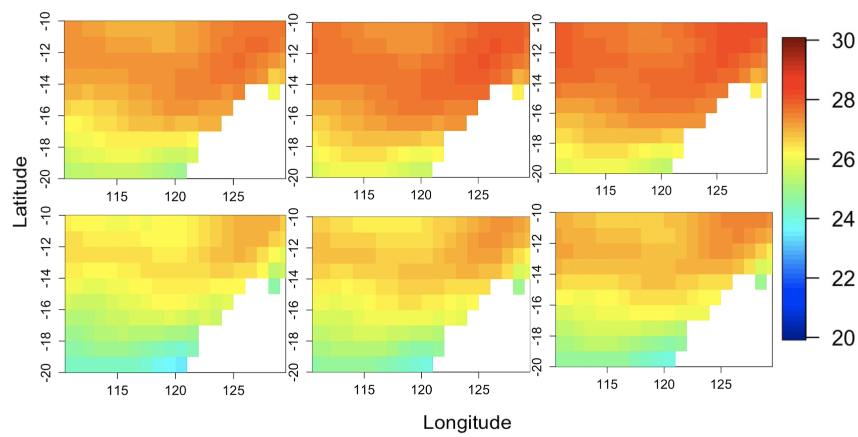

Figure 3: Average SST of northwest coast of Australia. From left to right: average SST from 1940 to 1981, average SST from 1982 to 1996, and average SST from 1997 to 2019. The top and bottom rows correspond to the June and July data respectively.

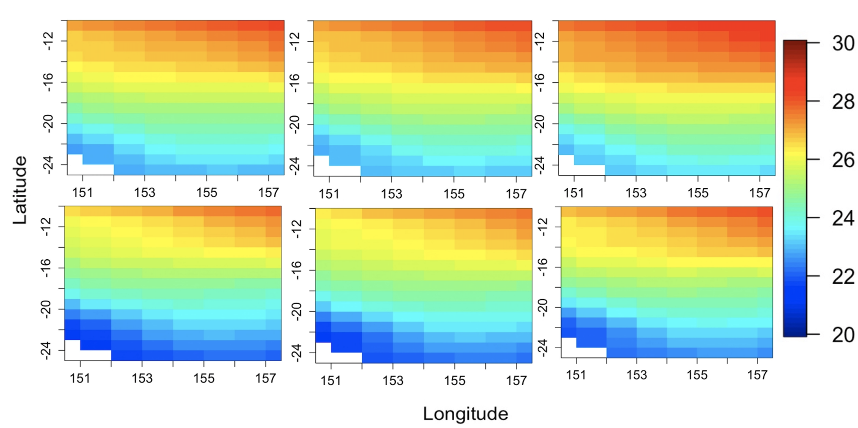

Figure 4: Average SST of northeast coast of Australia. From left to right: average SST from 1940 to 1981, average SST from 1982 to 1996, and average SST from 1997 to 2019. The top and bottom rows correspond to the June and July data respectively.

A.3 Sea surface temperature on Caribbean sea

We consider an additional real data example, also from the COBE-SSTE dataset [29], using data from June and July. FSBS is applied to estimate potential change points on a 1 degree latitude by 1 degree longitude grid (10×6), located at the Caribbean sea. In both months, FSBS identified the year 2004 as a change-point. This might be associated with the development of a Modoki El Niño – a rare type of El Niño in which unfavourable conditions are produced over the eastern Pacific instead of the Atlantic basin due to warmer sea surface temperatures farther west along the equatorial Pacific [38]. Variability in the climate of northeastern Caribbean is connected with this phenomenon, see for example [26].

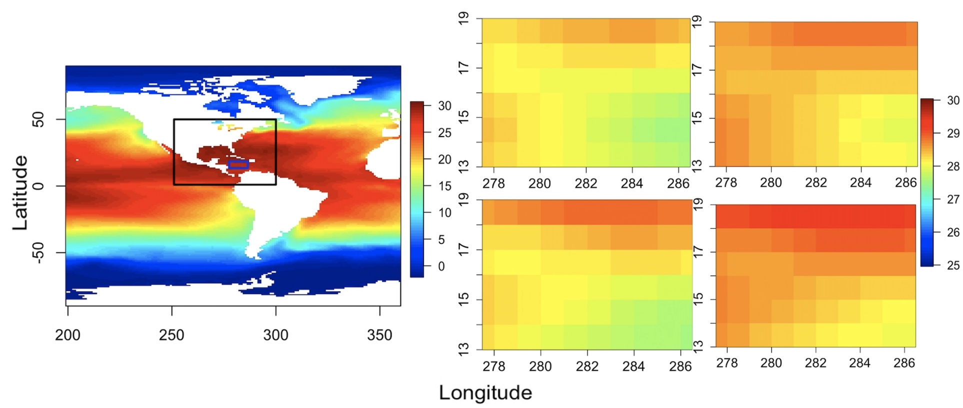

Figure 5: Average SST of Caribbean sea. From left to right: The first image shows the region chosen, the small blue rectangle into the black rectangle. The second image contains four different sub-images. Here, from left to right, the average SST from 1940 to 2003 and average SST from 2004 to 2019 is presented. The top and bottom rows correspond to the June and July data respectively.

A.4 On the dimension d

Recall that the localisation error rate of change-point estimation in Theorem 1 is

$$C_{\mathrm{FSBS}} \log^{\max\{1,10/q\}}(T) \left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right) \kappa_k^{-2},$$

which is an increasing function of d, i.e. a larger d will lead to a worse localization error rate.

In addition, Assumption 3 requires that the signal-to-noise ratio to be lower bounded by

$$C_{\mathrm{SNR}} \log^{\max\{1/2,5/q\}}(T) \left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)^{1/2},$$

which implies that a larger d will also require a stronger signal.

We conducted additional numerical results to further show the influence of d. Using the same setting as that in Scenario 4 in Section 4, we vary the dimension $d \in \{2, 3, 5, 10\}$. Results are collected in Table 6 and Appendix A.4, supporting our theoretical findings.

Table 6: **FSBS on Scenario 4** ($n = 10$, changes from 0-3$x^{(1)}x^{(2)}$-0)

Dimension	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
$d = 2$	0	0.92	0.08	0.08 (0.02)	5.02 (1.25)
$d = 3$	0.02	0.89	0.09	0.11 (0.03)	5.73 (1.22)
$d = 5$	0.18	0.82	0	0.18 (0.05)	5.92 (1.23)
$d = 10$	0.21	0.79	0	0.22 (0.08)	6.58 (1.24)

Performance of FSBS with different choices of dimension d is studied for **S4**. The mean over 100 repetitions is reported, and the numbers in parenthesis denote standard errors. It includes the proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive.

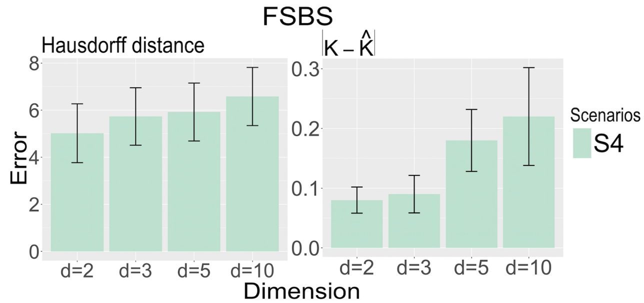

Figure 6: Bar plots for simulation results of FSBS performance on **S4** with respect to the dimension d. Each bar reports the mean and standard error computed based on 100 experiments.

A.5 Choice of kernels

The choice of kernels may affect the performance of kernel based methods. We choose Gaussian kernel in Section 4 and demonstrate the robustness against the choice of kernels of FSBS in this section, by choosing different kernels. Tables 7, 8 and Appendix A.5 collect results of the performance of the FSBS with different choices of kernels, based on the settings detailed in Scenarios 1 and 2 in Section 4, with Gaussian, Uniform, Epanechnikov and Quartic kernels.

Table 7: **FSBS in Scenario 1 (different kernels comparison)**

Kernel	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
Gaussian	0.05	0.86	0.09	0.17 (0.05)	16.15 (4.09)
Uniform	0.01	0.99	0	0.01 (0.01)	13.32 (0.42)
Epanechnikov	0.06	0.87	0.17	0.13 (0.03)	15.14 (2.40)
Quartic	0.07	0.84	0.09	0.20 (0.04)	18.28 (1.63)

The mean over 100 repetitions is reported together with the standard errors into parenthesis. The proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive are included.

Table 8: **FSBS in Scenario 2 (different kernels comparison)**

Kernel	$K - \hat{K} < 0$	$K - \hat{K} = 0$	$K - \hat{K} > 0$	$\lvert\hat{K} - K\rvert$	$d(\hat{\mathcal{C}}, \mathcal{C})$
Gaussian	0.05	0.95	0	0.05 (0.02)	3.32 (1)
Uniform	0	0.99	0.01	0.01 (0.01)	2.93 (1.03)
Epanechnikov	0	100	0	0 (0)	1.24 (0.28)
Quartic	0.01	0.99	0	0.01 (0.01)	2.3 (0.55)

The mean over 100 repetitions is reported together with the standard errors into parenthesis. The proportions of estimations for which the $\hat{K} - K$ distance is negative, zero, or positive are included.

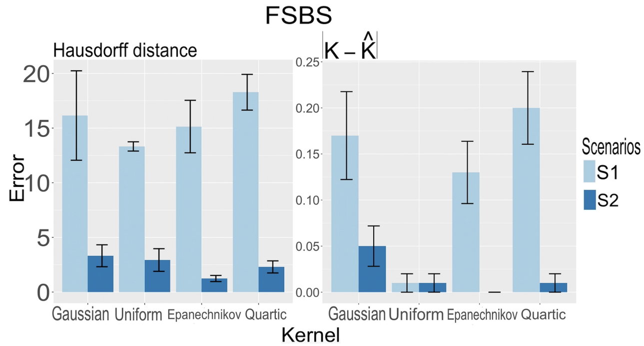

Figure 7: Bar plots for simulation results of FSBS performance on **S1** and **S2** with respect to different choices of kernels. Each bar reports the mean and standard error computed based on 100 experiments.

A.6 Computational costs

Our method is computationally efficient and its computational complexity is $O(nT\log T + T(\log T)^2)$. Specifically, as can be seen from Algorithm 1, we need to conduct kernel smoothing of the sampling distribution and mean function at $\log T$ measurement locations, which costs $O(nT\log T)$ operations. Once this is done, we conduct seeded binary segmentation (SBS) at the $\log T$ measurement locations/grids. It is known that SBS has a computational cost of $O(T\log T)$. Thus, this step costs $O(T(\log T)^2)$ computational complexity. In total, the computational complexity of our method is $O(nT\log T + T(\log T)^2)$.

As for existing methods in the literature, in terms of implementation, they all rely on the two-stage procedure. Specifically, the first stage is to register/estimate the discretely observed points into a functional curve on each time t. Taking the B-spline smoothing with p basis functions for example, this costs $O(n^2p + p^3)$ computational complexity for each time t due to a least square estimation. Thus this step costs $O(T(n^2p + p^3))$ computational complexity. Once the functional curves are registered, in the second stage, the existing methods conduct functional PCA to extract p' principle component scores from each function and then conduct mean change-point detection on the p'-dimension time series of principle component scores. Ignoring the computational cost of functional PCA, the change-point detection procedure costs at least $O(T\log T)$ computational complexity if a standard binary segmentation is used and could be more expensive if other segmentation algorithms are used to conduct change-point estimation. Thus, in total, the computational complexity of existing methods is at least $O(T(n^2p + p^3) + T\log T)$, which is more expensive unless $n \preceq \log T$.

B Proof of Theorem 1

In this section, we present the proofs of theorem Theorem 1. To this end, we will invoke the following well-known l_∞ bounds for kernel density estimation.

