Locally Adaptive Online Functional Data Analysis

Valentin Patilea* Jeffrey S. Racine†

April 25, 2024

Abstract

We consider the problem of building adaptive, rate optimal estimators for the mean and covariance functions of functional data objects in the context of streaming data. In general, functional data analysis requires nonparametric smoothing of noisy curves observed at a discrete set of design points, which may be measured with error. However, classical nonparametric smoothing methods (e.g., kernels, splines, etc.) assume that the degree of smoothness is known. In many applications functional data could be irregular (i.e., non-smooth and even perhaps nowhere differentiable) and, as well, the (ir)regularity of the underlying function could vary across its support. We contribute to the literature by providing estimators and inference procedures that use an iterative plug-in estimator of 'local regularity' which delivers a computationally attractive, recursive, online updating method that is well-suited to streaming data. We are also able to separate measurement error noise from irregularity. Theoretical support and Monte Carlo simulation evidence is provided, and code in the R language is available for the interested reader.

Key words: Adaptive estimator; Covariance function; Hölder exponent; Optimal smoothing
MSC2020: 62R10; 62G05; 62M99

*Ensai, CREST - UMR 9194, France; valentin.patilea@ensai.fr
†McMaster University, Department of Economics & Graduate Program in Statistics, racinej@mcmaster.ca

Contents

1 Introduction

Functional data consists of curves, surfaces, or essentially anything varying over a continuum. Often such data varies over time, though its domain is unrestricted and can include frequency, wavelength, etc. Functional data analysis (FDA) is the statistical analysis of samples of curves, i.e., samples of random variables taking values in spaces of functions. Functional data carry information along the curves and among the curves and, often, the goal is to estimate the mean and covariance functions of the sample of curves. However, we never know the curve values at all points, the points at which curves are available can differ across curves, and the curves may be measured with error, which makes nonparametric methods an essential component of sound statistical analysis in this setting. The mean and covariance functions are the building blocks of FDA, and are used in a variety of applications, such as in the analysis of functional magnetic resonance imaging (fMRI) data, in the analysis of gene expression data, and in the analysis of financial data. Functional data is pervasive, and methods for its analysis are rapidly evolving.

A number of recent literature surveys on FDA have been published including Wang et al. (2016), Aneiros et al. (2019), and Gertheiss et al. (2023). These surveys provide a comprehensive overview of the field, and highlight the importance of the mean and covariance functions in FDA. Recent textbooks on FDA include Ramsay and Silverman (2005) and Kokoszka and Reimherr (2017), which provide a detailed introduction to the field, and discuss the estimation of the mean and covariance functions in more detail.

Unlike survey settings where data is known, fixed, and available in its entirety, functional data streams are often 'online', which means that the data is arriving continuously in real-time. This stands in stark contrast to 'batch' data, where the data is known in advance and can be processed all at once. Online data presents a number of challenges, including the need to adapt to unknown smoothness and measurement error, and the need to estimate the mean and covariance functions in real-time. The challenge is to adapt existing methods to the online setting, and to develop new methods that are computationally efficient and can be implemented in real-time, ideally in one-pass of the online data.

Online computational concerns have been addressed in the recent literature, though mainly outside of the FDA setting. For example, Luo and Song (2020) explored generalized linear models for data streams using an incremental updating algorithm for the maximum likelihood estimator. In the related context of limited memory constraints, Chen et al. (2019) considered computationally efficient linear quantile regression analysis. Gu and Lafferty (2012) studied sequential local polynomial smoothing and pointed out that conventional bandwidth rules, such as cross-validation, are not practical in an online setting. They combine estimators using different bandwidths through an exponential weighting strategy, a form of model averaging used to deal with bandwidth uncertainty. Kong and Xia (2019) provide explicit relationships between the standard optimal (in the sense of AMSE) bandwidths of the Nadaraya-Watson and local linear estimators and their recursive (online) versions, presuming that the unknown curve is smooth and twice differentiable. Xue and Yao (2022) proposed a dynamic spline basis expansion method to estimate nonparametric regression functions which also presumes smoothness. Zhang and Simon (2023) proposed online nonparametric estimators applying the Mercer expansion of a kernel function and stochastic gradient descent, when the regression function is presumed to belong to a prespecified reproducing kernel Hilbert space. Quan and Lin (2024) proposed a one-pass estimator based on penalized orthogonal basis expansions and studied the trade-off between statistical efficiency and memory consumption of estimators.

In the FDA setting, however, computationally efficient online contributions are more limited. Recent contributions include Yang and Yao (2023), who studied mean and covariance estimation of functional data streams using nonparametric kernel methods. They focused on a pooling method, considered by Yao et al. (2005) and Zhang and Wang (2016), assuming that data is available in blocks (batches) of curves. Assuming that part of the information contained in historical data can be stored, they proposed generating a dynamic sequence of candidate bandwidths that includes the current on as well as candidates for the future globally optimal ones.

In this paper we are interested in the efficient estimation of FDA objects based on online (streaming) data feeds which, by its very nature, all but rules out methods that rely on batch computation including those of Yang and Yao (2023), Yao et al. (2005), Zhang and Wang (2016), etc. Furthermore, feasible estimation must be based upon recursive updating approaches whose complexity does not increase with the amount of data already processed. In this paper we shall focus on the problem of building rate optimal estimators for the mean and covariance functions of functional data objects in online settings. Obtaining rate optimal estimators requires methods that adapt to unknown smoothness and measurement error which, to the best of our knowledge, remains an unsolved problem.

Before proceeding, we should first ask why we should adapt to unknown smoothness and measurement error with online real-time data? Because nonparametric methods currently used for both classical and functional analysis rely crucially upon smoothness assumptions which requires the practitioner to assume something we do not know. Such assumptions manifest themselves in the bias terms present in assumed sampling distributions, and many working in the FDA field acknowledge that these smoothness assumptions are often at odds with the nature of the data being analyzed (Wang et al., 2016). Existing FDA implementations not only depend on such smoothness assumptions holding, but often assume that the functions being measured are free of measurement error. By way of illustration, the fdapace implementation (Zhou et al., 2022) is an open source and popular library for FDA analysis written in the R programming language, which presumes smoothness and for some functions (e.g., FCCor), no measurement error. As such, with surprisingly few exceptions, the current state of FDA involves two very strong assumptions that makes many practitioners uncomfortable, understandably so, which calls into question the soundness of much existing analysis.

Also, we should ask why online real-time estimation is necessary? Because existing implementations typically rely upon batch estimation of the objects of interest based on a set of pre-existing sample curves, which presents two issues, namely i) data-driven batch computation can be computationally demanding, and ii) streaming data necessitates a non-batch approach since the data naturally arrives in real-time. By way of illustration, using modern processors (e.g., 2024-era Intel chips) a popular current implementation in the R package fdapace will take roughly 4 days in order to compute estimators of the mean and covariance of 100 curves each having 100 measurements/curve using cross-validation for bandwidth selection on a modern desktop computer. But it is not uncommon to encounter functional settings in which the amount of data involved is orders of magnitude larger than this. We propose tackling computation issues using stochastic approximation algorithms such as Robbins and Monro (1951) and Kiefer and Wolfowitz (1952) which have withstood the test of time. Real-time estimation is crucial since much modern functional data is arriving continuously in real-time, hence any method relying on batch computation (e.g., all curves processed simultaneously as in the fdapace package) will be infeasible in many practical settings.

We propose 'adaptive', online, nonparametric kernel estimators in the context of functional data. The curves are observed on a discrete set of domain points and measured with error that can be heteroscedastic. Here, adaptiveness means that our approach is tailored to i) the purpose of the estimation (we focus on the mean and covariance functions), ii) the nature of the design (the curves can be sampled at the same domain points, or the sample points can can differ in number and location from curve to curve), iii) the conditional variance of the noise and, most importantly, iv) the regularity of the random process generating the curves. From a computational perspective, our approach does not require any storage whatsoever of past curves. Each new curve contributes to updating online estimators of local regularity along with bandwidth parameters used to smooth each new curve, and to updating the mean and covariance estimators. The local data-driven bandwidth is explicit and fully exploits the local regularity, an idea suggested by Golovkine et al. (2022) in the context of FDA. For non-differentiable sample paths, the local regularity is given by the local Hölder exponent and local Hölder constant defined using the mean squared increments of the random process generating the curves. With smooth sample paths, the definition extends from the increments of the sample paths to those of the highest order derivative they admit. A similar notion was studied by Hsing et al. (2016) in the context of dense spatial data obtained from one realization of a spatial process observed on a grid of points from a fixed domain. The asymptotic theory developed by Hsing et al. (2016) thus requires the grid size to tend to zero. The regularity of the processes generating the random curves is known to determine the minimax optimal rates of convergence for different FDA objects; see for example Cai and Yuan (2011), Cai and Yuan (2010) for the case of the mean and covariance functions, respectively.

As in Golovkine et al. (2022), the estimation of the local regularity we propose below exploits a key feature of functional data known as ‘replication’. For each domain point, using the local information available on past curves updated with the information on the current curve, we are able to construct simple and effective local regularity estimates. Knowing the local regularity allows us to formalize the risk bounds of the nonparametric estimators of the mean and covariance functions and to derive recursive, plug-in, bandwidth selection rules. Our recursive mean estimator achieves the optimal rate, see Cai and Yuan (2011), while the covariance function recursive estimator has the same rate as the adaptive ‘first smooth, then estimate’ batch estimator studied by Golovkine et al. (2023). However, unlike existing batch methods, our method allows us to process millions of curves on a laptop in essentially real-time. For simplicity, we focus on the case of non-differentiable sample paths, and there are now several prominent studies that advocate modelling functional data using non-differentiable curves, which we believe is fully justified; see, for example, Poß et al. (2020), Mohammadi et al. (2024), Mohammadi and Panaretos (2024), and Golovkine et al. (2023), by way of illustration.

The rest of this paper proceeds as follows. In section 2 we introduce the framework and some definitions, describe the data structure, and present the local regularity definition. Section 3 is dedicated to the online local regularity estimation where, for simplicity, we focus on the case of non-differentiable sample paths. The local regularity may change over the domain, which allows us to detect patterns that may be smoother on some parts of the domain and more irregular on others. We provide theoretical guarantees for our local regularity parameters, in the form of non-asymptotic exponential bounds for the uniform concentration of the Hölder exponent and Hölder constant estimators, respectively. The implementation of the local regularity estimators is based on an iterative procedure which appears to stabilize after only a few iterations and is computationally trivial. With local regularity in hand, in section 4 we propose a recursive kernel mean function estimator and derive its rate of convergence. The data-driven local bandwidth rule does not require any supplemental optimization and can be updated instantly. We also provide a pointwise asymptotic normality result, with a nearly optimal rate of convergence in distribution. In section 5 we propose a recursive kernel estimator of the covariance function, and the corresponding data-driven bandwidth rule. We conclude with a set of simulations in section 6 and offer a few summary remarks. The Appendix contains detailed proofs of the results on the mean function estimator, while further proofs and additional technical results are relegated to an auxiliary Supplementary Material document.

2 The framework

Consider a second-order stochastic process $X = (X_t : t \in \mathcal{T})$ with continuous trajectories, defined on a compact interval $\mathcal{T}$, for instance $\mathcal{T} = [0, 1]$. The mean and covariance functions are

$$\mu(t) = \mathbb{E}(X_t) \quad \text{and} \quad \Gamma(s,t) = \mathbb{E}\left\{[X_s - \mu(s)][X_t - \mu(t)]\right\} = \mathbb{E}\left(X_s X_t\right) - \mu(s)\mu(t), \qquad s,t \in \mathcal{T},$$

respectively. Our purpose is to build rate optimal estimators for $\mu(t)$ and $\Gamma(s,t)$ in the context of streaming data.

If independent realizations $X^{(1)}, \dots, X^{(i)}, \dots$ of X were observed, the ideal estimators would be

$$\overline{\mu}^{(i)}(t) = \frac{1}{i}\sum_{j=1}^{i} X_t^{(j)} \quad \text{and} \quad \overline{\Gamma}^{(i)}(s,t) = \frac{1}{i-1}\sum_{j=1}^{i}\{X_s^{(j)} - \overline{\mu}^{(i)}(s)\}\{X_t^{(j)} - \overline{\mu}^{(i)}(t)\}, \quad s,t \in \mathcal{T}. \tag{1}$$

The update of these estimators when new realizations are available is trivial. However, in applications, the curves are rarely observed without errors and never at each value $t \in \mathcal{T}$. In the following, we consider a more realistic setup, requiring more elaborate updating procedures for mean and covariance function estimation.

2.1 Data

We consider situations in which the independent sample paths (or realizations) $X^{(i)}$, $i = 1, 2 \dots$, of X are measured with errors at discrete times. More precisely, the data associated with the sample path

$X^{(i)}$ consists of the pairs $(Y_m^{(i)}, T_m^{(i)}) \in \mathbb{R} \times \mathcal{T}$, generated according to

$$Y_m^{(i)} = X^{(i)}(T_m^{(i)}) + \varepsilon_m^{(i)}, \qquad 1 \leq m \leq M_i. \tag{2}$$

Here, and in what follows, we use the notation $X_t^{(i)}$ for the value at a generic point $t \in \mathcal{T}$ of the realization $X^{(i)}$ of X, while $X^{(i)}(T_m^{(i)})$ denotes the measurement at $T_m^{(i)}$ of this realization. In (2), M_i is an integer which can be non random and common to several $X^{(i)}$, or can be a random copy of some variable M, drawn independently of X. The $T_m^{(i)}$ are the ordered measurement times for $X^{(i)}$. They can be randomly drawn from some distribution, independently of X and M. This case will be referred as the *independent design* case. The $T_m^{(i)}$ can also be restricted to a fixed grid of points, possibly the same for several curves $X^{(i)}$, which will be referred as the *fixed design* case. The $\varepsilon_m^{(i)}$ are the measurement errors and we assume that

$$\varepsilon_m^{(i)} = \sigma(T_m^{(i)}) e_m^{(i)}, \quad 1 \leq m \leq M_i,$$

where $e_m^{(i)}$ are independent copies of a centered variable e with unit variance, and $\sigma(t)$ is some unknown bounded function which accounts for possibly heteroscedastic measurement errors.

To build nonparametric estimates of $\mu(t)$ and $\Gamma(s,t)$ from the data points $(Y_m^{(i)}, T_m^{(i)})$, in particular when such observations arrive sequentially, one needs to use the local information carried by these pairs for filling in the missing information on $X_t^{(i)}$ and $X_s^{(i)}$. It is well-known in nonparametric statistics that the optimal localization depends on the regularity of the curves. We therefore consider a meaningful concept of regularity for functional data in the following section.

2.2 Local regularity

Our adaptive recursive approach relies on the notion of local regularity, considered by Golovkine et al. (2023) and Wang Guang Wei et al. (2023), which we describe below, while Hsing et al. (2016) considered a similar notion that they called *local intrinsic stationarity*.

For given Lipschitz continuous functions $H \in (0,1)$ and $L > 0$ defined on $\mathcal{T}$, the class $\mathcal{X}(H, L)$ is the set of second-order stochastic processes X defined on $\mathcal{T}$ for which constants $\Delta_0 > 0$, $S \geq 0$ and $\beta > 0$ exist such that

$$\left| \mathbb{E}\left[(X_u - X_v)^2\right] - L_t^2 |u-v|^{2H_t} \right| \leq S|u-v|^{2H_t + 2\beta}, \qquad \forall\, u \leq t \leq v \text{ with } |u-v| \leq \Delta_0, \forall t \in \mathcal{T}. \tag{3}$$

The function H is the local regularity exponent, while L determines the local Hölder constant. The class $\mathcal{X}(H, L)$ as defined by (3) is designed for processes with non-differentiable sample paths. The condition of a positive L function avoids (locally) constant sample paths, in which case H_t is not well defined. In the case of processes with sample paths admitting derivatives up to an integer order, say, $\alpha \geq 1$, the condition (3) has to be stated with the derivative of order α of the sample path of X at u and v in place of X_u and X_v, respectively. The local regularity is then $\alpha + H$. See Golovkine et al. (2022) for details. For simplicity, and because considering X with irregular sample paths seems appropriate in many applications, we focus on the case $X \in \mathcal{X}(H, L)$ with $\mathcal{X}(H, L)$ as defined by (3). The class $\mathcal{X}(H, L)$ is general, it does not require stationarity, stationary increments, or Gaussianity. Several examples are provided by Golovkine et al. (2022) and Wang Guang Wei et al. (2023), see also section 6 below.

The local regularity and the moment conditions on the increments of X determine the regularity of the sample paths which are Hölder continuous under mild conditions. For instance, the Brownian motion (Bm) has a constant local regularity equal to $1/2$ and moments of any order. With probability 1, the sample paths of the Bm belong to any Hölder space of local regularity less than $1/2$, but cannot be Hölder continuous with exponent greater than or equal to $1/2$. On the other hand, the sample paths regularity determines the minimax optimal rates for estimating the characteristics of X, such as their mean and covariance functions. See Cai and Yuan (2010), Cai and Yuan (2011), Wong and Zhang (2019). The replication feature of functional data allows us to estimate the local regularity, and thus devise adaptive, optimal rate, estimation methods. This idea was used by Golovkine et al. (2023) for mean and covariance estimation, and further exploited by Wang Guang Wei et al. (2023) in the context of functional principal component analysis. Here we reconsider the local regularity estimation and the adaptive mean and covariance function estimation in the context of streaming data, using a recursive data-driven approach.

3 Online local regularity estimation

We now introduce a recursive estimator of the functions H and L defining the local regularity of the processes generating the functional data. Let $X \in \mathcal{X}(H, L)$ and fix arbitrarily $t \in \mathcal{T}$. For $u, v \in \mathcal{T}$, let

$$\theta(u, v) = \mathbb{E}\left[(X_u - X_v)^2\right].$$

We first define a proxy function for H. Given the points $t_1, t_3 \in \mathcal{T}$, let $t_2 = (t_1 + t_3)/2$ and assume t belongs to the interval defined by t_1 and t_2. Define

$$\widetilde{H}_t = \frac{\log(\theta(t_1, t_3)) - \log(\theta(t_1, t_2))}{2\log(2)}, \qquad t \in \mathcal{T}. \tag{4}$$

If H is Lipschitz continuous and $|t_1 - t_3|$ is sufficiently small, the quantity $\widetilde{H}_t$ is a proxy for H_t. Moreover,

$$\widetilde{L}_t^2 = \frac{\theta(t_1, t_3)}{|t_1 - t_3|^{2\widetilde{H}_t}}, \qquad t \in \mathcal{T}, \tag{5}$$

is a proxy for L_t^2. We approach online estimation of the local regularity of X by plugging a recursive estimator of $\theta(\cdot, \cdot)$ into the proxy expression $\widetilde{H}_t$ (i.e., (4)), and we then plug this object and the recursive estimator of $\theta(\cdot, \cdot)$ into the proxy expression for the local Hölder constant L_t^2 (i.e., (5)).

Let $\kappa \geq 0$ and define

$$\gamma_i = \gamma_i(\kappa) = \frac{M_i^\kappa}{M_1^\kappa + \cdots + M_i^\kappa}.$$

When $\kappa = 0$ we get $\gamma_i = 1/i$, while $\kappa = 1$ yields $\gamma_i = M_i/(M_1 + \cdots + M_i)$. For any $t \in \mathcal{T}$ and u, v close to t, the online estimator of $\theta(u, v)$ is defined by the recurrence relationship

$$\widehat{\theta}^{(i)}(u, v) = (1 - \gamma_i)\widehat{\theta}^{(i-1)}(u, v) + \gamma_i \left(\widetilde{X}_u^{(i)} - \widetilde{X}_v^{(i)}\right)^2, \qquad i = 1, 2 \dots \tag{6}$$

Here, $\widetilde{X}_u^{(i)}$ is a pilot nonparametric estimator of $X_u^{(i)}$ built with $(Y_m^{(i)}, T_m^{(i)})$, $1 \leq m \leq M_i$. For the initial value $\widehat{\theta}^{(0)}(u, v)$, we use an estimate of the expected squared increment $X_u - X_v$ built using a batch sample (see Section B.7 for the details). Given the $\widehat{\theta}^{(i)}(\cdot, \cdot)$'s, a natural estimator of $\widetilde{H}_t$, and thus of H_t, is then

$$\widehat{H}_t^{(i)} = \frac{\log\left(\widehat{\theta}^{(i)}(t_1, t_3)\right) - \log\left(\widehat{\theta}^{(i)}(t_1, t_2)\right)}{2\log(2)}, \qquad t \in \mathcal{T},\ i = 1, 2 \dots \tag{7}$$

For implementation purposes, we confine the estimate of H_t to some interval $[\eta, 1 - \eta]$ for some small $\eta > 0$. That is, in applications we use $\min(\max(\widehat{H}_t^{(i)}, \eta), 1 - \eta)$ instead of $\widehat{H}_t^{(i)}$ defined in (7). At each stage i, with the estimate $\widehat{H}_t^{(i)}$ of H_t at hand, an estimate of L_t can be readily obtained as

$$\widehat{L}_t^{(i)} = \frac{\sqrt{\widehat{\theta}^{(i)}(t_1, t_3)}}{|t_1 - t_3|^{\widehat{H}_t^{(i)}}}, \qquad t \in \mathcal{T},\ i = 1, 2 \dots \tag{8}$$

The pilot estimator $\widetilde{X}^{(i)}$ can be built in many ways using, for example, kernel smoothing, splines, series estimators, and so forth. The only restriction is that the pilot estimate must satisfy some mild technical conditions which are stated in Section 3.2. We prefer the simplicity of (local constant) kernel regression for which the optimal bandwidth is easily available when the regularity is given (see Section 3.3 for the details).

3.1 Assumptions

To derive theoretical guarantees for the estimators $\widehat{H}^{(i)}$, $\widehat{L}^{(i)}$ of the functions H, L, we first impose the following assumptions on the class $\mathcal{X}(H,L)$ defined in (3).

(H1) The functions $H : u \mapsto H_u$ and $L : u \mapsto L_u$ are Lipschitz continuous, and

$$0 < \underline{H} := \min_{u\in\mathcal{T}} H_u \leq \max_{u\in\mathcal{T}} H_u =: \overline{H} < 1 \quad \text{and} \quad 0 < \underline{L} := \min_{u\in\mathcal{T}} L_u \leq \max_{u\in\mathcal{T}} L_u =: \overline{L} < \infty.$$

(H2) Constants $C > 0$, $\mathfrak{a} > 0$ and $\mathfrak{A} > 0$ exist such that

$$\mathbb{E}\left[|X_u - X_v|^4\right] \leq C\mathbb{E}^2\left[|X_u - X_v|^2\right], \quad \forall |u-v| \leq \Delta_0,$$

and

$$\sup_{u,v\in\mathcal{T},|u-v|\leq\Delta_0} \mathbb{E}\left[|X_u - X_v|^{2p}\right] \leq \frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2}, \qquad \forall p \geq 1.$$

To avoid technical problems with the definition of the local regularity, condition (H1) prevents H and L from approaching the boundary of their set of values. In (H2) we assume that the increments of X are sub-Gaussian. This allows us to derive simple exponential bounds for the concentration of the estimators of H_t and L_t, using Bernstein's inequality. Relaxing the sub-Gaussian condition is possible, at the expense of more intricate concentration bounds (see Maissoro et al., 2024). Imposing the condition that the increments of X have fourth order moments bounded by a constant times the square of the second order moments is also a mild simplifying assumption.

We next present the assumptions placed on the data generating process (DGP).

(D1) The integers $M_i \geq 2$ represent an independent sample of positive integer-valued random variables with finite expectations $\mathfrak{m}_i$.

(D2) For each i, the observation times $T_m^{(i)} \in \mathcal{T}$ are ordered. In the independent design case, all the $T_m^{(i)}$ are random realizations of a variable $T \in \mathcal{T}$, which admits a bounded, Hölder continuous density f_T, bounded away from zero on $\mathcal{T}$. In the fixed design case, for each i, the sequence of ratios $\max_m |T_m^{(i)} - T_{m-1}^{(i)}| / \min_m |T_m^{(i)} - T_{m-1}^{(i)}|$ is bounded.

(D3) The conditional standard deviation of the errors $t \mapsto \sigma(t)$, is a positive, Hölder continuous function on $\mathcal{T}$.

(D4) The realizations $X^{(1)}, \dots, X^{(i)}, \dots$ of X are independent, and the realizations of X, e, T and the M_i's are mutually independent.

Let $\mathcal{T}_{obs}^{(i)}$ denote the set of observation times $T_m^{(i)}$, $1 \leq m \leq M_i$, over the trajectory $X^{(i)}$. In the fixed design case, the $\mathcal{T}_{obs}^{(i)}$ is a grid of points, possibly the same for several i. If not stated otherwise, the issues discussed below apply to both independent design and fixed design cases. The theoretical arguments we provide in the following easily adapt to the case where the density of the independent $T_m^{(i)}$ changes with i, provided it remains bounded from above and below by the same positive constants. We claim that the recursive procedures we propose have also theoretical guarantees in the case of dependent functional data, for instance with weakly dependent functional time series as considered by Hörmann and Kokoszka (2010); see also see Maissoro et al. (2024). We defer investigation of this extension to future work.

3.2 Theoretical guarantees

The behavior of the estimates $\widehat{H}_t^{(i)}$ and $\widehat{L}_t^{(i)}$ depends on the quality of the nonparametric estimators $\widetilde{X}_u^{(i)}$ of $X_u^{(i)}$. To quantify their behavior, we consider the uniform $\mathbb{L}^p$-risk

$$R(p;m) = \mathbb{E}\left(\left|\widetilde{X}^{(i)} - X^{(i)}\right|_\infty^p \mid M_i = m\right), \quad p \geq 1.$$

Here, and in the following, $|\cdot|_\infty$ is the uniform norm for the continuous functions defined on $\mathcal{T}$. The uniform risk above is conditional on the number M_i of sample points observed on the ith curve. Since the realizations of X and the observation times T are independent of M, the risk $R(p;m)$ does not depend on i, but depends on the realization M_i. Any type of nonparametric estimator $\widetilde{X}$ can be used, provided that its uniform $\mathbb{L}^p$-risk is suitably bounded for $p=2$ and $p>2$, respectively. See Proposition 1 below for the required mild conditions.

We now state a non-asymptotic uniform concentration result for the estimators $\widehat{H}_t^{(i)}$ and $\widehat{L}_t^{(i)}$ obtained from an online sample of i independent curves. For this purpose, we need to define more precisely t_1 and t_3 in terms of $t\in\mathcal{T}$. Below, $a\vee b$ and $a\wedge b$ denote the maximum and the minimum of a and b, respectively. Following Wang Guang Wei et al. (2023), for some small $\Delta_*>0$, let

$$\underline{t}=0\vee(t-\Delta_*),\quad \bar{t}=(t+\Delta_*)\wedge 1,\quad \text{and}\quad (t_1,t_3)=(\underline{t},\bar{t}) \text{ if } t\leq 1/2 \text{ or } (t_1,t_3)=(\bar{t},\underline{t}) \text{ otherwise.}$$

Then, $t\in[t_1\wedge t_2, t_1\vee t_2]\subset[t_1\wedge t_3, t_1\vee t_3]$. Finally, we consider two mild and convenient technical conditions. First, we assume that a constant a exists such that:

$$0<a\leq\frac{\log(i)}{\log(M_i)}\leq a^{-1},\qquad \forall i>1. \tag{9}$$

At the expense of lengthy arguments, this condition can be relaxed, for instance replacing $\log(M_i)$ by $\log(\mathfrak{m}_i)$, where $\mathfrak{m}_i=\mathbb{E}(M_i)$, and adding some assumption on the probabilities in the tails of the distribution of $\log(M_i)/\log(\mathfrak{m}_i)$. However, for implementation purposes (9) does not represent a limiting constraint since in theory a is permitted to be arbitrarily small. The second technical condition helps to simplify the theoretical derivations: a constant C_κ exists such that

$$\frac{1}{i}\sum_{j=1}^{i}M_j^{2\kappa}\leq C_\kappa\left(\frac{1}{i}\sum_{j=1}^{i}M_j^{\kappa}\right)^2,\qquad \forall i>1. \tag{10}$$

Condition (10) is automatically fulfilled with $C_\kappa=1$ when $\kappa=0$. Conditions (9) and (10) are guaranteed, for instance, by the following one: a constant $c>0$ exists such that

$$0<\inf_{i\geq 1} i^{-c}M_i\leq\sup_{i\geq 1} i^{-c}M_i<\infty. \tag{11}$$

Proposition 1. *Assume that X belongs to $\mathcal{X}(H,L)$, and Assumptions (H1), (H2) and (D1) to (D4) hold true. Moreover,* (9) *holds true, and positive constants $\mathfrak{c}$ and $\mathfrak{C}$ exist such that, for any $p\geq 1$,*

$$R(2p;m)\leq\frac{p!}{2}\mathfrak{c}\mathfrak{C}^{p-2},\qquad \forall m\geq 1. \tag{12}$$

Assume also that constants $\tau>0, B>0$ exist such that

$$R(2;m)\leq Bm^{-\tau},\quad \forall m\geq 1. \tag{13}$$

Consider $\overline{\mathfrak{m}}_i=(\mathfrak{m}_1+\cdots+\mathfrak{m}_i)/i$, and let

$$\epsilon=\epsilon(\overline{\mathfrak{m}}_i)=\log^{-\varrho}(\overline{\mathfrak{m}}_i)\quad \text{and}\quad \Delta_*=\Delta_*(\overline{\mathfrak{m}}_i)=\exp(-\log^{\nu}(\overline{\mathfrak{m}}_i)), \tag{14}$$

for some $\varrho>1$ and $0<\nu<1$. Let $\widehat{H}^{(i)}$ and $\widehat{L}^{(i)}$ be defined as in (7) *and* (8)*, respectively, for some $\kappa\geq 0$ such that* (10) *holds true. Then, for any i larger than some i_0 depending on B, τ, ν, ϱ, H, β, L, κ, and for some positive constants $\mathfrak{f}$ and $\mathfrak{g}$ we have*

$$\mathbb{P}\left(\left|\widehat{H}^{(i)}-H\right|_\infty>\epsilon\right)\leq\exp\left(-\mathfrak{f}\times i\epsilon^2\Delta_*^4\right), \tag{15}$$

and

$$\mathbb{P}\left(\left|\widehat{L}^{(i)}-L\right|_\infty>\epsilon\right)\leq\exp\left(-\mathfrak{g}\times\frac{i\epsilon^2\Delta_*^4}{\log^2(\Delta_*)}\right). \tag{16}$$

A condition like (12) is satisfied by common nonparametric estimators given the realization of M_i. See for instance Theorem 1 in Gaïffas (2007) for the case of local polynomial smoothing. Condition (13) is also a mild uniform convergence condition satisfied by popular nonparametric regression estimators, given the number of points on a curve. See for instance Tsybakov (2009) and Belloni et al. (2015). In particular, the required conditions for the uniform risk of $\widetilde{X}^{(i)}$ can be obtained under general forms of heteroscedasticity and mild conditions on the distribution of the design variable T. Concerning the quantities $\Delta_*(m)$, $\epsilon(m)$ (defined as in (14) with m instead of $\overline{\mathfrak{m}}_i$), and the quantity $R(2;m)$, they are required to be such that, for some suitable $c>0$, $R(2;m)/\Delta_*^c(m)+\Delta_*^{1/c}(m)/\epsilon(m)$ becomes negligible as m increases. With our choices in (14), this holds true for any $c>0$. On the other hand, for the purpose of adaptive kernel smoothing for the mean and covariance function estimation, under the mild condition that $\log(i)/\log(\overline{\mathfrak{m}}_i)$ is bounded, which is implied by (9), the effect of estimating H is negligible as soon as ϵ is negligible compared to $\log^{-1}(\overline{\mathfrak{m}}_i)$. This explains our condition $\varrho>1$. Finally, the condition $\gamma<1$ combined with the fact that $\log(i)/\log(\overline{\mathfrak{m}}_i)$ stays away from zero, which is a consequence of (9), make the bounds for the concentration of $\widehat{H}$ and $\widehat{L}$ exponentially small when i increases. Concerning condition (10), by the Cauchy-Schwarz inequality, the mean of the $M_i^{2\kappa}$'s is always larger than the square of the mean of the M_i^{κ}'s. For simplicity, we impose that the ratio of these quantities remains bounded. If this is not the case, Proposition 1 remains true but with a different, more complicated expression of the exponential bounds. Detailed explanations are provided in the proof. In conclusion, the only practical choice we have to make is that of γ. Based on our extensive empirical studies, where the γ_i's in (6) are defined with $\kappa=1$ and $\overline{\mathfrak{m}}_i$ is simply estimated by $(M_1+\cdots+M_i)/i$, we set $\nu=1/3$.

3.3 Implementation aspects

The sequence of adaptive bandwidths we consider for building the local constant estimators $\widetilde{X}_u^{(i)}$ and $\widetilde{X}_v^{(i)}$, used to update $\widehat{\theta}^{(i)}(u,v)$ with u,v close to t, is defined as

$$h_t^{(i)}=\left[C_t^{(i)}\times\frac{1}{M_i}\right]^{\alpha_t^{(i)}},\qquad i=1,2,\ldots,\tag{17}$$

where

$$C_t^{(i)}=\frac{\{\widehat{\sigma}^{(i)}(t)\}^2\int K^2(u)du}{2\widehat{H}_t^{(i-1)}\{\widehat{L}_t^{(i-1)}\}^2\times\int|u|^{2\widehat{H}_t^{(i-1)}}|K(u)|du\times\widehat{f}_T^{(i)}(t)}\qquad\text{and}\qquad\alpha_t^{(i)}=\frac{1}{2\widehat{H}_t^{(i-1)}+1}.\tag{18}$$

Here, $\{\widehat{\sigma}^{(i)}(t)\}^2$ and $\widehat{f}_T^{(i)}(t)$ are positive estimates of $\sigma^2(t)$ and $f_T(t)$ respectively. The expression $h_t^{(i)}$ is a plug-in variant based upon a bandwidth that minimizes an average, pointwise quadratic risk bound. See Remark 2 in the Supplementary Material. The integrals involving the kernel function are easy to calculate. For instance, for the Epanechnikov kernel with support in $[-1,1]$,

$$\int K^2(u)du=\frac{3}{5}\qquad\text{and}\qquad\int|u|^{2H_t}|K(u)|du=\frac{3}{\{2H_t+1\}\{2H_t+3\}}.$$

The starting points $\widehat{H}_t^{(0)}$ and $\widehat{L}_t^{(0)}$ can be obtained from a batch sample, as we explain in the Supplementary Material. Let us point out that boundary points do not cause additional problems for the estimation of H and L when using a local constant pilot estimator $\widetilde{X}^{(i)}$, because its convergence rate does not deteriorate at those points.

Next, we turn to a discussion of the estimates $\widehat{\sigma}^{(i)}$ and $\widehat{f}_T^{(i)}$. For estimating the variance of the noise, without loss of generality we may assume that for every i, the design points $T_m^{(i)}$ are ordered. For $t\in\mathcal{T}$, we have

$$\mathbb{E}\left[\left(Y_m^{(i)} - Y_{m-1}^{(i)}\right)^2 \mid M_i, \mathcal{T}_{obs}^{(i)}\right] = \sigma^2(T_m^{(i)}) + \sigma^2(T_{m-1}^{(i)}) + \mathbb{E}\left[\left(X^{(i)}(T_m^{(i)}) - X^{(i)}(T_{m-1}^{(i)})\right)^2 \mid M_i, \mathcal{T}_{obs}^{(i)}\right]$$
$$\approx 2\sigma^2(t) + \left|T_m^{(i)} - T_{m-1}^{(i)}\right|^{2H_t},$$

provided that $T_m^{(i)}$ and $T_{m-1}^{(i)}$ are close to t. We then define

$$m_t^{(i)} = \arg\min_{2\le m\le M_i} \left\{\left|T_m^{(i)} - t\right| + \left|T_{m-1}^{(i)} - t\right|\right\}.$$

Moreover, let

$$\delta_t^{(i)} = 1 \quad \text{if} \quad \min_{2\le m\le M_i} \left\{\left|T_m^{(i)} - t\right| + \left|T_{m-1}^{(i)} - t\right|\right\} \le 1/\log(M_i), \qquad \text{and } \delta_t^{(i)} = 0 \quad \text{otherwise.}$$

Here, we introduce a threshold for deciding if $T_m^{(i)}$ and $T_{m-1}^{(i)}$ are sufficiently close to t, and thus $Y_m^{(i)}$ and $Y_{m-1}^{(i)}$ are useful for estimating $\sigma^2(t)$. The logarithmic decrease of the threshold is sufficient for guaranteeing the consistency of the variance estimator. Given the observations corresponding to $X^{(i)}$, the update of the recursive estimator of the noise variance is

$$\widehat{\sigma}^{(i)}(t)^2 = \frac{i-\delta_t^{(i)}}{i} \times \widehat{\sigma}^{(i-1)}(t)^2 + \frac{\delta_t^{(i)}}{i} \times \frac{1}{2}\left\{Y_{m_t^{(i)}}^{(i)} - Y_{m_t^{(i)}-1}^{(i)}\right\}^2, \qquad t\in\mathcal{T}, i = 1,2\dots \tag{19}$$

This estimator is updated with the information on curve i *only if* there are design points sufficiently close to t. For starting the recursion, we use an initial batch estimate of $\sigma^2(t)$; see the Supplementary Material for the details. The estimate $\widehat{f}_T^{(i)}(t)$ can be defined in many ways, the only requirement being that it is bounded from below and from above by positive constants with high probability. For computational simplicity and to avoid introducing unwanted boundary effects due to the bounded support of T, we use a computationally tractable non-smooth nonparametric histogram density estimator $\widehat{f}_T^{(j)}$ for each curve $X^{(j)}$, $1\le j\le i$, with data-driven selection of bin widths according to classical rules such as those provided in Scott (1979) or Freedman and Diaconis (1981). Then $\widehat{f}_T^{(i)}$ is simply the sum of all $j^{-1}\widehat{f}_T^{(j)}$, $1\le j\le i$.

4 Online mean function estimation

For the mean function, we consider a recursive version of the 'first smooth, then estimate' approach, as considered for instance by Golovkine et al. (2023) in the case of batch data. For any $t\in\mathcal{T}$, and bandwidth h, let

$$\widehat{X}_t^{(i)}(h) = \sum_{m=1}^{M_i} W_m^{(i)}(t;h) Y_m^{(i)}, \qquad \text{with} \qquad \sum_{m=1}^{M_i} W_m^{(i)}(t;h) = 1, \tag{20}$$

be a generic kernel-based nonparametric estimator of $X_t^{(i)}$. The weights $W_m^{(i)}$ are defined as functions of the elements in $\mathcal{T}_{obs}^{(i)}$, which depend on a bandwidth updated at each iteration that will be allowed to vary with t. In the case of local regularity smaller than 1 and thus non-differentiable sample paths $X^{(i)}$, the $\widehat{X}_t^{(i)}(h)$ we consider is the local constant (Nadaraya-Watson) estimator with the weights

$$W_m^{(i)}(t;h) = K\left(\frac{T_m^{(i)}-t}{h}\right)\left[\sum_{m'=1}^{M_i} K\left(\frac{T_{m'}^{(i)}-t}{h}\right)\right]^{-1}, \quad 1\le m\le M_i.$$

Here, K is a symmetric, non-negative, bounded kernel with compact support in $[-1, 1]$, and the convention $0/0 = 0$ applies. The pointwise indicator

$$w^{(i)}(t; h) = 1 \quad \text{if} \quad \sum_{m=1}^{M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\} > 1, \qquad \text{and} \qquad w^{(i)}(t; h) = 0 \quad \text{otherwise}, \tag{21}$$

will serve to count the non-degenerate estimates, by which we mean those estimates $\widehat{X}_t^{(i)}(h)$ for which at least one $W_m^{(i)}(t; h)$, $1 \leq m \leq M_i$ is not equal to zero.

We first investigate the independent design case, i.e., where the $T_m^{(i)}$ are independent realizations of T which has the support $\mathcal{T}$. Let $h_{\mu,t}^{(i)}$ denote the bandwidth for smoothing curve i at $t \in \mathcal{T}$. We here propose a simple, adaptive rule for the bandwidth $h_{\mu,t}^{(i)}$ based on the online estimates of the regularity proposed above, which does not require revisiting past observations. The bandwidth rule is

$$h_{\mu,t}^{(i)} = h_{\mu,t}^{(i)}(\zeta) = \left[\frac{C_{\mu,t}^{(i)}}{i}\right]^{\zeta}, \qquad i = 1, 2 \ldots, \tag{22}$$

where ζ is a positive exponent to be determined as a function of the rate of increase of M_i and the local regularity H_t, and $C_{\mu,t}^{(i)}$ is an element of a bounded sequence of positive numbers which stay away from zero. Practical choices for $C_{\mu,t}^{(i)}$ are discussed in Section 4.1.

The recursive estimator of the mean function is then

$$\widehat{\mu}^{(i)}(t; \zeta) = \frac{\mathcal{W}^{(i-1)}(t)}{\mathcal{W}^{(i)}(t)} \times \widehat{\mu}^{(i-1)}(t; \zeta) + \frac{1}{\mathcal{W}^{(i)}(t)} \times \widehat{X}_t^{(i)}\left(h_{\mu,t}^{(i)}\right), \tag{23}$$

with

$$\mathcal{W}^{(i)}(t) = \mathcal{W}^{(i-1)}(t) + w^{(i)}\left(t; h_{\mu,t}^{(i)}\right), \qquad t \in \mathcal{T}, i = 1, 2 \ldots, \tag{24}$$

with $\widehat{X}_t^{(i)}(\cdot)$ and $w^{(i)}(\cdot, \cdot)$ defined as in (20) and (21), respectively, and some suitable initial values $\mathcal{W}^{(0)}(t)$, $\widehat{\mu}^{(0)}(t)$. See the Supplementary Material for some proposals. Let us note that $\mathcal{W}^{(i)}(t) - \mathcal{W}^{(0)}(t)$ counts among the first i online curves those used for mean estimation at point t. By construction, $\mathcal{W}^{(i)}(t) - \mathcal{W}^{(0)}(t) \leq i$. The sparser are the curves being sampled around t, the smaller is $\mathcal{W}^{(i)}(t) - \mathcal{W}^{(0)}(t)$.

Typically, in nonparametric statistics, bandwidth selection proceeds by minimizing a quadratic risk between the estimator and the target, and thus balances squared bias, which depends on the regularity of the curve, and variance. This risk minimization results in a bandwidth sequence which decreases as fast as a negative power of the sample size, the negative power being determined by the regularity of the curve. See, for instance, Tsybakov (2009). In our context of online mean estimation, when H_t and L_t are given, we are able, up to some constant, to derive a bound for the quadratic risk

$$\mathcal{R}_\mu^{(i)}(t; \zeta) = \mathbb{E}\left[\left\{\widehat{\mu}^{(i)}(t; \zeta) - \mu(t)\right\}^2\right].$$

We will show that, in addition to the usual squared bias and variance terms, this bound contains a penalty for the number curves dropped in the mean estimation. See also Golovkine et al. (2023) for similar findings in the batch estimation of the mean. Finally, for each $t \in \mathcal{T}$, we select ζ_t^* to minimize the risk bound of $\mathcal{R}_\mu^{(i)}(t; \zeta)$. We next derive the pointwise rate of convergence for our recursive, nonparametric estimator of the mean function obtained with the bandwidth decreasing as fast as $i^{-\zeta_t^*}$. These facts lead us to the following result. Below, $\lesssim$ means that the left side is bounded by a positive constant times the right side.

Proposition 2. *Let $t \in \mathcal{T}$, assume that X belongs to $\mathcal{X}(H, L)$, and assume that the conditions of Proposition 1 hold true with an independent design. Moreover, assume that the condition (11) holds true.*

Let $\widehat{\mu}^{(i)}(t;\zeta)$ be defined as in (23), with the bandwidth of the local constant smoother defined in (22) for ζ such that

$$1 + c - \zeta > 0. \tag{25}$$

Then

$$\mathcal{R}_{\mu}^{(i)}(t;\zeta) \lesssim i^{-\min\{1,2\zeta H_t,1+c-\zeta)]} + i^{-(1+c-\zeta)} + i^{-[1+\min(0,c-\zeta)]} + i^{-1}, \tag{26}$$

provided $2cH_t \neq 1$. The minimum of the risk bound $\mathcal{R}_{\mu}^{(i)}(t;\zeta)$ is attained at

$$\zeta_t^* = \max\left\{c, \frac{c+1}{2H_t+1}\right\}, \tag{27}$$

which yields

$$\mathbb{E}\left[\left\{\widehat{\mu}^{(i)}(t;\zeta_t^*) - \mu(t)\right\}^2\right] \lesssim \min\left\{i^{-\{2(c+1)H_t/(2H_t+1)\}},\ i^{-2cH_t}\right\} + i^{-1}. \tag{28}$$

In the proof of Proposition 2, we decompose the risk $\mathcal{R}_{\mu}^{(i)}(t;\zeta)$ into four terms corresponding to the squared bias, the variance, the penalty for the number of curves dropped, and the quadratic risk on the infeasible empirical mean estimator, respectively. The rates of these terms yield the risk bound in (26). The condition (25) is needed to guarantee that the variance of the estimator tends to zero, and ζ_t^* satisfies this condition. We simplified the statement of Proposition 2 by avoiding the case in which $2cH_t = 1$, which represents the frontier between the so-called *sparse* regime ($2cH_t < 1$) and the *dense* regime ($2cH_t > 1$); see Zhang and Wang (2016) for this terminology. When $2cH_t = 1$, the squared bias term is inflated by a $\log(i)$ factor, and the optimal exponent ζ_t^* no longer has an explicit expression. The impact on the optimal risk rate in (28) would be a $\log(i)$ factor. The details are given in the proof of Proposition 2.

The risk bound we derive in Proposition 2 is sharp. In particular, our estimator achieves the minimax rates. More precisely, we have

$$2cH_t > 1 \Rightarrow i^{-2cH_t} \ll i^{-\frac{2(c+1)H_t}{2H_t+1}} \ll i^{-1} \qquad \text{and} \qquad 2cH_t < 1 \Rightarrow i^{-1} \ll i^{-\frac{2(c+1)H_t}{2H_t+1}} \ll i^{-2cH_t},$$

which implies

$$\mathbb{E}\left[\left\{\widehat{\mu}^{(i)}(t;\zeta_t^*) - \mu(t)\right\}^2\right] \lesssim i^{-\{2(c+1)H_t/(2H_t+1)\}} + i^{-1},$$

and the right-hand side has the minimax rate for the mean function estimation, as derived by Cai and Yuan (2011). Defining ζ_t^* as in (27), and not simply setting it equal to $(c+1)/(2H_t+1)$, does not change the asymptotic results, but will be helpful when discussing the implementation aspects.

Next, we derive the asymptotic distribution of our online mean function pointwise estimation procedure. Usually, the rate of convergence in distribution for an estimator is given by its standard deviation, which is in general a negative power of the sample size. For parametric models, the common power is $-1/2$. In nonparametric kernel regression, the standard deviation of pointwise estimation is the power $-1/2$ of the so-called effective sample size, that is, the expected number of observations in the neighborhood defined by the bandwidth. Assuming the univariate regression has continuous second order derivative, and selecting the bandwidth by balancing the squared bias and the variance, the standard deviation for kernel estimator has the rate given by the power $-2/5$ of the sample size. If instead the regression is just Hölder continuous with exponent H, the power obtained by balancing the squared bias and the variance becomes $-H/(2H+1)$. Let us point out that wrongly assuming a higher regularity than the true one breaks convergence in distribution, because the bias term times the square root of the effective sample size will tend to infinity. Disposing of a lower bound for the regularity of the curve to be estimated allows us to establish asymptotic distribution results which are robust to failures arising from a large bias.

In the functional data setting, the regularity of the mean function is necessarily equal to or larger than that of the sample paths. In particular, a consequence is that the minimax optimal rates for the mean function estimation are given by the sample path regularity; see Cai and Yuan (2011). This means that, from the minimax optimality perspective, the rate of convergence in distribution for the mean function estimation needs to take into account the sample path regularity. In what follows, we leverage this idea in the streaming data framework, and prove the asymptotic normality of our estimator with an adaptive, nearly optimal rate. More precisely, in order to make the bias term negligible and simplify the following proposition, we impose a slower rate on the bandwidth than the largest that would be required in order to obtain asymptotic normality. Moreover, we focus on the case $0 < 2cH_t < 1$, that is, the case we define as the *sparse regime* for streaming data where $M_i^{2H_t} \ll i$. This case corresponds to the most realistic situation with large streaming samples of curves. To the best of our knowledge, the next proposition is the first result providing adaptive inference for the mean function estimation with streaming data.

Proposition 3. *Let $t \in \mathcal{T}$ and assume that the conditions of Proposition 2 hold true, and $0 < c < (2H_t)^{-1}$. If $\zeta_t^* < \zeta < (2H_t)^{-1}$ for ζ_t^* in (27), then*

$$\sqrt{\mathcal{W}^{(i)}(t)}\left(\widehat{\mu}^{(i)}(t;\zeta) - \mu(t)\right) \xrightarrow{d} N\left(0, \operatorname{Var}(X_t) + \sigma^2(t)\right),$$

where $\widehat{\mu}^{(i)}(t;\zeta)$ and $\mathcal{W}^{(i)}(t)$ are defined as in (23) and (24), respectively, with $h_{\mu,t}^{(i)}(\zeta)$ defined in (22).

4.1 Implementation details

The sequence of constants $C_{\mu,t}^{(i)}$ can be traced to the numerous inequalities we derive in the proof of Proposition 2. These constants depend on H_t, L_t^2, the value of the density function $f_T(t)$, the variance of X_t and the error variance $\sigma^2(t)$, the upper and lower bounds in (11), and the kernel function. To simplify the procedure, in line with (18) and for the Epanechnikov kernel, we propose accounting for the density function value, the process and error variances and the regularity parameters, and to set

$$C_{\mu,t}^{(i)} = \frac{\left\{\widehat{\sigma}^{(i)}(t)\right\}^2 + \widehat{\operatorname{Var}(X_t)}}{\widehat{f}_T^{(i)}(t)} \times \frac{\left(2\widehat{H}_t + 1\right)\left(2\widehat{H}_t + 3\right)}{10} \times \frac{\log_2(iM_i)}{\widehat{H}_t^{(i)}\left\{\widehat{L}_t^{(i)}\right\}^2}, \qquad i \geq 1, \tag{29}$$

where $\{\widehat{\sigma}^{(i)}(t)\}^2$, $\widehat{H}_t^{(i)}$, $\{\widehat{L}_t^{(i)}\}^2$ and $\widehat{f}_T^{(i)}(t)$ are the estimates defined in Section 3.3, $\widehat{\operatorname{Var}(X_t)}$ is some nonparametric estimator of the variance of X_t, and $\log_2(\cdot) = \log\log(\cdot)$. Compared to (18), the additional term occurring in (29) corresponding to the estimate of $\operatorname{Var}(X_t)$ is due to the fact that here the target is the mean function, and for each data point $(Y_m^{(i)}, T_m^{(i)})$, $Y_m^{(i)} - \mu(T_m^{(i)})$ incorporates both the error term and the sample path variation. The $\log_2(iM_i)$ factor accounts for all the quantities which, for simplicity, are not included in the definition of $C_{\mu,t}^{(i)}$. The theoretical impact of this choice is a slight deterioration of the rate of convergence by a log-log factor when $2cH_t \leq 1$. The numerical experiments reveal that this simple choice performs well.

Given the optimal choice ζ_t^* in (27), let us note that

$$i^{-\zeta_t^*} = \left[\max(M_i^{2H_t}, i) \times M_i\right]^{-\frac{1}{2H_t+1}},$$

provided $M_i = i^c$. Thus the bandwidth sequence $h_{\mu,t}^{(i)}$ from (22) can be simply defined using the M_i's. At early stages of the recursions, $i^{-\zeta_t^*} = 1/M_i$ which basically means the bandwidth rule likely leads to using one or two data points from each curve. As i grows larger, the bandwidth becomes smaller and this guarantees a smaller bias at the cost of dropping more curves.

Combining results, our adaptive, online bandwidth rule becomes

$$\widehat{h}_{\mu,t}^{(i)} = \left[\frac{C_{\mu,t}^{(i)}}{\max\left(M_i^{2\widehat{H}_t}, i\right) \times M_i}\right]^{\frac{1}{2\widehat{H}_t+1}}, \qquad i = 1, 2\ldots, \tag{30}$$

with the online regularity estimate $\widehat{H}_t^{(i)}$ as defined in (7). Let us point out that, by Proposition 1, for any $a>1$, the event $\{|\widehat{H}_t^{(i)}/H_t - 1| > \log^{-a}(i)\}$ has exponentially small probability. Moreover, $i^{1/\log^a(i)} \to 1$, $\forall a>1$. As a consequence, the online mean estimator obtained with the bandwidth in (36) instead of the bandwidth in (22) with the optimal ζ_t^* in (27), has the same rate as stated in (28). The arguments are similar to those in the batch estimation, see Golovkine et al. (2023) and Wang Guang Wei et al. (2023), and are thus omitted. Finally, to determine a practical bandwidth allowing for inference using Proposition 3, it suffices to raise the bandwidth in (36) to some power larger than but close to 1, for instance 1.1.

4.2 The fixed design case

We now consider the fixed design case. Let us recall that by fixed design we mean that, for a positive integer j, the set of possible values for $T_m^{(j)}$ is a given grid, which may depend on j. For instance, it can be the equidistant grid of size M_j, or it can be a same fixed grid for several values of j. Here, for simplicity, we investigate the latter case with a fixed grid for all j less than or equal to i. For the theory, the size of the grid for $j \leq i$ is given by M_i satisfying (11) for some $c>0$.

The fixed design case with an equidistant grid on each curve is quite different from the independent design case. Regardless of the method used to smooth the curves, the bandwidth cannot be arbitrarily small as this will result in a degenerate smoother. Even in the case of a batch sample, arbitrarily small bandwidths will lead to degenerate mean function estimates. In our case, that means the optimization of the risk with respect to ζ defining the rate of the bandwidth in (22) is constrained to values $\zeta \leq c$, and we can thus consider $\mathcal{W}^{(i)}(t) \asymp_{\mathbb{P}} i$. The proof of the following result is similar to that of Proposition 2, and is thus omitted.

Proposition 4. *Assume the conditions of Proposition 2 in the case of fixed, equidistant design with $|T_m^{(j)} - T_{m-1}^{(j)}| \asymp i^{-c}$. Let $\widehat{\mu}^{(i)}(t;\zeta)$ be defined as in (23), with the bandwidth of the local constant smoother defined in (22) for some $\zeta \leq c$. Then*

$$\mathcal{R}_\mu^{(i)}(t;\zeta) \lesssim i^{-2\zeta H_t} + i^{-(1+c-\zeta)} + i^{-1}.$$

The minimum of the risk bound $\mathcal{R}_\mu^{(i)}(t;\zeta)$ is attained at

$$\zeta_t^* = \min\left\{c, \frac{c+1}{2H_t+1}\right\},$$

$$\mathbb{E}\left[\left\{\widehat{\mu}^{(i)}(t;\zeta_t^*) - \mu(t)\right\}^2\right] \lesssim \max\left\{i^{-2cH_t},\ i^{-\{2(c+1)H_t/(2H_t+1)\}}\right\} + i^{-1} = i^{-2cH_t} + i^{-1}.$$

5 Online covariance function estimation

For a generic bandwidth h, let

$$w^{(i)}(s,t;h) = w^{(i)}(s;h)w^{(i)}(t;h), \quad s,t \in \mathcal{T}. \tag{31}$$

To define a recursive nonparametric estimator of the covariance function, we first define the recursive estimator of

$$\gamma(s,t) = \mathbb{E}\left[X_s X_t\right] \qquad \text{for} \ \ s \neq t,$$

that is outside the diagonal subset of $\mathcal{T} \times \mathcal{T}$. With $\mathcal{W}^{(0)}(s,t)$ and $\widehat{\gamma}^{(0)}(s,t)$ some initial values, let

$$\widehat{\gamma}^{(i)}(s,t) = \frac{\mathcal{W}^{(i-1)}(s,t)}{\mathcal{W}^{(i)}(s,t)} \times \widehat{\gamma}^{(i-1)}(s,t) + \frac{1}{\mathcal{W}^{(i)}(s,t)} \times \widehat{X}_s^{(i)}\left(h_{\gamma,(s,t)}^{(i)}\right)\widehat{X}_t^{(i)}\left(h_{\gamma,(s,t)}^{(i)}\right), \tag{32}$$

where

$$\mathcal{W}^{(i)}(s,t) = \mathcal{W}^{(i-1)}(s,t) + w^{(i)}\left(s,t;h^{(i)}_{\gamma,(s,t)}\right), \qquad s,t\in\mathcal{T}, s\neq t, i=1,2\dots,$$

with $h^{(i)}_{\gamma,(s,t)}$ a sequence of bandwidths to be defined. Here, $\mathcal{W}^{(i)}(s,t)-\mathcal{W}^{(0)}(s,t)$ counts the number of online curves used for the estimation of $\widehat{\gamma}^{(i)}(s,t)$. Thus, by construction $\mathcal{W}^{(i)}(s,t)-\mathcal{W}^{(0)}(s,t)\leq i$, and the sparser are the curves sampled around (s,t), the smaller is $\mathcal{W}^{(i)}(s,t)-\mathcal{W}^{(0)}(s,t)$. Simple choices for the initial values $\mathcal{W}^{(0)}(s,t)$ and $\widehat{\gamma}^{(0)}(s,t)$ are discussed in the Supplementary Material.