Lemma 1. *Let $\{x_{t,i}\}_{i=1,t=1}^{n,T}$ be random grid points independently sampled from a common density function $u : [0,1]^d \to \mathbb{R}$. Under Assumption 2-**b**, the density estimator of the sampling distribution μ,*

$$\hat{p}(x) = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n_t} K_{\bar{h}}(x - x_{i,t}), \quad x \in [0,1]^d,$$

satisfies,

$$||\hat{p} - \mathbb{E}(\hat{p})||_\infty \leq C\sqrt{\frac{\log(nT) + \log(1/\bar{h})}{nT\bar{h}^d}} \tag{4}$$

*with probability at least $1 - \frac{1}{nT}$. Moreover, under Assumption 2-**a**, the bias term satisfies*

$$||\mathbb{E}(\hat{p}) - u||_\infty \leq C_2\bar{h}^r. \tag{5}$$

Therefore,

$$||\hat{p} - u||_\infty = O\Big(\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \tag{6}$$

with probability at least $1 - \frac{1}{nT}$.

The verification of these bounds can be found in many places in the literature. For equation (4) see for example [17], [31], [32] and [19]. For equation (5), [33] is a common reference.

Proof of Theorem 1. For any $(s,e] \subseteq (0,T]$, let

$$\widetilde{f}_t^{(s,e]}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}} \sum_{l=s+1}^{t} f_l^*(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}} \sum_{l=t+1}^{e} f_l^*(x), \ x \in [0,1]^d.$$

For any $\tilde{r} \in (\rho, T-\rho]$ and $x \in [0,1]$, we consider

$$\mathcal{A}_x((s,e],\rho,\lambda) = \Big\{ \max_{t=s+\rho+1}^{e-\rho} |\widetilde{F}_{t,h}^{s,e}(x) - \widetilde{f}_t^{s,e}(x)| \le \lambda \Big\};$$

$$\mathcal{B}_x(\tilde{r},\rho,\lambda) = \Big\{ \max_{N=\rho}^{T-\tilde{r}} \Big| \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}+1}^{\tilde{r}+N} F_{t,h}(x) - \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}+1}^{\tilde{r}+N} f_t(x) \Big| \le \lambda \Big\} \bigcup$$
$$\Big\{ \max_{N=\rho}^{\tilde{r}} \Big| \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}-N+1}^{\tilde{r}} F_{t,h}(x) - \frac{1}{\sqrt{N}} \sum_{t=\tilde{r}-N+1}^{\tilde{r}} f_t(x) \Big| \le \lambda \Big\}.$$

From Algorithm 1, we have that

$$\rho = \frac{\log(T)}{nh^d}.$$

We observe that, $\rho nh^d = \log(T)$ and for $T \ge 3$, we have that

$$\rho^{1/2-1/q} \ge (nh^d)^{1/2-(q-1)/q}.$$

Therefore, Proposition 1 and Corollary 1 imply that with

$$\lambda = C_\lambda \Big(\log^{5/q}(T) \sqrt{\frac{1}{nh^d} + 1} + \sqrt{\frac{\log(T)}{nh^d}} + \sqrt{T} h^r + \sqrt{T} \Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}} \Big), \tag{7}$$

for some diverging sequence C_λ, it holds that

$$P\Big\{ \mathcal{A}_x^c((s,e],\rho,\lambda) \Big\} \le 4C_1 \frac{\log(T)}{(\log^{5/q}(T))^q} + \frac{2}{T^5} + \frac{10}{Tn}$$

and

$$P\Big\{ \mathcal{B}_x^c(r,\rho,\lambda) \Big\} \le 2C_1 \frac{\log(T)}{(\log^{5/q}(T))^q} + \frac{1}{T^5} + \frac{5}{Tn}.$$

Then, using that $\log^4(T) = O(T)$, from above

$$P\Big\{ \mathcal{A}_x^c((s,e],\rho,\lambda) \Big\} = O(\log^{-4}(T)) \quad \text{and} \quad P\Big\{ \mathcal{B}_x^c(r,\rho,\lambda) \Big\} = O(\log^{-4}(T)).$$

Now, we notice that,

$$\sum_{k=1}^{\mathcal{K}} \tilde{n}_k = \sum_{k=1}^{\mathcal{K}} (2^k - 1) \le \sum_{k=1}^{\mathcal{K}} 2^k \le 2(2^{\lceil \log(2) C_{\mathcal{K}} (\log(\log(T)))/\log 2 \rceil} - 1)$$
$$\le 4(2^{(\log(\log(T)))/\log 2})^{\log(2) C_{\mathcal{K}}} = O(\log^{\log(2) C_{\mathcal{K}}}((T))).$$

In addition, there are $K = O(1)$ number of change-points. In consequence, it follows that

$$P\left\{\mathcal{A}_u(\mathcal{I},\rho,\lambda) \text{ for all } \mathcal{I} \in \mathcal{J} \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\right\} \geq 1 - \frac{1}{\log^2(T)}, \tag{8}$$

$$P\left\{\mathcal{B}_u(s,\rho,\lambda) \cup \mathcal{B}_u(e,\rho,\lambda) \text{ for all } (s,e] = \mathcal{I} \in \mathcal{J} \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\right\} \geq 1 - \frac{1}{\log(T)}, \tag{9}$$

$$P\left\{\mathcal{B}_u(\eta_k,\rho,\lambda) \text{ for all } 1 \leq k \leq K \text{ and all } u \in \{u_m\}_{m=1}^{\log(T)}\right\} \geq 1 - \frac{1}{\log^3(T)}. \tag{10}$$

The rest of the argument is made by assuming the events in equations (8), (9) and (10) hold.

Denote

$$\Upsilon_k = C\log^{\max\{1,10/q\}}(T)\left(1+T^{\frac{d}{2r+d}}n^{\frac{-2r}{2r+d}}\right)\kappa_k^{-2} \quad \text{and} \quad \Upsilon_{\max} = C\log^{\max\{1,10/q\}}(T)\left(1+T^{\frac{d}{2r+d}}n^{\frac{-2r}{2r+d}}\right)\kappa^{-2},$$

where $\kappa = \min\{\kappa_1,\ldots,\kappa_K\}$. Since Υ_k is the desired localisation rate, by induction, it suffices to consider any generic interval $(s,e] \subseteq (0,T]$ that satisfies the following three conditions:

$$\begin{aligned}
&\eta_{m-1} \leq s \leq \eta_m \leq \ldots \leq \eta_{m+q} \leq e \leq \eta_{m+q+1}, \quad q \geq -1;\\
&\text{either } \eta_m - s \leq \Upsilon_m \quad \text{or} \quad s - \eta_{m-1} \leq \Upsilon_{m-1};\\
&\text{either } \eta_{m+q+1} - e \leq \Upsilon_{m+q+1} \quad \text{or} \quad e - \eta_{m+q} \leq \Upsilon_{m+q}.
\end{aligned}$$

Here $q = -1$ indicates that there is no change-point contained in $(s,e]$.

Denote

$$\Delta_k = \eta_{k-1} - \eta_k \text{ for } k = 1,\ldots,K+1 \quad \text{and} \quad \Delta = \min\{\Delta_1,\ldots,\Delta_{K+1}\}.$$

Observe that since $\kappa_k > 0$ for all $1 \leq k \leq K$ and that $\Delta_k = \Theta(T)$, it holds that $\Upsilon_{\max} = o(\Delta)$. Therefore, it has to be the case that for any true change-point $\eta_m \in (0,T]$, either $|\eta_m - s| \leq \Upsilon_m$ or $|\eta_m - s| \geq \Delta - \Upsilon_{\max} \geq \Theta(T)$. This means that $\min\{|\eta_m - e|, |\eta_m - s|\} \leq \Upsilon_m$ indicates that η_m is a detected change-point in the previous induction step, even if $\eta_m \in (s,e]$. We refer to $\eta_m \in (s,e]$ as an undetected change-point if $\min\{\eta_m - s, \eta_m - e\} = \Theta(T)$. To complete the induction step, it suffices to show that FSBS $((s,e],h,\tau)$
(i) will not detect any new change point in $(s,e]$ if all the change-points in that interval have been previously detected, and
(ii) will find a point $D_{m*}^{\mathcal{I}^*}$ in $(s,e]$ such that $|\eta_m - D_{m*}^{\mathcal{I}^*}| \leq \Upsilon_m$ if there exists at least one undetected change-point in $(s,e]$.

In order to accomplish this, we need the following series of steps.

Step 1. We first observe that if $\eta_k \in \{\eta_k\}_{k=1}^K$ is any change-point in the functional time series, by Lemma 8, there exists a seeded interval $\mathcal{I}_k = (s_k, e_k]$ containing exactly one change-point η_k such that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \geq \frac{1}{16}\zeta_k, \quad \text{and} \quad \max\{\eta_k - s_k, e_k - \eta_k\} \leq \zeta_k$$

where,

$$\zeta_k = \frac{9}{10}\min\{\eta_{k+1} - \eta_k, \eta_k - \eta_{k-1}\}.$$

Even more, we notice that if $\eta_k \in (s,e]$ is any undetected change-point in $(s,e]$. Then it must hold that

$$s - \eta_{k-1} \leq \Upsilon_{\max}.$$

Since $\Upsilon_{\max} = O(\log^{\max\{1,10/q\}}(T)T^{\frac{d}{2r+d}})$ and $O(\log^a(T)) = o(T^b)$ for any positive numbers a and b, we have that $\Upsilon_{\max} = o(T)$. Moreover, $\eta_k - s_k \leq \zeta_k \leq \frac{9}{10}(\eta_k - \eta_{k-1})$, so that it holds that

$$s_k - \eta_{k-1} \geq \frac{1}{10}(\eta_k - \eta_{k-1}) > \Upsilon_{\max} \geq s - \eta_{k-1}$$

and in consequence $s_k \geq s$. Similarly $e_k \leq e$. Therefore

$$\mathcal{I}_k = (s_k, e_k] \subseteq (s, e].$$

Step 2. Consider the collection of intervals $\{\mathcal{I}_k = (s_k, e_k]\}_{k=1}^K$ in **Step 1.** In this step, it is shown that for each $k \in \{1, \ldots, K\}$, it holds that

$$\max_{t=s_k+\rho}^{t=e_k-\rho} \max_{m=1}^{m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \geq c_1 \sqrt{T} \kappa_k, \tag{11}$$

for some sufficient small constant c_1.

Let $k \in \{1, \ldots, K\}$. By **Step 1**, $\mathcal{I}_k$ contains exactly one change-point η_k. Since for every u_m, $f_t^*(u_m)$ is a one dimensional population time series and there is only one change-point in $\mathcal{I}_k = (s_k, e_k]$, it holds that

$$f_{s_k+1}^*(u_m) = \ldots = f_{\eta_k}^*(u_m) \neq f_{\eta_k+1}^*(u_m) = \ldots = f_{e_k}^*(u_m)$$

which implies, for $s_k < t < \eta_k$

$$\begin{aligned}
\widetilde{f}_t^{(s_k,e_k]}(u_m) =& \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}} \sum_{l=s_k+1}^{t} f_{\eta_k}^*(u_m) - \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} \sum_{l=t+1}^{\eta_k} f_{\eta_k}^*(u_m) \\
&- \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} \sum_{l=\eta_k+1}^{e_k} f_{\eta_k+1}^*(u_m) \\
=& (t - s_k) \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& \sqrt{\frac{(t - s_k)(e_k - t)}{(e_k - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - t) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} f_{\eta_k}^*(u_m) - (\eta_k - t) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k}^*(u_m) \\
&- (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} f_{\eta_k}^*(u_m) - (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - s_k)(e_k - t)}} f_{\eta_k+1}^*(u_m) \\
=& (e_k - \eta_k) \sqrt{\frac{t - s_k}{(e_k - t)(e_k - s_k)}} (f_{\eta_k}^*(u_m) - f_{\eta_k+1}^*(u_m)).
\end{aligned}$$

Similarly, for $\eta_k \leq t \leq e_k$

$$f_t^{(s_k,e_k]}(u_m) = \sqrt{\frac{e_k - t}{(e_k - s_k)(t - s_k)}} (\eta_k - s_k)(f_{\eta_k}^*(u_m) - f_{\eta_k+1}^*(u_m)).$$

Therefore,

$$\widetilde{f}_t^{(s_k,e_k]}(u_m) = \begin{cases} \sqrt{\frac{t-s_k}{(e_k-s_k)(e_k-t)}} (e_k - \eta_k)(f_{\eta_k}^*(u_m) - f_{\eta_k+1}^*(u_m)), & s_k < t < \eta_k; \\ \sqrt{\frac{e_k-t}{(e_k-s_k)(t-s_k)}} (\eta_k - s_k)(f_{\eta_k}^*(u_m) - f_{\eta_k+1}^*(u_m)), & \eta_k \leq t \leq e_k. \end{cases} \tag{12}$$

By Lemma 7, with probability at least $1 - o(1)$, there exists $u_{\tilde{k}} \in \{u_m\}_{m=1}^{\log(T)}$ such that

$$|f^*_{\eta_k}(u_{\tilde{k}}) - f^*_{\eta_k+1}(u_{\tilde{k}})| \geq \frac{3}{4}\kappa_k.$$

Since $\Delta = \Theta(T)$, $\rho = O(\log(T)T^{\frac{d}{2r+d}})$ and $\log^a(T) = o(T^b)$ for any positive numbers a and b, we have that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \geq \frac{1}{16}\zeta_k \geq c_2 T > \rho, \tag{13}$$

so that $\eta_k \in [s_k + \rho, e_k - \rho]$. Then, from (12), (13) and the fact that $|e_k - s_k| < T$ and $|\eta_k - s_k| < T$,

$$|\widetilde{f}^{(s_k,e_k]}_{\eta_k}(u_{\tilde{k}})| = \sqrt{\frac{e_k - \eta_k}{(e_k - s_k)(\eta_k - s_k)}}(\eta_k - s_k)|f^*_{\eta_k}(u_{\tilde{k}}) - f^*_{\eta_k+1}(u_{\tilde{k}})| \geq c_2\sqrt{T}\frac{3}{4}\kappa_k. \tag{14}$$

Therefore, it holds that

$$\begin{aligned}\max_{t=s_k+\rho}^{t=e_k-\rho}\max_{m=1}^{m=\log(T)}|\widetilde{F}^{(s_k,e_k]}_{t,h}(u_m)| \geq& |\widetilde{F}^{(s_k,e_k]}_{\eta_k,h}(u_{\tilde{k}})| \\ \geq& |\widetilde{f}^{(s_k,e_k]}_{\eta_k}(u_{\tilde{k}})| - \lambda \\ \geq& c_2\frac{3}{4}\sqrt{T}\kappa_k - \lambda,\end{aligned}$$

where the first inequality follows from the fact that $\eta_k \in [s_k + \rho, e_k - \rho]$, the second inequality follows from the good event in (8), and the last inequality follows from (14).
Next, we observe that $\log^{\frac{5}{q}}(T)\sqrt{\frac{1}{nh^d} + 1} = o(\sqrt{T^{\frac{2r+d}{d}}})O(\sqrt{T^{\frac{d}{2r+d}}}) = o(\sqrt{T})$, $\rho < c_2T$, $h^r = o(1)$ and $\left(\frac{\log nT}{nT}\right)^{\frac{2r}{2r+d}} = o(1)$. In consequence, since κ_k is a positive constant, by the upper bound of λ on Equation (7), for sufficiently large T, it holds that

$$\frac{c_2}{4}\sqrt{T}\kappa_k \geq \lambda.$$

Therefore,

$$\max_{t=s_k+\rho}^{t=e_k-\rho}\max_{m=1}^{m=\log(T)}|\widetilde{F}^{(s_k,e_k]}_{t,h}(u_m)| \geq \frac{c_2}{2}\sqrt{T}\kappa_k.$$

Therefore Equation (11) holds with $c_1 = \frac{c_2}{2}$.

Step 3. In this step, it is shown that FSBS$((s,e], h, \tau)$ can consistently detect or reject the existence of undetected change-points within $(s,e]$.

Suppose $\eta_k \in (s,e]$ is any undetected change-point. Then by the second half of **Step 1**, $\mathcal{I}_k \subseteq (s,e]$. Therefore

$$A^{\mathcal{I}^*}_{m^*} \geq \max_{t=s_k+\rho}^{t=e_k-\rho}\max_{m=1}^{m=\log(T)}|\widetilde{F}^{(s_k,e_k]}_{t,h}(u_m)| \geq c_1\sqrt{T}\kappa_k > \tau,$$

where the second inequality follows from Equation (11), and the last inequality follows from the fact that, $\log^a(T) = o(T^b)$ for any positive numbers a and b implies $\tau = C_\tau\left(\log^{\max\{1,10/q\}}(T)\sqrt{\frac{1}{nh^d} + 1}\right) = o(\sqrt{T})$.

Suppose there does not exist any undetected change-point in $(s,e]$. Then for any $\mathcal{I} = (\alpha,\beta] \subseteq (s,e]$, one of the following situations must hold,

(a) There is no change-point within $(\alpha,\beta]$;

(b) there exists only one change-point η_k within $(\alpha,\beta]$ and $\min\{\eta_k - \alpha, \beta - \eta_k\} \leq \Upsilon_k$;

(c) there exist two change-points η_k, η_{k+1} within $(\alpha,\beta]$ and

$$\eta_k - \alpha \leq \Upsilon_k \quad \text{and} \quad \beta - \eta_{k+1} \leq \Upsilon_{k+1}.$$

The calculations of (c) are provided as the other two cases are similar and simpler. Note that for any $x \in [0,1]^d$, it holds that

$$|f^*_{\eta_{k+1}}(x) - f^*_{\eta_{k+1}+1}(x)| \le \|f^*_{\eta_{k+1}} - f^*_{\eta_{k+1}+1}\|_\infty = \kappa_{k+1}$$

and similarly

$$|f^*_{\eta_k}(x) - f^*_{\eta_k+1}(x)| \le \kappa_k.$$

By Lemma 10 and the assumption that $(\alpha,\beta]$ contains only two change-points, it holds that for all $x \in [0,1]^d$,

$$\begin{aligned}\max_{t=\alpha}^{\beta} |\widetilde{f}^{(a,\beta]}_t(x)| \le& \sqrt{\beta - \eta_{r+1}}|f^*_{\eta_{r+1}}(x) - f^*_{\eta_{r+1}+1}(x)| + \sqrt{\eta_r - \alpha}|f^*_{\eta_r}(x) - f_{\eta^*_r+1}(x)| \\ \le& \sqrt{\Upsilon_{k+1}}\kappa_{k+1} + \sqrt{\Upsilon_k}\kappa_k \le 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}}.\end{aligned}$$

Thus

$$\max_{t=\alpha}^{\beta} \|\widetilde{f}^{(a,\beta]}_t\|_\infty \le 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}}. \tag{15}$$

Therefore in the good event in Equation (8), for any $1 \le m \le \log(T)$ and any $\mathcal{I} = (\alpha,\beta] \subseteq (s,e]$, it holds that

$$\begin{aligned}A^{\mathcal{I}}_m =& \max_{t=\alpha+\rho}^{\beta-\rho} |\widetilde{F}^{(\alpha,\beta]}_{t,h}(u_m)| \\ \le& \max_{t=\alpha+\rho}^{\beta-\rho} \|\widetilde{f}^{(\alpha,\beta]}_t\|_\infty + \lambda \\ \le& 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}} + \lambda,\end{aligned}$$

where the first inequality follows from Equation (8), and the last inequality follows from Equation (15). Then,

$$\begin{aligned}&2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}} + \lambda \\ =& 2\sqrt{C}\log^{\max\{1/2,5/q\}}(T)\sqrt{\frac{1}{nh^d} + 1} \\ &+ C_\lambda \log^{5/q}(T)\sqrt{\frac{1}{nh^d} + 1} + C_\lambda\sqrt{\frac{\log(T)}{nh^d}} + C_\lambda\sqrt{T}h^r + C_\lambda\sqrt{T}\Big(\frac{\log nT}{nT}\Big)^{\frac{2r}{2r+d}}.\end{aligned}$$

We observe that $\sqrt{\frac{\log(T)}{nh^d}} = O(\log(T)^{1/2}\sqrt{\frac{1}{nh^d} + 1})$. Moreover,

$$\sqrt{T}h^r = \sqrt{T}\Big(\frac{1}{nT}\Big)^{\frac{r}{2r+d}} \le (T^{\frac{1}{2} - \frac{r}{2r+d}})\frac{1}{n^{\frac{r}{2r+d}}},$$

and given that,

$$\frac{1}{2} - \frac{r}{2r+d} = \frac{d}{2(2r+d)},$$

we get,

$$\sqrt{T}h^r = o\Big(\log^{\max 1/2,5/q}(T)\sqrt{\frac{1}{nh^d} + 1}\Big).$$

Following the same line of arguments, we have that

$$\sqrt{T}\Big(\frac{\log nT}{nT}\Big)^{\frac{2r}{2r+d}} = T^{\frac{1}{2} - \frac{2r}{2r+d}} \log^{\frac{2r}{2r+d}}(T) = o\Big(\log T\sqrt{\frac{1}{nh^d} + 1}\Big).$$

Thus, by the choice of τ, it holds that with sufficiently large constant C_τ,

$$A^{\mathcal{I}}_m \le \tau \quad \text{for all } 1 \le m \le \log(T) \quad \text{and all} \quad \mathcal{I} \subseteq (s,e]. \tag{16}$$

As a result, FSBS $((s,e],h,\tau)$ will correctly reject if $(s,e]$ contains no undetected change-points.

Step 4. Assume that there exists an undetected change-point $\eta_{\tilde{k}} \in (s,e]$ such that

$$\min\{\eta_{\tilde{k}} - s, \eta_{\tilde{k}} - e\} = \Theta(T).$$

Let m^* and $\mathcal{I}^*$ be defined as in FSBS $((s, e], h, \tau)$ with

$$\mathcal{I}^* = (\alpha^*, \beta^*].$$

To complete the induction, it suffices to show that, there exists a change-point $\eta_k \in (s, e]$ such that $\min\{\eta_k - s, \eta_k - e\} = \Theta(T)$ and $|D_{m^*}^{\mathcal{I}^*} - \eta_k| \le \Upsilon_k$.

Consider the uni-variate time series

$$F_{t,h}(u_{m*}) = \frac{1}{n}\sum_{i=1}^{n} y_{t,i} K_h(u_{m*} - x_{t,i}) \quad \text{and} \quad f_t^*(u_{m*}) \quad \text{for all } 1 \le t \le T.$$

Since the collection of the change-points of the time series $\{f_t^*(u_{m*})\}_{t\in\mathcal{I}^*}$ is a subset of that of $\{\eta_k\}_{k=0}^{K+1} \cap (s, e]$, we may apply Lemma 9 to by setting

$$\mu_t = F_{t,h}(u_{m*}) \quad \text{and} \quad \omega_t = f_t^*(u_{m*})$$

on the interval $\mathcal{I}^*$. Therefore, it suffices to justify that all the assumptions of Lemma 9 hold.