We now consider the bandwidth rule for $s\neq t$, that is

$$h^{(i)}_{\gamma,(s,t)} = h^{(i)}_{\gamma,(s,t)}(\lambda) = \left[\frac{D^{(i)}_{\gamma,(s,t)}}{i}\right]^{\lambda}, \qquad i=1,2\dots, \tag{33}$$

with λ is a positive exponent to be determined as a function of the rate of increase of M_i and the local regularity H_s and H_t, and $D^{(i)}_{\mu,(s,t)}$ is an element of a bounded sequence of positive numbers which stay away from zero. Practical choices for $D^{(i)}_{\gamma,(s,t)}$ are discussed in Section 4.1.

Let $\widehat{\gamma}^{(i)}(s,t;\lambda)$ be the estimate defined in (32) with the bandwidths (33). As was the case for the mean function estimation, we are able, up to some constant, to derive a bound for the quadratic risk

$$\mathcal{R}^{(i)}_{\gamma}(s,t;\lambda) = \mathbb{E}\left[\left\{\widehat{\gamma}^{(i)}(s,t;\lambda)-\gamma(s,t)\right\}^2\right], \qquad i\geq 1, \quad s\neq t.$$

The new bound also includes a penalty for the number of curves dropped in the $\gamma(\cdot,\cdot)$ function estimation. For each $s\neq t$, we next select λ^*_t to minimize the risk bound of $\mathcal{R}^{(i)}_{\gamma}(s,t;\lambda)$. We derive the pointwise rate of convergence for this recursive, nonparametric estimator of the $\gamma(s,t)$ function outside the diagonal.

Proposition 5. *Let $s\neq t$, $H_{s,t}=\min\{H_s,H_t\}$ and assume the conditions of Proposition 2 hold true with an independent design. Let $\widehat{\gamma}^{(i)}(s,t;\lambda)$ be defined as in (32), with the bandwidth of the local constant smoother defined in (33) for λ such that*

$$1+2(c-\lambda)>0. \tag{34}$$

Then,

$$\mathcal{R}^{(i)}_{\gamma}(s,t;\lambda) \lesssim i^{-\min\{1,2\lambda H_{s,t},1+2(c-\lambda)\}} + i^{-[1+\min\{(c-\lambda),2(c-\lambda)\}]}i^{-[1+\min\{0,2(c-\lambda)\}]} + i^{-1}, \qquad i\geq 1,$$

provided $2cH_t\neq 1$. The minimum of the risk bound $\mathcal{R}^{(i)}_{\gamma}(s,t;\lambda)$ is attained at

$$\lambda^*_{s,t} = \max\left\{c, \frac{2c+1}{2(H_{s,t}+1)}\right\}, \tag{35}$$

which yields

$$\mathbb{E}\left[\left\{\widehat{\gamma}^{(i)}(s,t;\lambda^*_{s,t})-\gamma(s,t)\right\}^2\right] \lesssim \min\left\{i^{-(2c+1)H_{s,t}/(H_{s,t}+1)},\ i^{-2cH_{s,t}}\right\} + i^{-1}, \qquad s\neq t.$$

Finally, using this and the mean function estimator introduced in Section 4, we define the estimator of the covariance function $\Gamma(s,t)=\gamma(s)\gamma(t)-\mu(s)\mu(t)$, outside the diagonal, as

$$\widehat{\Gamma}^{(i)}(s,t) = \widehat{\Gamma}^{(i)}(s,t;\lambda^*_{s,t}) = \widehat{\gamma}^{(i)}(s,t;\lambda^*_{s,t}) - \widehat{\mu}^{(i)}(s;\lambda^*_{s,t})\widehat{\mu}^{(i)}(t;\lambda^*_{s,t}), \qquad s\neq t.$$

The mean estimators are computed as in (23) with the bandwidth $h^{(i)}_{\gamma,(s,t)}(\lambda^*_{s,t})$.

Corollary 1. *Assume the conditions of Proposition 5 hold true. Then*

$$\mathbb{E}\left[\left\{\widehat{\Gamma}^{(i)}(s,t;\lambda_t^*)-\Gamma(s,t)\right\}^2\right]\lesssim i^{-(2c+1)H_{s,t}/(H_{s,t}+1)}+i^{-1}.$$

Condition (34) is needed to guarantee that the variance of the estimator of the function $\gamma(s,t)$ tends to zero, and λ_t^* satisfies this condition. In the 'sparse' case ($2cH_{s,t}<1$), our optimal bandwidth for the estimation of $\gamma(s,t)$ has the same rate as the one of Golovkine et al. (2023) for batch estimation, and consequently the rate of the quadratic risk $\mathcal{R}_\gamma^{(i)}(s,t;\lambda)$ matches the one from the batch estimation. In the dense case, the rate of the quadratic risk is the parametric one i^{-1} in both the batch case of Golovkine et al. (2023) and for our online estimator. We thus conclude that our online covariance function estimator achieves the same rate as their batch estimator.

5.1 Implementation aspects

The suitable sequence of constants $D_{\gamma,(s,t)}^{(i)}$ can be traced to the numerous inequalities we derive in the proof of Proposition 5. These constants depend on $H_{s,t}$ and the corresponding L function value, on the density values $f_T(s)$, $f_T(t)$, the variance of X_sX_t and the error variances $\sigma^2(s),\sigma^2(t)$, the upper and lower bounds in (11), and the kernel. To simplify the procedure, we propose to account for the density value, the process and error variances and the regularity parameters, and in line with (18) and for the Epanechnikov kernel to set

$$D_{\gamma,(s,t)}^{(i)}=\frac{\left[\max\{\widehat{\sigma}^{(i)}(s),\widehat{\sigma}^{(i)}(t)\}\right]^2+\widehat{\operatorname{Var}(X_sX_t)}}{\min\left\{\widehat{f}_T^{(i)}(s),\widehat{f}_T^{(i)}(t)\right\}}\times\frac{\left(2\widehat{H}_{s,t}+1\right)\left(2\widehat{H}_{s,t}+3\right)}{10}\times\frac{\log_2(iM_i^2)}{\widehat{H}_{s,t}^{(i)}\left\{\widehat{L}_{s,t}^{(i)}\right\}^2},\qquad i\geq 1.$$

Here, $\widehat{L}_{s,t}^{(i)}$ is the estimator of the local Hölder constant corresponding to $\widehat{H}_{s,t}^{(i)}$, and $\widehat{\operatorname{Var}(X_sX_t)}$ is some nonparametric estimator of the variance of X_sX_t. The $\log_2(iM_i^2)$ factor accounts for all of the quantities which, for simplicity, are not included in the definition of $D_{\gamma,(s,t)}^{(i)}$. The theoretical impact of this choice is a slight deterioration of the rate of convergence by a log-log factor when $2cH_{s,t}\leq 1$. The numerical experiments reveal that this simple choice performs well.

Given the optimal choice $\lambda_{s,t}^*$ in (35), let us note that

$$i^{-\lambda_{s,t}^*}=\left[\max\left(M_i^{2H_{s,t}},i\right)\times M_i^2\right]^{-\frac{1}{2H_{s,t}+2}},$$

provided $M_i=i^c$. Combining results, our adaptive, online bandwidth rule becomes

$$\widehat{h}_{\gamma,(s,t)}^{(i)}=\left[\frac{D_{\gamma,(s,t)}^{(i)}}{\max\left(M_i^{2\widehat{H}_{s,t}^{(i)}},i\right)\times M_i^2}\right]^{1/\left(2\widehat{H}_{s,t}^{(i)}+2\right)},\qquad i=1,2\dots,\tag{36}$$

with the online regularity estimate $\widehat{H}_{s,t}^{(i)}=\min\{\widehat{H}_s^{(i)},\widehat{H}_t^{(i)}\}$ and $\widehat{H}_s^{(i)}$, $\widehat{H}_t^{(i)}$ defined as in (7).

5.2 Diagonal estimation

The asymptotic results for the online covariance estimator are obtained using a kernel with the support in $[-1,1]$ and the fact that $|s-t|$ is larger than twice the optimal bandwidth we propose, provided i is sufficiently large. What remains is to provide a rule for choosing a set $\mathcal{D}^{(i)}\subset\mathcal{T}\times\mathcal{T}$ shrinking to the diagonal segment $\{(u,u):u\in\mathcal{T}\}$ as i increases, and to propose an estimator for the covariance function on $\mathcal{D}^{(i)}$. It is well known that a 'first smooth, then estimate' approach for the covariance function induces a singularity when estimating the diagonal $\Gamma(u,u) = \mathrm{Var}(X_u)$. Below, we propose a simple, data-driven way to build the diagonal set $\mathcal{D}^{(i)}$. The idea is that, given $\mathcal{D}^{(i)}$, the estimate $\overline{\Gamma}^{(i)}(u,v)$ with $(u,v) \in \mathcal{D}^{(i)}$ is set equal to the estimate of the covariance function for the closest point on the boundary of $\mathcal{D}^{(i)}$.

Let us fix $t \in \mathcal{T}$, and consider $\mathfrak{d}_t \leq \Delta_0/2$, with Δ_0 from the condition (3). It can be easily shown (see Golovkine et al., 2023) that

$$\mathbb{E}\left[\left(\overline{\Gamma}^{(i)}(t-\mathfrak{u}_1, t+\mathfrak{u}_2) - \overline{\Gamma}^{(i)}(t,t)\right)^2\right] \lesssim \mathfrak{d}_t^{2H_t}, \qquad \forall\, 0 \leq \mathfrak{u}_1, \mathfrak{u}_2 \leq \mathfrak{d}_t,$$

where $\overline{\Gamma}^{(i)}(\cdot,\cdot)$ is the infeasible empirical covariance estimator in (1). The value $\mathfrak{d}_t$ determines the diagonal set where the covariance function estimator needs a specific definition. Our estimator of $\Gamma(t-\mathfrak{u}_1, t+\mathfrak{u}_2)$ and $\Gamma(t+\mathfrak{u}_2, t-\mathfrak{u}_1)$ is then

$$\widehat{\Gamma}^{(i)}(t-\mathfrak{u}_1, t+\mathfrak{u}_2) = \widehat{\Gamma}^{(i)}(t+\mathfrak{u}_2, t-\mathfrak{u}_1) = \widehat{\Gamma}^{(i)}(t-\mathfrak{d}_t, t+\mathfrak{d}_t), \qquad \text{for } 0 \leq \mathfrak{u}_1, \mathfrak{u}_2 \leq \mathfrak{d}_t.$$

In practice, the quantity $\mathfrak{d}_t$ depends on i and can be defined as the smallest value $d \geq 0$ on a grid such that $d \geq \widehat{h}^{(i)}_{\gamma,(t-d,t+d)}$, with $\widehat{h}^{(i)}_{\gamma,(t-d,t+d)}$ defined in (36).

6 Monte Carlo Study

In this section we report results from a series of simulation exercises undertaken to assess the finite sample performance of the proposed online estimators $\widehat{H}_t$, $\widehat{\mu}(t)$, and $\widehat{\Gamma}(s,t)$, and we also compare them with alternative data-driven approaches written in the R language (R Core Team, 2024) that are publicly available via CRAN. The DGP is described in Section 6.1, and in the following sections we report three sets of simulations, each with a different focus.

6.1 Summary of the DGP

The DGP is constructed with $\mathcal{T} = [0,1]$, using a multifractional Brownian motion (MfBm) with a time-varying Hurst index function. The MfBm, say $W = (W(t))_{t\geq 0}$, with Hurst index function, say $t \mapsto H_t \in (0,1)$, is a centered Gaussian process with $W(0) = 0$ and covariance function

$$C(s,t) = \mathbb{E}\left[W(s)W(t)\right] = D(H_s, H_t)\left[s^{H_s+H_t} + t^{H_s+H_t} - |t-s|^{H_s+H_t}\right], \qquad s,t \geq 0, \tag{37}$$

where

$$D(x,y) = \frac{\sqrt{\Gamma(2x+1)\Gamma(2y+1)\sin(\pi x)\sin(\pi y)}}{2\Gamma(x+y+1)\sin(\pi(x+y)/2)}, \qquad D(x,x) = 1/2, \qquad x,y > 0.$$

See, *e.g.*, Balança (2015) and the references therein for the formal definition. Wang Guang Wei et al. (2023) show that W satisfies (3) with the Hurst index as local regularity exponent and $L \equiv 1$. The fractional Brownian motion is an example of an MfBm with a constant Hurst index. Let us recall that while the *increments* for a Brownian Motion (i.e., when $H \equiv 1/2$) are stationary and independent, increments for an MfBm are neither.

Given $0 < \underline{H} \leq \overline{H} < 1$, a change point $t_c \in (0,1)$, and a slope $S > 0$, the Hurst index function we consider is given by

$$H_t = \underline{H} + \frac{\overline{H} - \underline{H}}{1 + \exp(-S(t - t_c))}, \qquad t \in [0,1].$$

In the simulations that follow we set $\underline{H} = 0.25$, $\overline{H} = 0.75$ and $t_c = 1/2$. The mean function is set to be the 'bump' function $\mu(t) = t + \exp(-200(t-0.5)^2)$. The covariance function is $\Gamma(s,t) = 1 + C(s,t)$, with $C(s,t)$ defined in (37). By adding 1 to the covariance function of the MfBm, the initial value $X(0)$ becomes a standard normal random variable.

Given the online sample size i, the data points $(Y_m^{(j)}, T_m^{(j)})$ are then generated as follows: for each $1 \leq j \leq i$,

- an integer M_j is generated as a realization of some random variable with integer mean $\mathfrak{m}_i$ (e.g., Poisson), or M_j could be simply set equal to $\mathfrak{m}_i$;
- next, M_j independent draws $T_1^{(j)}, \ldots, T_{M_j}^{(j)}$ from a uniform random variable on $[0,1]$ are generated;
- the M_j-dimensional mean vector $\mu^{(j)} = (\mu(T_1^{(j)}), \ldots, \mu(T_{M_j}^{(j)}))$ is computed, and using (37), the $M_j \times M_j$ covariance matrix $\Gamma^{(j)}$ with the entries $1 + C(T_m^{(j)}, T_{m'}^{(j)})$, $1 \le m, m' \le M_j$, is computed;
- an M_j-dimensional vector with components $X^{(j)}(T_m^{(j)})$, $1 \le m \le M_j$, is computed as the realization of a multivariate Gaussian distribution with mean $\mu^{(j)}$ and covariance matrix $\Gamma^{(j)}$;
- given the error's standard deviation function $t \mapsto \sigma(t)$, build

  $$Y_m^{(j)} = X^{(j)}(T_m^{(j)}) + \sigma(T_m^{(j)}) e_m^{(j)}, \quad 1 \le m \le M_j,$$

  where the $e_m^{(j)}$ are independent random draws from a standard normal distribution. In the simulations $\sigma(t)$ is constant and equal to 0.5.

Figure 1 presents four plots that summarize the DGP described above which we use in the simulations below. The first plot shows the Hurst index function H_t, the second plot shows the mean function $\mu(t)$, the third plot shows the covariance function $\Gamma(s,t)$, and the fourth plot shows a handful of sample random functions $X^{(j)}(t)$ and the data points $(Y_m^{(j)}, T_m^{(j)})$. By construction, the true L_t functions is constant and equal to 1.

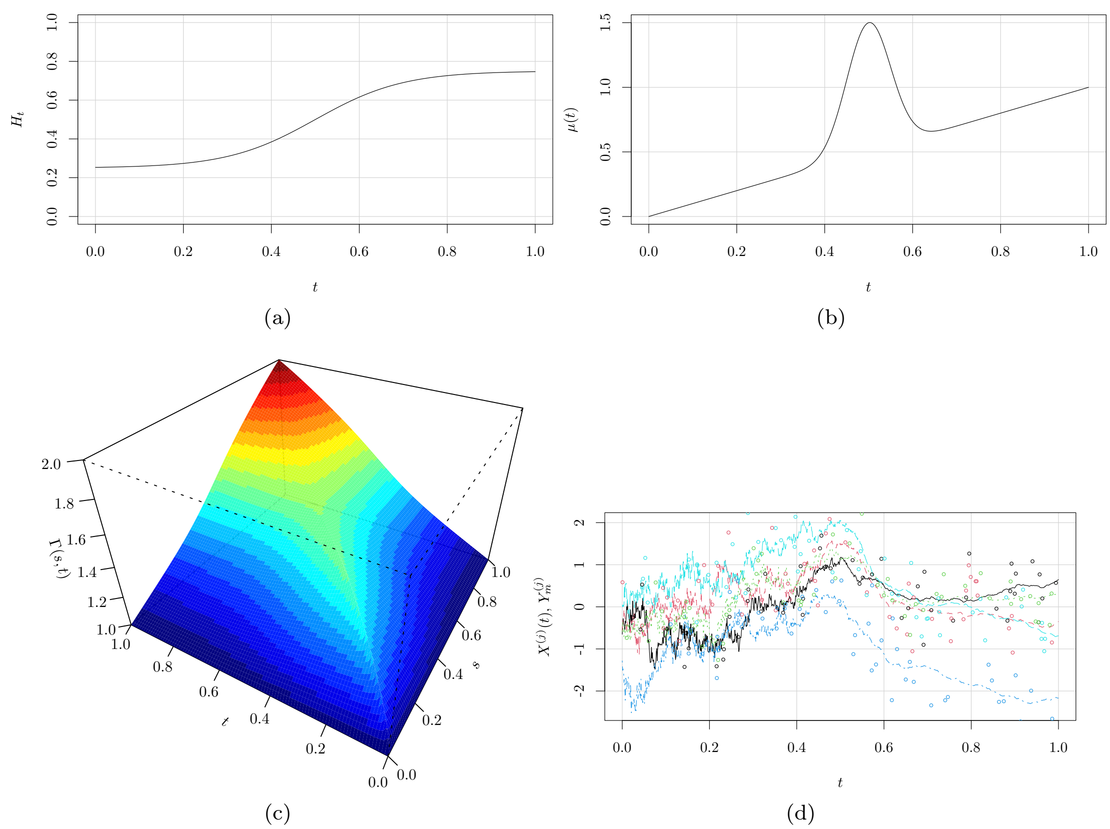

Figure 1: (a) Hurst Index Function H_t, (b) Mean Function $\mu(t)$, (c) Covariance Function $\Gamma(s,t)$, (d) Sample Random Functions $X^{(j)}(t)$ and Data Points $(Y_m^{(j)}, T_m^{(j)})$

For each simulation setup summarized below we conduct 1,000 Monte Carlo (MC) replications, and for each Monte Carlo replication we estimate H_t and L_t using (7) and (8), respectively. The pilot estimators $\widetilde{X}_u^{(i)}$ and $\widetilde{X}_v^{(i)}$, used to update $\widehat{\theta}^{(i)}(u,v)$ with u,v close to t when computing H_t and L_t using (7) and (8), are obtained using the sequence of bandwidths defined in (17) with $C_t^{(i)}$ and $\alpha_t^{(i)}$ defined in (18). The recursive estimator of the mean function $\mu(t)$ is constructed using (23) with bandwidth given in (22), and the recursive estimator of the covariance function $\Gamma(s,t)$ is constructed using (32) with the procedure described in Section 5.2 for the diagonal.

All code is written in R and is available upon request. We have conducted a fairly extensive range of simulations (not reported) that vary the signal-to-noise ratio, mean function, and the Hurst index function. These results are consistent with the ones reported below.

6.2 Monte Carlo evidence for the uniform convergence of $\widehat{H}_t$

We begin with the summary in Figure 2 which presents boxplots for the maximum deviation of the estimated Hurst exponent from its actual value, computed as the maximum of $|\widehat{H}_t - H_t|$ on an equidistant grid $\mathcal{T}_0 \subset [0,1]$ of $|\mathcal{T}_0| = 100$ values t. The boxplots are based on replications of samples of $i = 800$ curves. For each sample, the M_j, $1 \leq j \leq i$, are set equal to $\mathfrak{m}_i$. The numbers $\mathfrak{m}_i$ of data points per curve we consider are $\mathfrak{m}_i = 100, 200$ and 400.

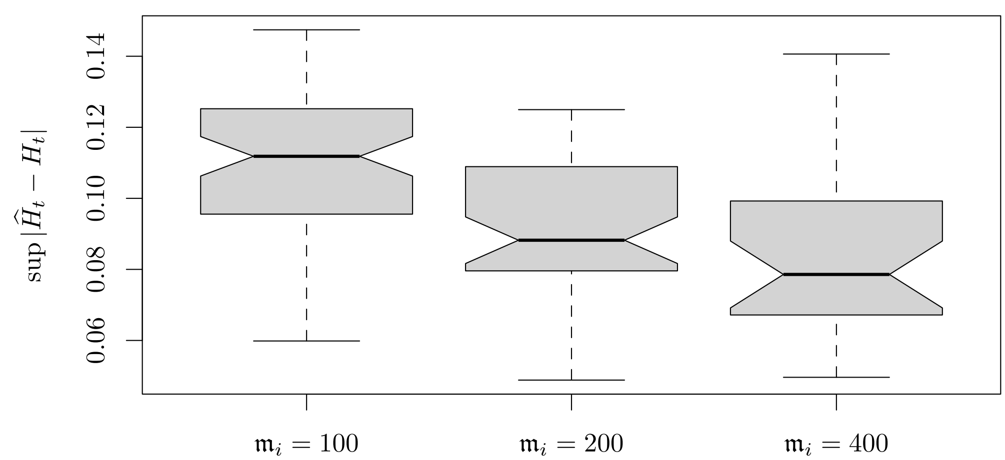

Figure 2: Boxplots for $\sup|\widehat{H}_t - H_t|$ for $\mathfrak{m}_i = 100, 200, 400$, $i = 800$ from 1,000 MC replications.

It is evident from Figure 2 that the distribution of $\sup|\widehat{H}_t - H_t|$ shifts towards zero as the number of observations per curve increases. This is consistent with the theory which indicates that the bias, induced by the pilot estimators and the higher order term $S|u-v|^{2H_t+2\beta}$ in (3), should decrease with $\mathfrak{m}_i$. On the other hand, the theory also indicates that even with very large $\mathfrak{m}_i$ the variance would stay away from zero as long as i is fixed, as is the case in our simulations.

6.3 Monte Carlo evidence for the mean and covariance functions estimators

Next, we assess the performance of the estimators of the mean and covariance functions, $\mu(t)$ and $\Gamma(s,t)$. We are interested in the performance of the estimators as the number of curves (i.e., i) and the average number of observations per curve (i.e., $\mathfrak{m}_i$) increases. For simplicity, we again set all the M_j, $1 \le j \le i$, equal to $\mathfrak{m}_i$. Moreover, we omit the subscript i and write $\mathfrak{m}$ instead of $\mathfrak{m}_i$.

We start with $\mathfrak{m} = 25$ and $i = 1000$ and continue to double $\mathfrak{m}$ and i up to maximum values of $\mathfrak{m} = 400$ and $i = 16000$. Results are summarized in Tables 1 and 2 which report the mean RMSE for the estimators of the mean and covariance functions evaluated on equidistant grids $s \in [0,1]$ and $t \in [0,1]$ of sizes $|\mathcal{T}_0| = 50$. RMSE is computed as

$$\mathrm{RMSE}_\mu = \sqrt{|\mathcal{T}_0|^{-1} \sum_{t=1}^{|\mathcal{T}_0|} \{\widehat{\mu}(t) - \mu(t)\}^2} \quad \text{and} \quad \mathrm{RMSE}_\Gamma = \sqrt{|\mathcal{T}_0|^{-2} \sum_{s,t=1}^{|\mathcal{T}_0|} \left\{\widehat{\Gamma}(s,t) - \Gamma(s,t)\right\}^2},$$

respectively.

Table 1: Mean RMSE_μ after i recursions. Results from 1,000 MC replications

	$\mathfrak{m} = 25$	$\mathfrak{m} = 50$	$\mathfrak{m} = 100$	$\mathfrak{m} = 200$	$\mathfrak{m} = 400$
$i = 1000$	0.0512	0.0429	0.0390	0.0368	0.0389
$i = 2000$	0.0375	0.0312	0.0279	0.0275	0.0298
$i = 4000$	0.0281	0.0230	0.0210	0.0201	0.0206
$i = 8000$	0.0209	0.0171	0.0150	0.0152	0.0153
$i = 16000$	0.0156	0.0127	0.0115	0.0113	0.0119

Table 2: Mean RMSE_Γ RMSE after i recursions. Results from 1,000 MC replications

	$\mathfrak{m} = 25$	$\mathfrak{m} = 50$	$\mathfrak{m} = 100$	$\mathfrak{m} = 200$	$\mathfrak{m} = 400$
$i = 1000$	0.0804	0.0704	0.0633	0.0652	0.0558
$i = 2000$	0.0620	0.0529	0.0455	0.0485	0.0403
$i = 4000$	0.0481	0.0415	0.0353	0.0352	0.0318
$i = 8000$	0.0381	0.0316	0.0272	0.0263	0.0256
$i = 16000$	0.0309	0.0244	0.0218	0.0207	0.0210

Tables 1 and 2 reveal that RMSE falls as both i and $\mathfrak{m}$ increase, consistent with the theory.

6.4 Monte Carlo Comparison with `fdapace` $\mu(t)$ and $\Gamma(s,t)$ Estimators

Finally, we compare the performance of the batch mean and covariance estimators from the R package `fdapace` (Zhou et al., 2022) with our proposed recursive online method. The approach in the R package `fdapace` is a batch approach, i.e., it processes all curves simultaneously, while the approach we propose begins with a subset of curves processed in a batch-like manner and then proceeds to process additional curves in an online manner. In what follows, when applying our proposed approach we select the first $i/2$ curves to be used as the batch curves which provide initial estimates of H_t, L_t, etc., and then process the remaining curves $i/2+1, i/2+2, \ldots, i$ in an online manner. Given the computational burden associated with the method in the `fdapace` package, we are forced to restrict both $\mathfrak{m}$ and i to quite small values. In the tables that follow, results denoted `fdapace` employ the R package `fdapace` (Zhou et al., 2022) with an Epanechnikov kernel and data-driven (cross-validated) bandwidth selection. Results denoted `fda.la` employ the R function `fda.la` (i.e., the name of the R function we wrote that implements our approach) with an Epanechnikov kernel and data-driven estimates of the local Hölder exponent and bandwidths.

Mean RMSE results are summarized in Table 3 and Table 4, and the last column reports the ratio of the RMSE of the proposed approach with that from the R package `fdapace` with values < 1 indicating superior RMSE performance of the proposed approach.