In the following, λ is used in Lemma 9. Then Equation (36) and Equation (37) are directly consequence of Equation (8), Equation (9), Equation (10).
We observe that, for any $\mathcal{I} = (\alpha, \beta] \subseteq (s, e]$,

$$\max_{t=\alpha^*+\rho}^{\beta^*-\rho} |\widetilde{F}_{t,h}^{(\alpha^*,\beta^*]}(u_{m^*})| = A_{m^*}^{\mathcal{I}^*} \ge A_m^{\mathcal{I}} = \max_{t=\alpha+\rho}^{\beta-\rho} |\widetilde{F}_{t,h}^{(\alpha,\beta]}(u_m)|$$

for all m. By **Step 1** with $\mathcal{I}_k = (s_k, e_k]$, it holds that

$$\min\{\eta_k - s_k, e_k - \eta_k\} \ge \frac{1}{16}\zeta_k \ge c_2 T,$$

Therefore for all $k \in \{\tilde{k} : \min\{\eta_{\tilde{k}} - s, e - \eta_{\tilde{k}}\} \ge c_2 T\}$,

$$\max_{t=\alpha^*+\rho}^{\beta^*-\rho} |\widetilde{F}_{t,h}^{(\alpha^*,\beta^*]}(u_{m^*})| \ge \max_{t=s_k+\rho, m=1}^{t=e_k-\rho, m=\log(T)} |\widetilde{F}_{t,h}^{(s_k,e_k]}(u_m)| \ge c_1\sqrt{T}\kappa_k,$$

where the last inequality follows from Equation (11). Therefore Equation (38) holds in Lemma 9. Finally, Equation (39) is a direct consequence of the choices that

$$h = C_h(Tn)^{\frac{-1}{2r+d}} \quad \text{and} \quad \rho = \frac{\log(T)}{nh^d}.$$

Thus, all the conditions in Lemma 9 are met. So that, there exists a change-point η_k of $\{f_t^*(u_{m*})\}_{t\in\mathcal{I}^*}$, satisfying

$$\min\{\beta^* - \eta_k, \eta_k - \alpha^*\} > cT, \tag{17}$$

and

$$\begin{aligned}|D_{m*}^{\mathcal{I}^*} - \eta_k| \le \max\{C_3\lambda^2\kappa_k^{-2}, \rho\} \le & C_4 \log^{\max\{10/q,1\}}(T)\left(1 + \frac{1}{nh^d} + Th^{2r} + T\left(\frac{\log(nT)}{nT}\right)^{\frac{4r}{2r+d}}\right)\kappa_k^{-2} \\ \le & C \log^{\max\{10/q,1\}}(T)\left(1 + T^{\frac{d}{2r+d}} n^{\frac{-2r}{2r+d}}\right)\kappa_k^{-2}\end{aligned}$$

for sufficiently large constant C, where we have followed the same line of arguments than for the conclusion of (16). Observe that
i) The change-points of $\{f_t^*(u_{m^*})\}_{t\in\mathcal{I}^*}$ belong to $(s, e] \cap \{\eta_k\}_{k=1}^K$; and
ii) Equation (17) and $(\alpha^*, \beta^*] \subseteq (s, e]$ imply that

$$\min\{e - \eta_k, \eta_k - s\} > cT \ge \Upsilon_{\max}.$$

As discussed in the argument before **Step 1**, this implies that η_k must be an undetected change-point of $\{f_t^*(u_{m^*})\}_{t\in\mathcal{I}^*}$. $\square$

C Deviation bounds related to kernels

In this section, we deal with all the large probability events occurred in the proof of Theorem 1. Recall that $F_{t,h}(x) = \frac{\frac{1}{n}\sum_{i=1}^{n} y_{t,i}K_h(x-x_{t,i})}{\hat{p}(x)}$, and

$$\widetilde{F}_{t,h}^{(s,e]}(x) = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{l=s+1}^{t} F_{l,h}(x) - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{l=t+1}^{e} F_{l,h}(x).$$

By assumption 2, we have $\max_{l=1}^{q}\|K^l\|_\infty = \max_{l=1}^{q}\|K\|_\infty^l < C_K$, where $C_K > 0$ is an absolute constant. Moreover, assumption 1b implies $|f_t^*(x)| < C_f$ for any $x \in [0,1]^d, t \in 1, ..., T$.

Proposition 1. *Suppose that Assumption 1 and 2 hold, that $\rho n h^d \geq \log(T)$ and that $T \geq 3$. Then for any $x \in [0,1]^d$*

$$\mathbb{P}\Big(\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\big(F_{t,h}(x) - f_t^*(x)\big)\Big| \geq \frac{2}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1} + \frac{\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r + \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big)$$
$$\leq 2C_1\frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}; \tag{18}$$
$$\mathbb{P}\Big(\max_{k=\rho}^{\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}-k+1}^{\tilde{r}}\big(F_{t,h}(x) - f_t^*(x)\big)\Big| \geq \frac{2}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1} + \frac{\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r + \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big)$$
$$\leq 2C_1\frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}. \tag{19}$$

Proof. The proofs of Equation (18) and Equation (19) are the same. So only the proof of Equation (18) is presented. We define the events $E_1 = \Big\{||\hat{p} - u||_\infty \leq \bar{C}\Big(\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}\Big)\Big\}$ and $E_2 = \Big\{\hat{p} \geq \bar{c},\ \bar{c} = \inf_x u(x) - \bar{C}\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}\Big\}$. Using Lemma 1, especifically by equation (6), we have that $P(E_1) \geq 1 - \frac{1}{nT}$. Then, we observe that in event E_1, for $x \in [0,1]^d$

$$\inf_s u(s) - \hat{p}(x) \leq u(x) - \hat{p}(x) \leq |u(x) - \hat{p}(x)| \leq \bar{C}\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}$$

which implies $E1 \subseteq E_2$. Therefore, $P(E_2^c) \leq \frac{1}{nT}$.
Now, for any x, observe that, by definition of $F_{t,h}$ and triangle inequality

$$\begin{aligned}
I &= \max_{k=\rho}^{T-\tilde{r}}\frac{1}{\sqrt{k}}\Big|\sum_{t=\tilde{r}+1}^{\tilde{r}+k}F_{t,h}(x) - \sum_{t=\tilde{r}+1}^{\tilde{r}+k}f_t^*(x)\Big| \\
&\leq \max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - f_t^*(x)\Big)\Big| \\
&+ \max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_t(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)}\Big| \\
&+ \max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_{t,i}K_h(x-x_{t,i})}{\hat{p}(x)}\Big| \\
&= I_1 + I_2 + I_3.
\end{aligned} \tag{20}$$

In the following, we will show that $I_1 \leq I_{1,1} + I_{1,2} + I_{1,3}$, and that

1. $\mathbb{P}\Big(I_{1,1} \geq \frac{\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)\Big) \leq \frac{1}{T^5} + \frac{1}{Tn}$,
2. $\mathbb{P}\Big(I_{1,2} \geq \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r\Big) \leq \frac{1}{Tn}$,

3. $\mathbb{P}\Big(I_{1,3} \geq \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \leq \frac{1}{Tn}$,

4. $\mathbb{P}\Big(I_2 \geq \frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\Big) \leq \frac{C_1 \log T}{z^q} + \frac{1}{Tn}$,

5. $\mathbb{P}\Big(I_3 \geq \frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\Big) \leq \frac{C_1 \log T}{z^q} + \frac{1}{Tn}$,

in order to conclude that,

$$
\begin{aligned}
&\mathbb{P}\Big(I \geq 2z\sqrt{\frac{1}{nh^d}+1} + \tilde{C}_1\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big) + \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r + \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \\
\leq&\mathbb{P}\Big(I_{1,1} \geq \tilde{C}_1\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)\Big) + \mathbb{P}\Big(I_{1,2} \geq \frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r\Big) + \mathbb{P}\Big(I_{1,3} \geq \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big) \\
+&\mathbb{P}\Big(I_2 \geq z\sqrt{\frac{1}{nh^d}+1}\Big) + \mathbb{P}\Big(I_3 \geq z\sqrt{\frac{1}{nh^d}+1}\Big) \\
\leq&2C_1\frac{\log(T)}{z^q} + T^{-5} + \frac{5}{Tn}.
\end{aligned}
$$

Step 1. The analysis for I_1 is done. We observe that,

$$
\begin{aligned}
&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - f_t^*(x)\Big)\Big| \\
\leq&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - \frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}\Big)\Big| \\
+&\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)} - f_t^*(x)\Big)\Big| = I_{1,1} + \tilde{I}_1.
\end{aligned}
$$

Step 1.1 The analysis for $I_{1,1}$ is done. We note that the random variables $\{f_t^*(x_{t,i})K_h(x - x_{t,i})\}_{1\leq i\leq n_t, 1\leq t\leq N}$ are independent distributed with mean $\int f_t^*(z)K_h(x-z)d\mu(z)$ and

$$
\begin{aligned}
\mathrm{Var}\big(f_t^*(x_{t,i})K_h(x-x_{t,i})\big) \leq& E\big\{(f_t^*)^2(x_{t,i})K_h^2(x-x_{t,i})\big\} \\
=&\int_{[0,1]^d}(f_t^*)^2(z)\frac{1}{h^{2d}}K^2\big(\frac{x-z}{h}\big)d\mu(z) \\
\leq&\frac{C_f^2}{h^d}\int_{[0,1]^d}\frac{1}{h^d}K^2\big(\frac{x-z}{h}\big)d\mu(z) \\
=&\frac{C_f^2}{h^d}\int_{[0,1]^d}K^2\big(u\big)d\mu(u) < \frac{C_f^2C_K^2}{h^d}.
\end{aligned}
$$

Since $|f_t^*(x_{t,i})K_h(x-x_{t,i})| \leq C_fC_Kh^{-d}$, by Bernstein inequality [34], we have that

$$
\mathbb{P}\Big(\Big|\frac{1}{kn}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\sum_{i=1}^{n}f_t^*(x_{t,i})K_h(x-x_{t,i}) - \int f_t^*(z)K_h(x-z)d\mu(z)\Big| \geq \tilde{C}_1\Big\{\sqrt{\frac{\log(T)}{knh^d}} + \frac{\log(T)}{knh^d}\Big\}\Big) \leq T^{-6}.
$$

Since $knh^d \geq \log(T)$ if $k \geq \rho$, with probability at most T^{-5}, it holds that

$$
\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}n}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\sum_{i=1}^{n}\Big(f_t^*(x_{t,i})K_h(x-x_{t,i}) - \int f_t^*(z)K_h(x-z)d\mu(z)\Big)\Big| \geq \tilde{C}_1\sqrt{\frac{\log(T)}{nh^d}}.
$$

Therefore, using that $P(E_2^c) \leq \frac{1}{Tn}$, we conclude

$$
\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{f_t^*(x_{t,i})K_h(x-x_{t,i})}{\hat{p}(x)} - \frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}\Big)\Big| \geq \frac{\tilde{C}_1}{\tilde{c}}\sqrt{\frac{\log(T)}{nh^d}}
$$

with probability at most $T^{-5}+\frac{1}{nT}$.

Step 1.2 The analysis for $I_{1,2}$ and $I_{1,3}$ is done. We observe that

$$\begin{aligned}
\tilde{I}_1 =& \max_{k=\rho}^{T-\tilde{r}}\bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg(\frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}-f_t^*(x)\bigg)\bigg|\\
\leq& \max_{k=\rho}^{T-\tilde{r}}\bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg(\frac{\int f_t^*(z)K_h(x-z)d\mu(z)}{\hat{p}(x)}-\frac{f_t^*(x)u(x)}{\hat{p}(x)}\bigg)\bigg| \qquad (21)\\
+& \max_{k=\rho}^{T-\tilde{r}}\bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg(\frac{f_t^*(x)u(x)}{\hat{p}(x)}-f_t^*(x)\bigg)\bigg| = I_{1,2}+I_{1,3}. \qquad (22)
\end{aligned}$$

Then, we observe that

$$\begin{aligned}
I_{1,2} =& \bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg(\int f_t^*(z)K_h(x-z)d\mu(z)-f_t^*(x)u(x)\bigg)\bigg|\\
\leq& \frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg|\int f_t^*(z)K_h(x-z)d\mu(z)-f_t^*(x)u(x)\bigg|\\
\leq& \frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\tilde{C}h^r\\
=& \frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\tilde{C}h^r\\
=& \sqrt{k}\tilde{C}h^r
\end{aligned}$$

where the second inequality follows from assumption 2. Therefore, using event E_2, we can bound (21) by $\frac{\tilde{C}}{\tilde{c}}\sqrt{T}h^r$ with probability at least $1-\frac{1}{nT}$. Meanwhile, for (22) we have that,

$$\begin{aligned}
I_{1,3} =& \max_{k=\rho}^{T-\tilde{r}}\bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\bigg(\frac{f_t^*(x)u(x)}{\hat{p}(x)}-f_t^*(x)\bigg)\bigg|\\
\leq& \max_{k=\rho}^{T-\tilde{r}}\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}|f_t^*(x)|\bigg|\frac{u(x)-\hat{p}(x)}{\hat{p}(x)}\bigg|. \qquad (23)
\end{aligned}$$

Then, since in the event E_1, it is satisfies that

$$||\hat{p}-u||_\infty \leq \bar{C}\Big(\Big(\frac{\log(Tn)}{Tn}\Big)^{\frac{2r}{2r+d}}\Big), \text{ and } \hat{p}\geq\bar{c};$$

we have that equation (23), is bounded by

$$\max_{k=\rho}^{T-\tilde{r}}\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{1}{n}\sum_{i=1}^{n}\frac{\bar{C}C_f}{\tilde{c}}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}} \leq \frac{\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}$$

with probability at least $1-\frac{1}{nT}$.

Step 2. The analysis for I_2 and I_3 is done. For $1\leq t\leq T$, let

$$Z_t = \frac{1}{n}\sum_{i=1}^{n}\xi_t(x_{t,i})K_h(x-x_{t,i}) \quad \text{and} \quad W_t=\frac{1}{n}\sum_{i=1}^{n}\delta_{t,i}K_h(x-x_{t,i}).$$

By Lemma 2 and event E_2, it holds that

$$\mathbb{P}\bigg\{\max_{k=\rho}^{T-\tilde{r}}\bigg|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{Z_t}{\hat{p}(x)}\bigg|\geq\frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\bigg\}\leq\frac{C_1\log(T)}{z^q}+\frac{1}{nT}$$

and

$$\mathbb{P}\Big\{\max_{k=\rho}^{T-\tilde{r}}\Big|\frac{1}{\sqrt{k}}\sum_{t=\tilde{r}+1}^{\tilde{r}+k}\frac{W_t}{\hat{p}(x)}\Big|\geq\frac{1}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}\Big\}\leq\frac{C_1\log(T)}{z^q}+\frac{1}{nT}.$$

The desired result follows from putting the previous steps together. □

Corollary 1. *Suppose that $\rho nh^d\geq\log(T)$ and that $T\geq 3$. Then for $z>0$*

$$\mathbb{P}\Big\{\max_{t=s+\rho+1}^{e-\rho}\Big|\widetilde{F}^{(s,e]}_{t,h}(x)-\widetilde{f}^{(s,e]}_{t}(x)\Big|\geq\frac{4}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}+\frac{2\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)+\frac{2\tilde{C}}{\tilde{c}}\sqrt{T}h^r+\frac{2\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}\Big\}$$
$$\leq 2T^{-5}+\frac{4C_1\log(T)}{z^q}+10\frac{1}{Tn}.$$

Proof. By definition of $\widetilde{F}^{(s,e]}_{t,h}$ and $\widetilde{f}^{(s,e]}_{t}$, we have that

$$\begin{aligned}\Big|\widetilde{F}^{(s,e]}_{t,h}(x)-\widetilde{f}^{(s,e]}_{t}(x)\Big|\leq&\Big|\sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{l=s+1}^{t}(F_{l,h}(x)-f^*_l(x))\Big|\\&+\Big|\sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{l=t+1}^{e}(F_{l,h}(x)-f^*_l(x))\Big|.\end{aligned}$$

Then, we observe that,

$$\sqrt{\frac{e-t}{(e-s)(t-s)}}\leq\sqrt{\frac{1}{t-s}}\text{ if }s\leq t,\text{ and }\sqrt{\frac{t-s}{(e-s)(e-t)}}\leq\sqrt{\frac{1}{e-t}}\text{ if }t\leq e.$$

Therefore,

$$\begin{aligned}X=\max_{t=s+\rho+1}^{e-\rho}\Big|\widetilde{F}^{(s,e]}_{t,h}(x)-\widetilde{f}^{(s,e]}_{t}(x)\Big|\leq&\max_{t=s+\rho+1}^{e-\rho}\Big|\sqrt{\frac{1}{t-s}}\sum_{l=s+1}^{t}\Big(F_{l,h}(x)-\{f^*_l(x)\Big)\Big|\\&+\max_{t=s+\rho+1}^{e-\rho}\Big|\sqrt{\frac{1}{e-t}}\sum_{l=t+1}^{e}\Big(F_{l,h}(x)-f^*_l(x)\Big)\Big|=X_1+X_2.\end{aligned}$$

Finally, letting $\lambda=\frac{4}{\tilde{c}}z\sqrt{\frac{1}{nh^d}+1}+\frac{2\tilde{C}_1}{\tilde{c}}\Big(\sqrt{\frac{\log(T)}{nh^d}}\Big)+\frac{2\tilde{C}}{\tilde{c}}\sqrt{T}h^r+\frac{2\bar{C}C_f}{\tilde{c}}\sqrt{T}\Big(\frac{\log(nT)}{nT}\Big)^{\frac{2r}{2r+d}}$, we get that

$$\begin{aligned}\mathbb{P}(X\geq\lambda)\leq&\mathbb{P}(X_1+X_2\geq\frac{\lambda}{2}+\frac{\lambda}{2})\\\leq&\mathbb{P}(X_1\geq\frac{\lambda}{2})+\mathbb{P}(X_2\geq\frac{\lambda}{2})\\\leq&2T^{-5}+\frac{4C_1\log(T)}{z^q}+10\frac{1}{Tn},\end{aligned}$$

where the last inequality follows from Proposition 1. □

C.1 Additional Technical Results

The following lemmas provide lower bounds for

$$Z_t = \frac{1}{n}\sum_{i=1}^{n} \xi_t(x_{t,i})K_h(x - x_{t,i}) \quad \text{and} \quad W_t = \frac{1}{n}\sum_{i=1}^{n} \delta_{t,i}K_h(x - x_{t,i}).$$

They are a direct consequence of the temporal dependence and heavy-tailedness of the data considered in Assumption 1.

Lemma 2. *Let $\rho \le T$ be such that $\rho n h^d \ge \log(T)$ and $T \ge 3$. Let $N \in \mathbb{Z}^+$ be such that $N \ge \rho$.*
a. *Suppose that for any $q \ge 3$ it holds that*

$$\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}\{\|\xi_t - \xi_t^*\|_\infty^q\}^{1/q} = \mathrm{O}(1). \tag{24}$$

Then for any $z > 0$,

$$\mathbb{P}\Big\{\max_{k=\rho}^{N}\Big|\Big\{\frac{1}{nh^d}+1\Big\}^{-1/2}\frac{1}{\sqrt{k}}\sum_{t=1}^{k} Z_t\Big| \ge z\Big\} \le \frac{C_1\log(T)}{z^q}.$$

b. *Suppose that for some $q \ge 3$,*

$$\sum_{t=1}^{\infty} t^{1/2-1/q}\max_{i=1}^{n}\{\mathbb{E}|\delta_{t,i} - \delta_{t,i}^*|^q\}^{1/q} < \mathrm{O}(1). \tag{25}$$

Then for any $w > 0$,

$$\mathbb{P}\Big\{\max_{k=\rho}^{N}\Big|\Big\{\frac{1}{nh^d}+1\Big\}^{-1/2}\frac{1}{\sqrt{k}}\sum_{t=1}^{k} W_t\Big| \ge w\Big\} \le \frac{C_1\log(T)}{w^q}.$$

Proof. The proof of part **b** is similar and simpler than that of part **a**. For conciseness, only the proof of **a** is presented.

By Lemma 4 and Equation (24), for all $J \in \mathbb{Z}^+$, it holds that

$$\mathbb{E}\Big\{\max_{k=1}^{J}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \le J^{1/2}C\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\} + J^{1/q}C''\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}.$$

As a result there exists a constant C_1 such that

$$\mathbb{E}\Big\{\max_{k=1}^{J}|\sum_{t=1}^{k} Z_t|^q\Big\} \le C_1 J^{q/2}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}^q + C_1 J\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q}+1\Big\}^q.$$

We observe that

$$J^{q/2} = \frac{q}{2}\int_0^J x^{q/2-1}dx \tag{26}$$

$$= \frac{q}{2}\Big(\int_0^1 x^{q/2-1}dx + \int_1^J x^{q/2-1}dx\Big) \tag{27}$$

$$\le \frac{q}{2}\Big(1 + \int_1^J x^{q/2-1}dx\Big) \tag{28}$$

$$= \frac{q}{2}\Big(1 + \int_1^2 x^{q/2-1}dx + \dots + \int_{J-1}^J x^{q/2-1}dx\Big) \tag{29}$$

$$\le \frac{q}{2}\Big(1 + \int_1^2 2^{q/2-1}dx + \dots + \int_{J-1}^J J^{q/2-1}dx\Big) \tag{30}$$

$$= \frac{q}{2}\sum_{k=1}^{J} k^{q/2-1}, \tag{31}$$

which implies, there is a constant C_2 such that

$$C_1 J^{q/2}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q + C_1 J\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q}+1\right\}^q \le C_2\sum_{k=1}^{J}\alpha_k,$$

where

$$\alpha_k = k^{q/2-1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q + \left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q}+1\right\}^q.$$

By theorem B.2 of Kirch (2006),

$$\begin{aligned}
\mathbb{E}\left\{\max_{k=1}^{N}\left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}Z_t\right|\right\}^q \le& 4C_2\sum_{l=1}^{N} l^{-q/2}\alpha_l\\
=&4C_2\sum_{l=1}^{N}\left(l^{-1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q + l^{-q/2}\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q}+1\right\}^q\right)\\
\le& C_3\log(N)\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q + C_3 N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q}+1\right\}^q
\end{aligned}$$

where the last inequality follows from the fact that $\int_1^N \frac{1}{x} = log(N)$ and that $\int_1^N x^{\frac{-q}{2}} = O(N^{-q/2+1})$. Since

$$N^{1/2-1/q} \ge \rho^{1/2-1/q} \ge (nh^d)^{1/2-(q-1)/q},$$

it holds that, $\frac{1}{nh^d} \le N$. Moreover,

$$\begin{aligned}
N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q}+1\right\}^q =& N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q+1/2-1/2}+1\right\}^q\\
=& N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q-1/2}\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\\
\le& N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{(q-1)/q-1/2}+1\right\}^q\\
=& N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{1/2-1/q}+1\right\}^q\\
=& N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{(q-2)/(2q)}+1\right\}^q\\
\le& C_3' N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{(q-2)/(2q)}\right\}^q\\
=& C_3' N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{(q-2)/(2)}\right\}\\
\le& C_3' N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q\left\{\left(\frac{1}{nh^d}\right)^{q/2-1}\right\}\\
\le& C_3' N^{-q/2+1}\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q N^{q/2-1}\\
=& C_3'\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q.
\end{aligned}$$