Table 3: Mean RMSE for $\mu(t)$ Estimates

i	$\mathtt{m}$	fda.la	fdapace	ratio
10	10	0.46	0.46	1.00
10	20	0.45	0.43	1.05
10	40	0.44	0.40	1.10
20	10	0.33	0.36	0.91
20	20	0.31	0.33	0.95
20	40	0.30	0.30	1.02
40	10	0.25	0.28	0.88
40	20	0.22	0.25	0.90
40	40	0.21	0.22	0.96

Table 4: Mean RMSE for $\Gamma(s, t)$ Estimates

i	$\mathtt{m}$	fda.la	fdapace	ratio
10	10	0.68	0.76	0.90
10	20	0.66	0.67	0.98
10	40	0.64	0.65	0.99
20	10	0.50	0.56	0.89
20	20	0.46	0.51	0.92
20	40	0.45	0.49	0.90
40	10	0.37	0.42	0.90
40	20	0.32	0.37	0.87
40	40	0.31	0.36	0.88

Even though one might expect that a batch approach would be more efficient than a recursive one (which is the case for small $\mathtt{m}$ and i as can be seen in Tables 3 and 4), it is apparent that as both $\mathtt{m}$ and i increase, the RMSE performance of our method is superior to that of the batch estimators in the `fdapace` package as $\mathtt{m}$ and i increase beyond some fairly small values. Furthermore, it is worth noting that our method can be deployed on millions of curves with thousands of observations per curve with minimal computational overhead (the computational cost of our approach is measured in milliseconds per curve, i.e., < 1 second per curve on a single desktop CPU).

A Proofs on the mean function estimation

Here we use the notation $\lesssim$ which means that the left side is bounded by an absolute constant times the right side. Moreover, $\asymp$ is used when the left side is bounded above and below by absolute constants times the right side. Finally, the notation $\asymp_{\mathbb{P}}$ (resp. $\lesssim_{\mathbb{P}}$) means that the left side is bounded above and below by absolute constants (resp. bounded by an absolute constant) times the right side with probability tending to 1. For simplicity, we provide the proofs for the mean function estimation under the additional assumption that the kernel K stays away from zero on the support. This assumption can be dropped at the price of intricate technical arguments for controlling the probabilities of the events where the denominators are close to zero.

Lemma 1. *Assume the conditions of Proposition 2 hold true, and the kernel $K(\cdot)$ is bounded and bounded away from zero on the support $[-1,1]$. Moreover, let $\zeta > 0$ and $0 < \underline{c} \leq \overline{c}$ be some constants, and, for each $i \geq 1$, define the bandwidth range $\mathcal{H}_i = [\underline{c}i^{-\zeta}, \overline{c}i^{-\zeta}]$. Then,*

$$0 \leq \max_{1 \leq m \leq M_i} W_m^{(i)}(t;h) \leq S_{i,W}(h) \times i^{\min(0,\zeta - c)}, \qquad \forall h \in \mathcal{H}_i,$$

where $S_{i,W}(h) \geq 1$ is a random variable with the mean and the variance bounded by constants which do not depend on h and i. Moreover, the variables $\{S_{i,W}(h), i \geq 1\}$ are independent.

Lemma 2. *Let $t \in \mathcal{T}$ and assume the conditions of Proposition 2 hold true. Let $\zeta > 0$ and set $h^{(i)} = i^{-\zeta}$, $i = 1, 2, \ldots$. Define*

$$\mathcal{W}^{(i)}(t) = \mathcal{W}^{(i-1)}(t) + w^{(i)}\left(t; h^{(i)}\right), \qquad i = 1, 2 \ldots,$$

with $w^{(i)}(\cdot;\cdot)$ defined as in (21), and $\mathcal{W}^{(0)}(t)$ some starting value. Then, constants $\underline{C}, \overline{C} \in (0,1]$ exist, depending on the bounds for the density f_T and $M_i/\mathfrak{m}_i$, on c from (11), such that

$$\underline{C} \times i^{1+\min(0,c-\zeta)} \times \{1 + o_{\mathbb{P}}(1)\} \leq \mathcal{W}^{(i)}(t) \leq \overline{C} \times i^{1+\min(0,c-\zeta)} \times \{1 + o_{\mathbb{P}}(1)\}.$$

The proofs of Lemmas 1 and 2 are given in the Supplementary Material.

Proof of Proposition 2. Since t is fixed, to simplify notation, below we write ζ instead of ζ_t. Moreover, we write $h^{(i)}$ instead of $h^{(i)}_{\mu,t}(\zeta)$ and $\widehat{\mu}^{(i)}(t)$ instead of $\widehat{\mu}^{(i)}(t;\zeta)$. We then have $h^{(i)} \asymp i^{-\zeta}$, for some $\zeta > 0$, and the purpose is then to determine a bound for the quadratic risk of the pointwise mean function estimator, and to find the minimum of the bound with respect to ζ, under the constraint

$$\zeta < c + 1. \tag{38}$$

The constraint (38) guarantees that the number $\mathcal{W}^{(i)}(t)$ of curves used in the estimation of the mean function increase as fast as some positive power of i, as is expected for an optimal choice of ζ.

Recall that $\mathcal{T}^{(j)}_{obs}$ denotes the set of observation times $T^{(j)}_m$, $1 \leq m \leq M_j$, over the trajectory $X^{(j)}$. Consider the notation $\mathbb{E}_{M,T}[\cdots] = \mathbb{E}[\cdots \mid M_j, \mathcal{T}^{(j)}_{obs}, j \geq 1]$. Let us now notice, that by construction,

$$\widehat{\mu}^{(i)}(t) = \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \widehat{X}^{(j)}_t\left(h^{(j)}\right).$$

Using the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we can bound the conditional quadratic risk of $\widehat{\mu}^{(i)}(t)$, given the M_j, $j \geq 1$, and the realizations of T, as follows:

$$\mathbb{E}_{M,T}\left[\left\{\widehat{\mu}^{(i)}(t) - \mu(t)\right\}^2\right] \lesssim \mathbb{E}_{M,T}\left[\left\{\widehat{\mu}^{(i)}(t) - \widetilde{\mu}^{(i)}(t)\right\}^2\right] + \mathbb{E}_{M,T}\left[\left\{\widetilde{\mu}^{(i)}(t) - \overline{\mu}^{(i)}(t)\right\}^2\right]$$
$$+ \mathbb{E}_{M,T}\left[\left\{\overline{\mu}^{(i)}(t) - \mu^{(i)}(t)\right\}^2\right] =: \mathcal{R}^{(i)}_1(\zeta) + \mathcal{R}^{(i)}_2(\zeta) + \mathcal{R}^{(i)}_3,$$

where $\widetilde{\mu}^{(i)}(t)$ and $\overline{\mu}^{(i)}(t)$ are infeasible estimators defined as

$$\widetilde{\mu}^{(i)}(t) = \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) X_t^{(j)} \quad \text{and} \quad \overline{\mu}^{(i)}(t) = \frac{1}{i} \sum_{j=1}^{i} X_t^{(j)},$$

respectively. Clearly, the $\mathcal{R}_3^{(i)}$ does not depend on the bandwidths $h^{(j)}$, the realizations of T and the M_i, and has the rate $O_{\mathbb{P}}(i^{-1})$. Concerning $\mathcal{R}_2^{(i)}(\zeta)$, we can write

$$\widetilde{\mu}^{(i)}(t) - \overline{\mu}^{(i)}(t) = \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} \left\{ w^{(j)}\left(t; h^{(j)}\right) - \frac{\mathcal{W}^{(i)}(t)}{i} \right\} \left\{ X_t^{(j)} - \mu(t) \right\},$$

and thus

$$\mathcal{R}_2^{(i)}(\zeta) = \frac{\operatorname{Var}(X_t)}{[\mathcal{W}^{(i)}(t)]^2} \sum_{j=1}^{i} \left\{ w^{(j)}\left(t; h^{(j)}\right) - \frac{\mathcal{W}^{(i)}(t)}{i} \right\}^2 = \operatorname{Var}(X_t) \left\{ \frac{1}{\mathcal{W}^{(i)}(t)} - \frac{1}{i} \right\}.$$

The term $\mathcal{R}_2^{(i)}(\zeta)$ is thus a penalty term for the amount of curves dropped because of a small bandwidth $h^{(i)}$. It ranges from zero (all curves used for the mean estimation) to infinity (all the curves are discarded).

Let

$$B_t^{(j)}(h) = \mathbb{E}_{M,T}\left[\widehat{X}_t^{(j)}(h)\right] - X_t^{(j)} \quad \text{and} \quad V_t^{(j)}(h) = \widehat{X}_t^{(j)}(h) - \mathbb{E}_{M,T}\left[\widehat{X}_t^{(i)}(h)\right], \quad t \in \mathcal{T},$$

be the bias and the stochastic part of $\widehat{X}_t^{(j)}(h)$, respectively, given M_j and the design points $T_m^{(j)}$. The pairs of random variables $(B_t^{(j)}, V_t^{(j)})$ are independent and square integrable. For the mean, we can then write

$$\widehat{\mu}^{(i)}(t) - \widetilde{\mu}^{(i)}(t) = \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) B_t^{(j)}\left(h^{(j)}\right) + \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) V_t^{(j)}\left(h^{(j)}\right) =: \mathcal{B}_t^{(i)} + \mathcal{V}_t^{(i)},$$

On the bias part, by Cauchy-Schwarz inequality, applied twice, and condition (3), we get

$$\begin{aligned}
\mathbb{E}_{M,T}\left[\left\{\mathcal{B}_t^{(i)}\right\}^2\right] &\leq \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \left\{ \sum_{m=1}^{M_j} \left| W_m^{(j)}(t; h^{(j)}) \right| \right. \\
&\qquad \left. \times \sum_{m=1}^{M_j} \mathbb{E}_{M,T}\left(\left\{ X^{(j)}(T_m^{(j)}) - X_t^{(j)} \right\}^2 \right) |W_m^{(j)}(t; h^{(j)})| \right\} \\
&= \frac{1 + o_{\mathbb{P}}(1)}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \left\{ \sum_{m=1}^{M_j} \left| W_m^{(j)}(t; h^{(j)}) \right| \times \sum_{m=1}^{M_j} L_t^2 \left| T_m^{(j)} - t \right|^{2H_t} |W_m^{(j)}(t; h^{(j)})| \right\} \\
&= L_t^2 \frac{1 + o_{\mathbb{P}}(1)}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \left\{ \{h^{(j)}\}^{2H_t} \times \sum_{m=1}^{M_j} \left| (T_m^{(j)} - t)/h^{(j)} \right|^{2H_t} W_m^{(j)}(t; h^{(j)}) \right\}.
\end{aligned}$$

For the last equality we eventually use the fact that the weights $W_m^{(j)}$ are positive and sum to 1, as is the case for the local constant smoother with a non-negative kernel. Since by the definition of the weights, that is with a kernel supported on $[-1, 1]$, we have $\mathbf{1}\{W_m^{(j)}(t; h) > 0\} |T_m^{(j)} - t| / h \leq 1$ for any h, we deduce

$$\mathbb{E}_{M,T}\left[\left\{\mathcal{B}_t^{(i)}\right\}^2\right] \lesssim_{\mathbb{P}} \frac{L_t^2}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \{h^{(j)}\}^{2H_t}.$$

By Lemma 2, on the one hand, $\mathcal{W}^{(i)}(t) \asymp_{\mathbb{P}} i^{1+\min(0,c-\zeta)}$, provided $h^{(j)} \asymp j^{-\zeta}$ and $M_j \asymp j^{-c}$. More precisely, $\mathcal{W}^{(i)}(t)$ concentrates around its conditional mean given the $M_1, \dots, M_i$, and $\mathbb{E}_M\left[\mathcal{W}^{(i)}(t)\right] \asymp i^{1+\min(0,c-\zeta)}$. (Recall the notation $\mathbb{E}_M[\cdots] = \mathbb{E}[\cdots \mid M_j, j \geq 1]$.) Note next that condition (38) implies $1 + \min(0, c - \zeta) > 0$ and thus $\mathcal{W}^{(i)}(t) \to \infty$ in probability. On the other hand, with $\pi^{(j)}(t;h) = \mathbb{E}_M\left[w^{(j)}(t;h)\right]$, we get

$$S^{(i)}(t;\zeta) := \mathbb{E}_M\left[\sum_{j=1}^{i} w^{(j)}\left(t;h^{(j)}\right)\{h^{(j)}\}^{2H_t}\right] \asymp \sum_{j=1}^{i} \pi^{(j)}(t;h^{(j)}) j^{-2\zeta H_t} \asymp \sum_{j=1}^{i} j^{\min(0,c-\zeta)-2\zeta H_t},$$

and

$$\begin{aligned}\Sigma^{(i)}(t;\zeta) := \mathrm{Var}_M\left[\sum_{j=1}^{i} w^{(j)}\left(t;h^{(j)}\right)\{h^{(j)}\}^{2H_t}\right] &= \sum_{j=1}^{i} \mathbb{E}_M\left[\left\{w^{(j)}\left(t;h^{(j)}\right) - \pi^{(j)}(t;h^{(j)})\right\}^2\right]\{h^{(j)}\}^{4H_t} \\ &\lesssim \sum_{j=1}^{i} \pi^{(j)}(t;h^{(j)}) j^{-4\zeta H_t} \asymp \sum_{j=1}^{i} j^{\min(0,c-\zeta)-4\zeta H_t}.\end{aligned}$$

By Lemma 5 in the Supplementary Material, we have

$$S^{(i)}(t,\zeta) \asymp_{\mathbb{P}} \begin{cases} i^{\min(1,1+c-\zeta)-2\zeta H_t} & \text{if } 2\zeta H_t < \min(1,1+c-\zeta) \\ \log(i) & \text{if } 2\zeta H_t = \min(1,1+c-\zeta) \\ 1 & \text{if } 2\zeta H_t > \min(1,1+c-\zeta) \end{cases},$$

$$\sqrt{\Sigma^{(i)}(t,\zeta)} \asymp_{\mathbb{P}} \begin{cases} i^{\min(1,1+c-\zeta)/2-2\zeta H_t} & \text{if } 4\zeta H_t < \min(1,1+c-\zeta) \\ \sqrt{\log(i)} & \text{if } 4\zeta H_t < \min(1,1+c-\zeta) \\ 1 & \text{if } 4\zeta H_t > \min(1,1+c-\zeta) \end{cases},$$

and, since by (38) we have $\min(1, 1 + c - \zeta) > 0$, we deduce that $\sqrt{\Sigma^{(i)}(t,\zeta)}$ is negligible compared to $S^{(i)}(t,\zeta)$. Gathering terms and using Lemma 2, we deduce

$$\mathbb{E}_M\left[\left\{\mathcal{B}_t^{(i)}\right\}^2\right] \lesssim \begin{cases} i^{-2\zeta H_t} & \text{if } \; 2\zeta H_t < \min(1,1+c-\zeta) \\ i^{-1-\min(0,c-\zeta)}\log(i) & \text{if } \; 2\zeta H_t = \min(1,1+c-\zeta) \\ i^{-1-\min(0,c-\zeta)} & \text{if } \; 2\zeta H_t > \min(1,1+c-\zeta) \end{cases} \longrightarrow 0, \tag{39}$$

provided (38) holds true.

For bounding the variance of the stochastic term $\mathcal{V}_t^{(i)}$, let us note that the variables $\varepsilon_m^{(j)}$ are centered and conditionally independent, with bounded conditional variance, given all M_j, $\mathcal{T}_{obs}^{(j)}$ and $X^{(j)}$. Moreover,

$$\mathbb{E}_{M,T}\left[V_t^{(j)}V_t^{(i)}\right] = \mathbb{E}_{M,T}\left[V_t^{(j)}B_t^{(j)}\right] = \mathbb{E}_{M,T}\left[V_t^{(j)}B_t^{(i)}\right] = 0, \qquad \forall i \neq j.$$

Using the Hölder continuity of the conditional variance $t \mapsto \sigma^2(t)$, we have

$$\begin{aligned}\mathbb{E}_{M,T}\left[\left\{\mathcal{V}_t^{(i)}\right\}^2\right] &\leq \frac{\sigma^2(t)\{1+o(1)\}}{\mathcal{W}^{(i)}(t)} \times \frac{1}{\mathcal{W}^{(i)}(t)}\sum_{j=1}^{i}\left\{\max_{1\leq m\leq M_j}\left|W_m^{(j)}\left(t;h^{(j)}\right)\right| \times \sum_{m=1}^{M_j}\left|W_m^{(j)}\left(t;h^{(j)}\right)\right|\right\} \\ &= \frac{\sigma^2(t)\{1+o(1)\}}{\mathcal{W}^{(i)}(t)} \times \frac{1}{\mathcal{W}^{(i)}(t)}\sum_{j=1}^{i}\max_{1\leq m\leq M_j} W_m^{(j)}\left(t;h^{(j)}\right).\end{aligned} \tag{40}$$

The inequality in the previous equation holds true for general weights since it uses the fact that the weights are positive and sum to 1, as is the case for the local constant smoother with non-negative kernel.

For simplicity, let us next consider the case where the local constant estimator is obtained with the uniform kernel. When the kernel is uniform, for fixed j, t and h, the value of $\max_{1\leq m\leq M_j} W_m^{(j)}(t;h)$ is the inverse of the number of $T_m^{(j)}$ in the interval $[t-h,t+h]$ (by definition $0/0=0$), that is

$$\max_{1\leq m\leq M_j} W_m^{(j)}(t;h) \in \{0, 1/M_j, 1/(M_j-1),\ldots,1/2,1\}, \qquad \forall j,h.$$

The value zero is obtained when there is no point $T_m^{(j)}$ in the interval $[t-h,t+h]$, in which case the rule $0/0=0$ applies. Next, by Lemma 1 and Lemma 5 in the Supplementary Material, we have

$$\begin{aligned}\mathbb{E}_M\left[\sum_{j=1}^{i}\max_{1\leq m\leq M_j} W_m^{(j)}\left(t;h^{(j)}\right)\right] &\lesssim \sum_{j=1}^{i} j^{-\max(0,c-\zeta)}\\ &\asymp i^{1+\min(0,\zeta-c)}\mathbf{1}\{c<\zeta+1\}+\log(i)\mathbf{1}\{c=\zeta+1\}+\mathbf{1}\{c>\zeta+1\}.\end{aligned}$$

Finally, using the fact that $\mathcal{W}^{(i)}(t)\asymp_{\mathbb{P}} i^{1+\min(0,c-\zeta)}$, we get

$$\mathbb{E}_M\left[\left\{\mathcal{V}_t^{(i)}\right\}^2\right] \lesssim \begin{cases} i^{-(1+c-\zeta)} & \text{if } \ 1+\zeta\neq c\\ i^{-2}\log(i) & \text{if } \ 1+\zeta=c\end{cases} \longrightarrow 0. \tag{41}$$

In the previous equation, the convergence to zero is guaranteed as soon as (38) holds true. Finally, we get

$$\begin{aligned}\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right] := \mathbb{E}_M\left[\left\{\widehat{\mu}^{(i)}(t)-\widetilde{\mu}^{(i)}(t)\right\}^2\right] &= \mathbb{E}_M\left[\left\{\mathcal{B}_t^{(i)}\right\}^2\right]+\mathbb{E}_M\left[\left\{\mathcal{V}_t^{(i)}\right\}^2\right]\\ &\lesssim \begin{cases} i^{-\min\{1,2\zeta H_t,1+c-\zeta)]} & \text{if } \ 2\zeta H_t\neq\min(1,1+c-\zeta)\\ i^{-\min(1,1+c-\zeta)}\log(i) & \text{if } \ 2\zeta H_t=\min(1,1+c-\zeta)\end{cases}\\ &\qquad\qquad + \begin{cases} i^{-\min(1,1+c-\zeta)} & \text{if } \ \zeta\neq c-1\\ i^{-2}\log(i) & \text{if } \ \zeta=c-1\end{cases}.\end{aligned}$$

- Sparse Case: $2cH_t<1$. We have to distinguish several sub-cases corresponding to the range of ζ.
  - If $0<1+\min(0,c-\zeta)=1+c-\zeta\leq 2\zeta H_t$, then necessarily $\zeta\geq c$ and $\zeta\geq(c+1)/(2H_t+1)$. We have

    $$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right]\lesssim i^{-(1+c-\zeta)}+i^{-(1+c-\zeta)} \quad \text{and} \quad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right]\asymp i^{-(1+c-\zeta)}, \tag{42}$$

    the optimum choice for $\mathcal{R}_1^{(i)}(\zeta)$ is then when λ takes the smallest admissible value, that is

    $$\zeta^*=\max\left\{c,\frac{c+1}{2H_t+1}\right\}=\frac{c+1}{2H_t+1}.$$

    The second equality in the previous equation is given by the sparse case condition $2cH_t\leq 1$. Moreover,

    $$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right]+\mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right]\lesssim i^{-(1+c-\zeta^*)}=i^{-2(c+1)H_t/(2H_t+1)}\gg i^{-1}. \tag{43}$$

  - If $0<2\zeta H_t\leq 1+\min(0,c-\zeta)=1+c-\zeta$, then necessarily $c\leq\zeta\leq(c+1)/(2H_t+1)$. Moreover,

    $$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right]\lesssim i^{-2\zeta H_t}+i^{-(1+c-\zeta)} \quad \text{and} \quad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right]\asymp i^{-(1+c-\zeta)},$$

    and the optimum choice of ζ for minimizing the expectation of $\mathcal{R}_1^{(i)}(\zeta)+\mathcal{R}_2^{(i)}(\zeta)$ is

    $$\zeta^*=\frac{c+1}{2H_t+1}=\max\left\{c,\frac{c+1}{2H_t+1}\right\},$$

    for which we get the same risk bound as in (43).

– If $1+\min(0,c-\zeta)=1$, in the sparse regime, we can only have $0<2\zeta H_t \le 1+\min(0,c-\zeta)=1$. Then we get

$$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right] \lesssim i^{-2\zeta H_t} + i^{-(1+c-\zeta)} \quad \text{and} \quad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right] \asymp i^{-1}.$$

The optimum rate for ζ would be $\zeta=(c+1)/(2H_t+1)$, but this is larger than or equal to c, which contradicts the condition $1+\min(0,c-\zeta)=1$ which means $c\ge\zeta$. The optimal value for ζ is the largest admissible one under the constraints, because it minimizes the bias term which is here dominant. Thus, $\zeta^*=c$. In this case,

$$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right] \lesssim i^{-2cH_t} + i^{-1} \asymp i^{-2cH_t} \qquad \text{and} \qquad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right] \lesssim i^{-1}.$$

- Dense Case: $2cH_t>1$.
  – If $0<1+\min(0,c-\zeta)=1+c-\zeta\le 2\zeta H_t$, then necessarily $\zeta\ge\max\{c,(c+1)/(2H_t+1)\}=c$. We then have

  $$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right] \lesssim i^{-(1+c-\zeta)} + i^{-(1+c-\zeta)} \quad \text{and} \quad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right] \asymp i^{-(1+c-\zeta)},$$

  the optimum choice for $\mathcal{R}_1^{(i)}(\zeta)$ is then when

  $$\zeta^* = c = \max\left\{c, \frac{c+1}{2H_t+1}\right\},$$

  and we have $\mathcal{R}_2^{(i)}(\zeta^*)$ and $\mathcal{R}_1^{(i)}(\zeta^*)$ of rate $O_{\mathbb{P}}(i^{-1})$, that is the parametric rate. If $0<2\zeta H_t< 1+\min(0,c-\zeta)=1+c-\zeta\le 1$, it means $\zeta\ge c$, and we thus get a contradiction with the dense regime condition $2cH_t>1$.

  – If $1+\min(0,c-\zeta)=1$, we can only have $0<1+\min(0,c-\zeta)=1<2\zeta H_t$ and thus

  $$\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right] \lesssim i^{-1} + i^{-[1+(c-\zeta)]} \asymp i^{-1} \quad \text{and} \quad \mathbb{E}_M\left[\mathcal{R}_2^{(i)}(\zeta)\right] \asymp i^{-1},$$

  and the variance term, the only one which still depends on ζ, is negligible as long as $c\ge\zeta$.

The conclusions for the optimal choice of ζ in the sparse and dense cases can be summarized by

$$\zeta^* = \max\left\{c, \frac{c+1}{2H_t+1}\right\}. \tag{44}$$

We next investigate the particular case where

$$\text{either } \ \zeta=c+1 \quad \text{or} \quad 2\zeta H_t = 1+\min(0,c-\zeta).$$

Form the calculation above, we deduce that the former cannot be satisfied by ζ^* derived in (44). The latter condition is met when $2cH_t=1$, that is at the frontier between sparse and dense regimes. In this case, we have

$$\zeta^* = c = \frac{c+1}{2H_t+1} \qquad \text{and thus} \qquad 2\zeta^* H_t = 1+\min(0,c-\zeta^*).$$

In this case, the optimal rate of $\mathbb{E}_M\left[\mathcal{R}_1^{(i)}(\zeta)\right]$ is inflated by a $\log(i)$ factor.

To summarize, whenever $2cH_t\ne 1$, we have

$$\mathbb{E}\left[\left\{\widehat{\mu}^{(i)}(t;\zeta_t^*) - \mu(t)\right\}^2\right] \lesssim \min\left\{i^{-\{2(c+1)H_t/(2H_t+1)\}},\ i^{-2cH_t}\right\} + i^{-1}. \tag{45}$$

In the case for which $2cH_t=1$, for simplicity, we propose keeping the same bandwidth rule, in which case the quadratic risk, (45), is inflated by a $\log(i)$ factor coming from $\mathbb{E}_M[\{\mathcal{B}_t^{(i)}\}^2]$; see (39). $\square$

Proof of Proposition 3. For simplicity, we assume that the kernel K stays away from zero on the support. The conditions in the statement of Proposition 3 correspond to Case 1 with $0 < c < \zeta < (2H_t + 1)^{-1}$ in the proof of Proposition 2. With the notation from that proof, and using (42) and the Hölder continuity of $t \mapsto \sigma^2(t)$, for any sequence of bandwidths $h^{(j)} = h^{(j)}_{\mu,t}(\zeta)$ as defined in (22), we get

$$\begin{aligned}
&\widehat{\mu}^{(i)}(t) - \mu(t) \\
&\quad = \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) B_t^{(j)}\left(h^{(j)}\right) + \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) V_t^{(j)}\left(h^{(j)}\right) + \left\{\widetilde{\mu}^{(i)}(t) - \mu^{(i)}(t)\right\} \\
&\qquad = \mathcal{B}_t^{(i)} + \mathcal{V}_t^{(i)} + \left\{\widetilde{\mu}^{(i)}(t) - \mu(t)\right\}.
\end{aligned}$$

Let us first note that using a kernel K which is bounded and bounded away from zero guarantees that

$$\frac{\max_{1\leq m\leq M_j} W_m^{(j)}}{\min_{1\leq m\leq M_j} W_m^{(j)}} \asymp 1, \qquad j \geq 1.$$

As a consequence, the inequality sign in (40) can be replaced by $\asymp$, and the same is true for the inequality sign $\lesssim$ in (41). Then, in view of the rates derived in the proof of Proposition 2, selecting ζ such that $\zeta_t^* < \zeta < (2H_t + 1)^{-1}$ makes the term $\mathcal{B}_t^{(i)}$ negligible compared to $\mathcal{V}_t^{(i)}$. We thus get

$$\begin{aligned}
\widehat{\mu}^{(i)}(t;\zeta) - \mu(t) \asymp_{\mathbb{P}} \frac{\sigma(t)}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \left\{\sum_{m=1}^{M_j} W_m^{(j)}\left(t; h^{(j)}\right) e_m^{(i)}\right\} \\
+ \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} w^{(j)}\left(t; h^{(j)}\right) \left\{X_t^{(j)} - \mu(t)\right\} = \mathfrak{T}_1^{(i)} + \mathfrak{T}_2^{(i)}.
\end{aligned}$$

By our assumptions, $\mathfrak{T}_1^{(i)}$ and $\mathfrak{T}_2^{(i)}$ are independent. For each of them, we can apply a classical CLT for independent random variables. For instance, conditionally given the M_j and the $T_m^{(j)}$, $1 \leq m \leq M_j$, $j \geq 1$, such that $\mathcal{W}^{(i)}(t) \to \infty$, by Lyapunov CLT,

$$\sqrt{\mathcal{W}^{(i)}(t)}\; \mathfrak{T}_2^{(i)} \xrightarrow{d} N\left(0, \operatorname{Var}(X_t)\right).$$

This implies that for any realization of the M_j and the $T_m^{(j)}$ such that $\mathcal{W}^{(i)}(t) \to \infty$, we get

$$\mathbb{E}_{M,T}\left[\exp\left\{-iu\sqrt{\mathcal{W}^{(i)}(t)}\; \mathfrak{T}_2^{(i)}\right\}\right] \longrightarrow \exp(-u^2 \operatorname{Var}(t)/2), \qquad \forall u \in \mathbb{R}.$$

Since $\mathcal{W}^{(i)}(t)/\mathbb{E}[\mathcal{W}^{(i)}(t)] = 1 + o_{\mathbb{P}}(1)$ and since our conditions imply $\mathcal{W}^{(i)}(t) \to \infty$, and applying Dominated Convergence Theorem arguments to a sequence of bounded random variables convergent in probability, we get

$$\mathbb{E}\left[\exp\left\{-iu\sqrt{\mathcal{W}^{(i)}(t)}\; \mathfrak{T}_2^{(i)}\right\}\right] \longrightarrow \exp(-u^2 \operatorname{Var}(t)/2), \qquad \forall u \in \mathbb{R}.$$

By similar arguments, we get

$$\mathbb{E}\left[\exp\left\{-iu\sqrt{\mathcal{W}^{(i)}(t)}\; \mathfrak{T}_1^{(i)}\right\}\right] \longrightarrow \exp(-u^2 \Sigma(t)/2), \qquad \forall u \in \mathbb{R},$$

provided

$$\mathcal{S}^{(i)}(t;\zeta) := \sigma^2(t) \times \frac{1}{\mathcal{W}^{(i)}(t)} \sum_{j=1}^{i} \sum_{m=1}^{M_j} \left\{W_m^{(j)}\left(t; h^{(j)}\right)\right\}^2 \longrightarrow \Sigma(t), \quad \text{in probability.}$$

By the proof of Lemma 1 given in the Supplementary Material, in the case for which $0 < c < \zeta < (2H_t + 1)^{-1}$, we get $\Sigma(t) = \sigma^2(t)$. This because the probability of having more than one design point $T_m^{(j)}$ in the intervals $[t - h^{(j)}, t + h^{(j)}]$ is negligible. If only one design point falls in $[t - h^{(j)}, t + h^{(j)}]$, then $W_m^{(j)}\left(t; h^{(j)}\right) = 1$. □

References

Aneiros, G., Cao, R., Fraiman, R., Genest, C., and Vieu, P. (2019). Recent advances in functional data analysis and high-dimensional statistics. *Journal of Multivariate Analysis*, 170:3–9. Special Issue on Functional Data Analysis and Related Topics.