It follows that,

$$\mathbb{E}\left\{\max_{k=1}^{N}\left|\frac{1}{\sqrt{k}}\sum_{t=1}^{k}Z_t\right|\right\}^q \le C_4\log(N)\left\{\left(\frac{1}{nh^d}\right)^{1/2}+1\right\}^q.$$

By Markov's inequality, for any $z > 0$ and the assumption that $T \geq N$,

$$\mathbb{P}\bigg\{ \max_{k=1}^{N} \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \bigg| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} Z_t \bigg| \geq z \bigg\} \leq \frac{C_1 \log(T)}{z^q}.$$

Since $N \geq \rho$, this directly implies that

$$\mathbb{P}\bigg\{ \max_{k=\rho}^{N} \bigg\{ \frac{1}{nh^d} + 1 \bigg\}^{-1/2} \bigg| \frac{1}{\sqrt{k}} \sum_{t=1}^{k} Z_t \bigg| \geq z \bigg\} \leq \frac{C_1 \log(T)}{z^q}.$$

□

Lemma 3. *Suppose Assumption 1* **c** *holds and $q \geq 2$. Then there exists absolute constants $C > 0$ so that*

$$\mathbb{E}|Z_t - Z_t^*|^q \leq C\mathbb{E}\big\{\|\xi_t - \xi_t^*\|_\infty^q\big\}\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{q-1} + 1 \bigg\}. \tag{32}$$

If in addition $\mathbb{E}\big\{\|\xi_t\|_\infty^q\big\} = O(1)$, then there exists absolute constants C' such that

$$\mathbb{E}|Z_t|^q \leq C'\bigg\{ \bigg(\frac{1}{nh^d}\bigg)^{q-1} + 1 \bigg\}. \tag{33}$$

Proof. The proof of the Equation (33) is simpler and simpler than Equation (32). So only the proof of Equation (32) is presented. Note that since $\{x_t\}_{t=1}^T$ and $\{\xi_t\}_{t=1}^T$ are independent, and that $\{x_t\}_{t=1}^T$ are independent identically distributed,

$$Z_t^* = \frac{1}{n} \sum_{i=1}^{n} \xi_t^*(x_{t,i}) K_h(x - x_{t,i}).$$

Step 1. Note that, by the Newton's binomial

$$\begin{aligned}
\mathbb{E}|Z_t - Z_t^*|^q =& \mathbb{E}\bigg\{ \bigg| \frac{1}{n} \sum_{i=1}^{n} \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \bigg|^q \bigg\} \\
\leq& \frac{1}{n^q} \mathbb{E}\bigg\{ \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta_1\geq 0,\ldots,\beta_n\geq 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\} \\
=& \frac{1}{n^q} \mathbb{E}\bigg\{ \sum_{k=1}^{q} \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta=(\beta_1,\ldots,\beta_n), \|\beta\|_0=k, \beta\geq 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\}.
\end{aligned}$$

Step 2. For a fixed $\beta = (\beta_1, \ldots, \beta_n)$ such that $\beta_1 + \ldots + \beta_n = q$ and that $\|\beta\|_0 = k$, consider

$$\mathbb{E}\bigg\{ \prod_{j=1}^{n} \big| \{\xi_t^* - \xi_t\}(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \bigg\}.$$

Without loss of generality, assume that $\beta_1, \ldots, \beta_k$ are non-zero. Then it holds that

$$\begin{aligned}
&\mathbb{E}\bigg\{ \big|(\xi_t^* - \xi_t)(x_{t,1})\big|^{\beta_1} \big|K_h(x - x_{t,1})\big|^{\beta_1} \cdots \big|(\xi_t^* - \xi_t)(x_{t,k})\big|^{\beta_k} \big|K_h(x - x_{t,k})\big|^{\beta_k} \bigg\} \\
=& \mathbb{E}_\xi \bigg\{ \int \big|(\xi_t^* - \xi_t)(r)\big|^{\beta_1} \big|K_h(x - r)\big|^{\beta_1} d\mu(r) \cdots \int \big|(\xi_t^* - \xi_t)(r)\big|^{\beta_k} \big|K_h(x - r)\big|^{\beta_k} d\mu(r) \bigg\} \\
=& \mathbb{E}_\xi \bigg\{ \int \big|(\xi_t^* - \xi_t)(x - sh)\big|^{\beta_1} \frac{\big|K(s)\big|^{\beta_1}}{h^{d(\beta_1 - 1)}} d\mu(s) \cdots \int \big|(\xi_t^* - \xi_t)(x - sh)\big|^{\beta_k} \frac{\big|K(s)\big|^{\beta_k}}{h^{d(\beta_k - 1)}} d\mu(s) \bigg\} \\
\leq& h^{-d\sum_{j=1}^{k}(\beta_j - 1)} \mathbb{E}_\xi \bigg\{ \|\xi_t^* - \xi_t\|_\infty^{\beta_1} C_K^{\beta_1} \cdots \|\xi_t^* - \xi_t\|_\infty^{\beta_k} C_K^{\beta_k} \bigg\} \\
\leq& h^{-d(q-k)} C_K^q \mathbb{E}_\xi \bigg\{ \|\xi_t^* - \xi_t\|_\infty^{\sum_{j=1}^{k} \beta_k} \bigg\} \\
\leq& h^{-d(q-k)} C_K^q \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\}
\end{aligned}$$

where the third equality follows by using the change of variable $s = \frac{x-r}{h}$, the first inequality by assumption 2.

Step 3. Let $k \in \{1, \ldots, q\}$ be fixed. Note that $\binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \le q!$. Consider set

$$\mathcal{B}_k = \Big\{ \beta \in \mathbb{N}^n : \beta \ge 0, \beta_1 + \ldots + \beta_n = q, |\beta|_0 = k \Big\}.$$

To bound the cardinality of the set $\mathcal{B}_k$, first note that since $|\beta|_0 = k$, there are $\binom{n}{k}$ number of ways to choose the index of non-zero entries of β.
Suppose $\{i_1, \ldots i_k\}$ are the chosen index such that $\beta_{i_1} \ne 0, \ldots, \beta_{i_k} \ne 0$. Then the constrains $\beta_{i_1} > 0, \ldots, \beta_{i_1} > 0$ and $\beta_{i_1} + \ldots + \beta_{i_k} = q$ are equivalent to that of diving q balls into k groups (without distinguishing each ball). As a result there are $\binom{q-1}{k-1}$ number of ways to choose the $\{\beta_{i_1}, \ldots, \beta_{i_k}\}$ once the index $\{i_1, \ldots i_k\}$ are chosen.

Step 4. Combining the previous three steps, it follows that for some constants $C_q, C_1 > 0$ only depending on q,

$$\begin{aligned}
\mathbb{E}|Z_t - Z_t^*|^q \le & \frac{1}{n^q} \mathbb{E} \Big\{ \sum_{k=1}^q \sum_{\substack{\beta_1 + \beta_2 + \ldots + \beta_n = q \\ \beta = (\beta_1, \ldots, \beta_n), |\beta|_0 = k, \beta \ge 0}} \binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \prod_{j=1}^n \big| (\xi_t^* - \xi_t)(x_{t,i}) K_h(x - x_{t,i}) \big|^{\beta_j} \Big\} \\
\le & \frac{1}{n^q} \sum_{k=1}^q \binom{n}{k} \binom{q-1}{k-1} q! h^{-d(q-k)} C_K^q \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \\
\le & \frac{1}{n^q} \sum_{k=1}^q n^k C_q C_K^q h^{-d(q-k)} \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \\
\le & C_1 \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} \Big\{ \Big(\frac{1}{nh^d} \Big)^{q-1} + \Big(\frac{1}{nh^d} \Big)^{q-2} + \ldots + \Big(\frac{1}{nh^d} \Big) + 1 \Big\} \\
\le & C_1 \mathbb{E}_\xi \big\{ \|\xi_t^* - \xi_t\|_\infty^q \big\} q \Big\{ \Big(\frac{1}{nh^d} \Big)^{q-1} + 1 \Big\},
\end{aligned}$$

where the second inequality is satisfied by step 3 and that $\binom{q}{\beta_1, \beta_2, \ldots, \beta_n} \le q!$, while the third inequality is achieved by using that $\binom{n}{k}\binom{q-1}{k-1} q! \le \binom{n}{k} C_q \le n^k C_q$. Moreover, given that $\frac{1}{n^q} n^k h^{-d(q-k)} = \left(\frac{1}{nh^d} \right)^{q-k}$ the fourth inequality is obtained. The last inequality holds because if $\frac{1}{nh^d} \le 1$, then $\left\{ \left(\frac{1}{nh^d} \right)^{q-1} + \ldots + \left(\frac{1}{nh^d} \right) + 1 \right\} \le q$, and if $\frac{1}{nh^d} \ge 1$, then $\left\{ \left(\frac{1}{nh^d} \right)^{q-1} + \ldots + \left(\frac{1}{nh^d} \right) + 1 \right\} \le q \left(\frac{1}{nh^d} \right)^{q-1}$. □

Lemma 4. *Suppose Assumption 1* **c** *holds. Let $\rho \le T$ be such that $\rho n h^d \ge \log(T)$ and $T \ge 3$. Let $N \in \mathbb{Z}^+$ be such that $N \ge \rho$. Then, it holds that*

$$\Big\{ \mathbb{E} \max_{k=1}^N \Big| \sum_{t=1}^k Z_t \Big|^q \Big\}^{1/q} \le N^{1/2} C \Big\{ \Big(\frac{1}{nh^d} \Big)^{1/2} + 1 \Big\} + N^{1/q} C' \Big\{ \Big(\frac{1}{nh^d} \Big)^{(q-1)/q} + 1 \Big\}.$$

Proof. We have that $q > 2$ and $\mathbb{E}|Z_1| < \infty$ by the use of Lemma 3. Then, making use of Theorem 1 of Liu et al. (2013), we obtain that

$$\begin{aligned}
\Big\{ \mathbb{E} \max_{k=1}^N \Big| \sum_{t=1}^k Z_t \Big|^q \Big\}^{1/q} \le & N^{1/2} C_1 \Big\{ \sum_{j=1}^N \Theta_{j,2} + \sum_{j=N+1}^\infty \Theta_{j,q} + \{\mathbb{E}|Z_1|^2\}^{1/2} \Big\} \\
& + N^{1/q} C_2 \Big\{ \sum_{j=1}^N j^{1/2 - 1/q} \Theta_{j,q} + \{\mathbb{E}|Z_1|^q\}^{1/q} \Big\},
\end{aligned}$$

where $\Theta_{j,q} = \{\mathbb{E}(|Z_j^* - Z_j|^q)\}^{1/q}$. Moreover, we observe that since $\Theta_{j,2} \leq \Theta_{j,q}$ for any $q \geq 2$, it follows

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1\Big\{\sum_{j=1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\} + N^{1/q}C_2\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\},$$

Next, by the first part of Lemma 3,

$$\Theta_{j,q}^q \leq C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}.$$

even more, we have that $N \geq \frac{1}{nh^d}$, implies that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq& N^{1/2}C_1^{'}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}\Big\}^{1/q} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}\\
\leq& N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}\\
\leq& N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(N\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\big\{\|\xi_j - \xi_j^*\|_\infty^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

From Assumption 1 **c**,

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{'''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + \{\mathbb{E}|Z_1|^2\}^{1/2}\Big\} + N^{1/q}C_2^{''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q} + \{\mathbb{E}|Z_1|^q\}^{1/q}\Big\}.$$

By the second part of Lemma 3, it holds that

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{''''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big) + 1\Big\}^{1/2}\Big\} + N^{1/q}C_2^{'''}\Big\{1 + \Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q}\Big\}.$$