Balança, P. (2015). Some sample path properties of multifractional Brownian motion. *Stochastic Processes Appl.*, 125(10):3823–3850.

Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Some new asymptotic theory for least squares series: pointwise and uniform results. *J. Econometrics*, 186(2):345–366.

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

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

Chao, M. T. and Strawderman, W. E. (1972). Negative moments of positive random variables. *Journal of the American Statistical Association*, 67(338):429–431.

Chen, X., Liu, W., and Zhang, Y. (2019). Quantile regression under memory constraint. *The Annals of Statistics*, 47(6):3244 – 3273.

Freedman, D. and Diaconis, P. (1981). On the histogram as a density estimator: L_2 theory. *Z. Wahrsch. Verw. Gebiete*, 57(4):453–476.

Gaïffas, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probability and Statistics*, 11:344–364.

Gertheiss, J., Rügamer, D., Liew, B. X. W., and Greven, S. (2023). Functional data analysis: An introduction and recent developments.

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2022). Learning the smoothness of noisy curves with application to online curve estimation. *Electronic Journal of Statistics*, 16(1):1485–1560.

Golovkine, S., Klutchnikoff, N., and Patilea, V. (2023). Adaptive optimal estimation of irregular mean and covariance functions. arxiv:2108.06507v2.

Gu, H. and Lafferty, J. (2012). Sequential nonparametric regression. In *Proceedings of the 29th International Coference on International Conference on Machine Learning*, ICML'12, page 387–394, Madison, WI, USA. Omnipress.

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

Hsing, T., Brown, T., and Thelen, B. (2016). Local intrinsic stationarity and its inference. *Ann. Statist.*, 44(5):2058 – 2088.

Kiefer, J. and Wolfowitz, J. (1952). Stochastic Estimation of the Maximum of a Regression Function. *The Annals of Mathematical Statistics*, 23(3):462 – 466.

Kokoszka, P. and Reimherr, M. (2017). *Introduction to functional data analysis*. Texts in Statistical Science Series. CRC Press, Boca Raton, FL.

Kong, E. and Xia, Y. (2019). On the efficiency of online approach to nonparametric smoothing of big data. *Statist. Sinica*, 29(1):185–201.

Luo, L. and Song, P. X.-K. (2020). Renewable estimation and incremental inference in generalized linear models with streaming data sets. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 82(1):69–97.

Maissoro, H., Patilea, V., and Vimond, M. (2024). Learning the smoothness of weakly dependent functional times series. arxiv 2403.13706.

Mohammadi, N. and Panaretos, V. M. (2024). Functional data analysis with rough sample paths? *Journal of Nonparametric Statistics*, 36(1):4–22.

Mohammadi, N., Santoro, L. V., and Panaretos, V. M. (2024). Nonparametric estimation for SDE with sparsely sampled paths: An FDA perspective. *Stochastic Processes and their Applications*, 167:104239.

Poß , D., Liebl, D., Kneip, A., Eisenbarth, H., Wager, T. D., and Barrett, L. F. (2020). Superconsistent Estimation of Points of Impact in Non-Parametric Regression with Functional Predictors. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 82(4):1115–1140.

Quan, M. and Lin, Z. (2024). Optimal one-pass nonparametric estimation under memory constraint. *J. Amer. Statist. Assoc.*, 119(545):285–296.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*. Springer Series in Statistics. Springer, New York, second edition.

Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407.

Scott, D. W. (1979). On optimal and data-based histograms. *Biometrika*, 66(3):605–610.

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York.

Wang, J. L., Chiou, J. M., and Müller, H. G. (2016). Functional Data Analysis. *Annual Review of Statistics and Its Application*, 3:257–295.

Wang Guang Wei, S., Patilea, V., and Klutchnikoff, N. (2023). Adaptive functional principal components analysis. arxiv 2306.16091.

Wong, R. K. and Zhang, X. (2019). Nonparametric operator-regularized covariance function estimation for functional data. *Comput. Statist. Data Anal.*, 131:131 – 144.

Xue, D. and Yao, F. (2022). Dynamic penalized splines for streaming data. *Statist. Sinica*, 32(3):1363–1380.

Yang, Y. and Yao, F. (2023). Online estimation for functional data. *J. Amer. Statist. Assoc.*, 118(543):1630–1644.

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

Zhang, T. and Simon, N. (2023). An online projection estimator for nonparametric regression in reproducing kernel Hilbert spaces. *Statist. Sinica*, 33(1):127–148.

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

Zhou, Y., Bhattacharjee, S., Carroll, C., Chen, Y., Dai, X., Fan, J., Gajardo, A., Hadjipantelis, P. Z., Han, K., Ji, H., Zhu, C., Müller, H.-G., and Wang, J.-L. (2022). *fdapace: Functional Data Analysis and Empirical Dynamics.* R package version 0.5.9.

B Supplementary Material

In what follows, $c, c_1, C, C_1, \underline{c}, \overline{c}, \mathfrak{c}, \mathfrak{C} \ldots$, are constants that assume different values from line to line. Recall the notation $\lesssim$ which means that the left side is bounded by an absolute constant times the right side. Moreover, $\asymp$ is used when the left side is bounded above and below by absolute constants times the right side. Finally, the notation $\asymp_{\mathbb{P}}$ (resp. $\lesssim_{\mathbb{P}}$) means that the left side is bounded above and below by absolute constants (resp. bounded by an absolute constant) times the right side with probability tending to 1.

B.1 Proofs on the local regularity

Proof of Proposition 1. Let us introduce the following notation:

$$\mathbb{P}_M[\cdots] = \mathbb{P}[\cdots \mid M_i, i \geq 1] \quad \text{and} \quad \mathbb{E}_M[\cdots] = \mathbb{E}[\cdots \mid M_i, i \geq 1].$$

Recall that $X \in \mathcal{X}(H, L)$, and thus a constant Δ_0 exists such that condition (3) holds true. Let

$$R^{(i)}(2p) = \mathbb{E}_M[R(2p; M_i)] = \mathbb{E}_M\left(\left|\widetilde{X}^{(i)} - X^{(i)}\right|_\infty^{2p}\right), \quad i, p \geq 1.$$

We prove below a slightly more general statement of Proposition 1 by assuming only that $\epsilon = \epsilon(\overline{\mathfrak{m}}_i)$, $\Delta_* = \Delta_*(\overline{\mathfrak{m}}_i)$ and $R^{(i)}(2)$ satisfy the following condition:

$$\frac{\sqrt{R^{(i)}(2)}}{\epsilon \Delta_*^{\alpha_1}} + \frac{\Delta_*^{\alpha_2}}{\epsilon} \xrightarrow[i \to \infty]{} 0, \qquad \text{for any } \alpha_1, \alpha_2 > 0. \tag{B.1.1}$$

For any $t \in \mathcal{T}$, we define $\Delta_t = |t_2 - t_1| = |t_3 - t_2|$ which satisfies $\Delta_*/2 \leq \Delta_t \leq \Delta_*$, with Δ_* as in (14). Recall that t_1, t_2, t_3 are defined with respect to t, but we omit this dependence in what follows. Note that, by construction, $\Delta_* \leq \Delta_0$ whenever $i \geq i_0$, if i_0 is sufficiently large. For simplicity, we assume $i_0 = 1$. Note also that by construction, for any $2 \leq j \leq i$,

$$P_{i,j} := (1 - \gamma_i)(1 - \gamma_{i-1}) \cdots (1 - \gamma_j) = \frac{M_1^\kappa + \cdots + M_{j-1}^\kappa}{M_1^\kappa + \cdots + M_i^\kappa}, \tag{B.1.2}$$

and

$$Q_{i,j} := (1 - \gamma_i)(1 - \gamma_{i-1}) \cdots (1 - \gamma_{j+1})\gamma_j = P_{i,j+1} - P_{i,j} = \frac{M_j^\kappa}{M_1^\kappa + \cdots + M_i^\kappa}, \tag{B.1.3}$$

where $P_{i,i+1} = 1$, and thus $Q_{i,i} = \gamma_i$. Moreover, following the lines of Golovkine et al. (2023), for $k = 2$ and $k = 3$, a constant C exists (and does not depend on i) such that we have

$$\begin{aligned}
\left|\mathbb{E}\left\{\left(\widetilde{X}_{t_1}^{(i)} - \widetilde{X}_{t_k}^{(i)}\right)^2 - \left(X_{t_1}^{(i)} - X_{t_k}^{(i)}\right)^2\right\}\right| &\leq 2\mathbb{E}\left\{\left(\widetilde{X}_{t_1}^{(i)} - X_{t_1}^{(i)}\right)^2\right\} + 2\mathbb{E}\left\{\left(\widetilde{X}_{t_k}^{(i)} - X_{t_k}^{(i)}\right)^2\right\} \\
&\quad + 2\mathbb{E}\left\{\left[\left|\widetilde{X}_{t_1}^{(i)} - X_{t_1}^{(i)}\right| + \left|\widetilde{X}_{t_k}^{(i)} - X_{t_k}^{(i)}\right|\right]\left|X_{t_1}^{(i)} - X_{t_k}^{(i)}\right|\right\} \\
&\leq 4R^{(i)}(2) + 2\Delta_*^{\frac{H}{2}}\overline{L}\sqrt{R^{(i)}(2)} \lesssim \sqrt{R^{(i)}(2)}, \qquad \forall i \geq 1.
\end{aligned} \tag{B.1.4}$$

For the last inequalities we use Cauchy-Schwartz inequality and the inequality $(a+b)^2 \leq 2(a^2+b^2)$. The remainder of the proof is decomposed into three steps.

Step 1: Uniform concentration of $\widehat{\theta}^{(i)}$. Let $\psi = \psi(\overline{\mathfrak{m}}_i) > 0$ be such that,

$$\text{for some suitable } b \geq 2,\ R^{(i)}(2)/\psi^b \text{ tends to 0 as } i \text{ increases.} \tag{B.1.5}$$

We then show in the following that, for $k = 2$ and $k = 3$, we have

$$\mathbb{P}\left(\sup_{t \in \mathcal{T}}\left|\widehat{\theta}^{(i)}(t_1, t_k) - \theta(t_1, t_k)\right| > \psi\right) \leq \exp\left(-\mathfrak{c} \times i \times \psi^2\right), \tag{B.1.6}$$

for some positive constant $\mathfrak{c}$. Indeed, we first decompose

$$\begin{aligned}\widehat{\theta}^{(i)}(t_1,t_k)-\theta(t_1,t_k) &= \widehat{\theta}^{(i)}(t_1,t_k)-\mathbb{E}_M\left[\widehat{\theta}^{(i)}(t_1,t_k)\right]+\mathbb{E}_M\left[\widehat{\theta}^{(i)}(t_1,t_k)\right]-\theta(t_1,t_k)\\ &=: \underbrace{\widehat{\theta}^{(i)}(t_1,t_k)-\mathbb{E}_M\left[\widehat{\theta}^{(i)}(t_1,t_k)\right]}_{\text{stochastic term}}+\underbrace{B_M^{(i)}(t_1,t_k)}_{\text{bias term}}.\end{aligned}$$

For the stochastic term, by the definition of $\widehat{\theta}^{(i)}$ and $Q_{i,j}$, we have

$$\widehat{\theta}^{(i)}(u,v)=(1-\gamma_i)\widehat{\theta}^{(i-1)}(u,v)+\gamma_i\left(\widetilde{X}_u^{(i)}-\widetilde{X}_v^{(i)}\right)^2=\cdots=\sum_{j=1}^{i}Q_{i,j}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2,$$

and thus

$$\widehat{\theta}^{(i)}(t_1,t_k)-\mathbb{E}_M\left[\widehat{\theta}^{(i)}(t_1,t_k)\right]=\sum_{j=1}^{i}Q_{i,j}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2-\mathbb{E}_M\left[\sum_{j=1}^{i}Q_{i,j}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2\right].$$

Taking the supremum, we then have

$$\begin{aligned}&\sup_{t\in\mathcal{T}}\left|\widehat{\theta}^{(i)}(t_1,t_k)-\mathbb{E}_M\left[\widehat{\theta}^{(i)}(t_1,t_k)\right]\right|\\ &\qquad\le\sup_{t\in\mathcal{T}}\sum_{j=1}^{i}Q_{i,j}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2+\sup_{t\in\mathcal{T}}\mathbb{E}_M\left[\sum_{j=1}^{i}Q_{i,j}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2\right]\\ &\qquad\le\sum_{j=1}^{i}Q_{i,j}\sup_{t\in\mathcal{T}}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2+\sum_{j=1}^{i}Q_{i,j}\mathbb{E}_M\left[\sup_{t\in\mathcal{T}}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2\right]\\ &\qquad=\sum_{j=1}^{i}Q_{i,j}Z_j+\sum_{j=1}^{i}Q_{i,j}R^{(j)}(2)=\sum_{j=1}^{i}Q_{i,j}\overline{Z}_j+2\sum_{j=1}^{i}Q_{i,j}R^{(j)}(2),\end{aligned}$$

where

$$Z_j=\sup_{t\in\mathcal{T}}\left(\widetilde{X}_{t_1}^{(j)}-\widetilde{X}_{t_k}^{(j)}\right)^2\qquad\text{and}\qquad\overline{Z}_j=Z_j-\mathbb{E}_M(Z_j)=Z_j-R^{(j)}(2).$$

We then obtain

$$\begin{aligned}\sup_{t\in\mathcal{T}}\left|\widehat{\theta}^{(i)}(t_1,t_k)-\theta(t_1,t_k)\right| &\le\sum_{j=1}^{i}Q_{i,j}\overline{Z}_j+2\sum_{j=1}^{i}Q_{i,j}R^{(j)}(2)+\sup_{t\in\mathcal{T}}\left|\mathbb{E}_M\left\{\widehat{\theta}^{(i)}(t_1,t_k)\right\}-\theta(t_1,t_k)\right|\\ &=\sum_{j=1}^{i}Q_{i,j}\overline{Z}_j+2\sum_{j=1}^{i}Q_{i,j}R^{(j)}(2)+\sup_{t\in\mathcal{T}}\left|B_M^{(i)}(t_1,t_k)\right|.\end{aligned}$$

Note that the bias term $B_M^{(i)}(t_1, t_k)$ depends on $M_1, \ldots, M_i$. By (B.1.2), (B.1.3) and (B.1.4), we have

$$\begin{aligned}\left|B_M^{(i)}(t_1,t_k)\right| = \left|\mathbb{E}_M\left\{\widehat{\theta}^{(i)}(t_1,t_k)\right\} - \theta(t_1,t_k)\right| &\leq (1-\gamma_i)\left|B_M^{(i-1)}(t_1,t_k)\right| \\ &+ \gamma_i \left|\mathbb{E}_M\left\{\left(\widetilde{X}_{t_1}^{(i)} - \widetilde{X}_{t_k}^{(i)}\right)^2 - \left(X_{t_1}^{(i)} - X_{t_k}^{(i)}\right)^2\right\}\right| \\ &\leq (1-\gamma_i)(1-\gamma_{i-1})\left|B_M^{(i-2)}(t_1,t_k)\right| \\ &+ \gamma_i \left|\mathbb{E}_M\left\{\left(\widetilde{X}_{t_1}^{(i)} - \widetilde{X}_{t_k}^{(i)}\right)^2 - \left(X_{t_1}^{(i)} - X_{t_k}^{(i)}\right)^2\right\}\right| \\ + (1-\gamma_i)\gamma_{i-1} &\left|\mathbb{E}_M\left\{\left(\widetilde{X}_{t_1}^{(i-1)} - \widetilde{X}_{t_k}^{(i-1)}\right)^2 - \left(X_{t_1}^{(i-1)} - X_{t_k}^{(i-1)}\right)^2\right\}\right| \\ &\leq \cdots \leq P_{i,2}\left|B_M^{(1)}(t_1,t_k)\right| + C\sum_{j=2}^{i} Q_{i,j}\sqrt{R^{(j)}(2)},\end{aligned}$$

with C the constant in the $\lesssim$ relationship in (B.1.4). Thus, with C_1 being a constant which uniformly bounds the absolute value of $B_M^{(1)}(t_1, t_k)$ times M_1^κ, we have

$$\left|B_M^{(i)}(t_1,t_k)\right| \leq \frac{C_1}{M_1^\kappa + \cdots + M_i^\kappa} + CB\frac{M_1^{\kappa-\tau/2} + \cdots + M_i^{\kappa-\tau/2}}{M_1^\kappa + \cdots + M_i^\kappa}, \tag{B.1.7}$$

with B being the constant in (13). We next show that the right-hand side of the last inequality can be further bounded by quantities which do not depend neither on the realizations M_j, $j \geq 1$, nor on $t \in \mathcal{T}$. By Jensen's inequality, for any $0 < \delta \leq \kappa$,

$$\frac{M_1^{\kappa-\delta} + \cdots + M_i^{\kappa-\delta}}{M_1^\kappa + \cdots + M_i^\kappa} \leq \left\{\left(M_1^\kappa + \cdots + M_i^\kappa\right)/i\right\}^{-\delta/\kappa}. \tag{B.1.8}$$

Here, we will use this inequality with $\delta = \tau/2$ and $\delta = \kappa$. On the other hand, by (9),

$$\frac{M_1^\kappa + \cdots + M_i^\kappa}{i} \geq \frac{1}{i}\sum_{j=2}^{i} j^{a\kappa} \geq \frac{1}{i}\frac{(i-1)^{a\kappa+1}}{a\kappa+1} \geq \frac{i^{a\kappa}}{2^{a\kappa+1}(a\kappa+1)} \gtrsim i^{a\kappa}. \tag{B.1.9}$$

In the case $\kappa = 0$, the bound on the right-hand side of (B.1.7) becomes

$$\frac{C_1}{i} + \frac{CB}{i}\sum_{j=1}^{i} M_j^{-\tau/2} \leq \frac{C_1}{i} + \frac{CB}{i}\sum_{j=1}^{i} j^{-a\tau/2} \lesssim i^{-\min(1, a\tau/2)}. \tag{B.1.10}$$

Finally, since the bounds for M_j in (9) are inherited by the expectations $\mathfrak{m}_j$, we have

$$i^a \lesssim \frac{1}{i}\sum_{j=1}^{i} i^a \leq \overline{\mathfrak{m}}_i = \frac{1}{i}\sum_{i=1}^{i}\mathfrak{m}_j \leq \frac{1}{i}\sum_{j=1}^{i} j^{1/a} \leq \frac{1}{i}\frac{i^{1+1/a}}{1+1/a} = \frac{i^{1/a}}{1+1/a} \lesssim i^{1/a}. \tag{B.1.11}$$

Applying the bounds in (B.1.8), (B.1.9), (B.1.10) and (B.1.11) in (B.1.7), we deduce that $c > 0$ exists such that, for ψ smaller than or equal some suitable negative power of $\overline{\mathfrak{m}}_i$, we have

$$\sup_{t\in\mathcal{T}}\left|B_M^{(i)}(t_1,t_k)\right| \lesssim \overline{\mathfrak{m}}_i^{\,-c} \ll \psi,$$

where the last inequality holds true if $b \geq \tau/c$ in the definition (B.1.5). We then deduce, for sufficiently large i,

$$\mathbb{P}_M\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}^{(i)}(t_1,t_k)-\theta(t_1,t_k)\right|>\psi\right)\leq \mathbb{P}_M\left(\sum_{j=1}^{i}Q_{i,j}\bar{Z}_j>\psi-2\sum_{j=1}^{i}Q_{i,i}R^{(j)}(2)-\sup_{t\in\mathcal{T}}\left|B_M^{(i)}(t_1,t_k)\right|\right)$$
$$\leq \mathbb{P}_M\left(\sum_{j=1}^{i}Q_{i,j}\bar{Z}_j>\frac{\psi}{2}\right).$$

In order to apply Bernstein's inequality for the last probability, we need to bound the moments of the stochastic term. For any $p\geq 1$ and $1\leq j\leq i$, we have

$$|Z_j|^p \leq 3^{2p-1}\left\{\sup_{t\in\mathcal{T}}\left|X_{t_1}^{(j)}-X_{t_k}^{(j)}\right|^{2p}+\sup_{t\in\mathcal{T}}\left|\widetilde{X}_{t_1}^{(j)}-X_{t_1}^{(j)}\right|^{2p}+\sup_{t\in\mathcal{T}}\left|\widetilde{X}_{t_k}^{(j)}-X_{t_k}^{(j)}\right|^{2p}\right\}$$
$$\leq 3^{2p-1}\left\{\sup_{t\in\mathcal{T}}\left|X_{t_1}^{(j)}-X_{t_k}^{(j)}\right|^{2p}+2\sup_{t\in\mathcal{T}}\left|\widetilde{X}_{t}^{(j)}-X_{t}^{(j)}\right|^{2p}\right\}.$$

Since $\mathbb{E}_M(|\bar{Z}_j|^{2p})\leq 2^p\mathbb{E}_M(|Z_j|^{2p})$, we obtain:

$$\mathbb{E}_M(|\bar{Z}_j|^{2p})\leq\frac{18^p}{3}\left[\mathbb{E}_M\left(\sup_{t\in\mathcal{T}}\left|X_{t_1}^{(j)}-X_{t_k}^{(j)}\right|^{2p}\right)+2R(2p)\right].$$

Now, by (12) and (H2), we have

$$\mathbb{E}_M\left(|\bar{Z}_j|^{2p}\right)\leq\frac{18^p}{3}\frac{p!}{2}\left[\mathfrak{a}\mathfrak{A}^{p-2}+2\mathfrak{c}\mathfrak{C}^{p-2}\right]\leq\frac{p!}{2}\mathfrak{d}\mathfrak{D}^{p-2},$$

for some constants $\mathfrak{d}$ and $\mathfrak{D}$. Finally, note that using (10) we have

$$\sum_{j=1}^{i}Q_{i,j}^2=\frac{1}{i}\frac{1}{\{(M_1^\kappa+\cdots+M_i^\kappa)/i\}^2}\frac{1}{i}\sum_{j=1}^{i}M_j^{2\kappa}\leq\frac{C_\kappa}{i}. \tag{B.1.12}$$

Gathering terms, by Bernstein's inequality and (B.1.1), we obtain

$$\mathbb{P}_M\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_k)-\theta(t_1,t_k)\right|>\psi\right)\leq\exp\left(-\mathfrak{e}\times i\times\psi^2\right), \tag{B.1.13}$$

with $\mathfrak{e}$ being a constant depending on $\mathfrak{d}$, $\mathfrak{D}$, C_κ. Integrating both sides with respect to the M_i's yields (B.1.6). Let us note that (10) can be relaxed, but this would increase the bound in (B.1.12) and result in a deteriorated exponential bound after applying Bernstein's inequality, that is, i would be replaced by a power between 0 and 1 on the right-hand side of (B.1.13).

Step 2: Proxy approximation accuracy. To prove (15), we have to control the uniform norm of the difference between H and the proxy value $\widetilde{H}$ defined in (4). Notice that

$$\theta(t_1,t_k)=L_t^2|t_k-t_1|^{2H_t}\{1+\rho_t(k)\},$$

where, by (3) and (B.1.1), for sufficiently large i,

$$\rho_t(k)=\frac{\theta(t_1,t_k)-L_t^2|t_k-t_1|^{2H_t}}{L_t^2|t_k-t_1|^{2H_t}}\quad\text{is such that}\quad|\rho_t(k)|\leq\left(\frac{S}{\underline{L}}\right)^2\Delta_t^{2\beta}\leq\frac{\epsilon\log(2)}{4}.$$

This implies:

$$|\widetilde{H}-H|_\infty=\sup_{t\in\mathcal{T}}\frac{|\log(1+\rho_t(3))-\log(1+\rho_t(2))|}{2\log(2)}\leq\frac{\sup_{t\in\mathcal{T}}|\rho_t(3)-\rho_t(2)|}{\log(2)}\leq\frac{2}{\log(2)}\left(\frac{S}{\underline{L}}\right)^2\Delta_t^{2\beta}\leq\frac{\epsilon}{2}.$$

We deduce that

$$\begin{aligned}\mathbb{P}(|\widehat{H}-H|_\infty>\epsilon) &\le \mathbb{P}(|\widehat{H}-\widetilde{H}|_\infty>\epsilon-|\widetilde{H}-H|_\infty)\\ &\le \mathbb{P}(|\widehat{H}-\widetilde{H}|_\infty>\epsilon/2)\\ &\le \mathbb{P}\left(\sup_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_3)}{\theta(t_1,t_3)}\sup_{t\in\mathcal{T}}\frac{\theta(t_1,t_2)}{\widehat{\theta}(t_1,t_2)}>2^{\epsilon}\right)+\mathbb{P}\left(\inf_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_3)}{\theta(t_1,t_3)}\inf_{t\in\mathcal{T}}\frac{\theta(t_1,t_2)}{\widehat{\theta}(t_1,t_2)}<2^{-\epsilon}\right)\\ &\le (A)+(B)+(C)+(D),\end{aligned}$$

where

$$(A)=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_3)}{\theta(t_1,t_3)}>2^{\epsilon/2}\right),\qquad (B)=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\frac{\theta(t_1,t_2)}{\widehat{\theta}(t_1,t_2)}>2^{\epsilon/2}\right),$$

$$(C)=\mathbb{P}\left(\inf_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_3)}{\theta(t_1,t_3)}<2^{-\epsilon/2}\right),\quad \text{and}\quad (D)=\mathbb{P}\left(\inf_{t\in\mathcal{T}}\frac{\theta(t_1,t_2)}{\widehat{\theta}(t_1,t_2)}<2^{-\epsilon/2}\right).$$

The terms (A) to (D) can be bounded with similar arguments. We therefore provide the details only for (A) and (B). For (A) we can write

$$\begin{aligned}(A)&=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_3)}{\theta(t_1,t_3)}>2^{\epsilon/2}\right)\\ &=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\left[\widehat{\theta}(t_1,t_3)-\theta(t_1,t_3)\right]>\left(2^{\epsilon/2}-1\right)\inf_{t\in\mathcal{T}}\theta(t_1,t_3)\right)\\ &\le \mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_3)-\theta(t_1,t_3)\right|>\frac{\epsilon}{4}\cdot\inf_{t\in\mathcal{T}}\theta(t_1,t_3)\right).\end{aligned}$$

For the term (B) we have, for sufficiently large i,

$$\begin{aligned}(B)&=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\frac{\theta(t_1,t_2)}{\widehat{\theta}(t_1,t_2)}>2^{\epsilon/2}\right)=\mathbb{P}\left(\inf_{t\in\mathcal{T}}\frac{\widehat{\theta}(t_1,t_2)}{\theta(t_1,t_2)}<2^{-\epsilon/2}\right)\\ &=\mathbb{P}\left(\inf_{t\in\mathcal{T}}\left[\widehat{\theta}(t_1,t_2)-\theta(t_1,t_2)\right]<\left(2^{-\epsilon/2}-1\right)\sup_{t\in\mathcal{T}}\theta(t_1,t_2)\right)\\ &=\mathbb{P}\left(\sup_{t\in\mathcal{T}}\left[\theta(t_1,t_2)-\widehat{\theta}(t_1,t_2)\right]>\left(1-2^{-\epsilon/2}\right)\sup_{t\in\mathcal{T}}\theta(t_1,t_2)\right)\\ &\le \mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_2)-\theta(t_1,t_2)\right|\ge\frac{\epsilon}{4}\cdot\inf_{t\in\mathcal{T}}\theta(t_1,t_2)\right).\end{aligned}$$

The last inequality is a consequence of (B.1.1). Now, using (3) and (B.1.1), we get

$$\theta(t_1,t_k)\ge\Delta_t^{2H_t}\left(L_t^2-S^2\Delta_t^{2\beta}\right)\gtrsim\frac{\underline{L}^2}{2}\Delta_t^{2\overline{H}}\ge\frac{\underline{L}^2}{2}\left(\frac{\Delta_*}{2}\right)^{2\overline{H}}\ge\frac{\underline{L}^2}{8}\Delta_*^{2\overline{H}}.$$

Since the smallest bound in the inequalities does not depend on $t\in\mathcal{T}$, gathering terms we obtain:

$$\mathbb{P}\left(|\widehat{H}-H|_\infty>\epsilon\right)\lesssim 4\max_{k=1,2}\mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_k)-\theta(t_1,t_k)\right|>\frac{\underline{L}^2}{32}\epsilon\Delta_*^{2\overline{H}}\right).$$

Taking $\psi=(\underline{L}^2/32)\epsilon\Delta_*^{2\overline{H}}$ in (B.1.6), we readily obtain that

$$\mathbb{P}(|\widehat{H}-H|_\infty>\epsilon)\lesssim\exp\left(-\mathfrak{f}N\epsilon^2\Delta_*^{4\overline{H}}\right), \tag{B.1.14}$$

where $\mathfrak{f}$ is a positive constant that depends only on $\underline{L}$, $\overline{H}$, $\mathfrak{d}$ and $\mathfrak{D}$. This completes the proof of (15).

Step 3. Next, we prove (16). Using the definition of $\widehat{L}_t$, see (8), and using that $\Delta_* \leq |t_1 - t_3| \leq 2\Delta_*$, we have

$$\begin{aligned}
\left|\widehat{L}_t^2 - L_t^2\right| &= \left|\frac{\widehat{\theta}(t_1,t_3)}{|t_1-t_3|^{2\widehat{H}_t}} - \frac{\theta(t_1,t_3)}{|t_1-t_3|^{2H_t}}\right| \\
&\leq \frac{\left|\widehat{\theta}(t_1,t_3) - \theta(t_1,t_3)\right|}{|t_1-t_3|^{2\widehat{H}_t}} + \frac{\theta(t_1,t_3)}{|t_1-t_3|^{2H_t}} \cdot \left(|t_1-t_3|^{-2(\widehat{H}_t - H_t)} - 1\right) \\
&\leq \frac{\left|\widehat{\theta}(t_1,t_3) - \theta(t_1,t_3)\right|}{\Delta_*^{2\widehat{H}_t}} + \left(\overline{L}^2 + S^2(2\Delta_*)^{2\beta}\right) \cdot \left(|t_1-t_3|^{-2(\widehat{H}_t - H_t)} - 1\right).
\end{aligned}$$

Set $\phi = \epsilon/(2|\log(\Delta_*)|)$ and let us consider the following events

$$\mathcal{A} = \left\{\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_3) - \theta(t_1,t_3)\right| \leq \frac{1}{2}\epsilon\Delta_*^{2\overline{H}+\frac{\phi}{4\overline{L}^2}}\right\} \quad \text{and} \quad \mathcal{B} = \left\{\sup_{t\in\mathcal{T}}\left|\widehat{H}_t - H_t\right| \leq \frac{\phi}{16\overline{L}^2}\right\}.$$

It is easily seen that

$$\begin{aligned}
\left|\widehat{L}^2 - L^2\right|_\infty \mathbb{I}_{\mathcal{A}\cap\mathcal{B}} &\leq \frac{\epsilon}{2} + \left(\overline{L}^2 + S^2(2\Delta_*)^{2\beta}\right)\left((2\Delta_*)^{-\phi/(8\overline{L}^2)} - 1\right) \\
&\leq \frac{\epsilon}{2} + \frac{\overline{L}^2 + S^2(2\Delta_*)^{2\beta}}{2\overline{L}^2}\phi|\log\Delta_*| \leq \epsilon,
\end{aligned}$$

provided i is sufficiently large such that $S^2(2\Delta_*)^{2\beta}\overline{L}^{-2} \leq 1$. We thus obtain

$$\mathbb{P}\left(\left|\widehat{L} - L\right|_\infty > \epsilon\right) \leq \mathbb{P}\left(\left|\widehat{L} - L\right|_\infty > \epsilon, \mathcal{A}, \mathcal{B}\right) + \mathbb{P}(\overline{\mathcal{A}}) + \mathbb{P}(\overline{\mathcal{B}}) = \mathbb{P}(\overline{\mathcal{A}}) + \mathbb{P}(\overline{\mathcal{B}}).$$

Now, notice that, for sufficiently large i,

$$\begin{aligned}
\mathbb{P}(\overline{\mathcal{A}}) &= \mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_3) - \theta(t_1,t_3)\right| > \frac{1}{2}\epsilon\Delta_*^{2\overline{H}+\frac{\phi}{4\overline{L}^2}}\right) \\
&\leq \mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{\theta}(t_1,t_3) - \theta(t_1,t_3)\right| \geq \frac{1}{4}\times\epsilon\Delta_*^{2\overline{H}}\right) \leq \exp\left(-\frac{\mathfrak{e}}{16}\times i\times\epsilon^2\Delta_*^{4\overline{H}}\right),
\end{aligned}$$

where the last inequality is a consequence of (B.1.6) with $\psi = \epsilon\Delta_*^{2\overline{H}}/4$. Moreover:

$$\mathbb{P}(\overline{\mathcal{B}}) = \mathbb{P}\left(\sup_{t\in\mathcal{T}}\left|\widehat{H}_t - H_t\right| > \frac{\phi}{16\overline{L}^2}\right) \leq \exp\left(-\frac{\mathfrak{f}}{256\overline{L}^4}\times i\times\phi^2\Delta_*^{4\overline{H}}\right),$$

where the last inequality comes from (B.1.14) where ϵ is replaced by ϕ which also satisfies (B.1.1). We thus obtain

$$\mathbb{P}(\overline{\mathcal{B}}) \lesssim \exp\left(-\frac{\mathfrak{f}}{256\overline{L}^4}\times i\times\left(\frac{\epsilon}{\log(\Delta_*)}\right)^2\Delta_*^{4\overline{H}}\right).$$

Gathering terms, we have

$$\mathbb{P}\left(\left|\widehat{L} - L\right|_\infty > \epsilon\right) \lesssim \exp\left(-\mathfrak{g}\times i\times\left(\frac{\epsilon}{\log(\Delta_*)}\right)^2\Delta_*^{4\overline{H}}\right),$$

for some positive constant $\mathfrak{g}$. This ends the proof of Proposition 1. $\square$

B.2 Proofs on the covariance function estimation

Again, for simplicity, we provide the proofs for the covariance function estimation under the additional assumption that the kernel K stays away from zero on the support.

Lemma 3. *Let $s, t \in \mathcal{T}$, $s \neq t$, and assume the conditions of Proposition 5 hold true. Let $\lambda > 0$ and set $h^{(i)} = i^{-\lambda}$, $i = 1, 2, \ldots$. Define*

$$\mathcal{W}^{(i)}(s,t) = \mathcal{W}^{(i-1)}(s,t) + w^{(i)}\left(s, t; h^{(i)}\right), \qquad i = 1, 2 \ldots,$$

with $w^{(i)}(\cdot, \cdot; \cdot)$ defined as in (31), and $\mathcal{W}^{(0)}(s;t)$ some starting value. Then, constants $\underline{\mathfrak{C}}, \overline{\mathfrak{C}} \in (0,1]$ exist, depending on the bounds for the density f_T and $M_i/\mathfrak{m}_i$, on c from (11), such that

$$\underline{\mathfrak{C}} \times i^{1+2\min(0,c-\lambda)} \times \{1 + o_{\mathbb{P}}(1)\} \leq \mathcal{W}^{(i)}(s,t) \leq \overline{\mathfrak{C}} \times i^{1+2\min(0,c-\lambda)} \times \{1 + o_{\mathbb{P}}(1)\}. \qquad \text{(B.2.1)}$$

Some additional notation is required for the next lemma. Let $s \neq t$ be fixed in the following, and let $\lambda > 0$ and set $h^{(j)} = j^{-\lambda}$, $j \geq 1$. Let

$$\mathcal{N}^{(j)}(t;\lambda) = \frac{w^{(j)}(t; h^{(j)})}{\max_{1 \leq m \leq M_j} |W_m^{(j)}(t; h^{(j)})|}, \qquad j \geq 1, \qquad \text{(B.2.2)}$$

and

$$\frac{1}{\mathcal{N}_\gamma^{(i)}(t|s;\lambda)} = \frac{1}{\mathcal{W}^{(i)}(s,t)^2} \sum_{j=1}^{i} w^{(j)}(s; h^{(j)}) w^{(j)}(t; h^{(j)}) \max_{1 \leq m \leq M_j} |W_m^{(j)}(s; h^{(j)})|$$
$$= \frac{1}{\mathcal{W}^{(i)}(s,t)^2} \sum_{j=1}^{i} \frac{w^{(j)}(s,t; h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)}.$$

Here, we use the rule

$$\mathcal{N}^{(j)}(t;\lambda) = 0 \quad \Leftrightarrow \quad w^{(j)}(t; h^{(j)}) = 0 \quad \Leftrightarrow \quad \frac{w^{(j)}(t; h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)} = 0.$$

Lemma 4. *Assume that the conditions of Lemma 3 hold true. Then, with $h^{(j)} = j^{-\lambda}$, $j \geq 1$, we have*

$$\{1 + o_{\mathbb{P}}(1)\}\underline{\mathfrak{C}} \times i^{-\{1+\min[c-\lambda, 2(c-\lambda)]\}} \leq \frac{1}{\mathcal{N}_\gamma(s|t;\lambda)} \leq \{1 + o_{\mathbb{P}}(1)\}\overline{\mathfrak{C}} \times i^{-\{1+\min[c-\lambda, 2(c-\lambda)]\}}. \qquad \text{(B.2.3)}$$

The proofs of Lemmas 3 and 4 are provided in Section B.4. Before proceeding with the proof of Proposition 5, let us notice that, by construction,

$$\widehat{\gamma}^{(i)}(s,t) = \frac{1}{\mathcal{W}^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}\left(s, t; h^{(j)}\right) \widehat{X}_s^{(j)}\left(h^{(j)}\right) \widehat{X}_t^{(j)}\left(h^{(j)}\right).$$

Here, $h^{(j)} = j^{-\lambda}$ is a simplified version of $h^{(j)}_{\gamma,(s,t)}(\lambda)$. Using the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, we can bound the conditional quadratic risk of $\widehat{\gamma}^{(i)}(t)$, given the M_j, $j \geq 1$, and the realizations of T, as follows :

$$\mathbb{E}_{M,T}\left[\left\{\widehat{\gamma}^{(i)}(s,t) - \gamma(s,t)\right\}^2\right] \lesssim \mathbb{E}_{M,T}\left[\left\{\widehat{\gamma}^{(i)}(s,t) - \widetilde{\gamma}^{(i)}(s,t)\right\}^2\right] + \mathbb{E}_{M,T}\left[\left\{\widetilde{\gamma}^{(i)}(s,t) - \overline{\gamma}^{(i)}(s,t)\right\}^2\right]$$
$$+ \mathbb{E}_{M,T}\left[\left\{\overline{\gamma}^{(i)}(s,t) - \gamma^{(i)}(s,t)\right\}^2\right] =: \mathfrak{R}_1^{(i)}(\lambda) + \mathfrak{R}_2^{(i)}(\lambda) + \mathfrak{R}_3^{(i)},$$

where $\widetilde{\gamma}^{(i)}(s,t)$ and $\overline{\gamma}^{(i)}(s,t)$ are infeasible estimators defined as

$$\widetilde{\gamma}^{(i)}(s,t) = \frac{1}{\mathcal{W}^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}\left(s,t;h^{(j)}\right) X_s^{(j)} X_t^{(j)} \quad \text{and} \quad \overline{\gamma}^{(i)}(s,t) = \frac{1}{i}\sum_{j=1}^{i} X_s^{(j)} X_t^{(j)},$$

respectively. Clearly, the $\mathcal{R}_3^{(i)}$ does not depend on the bandwidths $h^{(j)}$, the realizations of T and the M_i, and has the rate $O_{\mathbb{P}}(i^{-1})$.

Proof of Proposition 5. Recall the notation

$$B_t^{(j)}(h) = \mathbb{E}_{M,T}\left[\widehat{X}_t^{(j)}(h)\right] - X_t^{(j)} \quad \text{and} \quad V_t^{(j)}(h) = \widehat{X}_t^{(j)}(h) - \mathbb{E}_{M,T}\left[\widehat{X}_t^{(i)}(h)\right], \quad t \in \mathcal{T},$$

that are the bias and the stochastic part of $\widehat{X}_t^{(j)}(h)$, respectively, given M_j and the design points $T_m^{(j)}$. Since $\widehat{X}_t^{(j)}(h^{(j)}) - X_t^{(j)}(h^{(j)}) = B_t^{(j)}(h^{(j)}) + V_t^{(j)}(h^{(j)})$, $\forall t$, and simplifying the expression by omitting the argument $h^{(j)}$, we have

$$\begin{aligned}
\widehat{\gamma}^{(i)}(s,t) - \widetilde{\gamma}^{(i)}(s,t) &= \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\{\widehat{X}_s^{(j)} - X_s^{(j)}\}X_t^{(j)} \\
&\quad + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t) X_s^{(j)}\{\widehat{X}_t^{(j)} - X_t^{(j)}\} \\
&\quad + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\{\widehat{X}_s^{(j)} - X_s^{(j)}\}\{\widehat{X}_t^{(j)} - X_t^{(j)}\} \\
&= \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\left\{B_s^{(j)} X_t^{(j)} + X_s^{(j)} B_t^{(j)}\right\} \\
&\quad + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\left\{V_s^{(j)} X_t^{(j)} + X_s^{(j)} V_t^{(j)}\right\} \\
&\quad + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\left\{B_s^{(j)} B_t^{(j)} + V_s^{(j)} V_t^{(j)}\right\} \\
&\quad + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t)\left\{B_s^{(j)} V_t^{(j)} + V_s^{(j)} B_t^{(j)}\right\}.
\end{aligned}$$

By construction,

$$\mathbb{E}_{M,T}\left\{V_s^{(j')} B_t^{(j)}\right\} = \mathbb{E}_{M,T}\left\{B_s^{(j')} V_t^{(j)}\right\} = 0, \quad \forall s,t,\ \forall 1 \le j,j' \le i.$$

Moreover, whenever $2h < |s-t|$, we have

$$\mathbb{E}_{M,T}\left\{V_s^{(j)} V_t^{(j')}\right\} = 0, \quad \forall 1 \le j,j' \le i.$$

Using these properties, the inequality $(a+b)^2 \le 2(a^2+b^2)$, and by repeated application of Cauchy-Schwarz inequality to verify the negligible terms, we deduce (with $w^{(j)}(s,t)$ a short notation for $w^{(j)}(s,t;h)$)

$$\begin{aligned}
\mathfrak{R}_1^{(i)}(\lambda) := \mathbb{E}_{M,T}&\left[\left\{\frac{1}{W^{(i)}(s,t)}\sum_{j=1}^{i} w^{(j)}(s,t)\left\{B_s^{(j)}X_t^{(j)} + X_s^{(j)}B_t^{(j)}\right\}\right\}^2\right] \\
&+ \mathbb{E}_{M,T}\left[\left\{\frac{1}{W^{(i)}(s,t)}\sum_{j=1}^{i} w^{(j)}(s,t)\left\{V_s^{(j)}X_t^{(j)} + X_s^{(j)}V_t^{(j)}\right\}\right\}^2\right] + \text{ negligible terms} \\
\leq 2\mathbb{E}_{M,T}&\left[\left\{\frac{1}{W^{(i)}(s,t)}\sum_{j=1}^{i} w^{(j)}(s,t)B_s^{(j)}X_t^{(j)}\right\}^2\right] + 2\mathbb{E}_{M,T}\left[\left\{\frac{1}{W^{(i)}(s,t)}\sum_{j=1}^{i} w^{(j)}(s,t)X_s^{(j)}B_t^{(j)}\right\}^2\right] \\
&+ \frac{1}{W^{(i)}(s,t)^2}\sum_{j=1}^{i} w^{(j)}(s,t)\mathbb{E}_{M,T}\left[\left\{V_s^{(j)}X_t^{(j)}\right\}^2 + \left\{X_s^{(j)}V_t^{(j)}\right\}^2\right] + \text{ negligible terms} \\
&=: \{\mathfrak{G}_1(s,t;\lambda) + \mathfrak{G}_2(s,t;\lambda)\}\{1 + o_{\mathbb{P}}(1)\}.
\end{aligned}$$

Using the fact that the $\varepsilon_m^{(j)}$ are independent, we can now write

$$\begin{aligned}
\mathbb{E}_{M,T}\left[\left\{X_s^{(j)}V_t^{(j)}\right\}^2\right] &= \mathbb{E}_{M,T}\left[\left\{X_s^{(j)}\right\}^2\left\{\sum_{m=1}^{M_j}\varepsilon_m^{(j)}W_m^{(j)}(t;h^{(j)})\right\}^2\right] \\
&= \mathbb{E}_{M,T}\left[\left\{X_s^{(j)}\right\}^2\sum_{m=1}^{M_j}\mathbb{E}\left[\left\{\left|\varepsilon_m^{(j)}\right|^2\right\} \mid X^{(j)}\right]\left|W_m^{(j)}(t;h^{(j)})\right|^2\right] \\
&\leq \{1+o(1)\}\sigma^2(t)m_2(s)\left\{\max_m\left|W_m^{(j)}(t;h^{(j)})\right| \times \sum_{m=1}^{M_j}\left|W_m^{(j)}(t;h^{(j)})\right|\right\},
\end{aligned}$$

where $m_2(s) = \mathbb{E}\left[X_s^2\right] = \mathbb{E}_{M,T}\left[X_s^2\right]$. We deduce

$$\mathfrak{G}_2(s,t;\lambda) \lesssim_{\mathbb{P}} \frac{\sigma^2(t)}{W^{(i)}(s,t)^2}\sum_{i=1}^{N} w^{(j)}(s,t;h^{(j)})\left[\frac{m_2(t)}{\mathcal{N}^{(j)}(s;\lambda)} + \frac{m_2(s)}{\mathcal{N}^{(j)}(t;\lambda)}\right],$$

where the $\mathcal{N}^{(j)}(s;\lambda)$ and $\mathcal{N}^{(j)}(t;\lambda)$ are defined as in (B.2.2).

To bound the terms related to the bias of $\widehat{X}_t^{(j)}(h^{(j)})$, by the law of large numbers and the Dominated

Convergence Theorem, we can write

$$\begin{aligned}
\mathfrak{G}_1(s,t;\lambda) \leq 2\mathbb{E}_{M,T} & \left[\frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \left| B_t^{(i)} \right|^2 \right. \\
& \left. \times \left\{ m_2(s) + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \left(\left| X_s^{(j)} \right|^2 - m_2(s) \right) \right\} \right] \\
& + 2\mathbb{E}_{M,T} \left[\frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \left| B_s^{(i)} \right|^2 \right. \\
& \left. \times \left\{ m_2(t) + \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \left(\left| X_t^{(j)} \right|^2 - m_2(t) \right) \right\} \right] \\
= \frac{2m_2(s)}{W^{(i)}(s,t)} & \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \mathbb{E}_{M,T} \left\{ \left| B_t^{(i)} \right|^2 \right\} + \frac{2m_2(t)}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) \mathbb{E}_{M,T} \left\{ \left| B_s^{(i)} \right|^2 \right\} \\
& + \text{ negligible terms} \\
& \leq 2 \left\{ m_2(s) L_t^2 \times \overline{\mathfrak{C}}(t,s;h,2H_t) + m_2(t) L_s^2 \times \overline{\mathfrak{C}}(s,t;h,2H_s) \right\} \{1 + o_{\mathbb{P}}(1)\},
\end{aligned}$$

where

$$\overline{\mathfrak{C}}(t,s;\lambda,\alpha) = \frac{1}{W^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)}) c^{(j)}(t;h^{(j)},\alpha) \{h^{(j)}\}^{2\alpha},$$

and

$$c^{(j)}(t;h,\alpha) = \sum_{m=1}^{M_j} \left| (T_m^{(j)} - t)/h \right|^{\alpha} \left| W_m^{(j)}(t;h) \right|.$$

Let us note that with the weights of the local constant smoother with the kernel support on $[-1,1]$, $0 \leq c^{(j)}(t;h,\alpha) \leq 1$. Moreover, $c^{(j)}(t;h,\alpha) \approx \int |u|^{\alpha} K(u) du$ if M_j is large. In the proof or Lemma 3 we prove the following rate

$$\mathbb{E}_M \left[w^{(j)} \left(s,t;h^{(j)} \right) \right] = \pi^{(j)}(s,t;h^{(j)}) \asymp \underline{\pi}^{(j)}(s,t;h^{(j)}) \asymp \min\{1, (j^c h^{(j)})^2\}, \qquad \forall s \neq t,$$

provided j is sufficiently large such that $|s-t| > 2h^{(j)}$. By the arguments used in the proof of Proposition 2, we deduce

$$\mathbb{E}_M \left[\mathfrak{G}_1(s,t;\lambda) \right] \lesssim i^{-\min\{1, 2\lambda H_{s,t}, 1+2(c-\lambda)\}},$$

provided $2\lambda H_{s,t} \neq 1 + \min\{0, 2(c-\lambda)\}$. In the case $2\lambda H_{s,t} = 1 + \min\{0, 2(c-\lambda)\}$, the bound of $\mathbb{E}_M \left[\mathfrak{G}_1(s,t;\lambda) \right]$ should be multiplied by $\log(i)$. On the other hand, by Lemma 4

$$\mathbb{E}_M \left[\mathfrak{G}_2(s,t;\lambda) \right] \lesssim i^{-[1+\min\{(c-\lambda), 2(c-\lambda)\}]}.$$

Gathering terms, we obtain

$$\mathbb{E}_M \left[\mathfrak{R}_1^{(i)}(\lambda) \right] \lesssim i^{-\min\{1, 2\lambda H_{s,t}, 1+2(c-\lambda)\}} + i^{-[1+\min\{(c-\lambda), 2(c-\lambda)\}]}.$$

Finally, we have

$$\begin{aligned}
\mathbb{E}_{M,T} \left[\mathfrak{R}_2^{(i)}(\lambda) \right] = \mathbb{E}_{M,T} \left[\left\{ \widetilde{\gamma}^{(i)}(s,t) - \overline{\gamma}^{(i)}(s,t) \right\}^2 \right] = \frac{\operatorname{Var}(X_s X_t)}{W^{(i)}(s,t)^2} \sum_{j=1}^{i} \left\{ w^{(j)}(s,t;h^{(j)}) - \frac{W^{(i)}(s,t)}{i} \right\}^2 \\
= \operatorname{Var}(X_s X_t) \left\{ \frac{1}{W^{(i)}(s,t)} - \frac{1}{i} \right\}.
\end{aligned}$$

By Lemma 3, we get
$$\mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda)\right] \lesssim i^{-[1+\min\{0,2(c-\lambda)\}]}.$$

We conclude that

$$\begin{aligned}\mathcal{R}_\gamma^{(i)}(s,t;\lambda) &= \mathbb{E}\left[\left\{\widehat{\gamma}^{(i)}(s,t)-\gamma(s,t)\right\}^2\right] \\ &\lesssim i^{-\min\{1,2\lambda H_{s,t},1+2(c-\lambda)\}} + i^{-[1+\min\{(c-\lambda),2(c-\lambda)\}]} + i^{-[1+\min\{0,2(c-\lambda)\}]} + i^{-1}.\end{aligned}$$

In the case $2\lambda H_t = 1+\min\{0,2(c-\lambda)$, the first of the four terms in the risk bound has to be multiplied by $\log(i)$. Abusing notation somewhat for simplicity, we omit this factor from our results.

- Sparse Case: $2cH_{s,t} < 1$. As in the case of the mean estimator, we have to distinguish several sub-cases corresponding to the range of λ.
  - If $0 < 1+\min\{0,2(c-\lambda)\} = 1+2(c-\lambda) \leq 2\lambda H_{s,t}$, which in particular means the constraints $c \leq \lambda$ and $(2c+1)/\{2(H_{s,t}+1)\} \leq \lambda$, we have

    $$\mathbb{E}_M\left[\mathfrak{R}_1^{(i)}(\lambda)\right] \lesssim i^{-[1+2(c-\lambda)]} + i^{-[1+2(c-\lambda)]} \quad \text{and} \quad \mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda)\right] \asymp i^{-[1+2(c-\lambda)]},$$

    the optimum choice for $\mathcal{R}_1^{(i)}(\lambda)$ is then when λ takes the smallest admissible value given the constraints, that is
    $$\lambda^* = \max\left\{c, \frac{2c+1}{2(H_{s,t}+1)}\right\} = \frac{2c+1}{2(H_{s,t}+1)}.$$
    (The second equality in the last display is given by the sparse case condition $2cH_{s,t} \leq 1$.) We then have
    $$\mathbb{E}_M\left[\mathfrak{R}_1^{(i)}(\lambda^*)\right] + \mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda^*)\right] \lesssim i^{-(2c+1)H_{s,t}/(H_{s,t}+1)}.$$
  - If $0 < 2\lambda H_{s,t} < 1+\min\{0,2(c-\lambda)\} = 1+2(c-\lambda)$, then necessarily $\lambda \geq c$ and $1+2c > 2\lambda(H_{s,t}+1)$, and these two conditions can be satisfied only if

    $$\frac{2c+1}{2(H_{s,t}+1)} \leq c \quad \text{which is equivalent to } 2cH_{s,t} \geq 1.$$

    We then get a contradiction with the condition $2cH_{s,t} < 1$ of the sparse case.
  - If $1+\min\{0,2(c-\lambda)\} = 1$, and thus $c \geq \lambda$, in the sparse regime, we can only have $0 < 2\lambda H_{s,t} \leq 1+\min\{0,2(c-\lambda)\} = 1$. Then we get

    $$\mathbb{E}_M\left[\mathfrak{R}_1^{(i)}(\lambda)\right] \lesssim i^{-2\lambda H_{s,t}} + i^{-(1+c-\lambda)} \quad \text{and} \quad \mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda)\right] \asymp i^{-1}.$$

    The optimum rate for λ would be $\lambda = (c+1)/(2H_{s,t}+1)$, but this is larger than or equal to c when $2cH_{s,t} < 1$, which contradicts the condition $1+\min(0,c-\lambda) = 1$ which means $c \geq \lambda$. The optimal value for λ is the largest admissible one under the constraints, because it minimizes the bias term which is here dominant. Thus, $\lambda^* = c$.

  The conclusion for the sparse case is the following: the optimal value of λ can be described as

  $$\lambda^* = \max\left[\frac{2c+1}{2(H_{s,t}+1)}, c\right] \geq c.$$

- Dense Case: $2cH_{s,t} > 1$.