This immediately implies the desired result. □

Lemma 5. *Suppose Assumption 1 holds. Then there exists absolute constants C_1 such that*

$$\mathbb{E}|W_t - W_t^*|^q \leq C_1\max_{i=1}^{n}\mathbb{E}\big\{|\delta_{t,i} - \delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}. \tag{34}$$

If in addition $\mathbb{E}\big\{|\delta_{t,i}|_q^q\big\} = O(1)$ for all $1 \leq i \leq n$, then there exists absolute constants C' such that

$$\mathbb{E}(|W_t|^q)^{1/q} \leq C'\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}^{1/q}. \tag{35}$$

Proof. The proof is similar to that of Lemma 3. The proof of the Equation (35) is simpler and simpler than Equation (34). So only the proof of Equation (34) is presented. Note that since $\{x_t\}_{t=1}^T$ and $\{\delta_t\}_{t=1}^T$ are independent, and that $\{x_t\}_{t=1}^T$ are independent identically distributed,

$$\delta_t^* = \frac{1}{n}\sum_{i=1}^n \delta_{t,i}^* K_h(x - x_{t,i}).$$

Step 1. Note that, by the Newton's binomial

$$\begin{aligned}
\mathbb{E}|\delta_t - \delta_t^*|^q =&\mathbb{E}\Big\{\Big|\frac{1}{n}\sum_{i=1}^n(\delta_{t,i}^* - \delta_{t,i})K_h(x - x_{t,i})\Big|^q\Big\} \\
\le&\frac{1}{n^q}\mathbb{E}\Big\{\sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta_1\ge 0,\ldots,\beta_n\ge 0}}\binom{q}{\beta_1,\beta_2,\ldots,\beta_n}\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x - x_{t,i})\big|^{\beta_j}\Big\} \\
=&\frac{1}{n^q}\mathbb{E}\Big\{\sum_{k=1}^q\sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta=(\beta_1,\ldots,\beta_n),|\beta|_0=k,\beta\ge 0}}\binom{q}{\beta_1,\beta_2,\ldots,\beta_n}\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x - x_{t,i})\big|^{\beta_j}\Big\}.
\end{aligned}$$

Step 2. For a fixed $\beta = (\beta_1, \ldots, \beta_n)$ such that $\beta_1 + \ldots + \beta_n = q$ and that $|\beta|_0 = k$, consider

$$\mathbb{E}\Big\{\prod_{j=1}^n\big|(\delta_{t,i}^* - \delta_{t,i})K_h(x - x_{t,i})\big|^{\beta_j}\Big\}.$$

Without loss of generality, assume that $\beta_1, \ldots, \beta_k$ are non-zero. Then it holds that

$$\begin{aligned}
&\mathbb{E}\Big\{\big|(\delta_{t,1}^* - \delta_{t,1})\big|^{\beta_1}\big|K_h(x - x_{t,1})\big|^{\beta_1}\cdots\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\big|K_h(x - x_{t,k})\big|^{\beta_k}\Big\} \\
=&\mathbb{E}_\delta\Big\{\int\big|(\delta_{t,1}^* - \delta_{t,1}\big|^{\beta_1}\big|K_h(x - r)\big|^{\beta_1}d\mu(r)\cdots\int\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\big|K_h(x - r)\big|^{\beta_k}d\mu(r)\Big\} \\
=&\mathbb{E}_\delta\Big\{\int\big|(\delta_{t,1}^* - \delta_{t,1}\big|^{\beta_1}\frac{\big|K(s)\big|^{\beta_1}}{h^{d(\beta_1-1)}}d\mu(s)\cdots\int\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}\frac{\big|K(s)\big|^{\beta_k}}{h^{d(\beta_k-1)}}d\mu(s)\Big\} \\
\le&h^{-d\sum_{j=1}^k(\beta_j-1)}\mathbb{E}_\delta\Big\{\big|(\delta_{t,1}^* - \delta_{t,1})\big|^{\beta_1}C_K^{\beta_1}\cdots\big|(\delta_{t,k}^* - \delta_{t,k})\big|^{\beta_k}C_K^{\beta_k}\Big\} \\
\le&h^{-d(q-k)}C_K^q\mathbb{E}_\delta\Big\{\max_{i=1}^n|\delta_{t,i} - \delta_{t,i}^*|^{\sum_{j=1}^k\beta_k}\Big\} \\
\le&h^{-d(q-k)}C_K^q\mathbb{E}_\delta\big\{\max_{i=1}^n|\delta_{t,i} - \delta_{t,i}^*|^q\big\}
\end{aligned}$$

where the third equality follows by using the change of variable $s = \frac{x-r}{h}$, the first inequality by assumption 2.
Step 3. Let $k \in \{1, \ldots, q\}$ be fixed. Note that $\binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \le q!$. Consider set

$$\mathcal{B}_k = \Big\{\beta \in \mathbb{N}^n : \beta \ge 0, \beta_1 + \ldots + \beta_n = q, |\beta|_0 = k\Big\}.$$

To bound the cardinality of the set $\mathcal{B}_k$, first note that since $|\beta|_0 = k$, there are $\binom{n}{k}$ number of ways to choose the index of non-zero entries of β.
Suppose $\{i_1, \ldots i_k\}$ are the chosen index such that $\beta_{i_1} \ne 0, \ldots, \beta_{i_k} \ne 0$. Then the constrains $\beta_{i_1} > 0, \ldots, \beta_{i_1} > 0$ and $\beta_{i_1} + \ldots + \beta_{i_k} = q$ are equivalent to that of diving q balls into k groups (without distinguishing each ball). As a result there are $\binom{q-1}{k-1}$ number of ways to choose the $\{\beta_{i_1}, \ldots, \beta_{i_k}\}$ once the index $\{i_1, \ldots i_k\}$ are chosen.

Step 4. Combining the previous three steps, it follows that for some constants $C_q, C_1 > 0$

only depending on q,

$$
\begin{aligned}
\mathbb{E}|W_t - W_t^*|^q \leq & \frac{1}{n^q}\mathbb{E}\Big\{\sum_{k=1}^{q} \sum_{\substack{\beta_1+\beta_2+\ldots+\beta_n=q \\ \beta=(\beta_1,\ldots,\beta_n),|\beta|_0=k,\beta\geq 0}} \binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \prod_{j=1}^{n} \big|(\delta_{t,i}^* - \delta_{t,i})K_h(x-x_{t,i})\big|^{\beta_j}\Big\} \\
\leq & \frac{1}{n^q}\sum_{k=1}^{q}\binom{n}{k}\binom{q-1}{k-1} q! h^{-d(q-k)} C_K^q \mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} \\
\leq & \frac{1}{n^q}\sum_{k=1}^{q} n^k C_q C_K^q h^{-d(q-k)}\mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} \\
\leq & C_1\mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \Big(\frac{1}{nh^d}\Big)^{q-2} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \\
\leq & C_1\mathbb{E}_\delta\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\} q\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\},
\end{aligned}
$$

where the second inequality is satisfied by step 3 and that $\binom{q}{\beta_1,\beta_2,\ldots,\beta_n} \leq q!$, while the third inequality is achieved by using that $\binom{n}{k}\binom{q-1}{k-1}q! \leq \binom{n}{k}C_q \leq n^k C_q$. Moreover, given that $\frac{1}{n^q}n^k h^{-d(q-k)} = \Big(\frac{1}{nh^d}\Big)^{q-k}$ the fourth inequality is obtained. The last inequality holds because if $\frac{1}{nh^d} \leq 1$, then $\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \leq q$, and if $\frac{1}{nh^d} \geq 1$, then $\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + \ldots + \Big(\frac{1}{nh^d}\Big) + 1\Big\} \leq q\Big(\frac{1}{nh^d}\Big)^{q-1}$. □

Lemma 6. *Suppose Assumption 1* **d** *holds. Let $\rho \leq T$ be such that $\rho n h^d \geq \log(T)$ and $T \geq 3$. Let $N \in \mathbb{Z}^+$ be such that $N \geq \rho$. Then, it holds that*

$$
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} W_t|^q\Big\}^{1/q} \leq N^{1/2}C\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2} + 1\Big\} + N^{1/q}C'\Big\{\Big(\frac{1}{nh^d}\Big)^{(q-1)/q} + 1\Big\}.
$$

Proof. We have that $q > 2$ and $E|W_1| < \infty$ by the use of Lemma 5. Then, making use of Theorem 1 of Liu et al. (2013), we obtain that

$$
\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} Z_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1\Big\{\sum_{j=1}^{N}\Theta_{j,2} + \sum_{j=N+1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& + N^{1/q}C_2\Big\{\sum_{j=1}^{N} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|W_1|^q\}^{1/q}\Big\},
\end{aligned}
$$

where $\Theta_{j,q} = \{\mathbb{E}(|W_j^* - W_j|^q)\}^{1/q}$. Moreover, we observe that since $\Theta_{j,2} \leq \Theta_{j,q}$ for any $q \geq 2$, it follows

$$
\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k} W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1\Big\{\sum_{j=1}^{\infty}\Theta_{j,q} + \{\mathbb{E}|W_1|^2\}^{1/2}\Big\} \\
& + N^{1/q}C_2\Big\{\sum_{j=1}^{\infty} j^{1/2-1/q}\Theta_{j,q} + \{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}
$$

Next, by the first part of Lemma 3,

$$
\Theta_{j,q}^q \leq C\mathbb{E}\big\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1} + 1\Big\}.
$$

Since we have that $N \geq \frac{1}{nh^d}$, the above inequality further implies that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1^{'}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}\Big\}^{1/q}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}\\
\leq & N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}\\
\leq & N^{1/2}C_1^{''}\Big\{\sum_{j=1}^{\infty}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(N\Big)^{1/2-1/q}\Big\}\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{'}\Big\{\sum_{j=1}^{\infty}j^{1/2-1/q}C\mathbb{E}\{\max_{i=1}^{n}|\delta_{t,i}-\delta_{t,i}^*|^q\}\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

From Assumption 1 **d**, the above inequality further implies that

$$\begin{aligned}
\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}W_t|^q\Big\}^{1/q} \leq & N^{1/2}C_1^{'''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{1/2}+1\Big\}+\{\mathbb{E}|W_1|^2\}^{1/2}\Big\}\\
&+N^{1/q}C_2^{''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}+\{\mathbb{E}|W_1|^q\}^{1/q}\Big\}.
\end{aligned}$$

By the second part of Lemma 3, it holds that

$$\Big\{\mathbb{E}\max_{k=1}^{N}|\sum_{t=1}^{k}Z_t|^q\Big\}^{1/q} \leq N^{1/2}C_1^{''''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)+1\Big\}^{1/2}\Big\}+N^{1/q}C_2^{'''}\Big\{1+\Big\{\Big(\frac{1}{nh^d}\Big)^{q-1}+1\Big\}^{1/q}\Big\}.$$

This immediately implies the desired result. □

D Additional Technical Results

Lemma 7. *Suppose that $f, g : [0,1]^d \to \mathbb{R}$ such that $f, g \in \mathcal{H}^r(L)$ for some $r \geq 1$ $L > 0$. Suppose in addition that $\{x_m\}_{m=1}^M$ is a collection of grid points randomly sampled from a density $u : [0,1]^d \to \mathbb{R}$ such that $\inf_{x\in[0,1]^d} u(x) \geq c_u > 0$. If $\|f-g\|_\infty \geq \kappa$ for some parameter $\kappa > 0$, then*

$$\mathbb{P}\Big\{\max_{m=1}^{M}|f(x_m)-g(x_m)| \geq \frac{3}{4}\kappa\Big\} \geq 1-\exp\big(-cM\kappa^d\big),$$

where c is a constant only depending on d.