– If $0 < 1 + \min\{0, 2(c-\lambda)\} = 1 + 2(c-\lambda) < 2\lambda H_{s,t}$, then necessarily $\lambda \geq \max\{c, (2c+1)/\{2(H_{s,t}+1)\}\} = c$ and $2\lambda H_{s,t} > 1$. We then have

$$\mathbb{E}_M\left[\mathfrak{R}_1^{(i)}(\lambda)\right] \lesssim i^{-(1+c-\lambda)} + i^{-[1+2(c-\lambda)]} \quad \text{and} \quad \mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda)\right] \asymp i^{-[1+2(c-\lambda)]},$$

the optimum choice for $\mathfrak{R}_1^{(i)}(\lambda)$ is then when

$$\lambda^* = c = \max\left\{c, \frac{2c+1}{2(H_{s,t}+1)}\right\},$$

and we have $\mathfrak{R}_2^{(i)}(\lambda^*)$ and $\mathfrak{R}_1^{(i)}(\lambda^*)$ of rate $O_{\mathbb{P}}(i^{-1})$, that is the parametric rate.

– If $0 < 2\lambda H_{s,t} < 1 + \min\{0, 2(c-\lambda)\} = 1 + 2(c-\lambda) \leq 1$, it means $\lambda \geq c$, and we thus get a contradiction with the dense regime condition $2cH_{s,t} > 1$.

– If $1 + \min\{0, 2(c-\lambda)\} = 1$, we can only have $0 < 1 + 2\min(0, c-\lambda) = 1 < 2\lambda H_{s,t}$ and thus

$$\mathbb{E}_M\left[\mathfrak{R}_1^{(i)}(\lambda)\right] \lesssim i^{-1} + i^{-[1+(c-\lambda)]} \asymp i^{-1} \quad \text{and} \quad \mathbb{E}_M\left[\mathfrak{R}_2^{(i)}(\lambda)\right] \asymp i^{-1},$$

and the variance term, the only one which still depends on λ, is negligible as long as $c \geq \lambda$. Let us point out that this case is different from what happens with batch samples. Here, the bias term no longer depends on the bandwidth, and this induces a different optimal bandwidth compared to that found by Golovkine et al. (2023). The reason comes from the recursive nature of our estimator and corresponds to the case where

$$\mathbb{E}_M\left[\sum_{j=1}^{i} w^{(j)}(s,t;h^{(j)})\left|B_t^{(i)}\right|^2\right] \asymp \mathbb{E}_M\left[W^{(i)}(s,t)\right] \asymp i,$$

while in the batch samples case the rate of the first expectation always depends on the bandwidth and has a faster decrease than the rate of the second expectation.

□

B.3 Technical proofs for mean function estimation

Here, we provide the proofs of Lemmas 1 and 2. For the sake of readability we reintroduce some previous notation and definitions. For simplicity, we provide the proofs for the mean function estimation under the additional assumption that the kernel K stays away from zero on the support. This assumption can be dropped at the price of intricate technical arguments for controlling denominators close to zero.

Lemma 1. *Assume the conditions of Proposition 2 hold true, and the kernel $K(\cdot)$ is bounded and bounded away from zero on the support $[-1,1]$. Moreover, let $\zeta > 0$ and $0 < \underline{c} \leq \overline{c}$ be some constants, and, for each $i \geq 1$, define the bandwidth range $\mathcal{H}_i = [\underline{c}i^{-\zeta}, \overline{c}i^{-\zeta}]$. Then,*

$$0 \leq \max_{1\leq m\leq M_i} W_m^{(i)}(t;h) \leq S_{i,W}(h) \times i^{\min(0,\zeta-c)}, \qquad \forall h \in \mathcal{H}_i,$$

where $S_{i,W}(h) \geq 1$ is a random variable with the mean and the variance bounded by constants which do not depend on h and i. Moreover, the variables $\{S_{i,W}(h), i \geq 1\}$ are independent.

Proof of Lemma 1. Without loss of generality, set $\mathcal{T} = [0,1]$. By construction, the weights of the local-constant smoother with a non-negative kernel satisfy

$$0 \leq \min_{1\leq m\leq M_i} W_m^{(i)}(t;h) \leq \max_{1\leq m\leq M_i} W_m^{(i)}(t;h) \leq 1, \qquad \forall |t| \leq 1, \forall h > 0.$$

Next, we study more carefully the case where $M_i h$ is sufficiently large, that means in our case $i^{c-\zeta} \geq C$ for some constant $C > 0$. Here, $c > 0$ is the power appearing in condition (11). Using the fact that the kernel is bounded and bounded away from zero on $[-1, 1]$, for each $i \geq 1$ we have

$$\max_{1 \leq m \leq M_i} W_m^{(i)}(t; h) \leq \frac{\|K\|_\infty}{\inf_{|t| \leq 1} K(t)} \frac{\mathbf{1}\{S(M_i, t, h) > 0\}}{S(M_i, t, h)},$$

where

$$S(M_i, t, h) = \sum_{m=1}^{M_i} \mathbf{1}\{|T_m^{(i)} - t| \leq h\} \leq M_i \leq i^c \times \sup_{j \geq 1} j^{-c} M_j.$$

Condition (11) guarantees $\sup_{j \geq 1} j^{-c} M_j < \infty$. Here, $S(M_i, t, h)$ is an integer-valued variable, non-decreasing as a function of h. Given the integer realization M_i, the variable $S(M_i, t, h)$ has a conditional Binomial distribution with parameters M_i and

$$\mathbb{P}(|T_m^{(i)} - t| \leq h) = \int_{(t-h) \vee 1}^{(t+h) \wedge 1} f_T(u) du \geq h \times \inf_{u \in \mathcal{T}} f_T(u).$$

Let us now recall a result of Chao and Strawderman (1972): if S is a non-degenerate Binomial random variable $B(n, p)$, then

$$\mathbb{E}[(1 + S)^{-1}] = \frac{1 - q^{n+1}}{(n+1)p}, \qquad \text{where } q = 1 - p. \tag{B.3.1}$$

We here have $S = S(M_i, t, h)$, $n = M_i$ and $p = \mathbb{P}(|T_m^{(i)} - t| \leq h)$. From (B.3.1), by some elementary algebra, we deduce that constants $\mathfrak{c}_1, \mathfrak{c}_2 > 0$ exist such that

$$\frac{\mathfrak{c}_1}{np} \leq \mathbb{E}\left[\frac{\mathbf{1}\{S > 0\}}{S}\right] \leq \frac{\mathfrak{c}_2}{np},$$

provided $i^{c-\zeta} \geq C$ and the constant C is sufficiently large. See also Maissoro et al. (2024) for the details. Replacing S by $S(M_i, t, h)$, and using the independence between the M_i and $T_m^{(i)}$, we get

$$\frac{c_1}{i^{c-\zeta}} \leq \mathbb{E}_M\left[\frac{\mathbf{1}\{S(M_i, t, h) > 0\}}{S(M_i, t, h)}\right] \leq \frac{c_2}{i^{c-\zeta}}, \qquad \forall h \in \mathcal{H}_i, \forall t \in \mathcal{T},$$

where c_1, c_2 are positive constants, depending only on C, the bounds of the density f_T, and $\inf_{j \geq 1} j^{-c} M_j$, $\sup_{j \geq 1} j^{-c} M_j$. Similarly, it can be shown that a constant c' exists such that

$$\mathbb{E}_M\left[\frac{\mathbf{1}\{S(M_i, t, h) > 0\}}{S^2(M_i, t, h)}\right] \leq \frac{c'}{(i^{c-\zeta})^2}, \qquad \forall h \in \mathcal{H}_i, \forall t \in \mathcal{T}.$$

The result then follows by defining

$$S_{i,W}(h) = \max\left\{1,\ i^{c-\zeta}\left(\|K\|_\infty / \inf_{|t| \leq 1} K(t)\right) \mathbf{1}\{S(M_i, t, h) > 0\} S^{-1}(M_i, t, h)\right\}.$$

Indeed, under our assumptions, the $\{S_{i,W}(h), i \geq 1\}$ are independent, and we showed above that their mean and variance are uniformly bounded with respect to h and i. $\square$

Lemma 2. *Let $t \in \mathcal{T}$ and assume the conditions of Proposition 2 hold true. Let $\zeta > 0$ and set $h^{(i)} = i^{-\zeta}$, $i = 1, 2, \ldots$. Define*

$$\mathcal{W}^{(i)}(t) = \mathcal{W}^{(i-1)}(t) + w^{(i)}\left(t; h^{(i)}\right), \qquad i = 1, 2 \ldots,$$

with $w^{(i)}(\cdot, \cdot)$ defined as in (21), and $\mathcal{W}^{(0)}(t)$ some starting value. Then, constants $\underline{C}, \overline{C} \in (0, 1]$ exist, depending on the bounds for the density f_T and $M_i/\mathfrak{m}_i$, on c from (11), such that

$$\underline{C} \times i^{1+\min(0, c-\zeta)} \times \{1 + o_{\mathbb{P}}(1)\} \leq \mathcal{W}^{(i)}(t) \leq \overline{C} \times i^{1+\min(0, c-\zeta)} \times \{1 + o_{\mathbb{P}}(1)\}.$$

Proof of Lemma 2. Without loss of generality, the initial value $\mathcal{W}^{(0)}(t)$ can be simply set equal to 0. In view of (11), constants $\underline{C}_{\mathfrak{m}}, \overline{C}_{\mathfrak{m}}$ exist such that

$$0 < \underline{C}_{\mathfrak{m}} \leq \min\{M_i/\mathfrak{m}_i, i^c/\mathfrak{m}_i\} \leq \max\{M_i/\mathfrak{m}_i, i^c/\mathfrak{m}_i\} \leq \overline{C}_{\mathfrak{m}}, \qquad \forall i. \tag{B.3.2}$$

By our continuity and boundedness assumptions on the density of the design, constants $C_{f,L}$ and $C_{f,U}$ exist such that

$$0 < C_{f,L} \leq f_T(t) \leq C_{f,U}, \qquad \forall t \in \mathcal{T}.$$

Next, by definition, for any t, $w^{(i)}(t;h)$ is a Bernoulli variable with parameter

$$\pi^{(i)}(t;h) = \left[1 - \{1 - p(t;h)\}^{M_i}\right] = 1 - \{1 - p(t;h)\}^{M_i},$$

where $p(t;h) = \int_{t-h}^{t+h} f_T(u)du = 2hf_T(t)\{1 + o(1)\}$. Here, the $o(1)$ term is bounded by a constant times h at some power given by the Hölder exponent of f_T. We thus have

$$\mathbb{E}_M\left[\mathcal{W}^{(i)}(t)\right] = \sum_{j=1}^{i} \mathbb{E}_M\left[w^{(j)}\left(t;h^{(j)}\right)\right] = \sum_{j=1}^{i} \pi^{(j)}\left(t;h^{(j)}\right),$$

and we therefore need to suitably bound the $\pi^{(j)}(t;h)$'s. We have the following elementary inequality: for any $y \in (0,1)$,

$$-\frac{y}{1-y} < \log(1-y) < -\frac{y}{1-y/2}. \tag{B.3.3}$$

Since

$$1 - \{1 - p(t;h)\}^{M_j} = 1 - \exp\left(M_j \log\{1 - p(t;h)\}\right),$$

we can write

$$1 - \exp\left(-M_j \frac{p(t;h)}{1 - p(t;h)/2}\right) \leq \pi^{(j)}(t;h) \leq 1 - \exp\left(-M_j \frac{p(t;h)}{1 - p(t;h)}\right), \qquad \forall h, \forall j.$$

Since, without loss of generality, we can consider $p(t;h) < 1/8$ and (11) holds true,

$$\begin{aligned} 2C_{f,L}\underline{C}_{\mathfrak{m}}\mathfrak{m}_j h \leq (16/15)\underline{C}_{\mathfrak{m}}\mathfrak{m}_j p(t;h) \leq M_j \frac{p(t;h)}{1 - p(t;h)/2} \\ < M_j \frac{p(t;h)}{1 - p(t;h)} \leq \overline{C}_{\mathfrak{m}}\mathfrak{m}_j p(t;h) \leq 2C_{f,U}\overline{C}_{\mathfrak{m}}\mathfrak{m}_j h, \qquad \forall h, \end{aligned} \tag{B.3.4}$$

and it thus suffices to conveniently bound $\exp\left(-C\mathfrak{m}_j h^{(j-1)}\right)$, for some $C > 0$. On the one hand, we have the following elementary inequality: if $C > 0$ then

$$\{1 - \exp(-C)\}u \leq 1 - \exp(-Cu) \leq Cu, \qquad \forall u \in (0,1].$$

On the other hand, whenever $\mathfrak{m}_j h > 1$, constants c_1, c_2 exist such that

$$0 < c_1 \leq \pi^{(j)}(t;h) \leq c_2.$$

Moreover, by definition $\mathcal{W}^{(i)}(t) \leq i$, and by (B.3.4) and (B.3.2) we then deduce that constants c_3, c_4 exist such that

$$\begin{aligned} 0 < c_3 \min\left\{i, \sum_{j=1}^{i} j^{c-\zeta}\right\} = c_3 \sum_{j=1}^{i} \min\{1, j^c h^{(j)}\} \leq \mathbb{E}_M[\mathcal{W}^{(i)}(t)] \\ = \sum_{j=1}^{i} \pi^{(j)}(t;h^{(j)}) \leq c_4 \sum_{j=1}^{i} \min\{1, j^c h^{(j)}\} = c_4 \min\left\{i, \sum_{j=1}^{i} j^{c-\zeta}\right\}, \end{aligned}$$

with $c > 0$ the constant from condition (11). By Lemma 5 in the Section B.6, and the condition $c+1 > \zeta$, $\sum_{j=1}^{i} j^{c-\zeta} \asymp i^{1+c-\zeta}$, we thus deduce

$$\mathbb{E}_M[\mathcal{W}^{(i)}(t)] \asymp i^{1+\min(0,c-\zeta)} \to \infty.$$

Finally, since $\mathcal{W}^{(i)}(t)$, which is a sum of independent Bernoulli random variables, using the Chernoff bounds, we deduce that $\mathcal{W}^{(i)}(t)$ concentrates around its expectation with high probability. Now, the result follows. □

B.4 Technical proofs for the covariance function estimation

Here, we provide the proofs of Lemmas 3 and 4. For the sake of readability we reintroduce some notation and definitions. Again, for simplicity, we provide the proofs for the covariance function estimation under the additional assumption that the kernel K stays away from zero on the support.

Lemma 3. *Let $s,t \in \mathcal{T}$, $s \neq t$, and assume the conditions of Proposition 5 hold true. Let $\lambda > 0$ and set $h^{(i)} = i^{-\lambda}$, $i = 1, 2, \ldots$. Define*

$$\mathcal{W}^{(i)}(s,t) = \mathcal{W}^{(i-1)}(s,t) + w^{(i)}\left(s,t;h^{(i)}\right), \qquad i = 1, 2 \ldots,$$

with $w^{(i)}(\cdot,\cdot;\cdot)$ defined as in (31), and $\mathcal{W}^{(0)}(s;t)$ some starting value. Then, constants $\underline{\mathfrak{C}}, \overline{\mathfrak{C}} \in (0,1]$ exist, depending on the bounds for the density f_T and $M_i/\mathfrak{m}_i$, on c from (11), such that

$$\underline{\mathfrak{C}} \times i^{1+2\min(0,c-\lambda)} \times \{1 + o_{\mathbb{P}}(1)\} \leq \mathcal{W}^{(i)}(s,t) \leq \overline{\mathfrak{C}} \times i^{1+2\min(0,c-\lambda)} \times \{1 + o_{\mathbb{P}}(1)\}. \tag{B.4.1}$$

Proof of Lemma 3. Without loss of generality, the initial value $\mathcal{W}^{(0)}(s,t)$ can be simply set equal to 0. Let $C_{f,L}, C_{f,U} > 0$ be the lower and upper bounds of the design density f_T, respectively. Moreover, by (11), constants $\underline{C}_\mathfrak{m}, \overline{C}_\mathfrak{m}$ exist such that

$$0 < \underline{C}_\mathfrak{m} \leq \min_{i\geq 1}\{M_i i^{-c}\} \leq \max_{i\geq 1}\{M_i i^{-c}\} \leq \overline{C}_\mathfrak{m}.$$

By definition, for any s, t, the weight $w^{(i)}(s,t;h)$ is a Bernoulli variable. Let $\pi^{(i)}(s,t;h)$ denote its success probability for which we next derive lower and upper bounds. Since $w^{(i)}(s,t;h) = w^{(i)}(s;h)w^{(i)}(t;h)$ and $w^{(i)}(s;h)$, $w^{(i)}(t;h)$ are indicator functions, we have

$$w^{(i)}(s;h)w^{(i)}(t;h) \geq 1 - \{1 - w^{(i)}(s;h)\} - \{1 - w^{(i)}(t;h)\},$$

and, with $p(t;h) = \int_{t-h}^{t+h} f_T(u)du = 2hf_T(t)\{1 + o(1)\}$, we have

$$\pi^{(i)}(s,t;h) \geq 1 - \{1 - p(s;h)\}^{M_i} - \{1 - p(t;h)\}^{M_i}.$$

From the arguments presented in the proof of Lemma 2, when $c > \lambda$ (that means when $M_i h \to \infty$), without loss of generality we can assume that h is such that $\{1 - p(s;h)\}^{M_i} + \{1 - p(t;h)\}^{M_i} \leq 1/2$, $\forall s,t$ and $\forall i$. On the other hand, by definition of the multinomial distribution, whenever $|s-t| > 2h$, we have

$$\pi^{(i)}(s,t;h) = \sum_{l+l'=0}^{M_i-2} \frac{M_i!}{(l+1)!(l'+1)!(M_i - 2 - (l+l'))!} \times p(s;h)^{l+1}p(t;h)^{l'+1}\{1 - p(s;h) - p(t;h)\}^{M_i-2-(l+l')}.$$

Let

$$\underline{\pi}^{(i)}(s,t;h) := \frac{M_i!}{(M_i-2)!}p(s;h)p(t;h)\{1 - p(s;h) - p(t;h)\}^{M_i-2}, \tag{B.4.2}$$

such that

$$\underline{\pi}^{(i)}(s,t;h) \leq \pi^{(i)}(s,t;h), \qquad \forall |s-t| > 2h.$$

In the case $c = \lambda$, that means when $M_i h \asymp 1$, a constant $\underline{c}$ exists such that

$$\inf_{|s-t|>2h} \underline{\pi}^{(i)}(s,t;h) \geq \underline{c} > 0, \qquad \forall i.$$

Gathering all of these terms, it is clear that when $c \geq \lambda$, $\pi^{(i)}(s,t;h)$ is uniformly bounded from below and above by positive constants.

Let us now focus on the case $c < \lambda$ (that means, when $M_i h \to 0$). More precisely, whenever $c < \lambda$ and $|s-t| > 2h$ (the latter is guaranteed when i is sufficiently large), we bound $\pi^{(i)}(s,t;h)$ from below and from above using $\underline{\pi}^{(i)}(s,t;h)$. Next, we bound $\underline{\pi}_i(s,t;h)$ from below and from above by $(M_i h)^2$ provided h is such $1 \gtrsim (M_i h)^2 \asymp i^{2(c-\lambda)} \to 0$. Definition (B.4.2) clearly entails

$$\underline{\pi}^{(i)}(s,t;h) \lesssim (M_i h)^2.$$

Next, let

$$4C_{f,L}h \leq \underline{g}(s,t;h) := p(s;h) + p(t;h) \leq 4C_{f,U}h,$$

and note that, by the elementary inequality (B.3.3), we have

$$\begin{aligned}\{1 - p(s;h) - p(t;h)\}^{M_i - 2} &= \exp\big((M_i - 2)\log\{1 - \underline{g}(s,t;h)\}\big) \geq \exp\left(-(M_i-2)\frac{\underline{g}(s,t;h)}{1-\underline{g}(s,t;h)}\right) \\ &\geq \exp\left(-(C_{\mathfrak{m}} i^c - 2)\frac{4C_{f,U}h}{1-4C_{f,L}h}\right) \geq \exp\left(-C_1 i^c h\right), \qquad \forall i \geq 1,\end{aligned}$$

for some constant $C_1 > 0$. Uniformly bounding $p(s;h)$ and $p(t;h)$ from below, we deduce that a constants $C_2, C_3 > 0$ exit such that

$$C_3 \times (i^c h)^2 \leq C_2 (i^c h)^2 \exp\left(-C_1 i^c h\right) \leq \underline{\pi}^{(i)}(s,t;h) \leq \pi^{(i)}(s,t;h), \qquad \forall s \neq t, \forall i \geq 1.$$

We now show that when $M_i h \lesssim 1$ (which, up to constants, means $c < \lambda$), modulo a constant, $\underline{\pi}^{(i)}(s,t;h)$ is also an upper bound for $\pi^{(i)}(s,t;h)$. We rewrite

$$\begin{aligned}\pi^{(i)}(s,t;h) &= \underline{\pi}^{(i)}(s,t;h) \times \sum_{l+l'=0}^{M_i-2} C(l,l') \times \frac{(M_i-2)!}{l!l'!(M_i-2-(l+l'))!} \times q_1^l \times q_2^{l'} \times q_3^{M_i-2-(l+l')} \\ &\leq \underline{\pi}^{(i)}(s,t;h) \times \sup_{|s-t|>2h} \max_{l,l'} C(l,l'),\end{aligned}$$

where

$$q_1 = \frac{p(s;h)}{1-\{p(s;h)+p(t;h)\}}, \qquad q_2 = \frac{p(t;h)}{1-\{p(s;h)+p(t;h)\}} \qquad q_3 = \frac{1-2\{p(s;h)+p(t;h)\}}{1-\{p(s;h)+p(t;h)\}},$$

$q_1, q_2, q_3 \in (0,1)$, $q_1 + q_2 + q_3 = 1$, and

$$C(l,l') = \frac{1}{(l+1)(l'+1)} \times \frac{\{1-\{p(s;h)+p(t;h)\}\}^{M_i-2-(l+l')}}{\{1-2\{p(s;h)+p(t;h)\}\}^{M_i-2-(l+l')}} \leq \frac{\{1-\{p(s;h)+p(t;h)\}\}^{M_i}}{\{1-2\{p(s;h)+p(t;h)\}\}^{M_i}}.$$

Next, we bound the right-hand side in the inequality, provided $M_i h \lesssim 1$ (*i.e.*, $i^c h \lesssim 1$). Let

$$\overline{g}(s,t;h) = \frac{p(s;h)+p(t;h)}{1-2\{p(s;h)+p(t;h)\}} \leq 2\{p(s;h)+p(t;h)\},$$

such that

$$C(l,l') \le \{1+\overline{g}(s,t;h)\}^{M_i} = \exp\left(M_i \log\{1+\overline{g}(s,t;h)\}\right)$$
$$\le \exp\left(\overline{C}_{\mathfrak{m}} i^c \log\{1+\overline{g}(s,t;h)\}\right) \le \exp\left(\overline{C}_{\mathfrak{m}} i^c \overline{g}(s,t;h)\right)$$
$$\le \exp\left(8C_{f,U}\overline{C}_{\mathfrak{m}} i^c h\right), \qquad \forall l,l' \le M_i - 2, \forall s \neq t, \forall i \ge 1.$$

Thus, $C(l,l')$ is uniformly bounded. Gathering terms, we deduce

$$\mathbb{E}_M\left[w^{(j)}\left(s,t;h^{(j)}\right)\right] = \pi^{(j)}(s,t;h^{(j)}) \asymp \underline{\pi}^{(j)}(s,t;h^{(j)}) \asymp \min\{1,(j^c h^{(j)})^2\}, \qquad \forall s \neq t,$$

provided $j \ge 1$ is sufficiently large such that $|s-t| > 2h^{(j)}$. Since

$$\mathbb{E}_M\left[\mathcal{W}^{(i)}(s,t)\right] = \sum_{j=1}^{i} \mathbb{E}_M\left[w^{(j)}\left(s,t;h^{(j)}\right)\right] = \sum_{j=1}^{i} \pi^{(j)}\left(s,t;h^{(j)}\right),$$

with $h^{(j)} = j^{-\lambda}$, we thus proved (B.4.1) with $\mathbb{E}_M[\mathcal{W}^{(i)}(s,t)]$ instead of $\mathcal{W}^{(i)}(s,t)$. Finally, applying Chernoff's bounds with $\mathcal{W}^{(i)}(s,t)$, which is a sum of independent Bernoulli random variables, we deduce (B.4.1). □

Let $s \neq t$ be fixed in the following, and let $\lambda > 0$ and set $h^{(j)} = j^{-\lambda}$, $j \ge 1$. Recall that

$$\mathcal{N}^{(j)}(t;\lambda) = \frac{w^{(j)}(t;h^{(j)})}{\max_{1\le m\le M_j} |W_m^{(j)}(t;h^{(j)})|}, \qquad j \ge 1, \tag{B.4.3}$$

and

$$\frac{1}{\mathcal{N}_\gamma^{(i)}(t|s;\lambda)} = \frac{1}{\mathcal{W}^{(i)}(s,t)^2} \sum_{j=1}^{i} w^{(j)}(s;h^{(j)}) w^{(j)}(t;h^{(j)}) \max_{1\le m\le M_j} |W_m^{(j)}(s;h^{(j)})|$$
$$= \frac{1}{\mathcal{W}^{(i)}(s,t)^2} \sum_{j=1}^{i} \frac{w^{(j)}(s,t;h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)}.$$

Here, we use the rule

$$\mathcal{N}^{(j)}(t;\lambda) = 0 \quad \Leftrightarrow \quad w^{(j)}(t;h^{(j)}) = 0 \quad \Leftrightarrow \quad \frac{w^{(j)}(t;h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)} = 0.$$

Lemma 4. *Assume that the conditions of Lemma 3 hold true. Then, with $h^{(j)} = j^{-\lambda}$, $j \ge 1$, we have*

$$\{1+o_{\mathbb{P}}(1)\}\underline{\mathfrak{C}} \times i^{-\{1+\min[c-\lambda,2(c-\lambda)]\}} \le \frac{1}{\mathcal{N}_\gamma(s|t;\lambda)} \le \{1+o_{\mathbb{P}}(1)\}\overline{\mathfrak{C}} \times i^{-\{1+\min[c-\lambda,2(c-\lambda)]\}}. \tag{B.4.4}$$

Proof of Lemma 4. For simplicity, we consider the case of a uniform kernel K. With a uniform kernel, for fixed t and j, all $W_m^{(j)}(t;h)$ are equal, and equal to the inverse of the number of $T_m^{(j)}$ in the interval $[t-h,t+h]$. (The case of a kernel bounded and bounded away from zero on the support can be handled similarly, modulo some more constants in the equations.) Thus $\mathcal{N}^{(j)}(t;\lambda) = w^{(j)}(t;h^{(j)})\mathcal{N}^{(j)}(t;\lambda)$ is a Binomial variable with M_j trials and success probability $p(t;h) = \int_{t-h}^{t+h} f_T(u)du$, and

$$\mathbb{E}_M\left[\mathcal{N}^{(j)}(t;\lambda)\right] = M_j p(t;h^{(j)}).$$

Moreover,

$$\frac{w^{(j)}(t;h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)} \leq w^{(j)}(t;h^{(j)}).$$

Let $\mathbb{P}_M$, $\mathbb{E}_M$, and Var_M, denote the conditional probability, expectation and variance, respectively, given $M^{(j)}$, $j \geq 1$. Finally, recall that, for any t and j, $w^{(j)}(t;h)$ is a Bernoulli variable with parameter $\pi^{(j)}(t;h)$. By Lemma 2, constants c_3, c_4 exist such that

$$0 < c_3 j^{\min\{0,c-\lambda\}} \leq \pi^{(j)}(t;h^{(j)}) \leq c_4 j^{\min\{0,c-\lambda\}}, \quad \forall j.$$

Because we here consider $s \neq t$, without loss of generality we can assume $|s-t| > 2h^{(j)}$ (possibly for $j \geq j_0$ for some fixed j_0, but the value of j_0 does not influence the asymptotic results, and we can thus consider $j_0 = 1$.). Using the fact that the harmonic mean is less than or equal to the mean, we have

$$\frac{1}{\mathcal{N}_\gamma^{(i)}(t|s;\lambda)} \geq \frac{1}{\sum_{j=1}^{i} w^{(j)}(s;h^{(j)}) w^{(j)}(t;h^{(j)}) \mathcal{N}^{(j)}(t;\lambda)}.$$

To justify the lower bound in (B.4.4), it suffices to prove that a positive constant $\underline{c}_{\mathcal{N}}$ exists such that

$$\frac{\sum_{j=1}^{i} w^{(j)}(s;h^{(j)}) w^{(j)}(t;h^{(j)}) \mathcal{N}^{(j)}(t;\lambda)}{i^{1+\min(c-\lambda, 2(c-\lambda))}} \leq \underline{\mathfrak{C}}^{-1}\{1 + o_{\mathbb{P}}(1)\}, \tag{B.4.5}$$

provided $|s-t| > 2h^{(j)}$.