Proof. Let $h = f - g$. Since $f, g \in \mathcal{H}^r(L)$, $h \in \mathcal{H}^r(L)$. Since $r \geq 1$, we have that

$$|h(x)-h(x')| \leq L|x-x'| \quad \text{for all} \quad x, x' \in [0,1]^d.$$

for some absolute constant $L > 0$. Let $x_0 \in [0,1]^d$ be such that

$$|h(x_0)| = \|h\|_\infty.$$

Then for all $x' \in B(x_0, \frac{\kappa}{4L}) \cap [0,1]^d$,

$$|h(x')| \geq |h(x_0)| - L|x_0 - x'| \geq \frac{3}{4}\kappa.$$

Therefore

$$\mathbb{P}\Big\{\max_{m=1}^{M}|f(x_m)-g(x_m)| < \frac{3}{4}\kappa\Big\} \leq P\Big(\{x_m\}_{m=1}^M \not\in B\big(x_0, \frac{\kappa}{4L}\big)\Big).$$

Since

$$P\Big(\{x_m\}_{m=1}^M \notin B(x_0, \frac{\kappa}{4L})\Big) = \Big\{1 - P\Big(x_1 \in B(x_0, \frac{\kappa}{4L})\Big)\Big\}^M \leq \Big(1 - \{\frac{c_u\kappa}{4L}\}^d\Big)^M \leq \exp\big(-Mc\kappa^d\big),$$

the desired result follows. □

Lemma 8. *Let $\mathcal{J}$ be defined as in Definition 1 and suppose Assumption 1* **e** *holds. Denote*

$$\zeta_k = \frac{9}{10}\min\{\eta_{k+1} - \eta_k, \eta_k - \eta_{k-1}\}\ k \in \{1, ..., K\}.$$

Then for each change-point η_k there exists a seeded interval $\mathcal{I}_k = (s_k, e_k]$ such that
a. *$\mathcal{I}_k$ contains exactly one change-point η_k;*
b. $\min\{\eta_k - s_k, e_k - \eta_k\} \geq \frac{1}{16}\zeta_k$*; and*
c. $\max\{\eta_k - s_k, e_k - \eta_k\} \leq \zeta_k$*;*

Proof. These are the desired properties of seeded intervals by construction. The proof is the same as theorem 3 of Kovács et al. (2020) and is provided here for completeness.

Since $\zeta_k = \Theta(T)$, by construction of seeded intervals, one can find a seeded interval $(s_k, e_k] = (c_k - r_k, c_k + r_k]$ such that $(c_k - r_k, c_k + r_k] \subseteq (\eta_k - \zeta_k, \eta_k + \zeta_k]$, $r_k \geq \frac{\zeta_k}{4}$ and $|c_k - \eta_k| \leq \frac{5r_k}{8}$. So $(c_k - r_k, c_k + r_k]$ contains only one change-point η_k. In addition,

$$e_k - \eta_k = c_k + r_k - \eta_k \geq r_k - |c_k - \eta_k| \geq \frac{3r_k}{8} \geq \frac{3\zeta_k}{32},$$

and similarly $\eta_k - s_k \geq \frac{3\zeta_k}{32}$, so **b** holds. Finally, since $(c_k - r_k, c_k + r_k] \subseteq (\eta_k - \zeta_k, \eta_k + \zeta_k]$, it holds that $c_k + r_k \leq \eta_k + \zeta_k$ and so

$$e_k - \eta_k = c_k + r_k - \eta_k \leq \zeta_k.$$

□

D.1 Univariate CUSUM

We introduce some notation for one-dimensional change-point detection and the corresponding CUSUM statistics. Let $\{\mu_i\}_{i=1}^n, \{\omega_i\}_{i=1}^n \subseteq \mathbb{R}$ be two univariate sequences. We will make the following assumptions.

Assumption 4 (Univariate mean change-points)**.** *Let $\{\eta_k\}_{k=0}^{K+1} \subseteq \{0, \dots, n\}$, where $\eta_0 = 0$ and $\eta_{K+1} = T$, and*

$$\omega_t \neq \omega_{t+1} \text{ if and only if } t \in \{\eta_1, ...\eta_K\},$$

Assume

$$\min_{k=1}^{K+1}(\eta_k - \eta_{k-1}) \geq \Delta > 0,$$
$$0 < |\omega_{\eta_{k+1}} - \omega_{\eta_k}| = \kappa_k \text{ for all } k = 1, \dots, K.$$

We also have the corresponding CUSUM statistics over any generic interval $[s, e] \subseteq [1, T]$ defined as

$$\widetilde{\mu}_t^{s,e} = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{i=s+1}^t \mu_i - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{i=t+1}^e \mu_i,$$
$$\widetilde{\omega}_t^{s,e} = \sqrt{\frac{e-t}{(e-s)(t-s)}}\sum_{i=s+1}^t \omega_i - \sqrt{\frac{t-s}{(e-s)(e-t)}}\sum_{i=t+1}^e \omega_i.$$

Throughout this section, all of our results are proven by regarding $\{\mu_i\}_{i=1}^T$ and $\{\omega_i\}_{i=1}^T$ as two deterministic sequences. We will frequently assume that $\widetilde{\mu}_t^{s,e}$ is a good approximation of $\widetilde{\omega}_t^{s,e}$ in ways that we will specify through appropriate assumptions.

Consider the following events

$$\mathcal{A}((s,e],\rho,\gamma)=\left\{\max_{t=s+\rho+1}^{e-\rho}|\widetilde{\mu}_t^{s,e}-\widetilde{\omega}_t^{s,e}|\leq\gamma\right\};$$

$$\mathcal{B}(r,\rho,\gamma)=\left\{\max_{N=\rho}^{T-r}\left|\frac{1}{\sqrt{N}}\sum_{t=r+1}^{r+N}(\mu_t-\omega_t)\right|\leq\gamma\right\}\bigcup\left\{\max_{N=\rho}^{r}\left|\frac{1}{\sqrt{N}}\sum_{t=r-N+1}^{r}(\mu_t-\omega_t)\right|\leq\gamma\right\}.$$

Lemma 9. *Suppose Assumption 4 holds. Let $[s,e]$ be an subinterval of $[1,T]$ and contain at least one change-point η_r with $\min\{\eta_r-s,e-\eta_r\}\geq cT$ for some constant $c>0$. Let $\kappa_{\max}^{s,e}=\max\{\kappa_p:\min\{\eta_p-s,e-\eta_p\}\geq cT\}$. Let*

$$b\in\arg\max_{t=s+\rho}^{e-\rho}|\widetilde{\mu}_t^{s,e}|.$$

For some $c_1>0$, $\lambda>0$ and $\delta>0$, suppose that the following events hold

$$\mathcal{A}((s,e],\rho,\gamma), \tag{36}$$

$$\mathcal{B}(s,\rho,\gamma)\cup\mathcal{B}(e,\rho,\gamma)\cup\bigcup_{\eta\in\{\eta_k\}_{k=1}^K}\mathcal{B}(\eta,\rho,\gamma) \tag{37}$$

and that

$$\max_{t=s+\rho}^{e-\rho}|\widetilde{\mu}_t^{s,e}|=|\widetilde{\mu}_b^{s,e}|\geq c_1\kappa_{\max}^{s,e}\sqrt{T} \tag{38}$$

If there exists a sufficiently small $c_2>0$ such that

$$\gamma\leq c_2\kappa_{\max}^{s,e}\sqrt{T}\quad\textit{and that}\quad\rho\leq c_2T, \tag{39}$$

then there exists a change-point $\eta_k\in(s,e)$ such that

$$\min\{e-\eta_k,\eta_k-s\}>c_3T\quad\textit{and}\quad|\eta_k-b|\leq C_3\max\{\gamma^2\kappa_k^{-2},\rho\},$$

where c_3 is some sufficiently small constant independent of T.

Proof. The proof is the same as that for Lemma 22 in Wang et al. (2020). □

Lemma 10. *If $[s,e]$ contain two and only two change-points η_r and η_{r+1}, then*

$$\max_{t=s}^{e}|\widetilde{\omega}_t^{s,e}|\leq\sqrt{e-\eta_{r+1}}\kappa_{r+1}+\sqrt{\eta_r-s}\kappa_r.$$

Proof. This is Lemma 15 in Wang et al. (2020). □

E Common Stationary Processes

Basic time series models which are widely used in practice, can be incorporated by Assumption 1b and c. Functional autoregressive model (FAR) and functional moving average model (FMA) are presented in examples 1 below. The vector autoregressive (VAR) model and vector moving average (VMA) model can be defined in similar and simpler fashions.

Example 1 (FMA and FAR)**.** *Let $\mathcal{L}=\mathcal{L}(H,H)$ be the set of bounded linear operators from H to H, where $H=\mathcal{L}_\infty$. For $A\in\mathcal{L}$, we define the norm operator $||A||_{\mathcal{L}}=\sup_{||\varepsilon||_H\leq1}||A\varepsilon||_H$. Suppose $\theta_1,\Psi\in\mathcal{L}$ with $\|\Psi\|_{\mathcal{L}}<1$ and $\|\theta_1\|_{\mathcal{L}}<\infty$.*

a) *For FMA model, let $(\varepsilon_t:t\in\mathbb{Z})$ be a sequence of independent and identically distributed random $\mathcal{L}_\infty$ functions with mean zero. Then the FMA time series $(\xi_j:j\in\mathbb{Z})$ of order 1 is given by the equation*

$$\xi_t=\theta_1(\varepsilon_{t-1})+\varepsilon_t=g(\ldots,\varepsilon_{-1},\varepsilon_0,\varepsilon_1,\ldots,\varepsilon_{t-1},\varepsilon_t). \tag{40}$$

For any $t\geq2$, by (40) *we have that*

$$\xi_t-\xi_t^*=0$$

and $\xi_1 - \xi_1^* = \theta_1(\varepsilon_0) - \theta_1(\varepsilon_0^{'})$*. As a result*

$$\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\xi_t - \xi_t^*||_\infty^q)^{1/q} = \mathbb{E}(||\xi_1 - \xi_1^*||_\infty^q)^{1/q} = \mathbb{E}(\|\theta_1(\varepsilon_0) - \theta_1(\varepsilon_0^{'})\|_\infty^q)^{1/q} < \infty.$$

Therefore Assumption 1b is satisfied by FMA models.

b) *We can define a FAR time series as*

$$\xi_t = \Psi(\xi_{t-1}) + \varepsilon_t. \tag{41}$$

It admits the expansion,

$$\begin{aligned}\xi_t =&\sum_{j=0}^{\infty} \Psi^j(\varepsilon_{t-j})\\ =&\Psi(\varepsilon_t) + \Psi^1(\varepsilon_{t-1}) + ... + \Psi^t(\varepsilon_0) + \Psi^{t+1}(\varepsilon_{-1}) + ...\\ =&g(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{t-1}, \varepsilon_t).\end{aligned}$$

Then for any $t \geq 1$*, we have that* $\xi_t - \xi_t^* = \Psi^t(\varepsilon_0) - \Psi^t(\varepsilon_0^{'})$*. Thus,*

$$\begin{aligned}\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\xi_t - \xi_t^*||_\infty^q)^{1/q} =&\sum_{t=1}^{\infty} t^{1/2-1/q}\mathbb{E}(||\Psi^t(\varepsilon_0) - \Psi^t(\varepsilon_0')||_\infty^q)^{1/q}\\ \leq&\sum_{t=1}^{\infty} t^{1/2-1/q}||\Psi||_{\mathcal{L}}^t\mathbb{E}(||\varepsilon_0 - \varepsilon_0'||_\infty^q)^{1/q} < \infty.\end{aligned}$$

Assumption 1b incorporates FAR time series.