In the case of the Nadaraya-Watson estimator with a uniform kernel and bandwidth h,

$$\sum_{j=1}^{i} w^{(j)}(s;h) w^{(j)}(t;h) \mathcal{N}^{(j)}(t;h) = \sum_{j=1}^{i} w^{(j)}(s;h) \sum_{m=1}^{M_j} \mathbf{1}\{|T_m^{(j)} - t| \leq h\} = \sum_{j=1}^{i} S^{(j)}(s,t;h),$$

where $\mathcal{N}^{(i)}(t;h)$ is defined as in (B.4.3) with a generic bandwidth h, and

$$S^{(j)}(s,t;h) = w^{(j)}(s;h) \sum_{1 \leq m \leq M_j} \mathbf{1}\{|T_m^{(j)} - t| \leq h\}.$$

We thus need to suitably bound the sum of the $S^{(j)}(s,t;h)$'s from above. We have

$$\begin{aligned}
\mathbb{E}_M[S^{(j)}(s,t;h)] &= \sum_{m=1}^{M_j} \mathbb{E}_M\left[w^{(j)}(s;h)\mathbf{1}\{|T_m^{(j)} - t| \leq h\}\right] \\
&= \sum_{m=1}^{M_j} \mathbb{E}_M\left[\mathbf{1}\left\{\sum_{1 \leq m' \neq m \leq M_j} \mathbf{1}\{|T_{m'}^{(j)} - s| \leq h\} \geq 1\right\}\mathbf{1}\{|T_m^{(j)} - t| \leq h\}\right] \\
&= \sum_{m=1}^{M_j} \mathbb{E}_M\left[\mathbf{1}\left\{\sum_{1 \leq m' \neq m \leq M_j} \mathbf{1}\{|T_{m'}^{(j)} - s| \leq h\} \geq 1\right\}\right] \times \mathbb{E}_M\left[\mathbf{1}\{|T_m^{(j)} - t| \leq h\}\right] \\
&= \left[1 - \{1 - p(t;h)\}^{M_j - 1}\right] \times M_j p(t;h) \\
&= \{1 + o(1)\} \times \pi^{(j)}(s;h) \times M_j p(t;h),
\end{aligned}$$

where $p(t;h) = \int_{t-h}^{t+h} f_T(u)du$. Moreover,

$$\begin{aligned}
\{S^{(j)}(s,t;h)\}^2 - S^{(i)}(s,t;h) &= w_i(s;h) \sum_{1 \leq m' \neq m \leq M_j} \mathbf{1}\{|T_m^{(j)} - t| \leq h\}\mathbf{1}\{|T_{m'}^{(j)} - t| \leq h\} \\
&= \mathbf{1}\left\{\sum_{\substack{1 \leq m'' \leq M_j \\ m'' \notin \{m,m'\}}} \mathbf{1}\{|T_{m''}^{(j)} - s| \leq h\} \geq 1\right\} \sum_{1 \leq m' \neq m \leq M_i} \mathbf{1}\{|T_m^{(j)} - t| \leq h\}\mathbf{1}\{|T_{m'}^{(j)} - t| \leq h\},
\end{aligned}$$

and thus,

$$\mathbb{E}_M[S^{(j)}(s,t;h)^2] = \mathbb{E}_M[S^{(j)}(s,t;h)] + \left[1 - \{1 - p^{(j)}(t;h)\}^{M_j - 2}\right] \times M_j(M_j - 1)p^2(t;h)$$
$$= \{1 + o(1)\} \times \pi^{(j)}(s;h) \times M_j p(t;h) \times \{1 + M_j p(t;h)\}.$$

We deduce that

$$\begin{aligned}\mathrm{Var}_M[S^{(j)}(s,t;h)] &= \mathbb{E}_M[S^{(j)}(s,t;h)^2] - \mathbb{E}_M^2[S^{(j)}(s,t;h)] \\ &= \{1 + o(1)\} \times \pi^{(j)}(s;h) M_j p^{(j)}(t;h) \times \left[1 + M_j p(t;h) - \pi^{(j)}(s;h) M_j p(t;h)\right] \\ &= \{1 + o(1)\} \times \pi^{(j)}(s;h) M_j p(t;h) + \{1 + o(1)\} \times \pi^{(j)}(s;h)\{1 - \pi^{(j)}(s;h)\}\{M_j p(t;h)\}^2.\end{aligned}$$

Replacing h by $h^{(j)} = j^{-\lambda}$ and recalling that $M_j \asymp j^c$, we get

$$\mathbb{E}_M[S^{(j)}(s,t;h^{(j)})] \asymp j^{\min\{0,c-\lambda\}+c-\lambda} = j^{\min\{c-\lambda,2(c-\lambda)\}},$$

and

$$\mathbb{E}_M\left[\sum_{j=1}^{i} S^{(j)}(s,t;h^{(j)})\right] \asymp \sum_{j=1}^{i} j^{\min\{c-\lambda,2(c-\lambda)\}} \asymp i^{1+\min\{c-\lambda,2(c-\lambda)\}}. \tag{B.4.6}$$

On the other hand,

$$\begin{aligned}\mathrm{Var}_M[S^{(j)}(s,t;h^{(j)})] &\asymp \pi^{(j)}(s;h^{(j)}) \times M_j h^{(j)} + \pi^{(j)}(s;h^{(j)})\{1 - \pi^{(j)}(s;h^{(j)})\} \times (M_j h^{(j)})^2 \\ &\asymp \pi^{(j)}(s;h^{(j)}) \max\{M_j h^{(j)}, (M_j h^{(j)})^2\} \asymp j^{\min\{0,c-\lambda\}} \times j^{\max\{c-\lambda,2(c-\lambda)\}} = j^{2(c-\lambda)}.\end{aligned}$$

For the last equivalence, we use the facts

$$0 \le (M_j h)^2 \pi^{(j)}(s;h)\{1 - \pi^{(j)}(s;h)\} \lesssim (M_j h)^2 \pi^{(j)}(s;h) \lesssim (M_j h)^2 \min\{1, M_j h\}$$
$$= \min\{(M_j h)^2, (M_j h)^3\}, \qquad \forall h.$$

We then deduce

$$\mathrm{Var}_M\left[\sum_{j=1}^{i} S^{(j)}(s,t;h^{(j)})\right] \asymp \sum_{j=1}^{i} j^{2(c-\lambda)} \asymp i^{1+2(c-\lambda)} \to \infty. \tag{B.4.7}$$

Since $1 + \min\{c-\lambda, 2(c-\lambda)\} > 1/2 + (c-\lambda)$, we deduce that the square root of the variance in (B.4.7) is negligible with respect to the expectation in (B.4.6). Therefore, since Bernstein's inequality guarantees that $S^{(j)}(s,t;h^{(j)})$ concentrates around its expectation, we have

$$\sum_{j=1}^{i} S^{(j)}(s,t;h^{(j)}) \asymp_{\mathbb{P}} i^{1+\min\{c-\lambda,2(c-\lambda)\}},$$

and (B.4.5) follows, and the lower bound in (B.4.4) is justified.

To bound $\mathcal{N}_\gamma^{(i)}(t|s;\lambda)^{-1}$ from above, let us first notice that by definition

$$\frac{\mathcal{W}^{(i)}(s,t)}{\mathcal{N}_\gamma^{(i)}(t|s;\lambda)} = \frac{1}{\mathcal{W}^{(i)}(s,t)} \sum_{j=1}^{i} \frac{w^{(j)}(s;h^{(j)}) w^{(j)}(t;h^{(j)})}{\mathcal{N}^{(j)}(t;\lambda)} \le \frac{1}{\mathcal{W}^{(i)}(s,t)} \sum_{j=1}^{i} w^{(j)}(s;h^{(j)}) w^{(j)}(t;h^{(j)}) \le 1.$$

In the case $\max\{1, M_j h^{(j)}\} = 1$, that is when $c \le \lambda$, it thus suffices to apply Lemma 3. We now investigate the case where $M_j h^{(j)} > 1$, that is $c > \lambda$. Let us recall that, if X is a non-degenerate Binomial random variable $B(n,p)$, then

$$\mathbb{E}[(1+X)^{-1}] = \frac{1 - q^{n+1}}{(n+1)p}, \qquad \text{where } q = 1 - p. \tag{B.4.8}$$

See Chao and Strawderman (1972). We next deduce

$$\mathbb{E}\left[\frac{\mathbf{1}\{X>0\}}{X}\right] \leq \frac{2}{(n+1)p} - \frac{2(1+np)}{(n+1)p}q^n \leq \frac{2}{(n+1)p} - \frac{2n}{n+1}q^n.$$

With $h>0$, let us now compute the expectation of $w^{(j)}(s;h)w^{(j)}(t;h)/\mathcal{N}^{(j)}(t;h)$ given M_j. Let us notice that, with an uniform kernel K, the distribution of $\mathcal{N}^{(j)}(t;h)$ given

$$n^{(j)}(s;h) := \sum_{1\leq m\leq M_j} \mathbf{1}\{|T_m^{(j)}-s|\leq h\} = k, \qquad k=1,2\ldots,M_j-1,$$

is Binomial $B(M_j-k, p(t;h)/\{1-p(s;h)\})$, while $n^{(j)}(s;h)$ is Binomial $B(M_j,p(s;h))$. We can write

$$\begin{aligned}\mathbb{E}_M\left[\frac{w^{(j)}(s;h)w^{(j)}(t;h)}{\mathcal{N}^{(j)}(t;h)}\right] &= \mathbb{E}_M\left\{\mathbb{E}_M\left[\frac{w^{(j)}(t;h)}{\mathcal{N}^{(j)}(t;h)}\mid n^{(j)}(s;h)\right]\right\}\\ &= \sum_{k=1}^{M_j-1}\mathbb{E}_M\left[\frac{w^{(j)}(t;h)}{\mathcal{N}^{(j)}(t;h)}\mid n^{(j)}(s;h)=k\right]\times\mathbb{P}_M\left(n^{(j)}(s;h)=k\right).\end{aligned}$$

Letting $\mathcal{N}^{(j)}(t;h)$ play the role of X in the equations above, we have

$$\begin{aligned}\mathbb{E}_M\left[\frac{w^{(j)}(t;h)}{\mathcal{N}^{(j)}(t;h)}\mid n^{(j)}(s;h)=k\right] &\leq 2\frac{1-p(s;h)}{(M_j-k+1)p(t;h)} - 2\frac{M_j-k}{M_j-k+1}\left\{1-\frac{p(t;h)}{1-p(s;h)}\right\}^{M_j-k}\\ &\leq \frac{2}{(M_j-k+1)p(t;h)}, \qquad \forall 1\leq k\leq M_j-1.\end{aligned}$$

Let us note that

$$\sum_{k=0}^{M_j}\frac{1}{M_j-k+1}\times\mathbb{P}_M\left(n^{(j)}(s;h)=k\right) = \mathbb{E}\left[\frac{1}{1+Y}\right],$$

where $Y=M_j-n^{(j)}(s;h)$ is Binomial $B(M_j,1-p(s;h))$. Using (B.4.8), we deduce that a constant C exists such that

$$\mathbb{E}_M\left[\frac{w^{(j)}(s;h^{(j)})w^{(j)}(t;h^{(j)})}{\mathcal{N}^{(j)}(t;h^{(j)})}\right] \leq \frac{2}{p(t;h^{(j)})}\times\frac{1-\{1-p(s;h^{(j)})\}^{M_j+1}}{(M_j+1)\{1-p(s;h^{(j)})\}} \leq \frac{C}{M_jh^{(j)}} = C\times i^{-(c-\lambda)},$$

provided $|s-t|>2h^{(j)}$. Next, since $\mathcal{N}^{(j)}(t;h)$ is a non-negative integer, the conditional variance of $w^{(j)}(s;h)w^{(j)}(t;h)/\mathcal{N}^{(j)}(t;h)$ given M_j is clearly negligible compared to its conditional expectation given M_j.

Gathering terms and using Lemma 3, we then deduce that a constant $\overline{\mathfrak{C}}$ exists such that

$$\mathbb{E}_M\left[\frac{1}{\mathcal{N}_\gamma^{(i)}(t|s;\lambda)}\right] \leq \overline{\mathfrak{C}}\times i^{-(1+\min\{(c-\lambda),2(c-\lambda)\}}\to 0.$$

Finally, since, by Bernstein's inequality, $1/\mathcal{N}_\gamma^{(i)}(t|s;\lambda)$ deviates from its mean with exponentially small probability, the arguments for the upper bound in (B.4.4) are now complete. $\square$

B.5 Details on the implementation aspects described in Section 3.3

For simplicity, we impose condition (11). In the case of independent design, using the design points on each curve $X^{(j)}$, $1\leq j\leq i$, we consider the case in which the density estimator $\widehat{f}_T^{(j)}$ is a histogram estimator with equal bin widths, having bin widths $cM_j^{-\alpha}$, for some fixed $\alpha\in(0,1/2)$, for instance $\alpha=1/3$ which was shown to have some optimality properties. The constant c can be determined, for instance, by the standard deviation or the interquartile range of the distribution of T.

Proposition 6. *Assume that X belongs to $\mathcal{X}(H,L)$, and Assumptions (H1), (H2), (D1) to (D4), and condition* (9) *hold true.*

1. *Assume that the error variable e is sub-Gaussian. If $\widehat{\sigma}^{(i)}$ is defined according to* (19)*, constants $c_1, C_1 > 0$ exist such that, for i sufficiently large,*

$$\mathbb{P}\left(\left| \widehat{\sigma}^{(i)}/\sigma - 1 \right|_\infty \geq \log^{-c_1}(i) \right) \leq \exp(-i^{C_1}). \tag{B.5.1}$$

2. *Constants $c_2, C_2 > 0$ exist such that, for j sufficiently large,*

$$\mathbb{P}\left(\left| f_T / \widehat{f}_T^{(j)} - 1 \right|_\infty \geq \log^{-c_2}(j) \right) \leq \exp(-j^{C_2}). \tag{B.5.2}$$

The concentration rate for the errors' conditional standard deviation estimator $\widehat{\sigma}^{(i)}$ could be improved to a polynomial rate by a refined analysis of the length of the spacing between consecutive $T_m^{(i)}$. However, for our purposes, we only require consistency of this estimator.

Proof of Proposition 6. 1) Recall that

$$m_t^{(i)} = \arg \min_{2 \leq m \leq M_i} \left\{ \left| T_m^{(i)} - t \right| + \left| T_{m-1}^{(i)} - t \right| \right\},$$

and

$$\delta_t^{(i)} = \mathbf{1}_{\mathcal{E}_t^{(i)}} \qquad \text{where } \mathcal{E}_t^{(i)} = \left\{ \min_{2 \leq m \leq M_i} \left\{ \left| T_m^{(i)} - t \right| + \left| T_{m-1}^{(i)} - t \right| \right\} \leq 1/\log(M_i) \right\}.$$

Recall also the notation $\mathcal{T}_{obs}^{(i)}$ which represents the set of observation times $T_m^{(i)}$, $1 \leq m \leq M_i$, over the trajectory $X^{(i)}$. We focus on the case of independent design, the common design case being much easier.

Let us first note that, for any $i \geq 1$,

$$\mathbb{P}\left(\delta_t^{(i)} = 0 \right) \leq \left\{ 1 - \log^{-1}(M_i) \right\}^{M_i} + M_i \log^{-1}(M_i) \left\{ 1 - \log^{-1}(M_i) \right\}^{M_i - 1}.$$

Using (9), we deduce that a constant $c_1 > 0$ exists such that

$$\mathbb{P}\left(\sum_{j=1}^{i} \delta_t^{(j)} = i \right) \geq 1 - \exp(-c_1 i / \log(i+1)), \qquad \forall i \geq 1.$$

Moreover, for each j given the points in the set $\mathcal{T}_{obs}^{(j)}$, the map $t \mapsto \delta_t^{(j)}$, is piecewise constant with at most $M_j + 1$ discontinuities. This means that given all the sets $\mathcal{T}_{obs}^{(j)}$, $1 \leq j \leq i$, the map $t \mapsto \sum_{j=1}^{i} \delta_t^{(j)}$, is piecewise constant with at most $i + \sum_{1 \leq j \leq i} M_j$ discontinuities. Since for any sequence of sets $A_1, \ldots, A_K$,

$$\mathbb{P}\left(\cap_{1 \leq l \leq K} A_l \right) = 1 - \mathbb{P}\left(\cup_{1 \leq l \leq K} A_l^c \right) \geq 1 - \sum_{l=1}^{K} \mathbb{P}(A_l^c),$$

we then deduce that a constant $c_2 > 0$ exists such that

$$\begin{aligned} \mathbb{P}\left(\inf_{t \in \mathcal{T}} \sum_{j=1}^{i} \delta_t^{(j)} = i \right) &= \mathbb{E}\left[\mathbb{P}\left(\inf_{t \in \mathcal{T}} \sum_{j=1}^{i} \delta_t^{(j)} = i \;\middle|\; M_1, \ldots, M_i \right) \right] \\ &\geq 1 - \exp(-c_1 i / \log(i+1)) \times \left\{ i + \sum_{1 \leq j \leq i} \mathbb{E}(M_j) \right\} \leq 1 - \exp(-c_2 i / \log(i+1)). \end{aligned} \tag{B.5.3}$$

In the following, to simply expressions whenever this is not confusing, we drop the subscript t and simply write $m^{(i)}$, $\delta^{(i)}$ and $\mathcal{E}^{(i)}$ instead of $m_t^{(i)}$, $\delta_t^{(i)}$ and $\mathcal{E}_t^{(i)}$, respectively. Let us next consider the estimator

$$\widetilde{\sigma}^{(i)}(t)^2 = \frac{i-1}{i} \times \widetilde{\sigma}^{(i-1)}(t)^2 + \frac{1}{i} \times \frac{\delta^{(i)}}{2} \left\{ Y_{m^{(i)}}^{(i)} - Y_{m^{(i)}-1}^{(i)} \right\}^2 = \frac{1}{i} \sum_{j=1}^{i} \frac{\delta^{(j)}}{2} \left\{ Y_{m^{(j)}}^{(j)} - Y_{m^{(j)}-1}^{(j)} \right\}^2,$$

and decompose it

$$\begin{aligned}
\widetilde{\sigma}^{(i)}(t)^2 = \frac{1}{i} \sum_{j=1}^{i} \frac{\delta^{(j)}}{2} &\left\{ \sigma(T_{m^{(j)}}^{(j)}) e_{m^{(j)}}^{(j)} - \sigma(T_{m^{(j)}-1}^{(j)}) e_{m^{(j)}-1}^{(j)} \right\}^2 \\
&+ \frac{1}{i} \sum_{j=1}^{i} \frac{\delta^{(j)}}{2} \left\{ X^{(j)}(T_{m^{(j)}}^{(j)}) - X^{(j)}(T_{m^{(j)}-1}^{(j)}) \right\}^2 \\
+ \frac{1}{i} \sum_{j=1}^{i} \delta^{(j)} &\left\{ X^{(j)}(T_{m^{(j)}}^{(j)}) - X^{(j)}(T_{m^{(j)}-1}^{(j)}) \right\} \left\{ \sigma(T_{m^{(j)}}^{(j)}) e_{m^{(j)}}^{(j)} - \sigma(T_{m^{(j)}-1}^{(j)}) e_{m^{(j)}-1}^{(j)} \right\} \\
&=: S_1^{(i)}(t) + S_2^{(i)}(t) + S_3^{(i)}(t), \qquad t \in \mathcal{T},\ i = 1, 2 \ldots
\end{aligned}$$

Since $t \mapsto \sigma(t)$ is Hölder continuous with some exponent, say, $a_\sigma > 0$, and condition (9) holds true, using (B.5.3) we get

$$S_1^{(i)}(t) = \sigma^2(t) + S_{11}^{(i)}(t) + S_{12}^{(i)}(t) + S_{13}^{(i)}(t) \quad \text{where} \quad S_{11}^{(i)}(t) = \sigma^2(t) \left[\frac{1}{i} \sum_{j=1}^{i} \delta^{(j)} - 1 \right],$$

and

$$S_{12}^{(i)}(t) = \frac{\sigma^2(t)}{2i} \sum_{j=1}^{i} \delta^{(j)} \left\{ \{e_{m^{(j)}}^{(j)}\}^2 + \{e_{m^{(j)}-1}^{(j)}\}^2 - 2 e_{m^{(j)}}^{(j)} e_{m^{(j)}-1}^{(j)} - 2 \right\}.$$

Here, $|S_{13}^{(i)}(t)| = O(\log^{-a_\sigma}(i+1)) \times S_{14}^{(i)}(t)$ is a remainder with

$$S_{14}^{(i)}(t) = \frac{\sigma^2(t)}{2i} \sum_{j=1}^{i} \delta^{(j)} \left\{ \{e_{m^{(j)}}^{(j)}\}^2 + \{e_{m^{(j)}-1}^{(j)}\}^2 + 2 \left| e_{m^{(j)}}^{(j)} e_{m^{(j)}-1}^{(j)} \right| \right\}.$$

The $O(\log^{-a_\sigma}(i+1))$ term is given by the Hölder continuity of $t \mapsto \sigma(t)$ and the sums occurring in the expressions of $S_{12}^{(i)}$ and $S_{13}^{(i)}$ do not depend on t. Then using the fact that the error variable e is sub-Gaussian, that the square of a sub-Gaussian variable is sub-exponential, and Bernstein's inequality, we deduce exponential bounds for the concentration (faster than any negative power of $\log(i)$) of the normalized sums occurring in $S_{12}^{(i)}$ and $S_{13}^{(i)}$ around their means, which are zero and a constant, respectively. From this and (B.5.3), we deduce the property (B.5.1) with the square root of $S_1^{(i)}$ in place of $\widehat{\sigma}$, and for some $0 < c_1 < a_\sigma$ and $0 < C_1 < 1$. The property (B.5.1) with $\widetilde{\sigma}$ in place of $\widehat{\sigma}$ follows next from the fact that $S_2^{(i)}(t)$ and $S_3^{(i)}(t)$ are negligible compared to $S_1^{(i)}(t)$. We sketch the arguments for this in the following. For showing that $S_2^{(i)}(t)$ is negligible, let us first bound its expectation using the local regularity of X. More precisely, for any $1 \leq j \leq i$ we have

$$\begin{aligned}
\mathbb{E}\left[s_2^{(j)}(t) \right] &:= \mathbb{E}\left[\delta^{(j)} \left\{ X^{(j)}(T_{m^{(j)}}^{(j)}) - X^{(j)}(T_{m^{(j)}-1}^{(j)}) \right\}^2 \right] \\
&= \mathbb{E}\left[\delta^{(j)} \mathbb{E}\left(\left\{ X^{(j)}(T_{m^{(j)}}^{(j)}) - X^{(j)}(T_{m^{(j)}-1}^{(j)}) \right\}^2 \mid M_j, T_1^{(j)}, \ldots, T_{M_j}^{(j)} \right) \right] \\
&\leq \mathbb{E}\left[\delta^{(j)} L_t^2 \left| T_{m^{(j)}}^{(j)} - T_{m^{(j)}-1}^{(j)} \right|^{H_t} \right] \leq \frac{C}{\log^{\underline{H}}(j+1)},
\end{aligned}$$

for some constant depending only on $\overline{L}$ and a in condition (9). We deduce that

$$\mathbb{E}\left[S_2^{(i)}(t)\right] = \frac{1}{2i}\sum_{j=1}^{i}\mathbb{E}\left[s_2^{(j)}(t)\right] \leq \frac{C}{2i}\sum_{j=1}^{i}\log^{-\underline{H}}(j+1) \leq C'\log^{-\underline{H}}(i+1) \to 0, \tag{B.5.4}$$

for some constant $C' > C/2$. By (H2), the standard deviation of $s_2^{(j)}(t)$ is bounded by a constant times its expectation. We next apply Bernstein's inequality to derive the concentration (faster than any negative power of $\log(i)$) of $S_2^{(i)}(t)$ around the mean, given all the design points generated by T. The uniformity with respect to t is obtained similarly to (B.5.3), that is by noting that, given the set of design points $\mathcal{T}_{obs}^{(j)}$, the functions $t \mapsto s_2^{(j)}(t)$ are piecewise constant. Finally, $S_3^{(i)}(t)$ is a sum of independent, centered random variables, with the variance tending to zero. Its concentration (faster than any negative power of $\log(i)$) around zero can be derived by the same arguments as mentioned for $S_3^{(i)}(t)$, the details are thus omitted.

Next, we note that the difference between $\widetilde{\sigma}^{(i)}$ and $\widehat{\sigma}^{(i)}$ is uniformly negligible. This is a direct consequence of the fact that, with exponentially small probability, $\inf_{t\in\mathcal{T}}\sum_{j=1}^{i}\delta_t^{(j)} = i$, see (B.5.3).

2) Let b_j be the bin width used for the $j-$th curve. Let $I_j(t)$ be the bin interval containing $t \in \mathcal{T}$, and let $B_j(t)$ be the random variable defined as the number of $T_m^{(j)}$ falling in $I_j(t)$. The histogram estimator obtained from the design points of the $j-$th curve is then $\widehat{f}_T^{(j)}(t) = B_j(t)\{M_j b_j\}^{-1}$. The Hölder continuity of f_T entails the bias of $\widehat{f}^{(j)}(t)$ to be bounded by a constant times b_j, with the constant independent of t. An exponential bound for the concentration of $\widehat{f}^{(j)}(t) - \mathbb{E}[\widehat{f}^{(j)}(t)]$ is then easily obtained from the concentration of a Binomial variable with M_j trials and success probability equal to the theoretical probability that T belongs to the bin. Since the histogram estimator is constant over a bin, the uniform concentration with respect to t is obtained by applying the exponential bound $\lceil 1/b_j \rceil$ times. Finally, the bound (B.5.2) follows from the exponential bound for the uniform concentration of $\widehat{f}^{(j)}(t) - \mathbb{E}[\widehat{f}^{(j)}(t)]$, the uniform bound on the bias of $\widehat{f}^{(j)}$, and the fact that f_T stays away from zero. The arguments are standard and the details are thus omitted. □

B.6 Additional technical results

Remark 1. *The condition*

$$\exists a \text{ such that } 0 < a \leq \frac{\log(i)}{\log(\mathfrak{m}_i)} \leq a^{-1}, \qquad \forall i > 1, \tag{B.6.1}$$

introduced in (9)*, implies the following one:*

$$\exists b \text{ such that } 0 < b \leq \frac{\log(i)}{\log(\overline{\mathfrak{m}}_i)} \leq b^{-1}, \qquad \forall i > 1, \tag{B.6.2}$$

where $\overline{\mathfrak{m}}_i = (\mathfrak{m}_1 + \cdots + \mathfrak{m}_i)/i$. *Indeed,* (B.6.1) *means*

$$i^a \leq \mathfrak{m}_i \leq i^{1/a}, \qquad \forall i \geq 1.$$

and thus

$$i^{-1}\sum_{j=1}^{i} j^a \leq \overline{\mathfrak{m}}_i \leq i^{-1}\sum_{j=1}^{i} j^{1/a}. \tag{B.6.3}$$

On the other hand, for any $c > 0$*,*

$$\frac{1}{c+1}i^{c+1} = \int_0^i x^c dx \leq \sum_{j=1}^{i} j^c \leq \int_1^{i+1} x^c dx = \frac{1}{c+1}\left\{(i+1)^{c+1} - 1\right\} \leq \frac{2^{c+1}}{c+1}i^{c+1},$$

and thus

$$\frac{1}{c+1}i^c \le i^{-1}\sum_{j=1}^{i} j^c \le \frac{2^{c+1}}{c+1}i^c.$$

From this and (B.6.3), taking logarithms, we deduce (B.6.2).

Conversely, by (B.6.2) we have

$$i^{1+b} \le \mathfrak{m}_1 + \cdots + \mathfrak{m}_i \le i^{1+1/b},$$

and this implies

$$\frac{b}{b+1} \le \frac{\log(i)}{\log(\mathfrak{m}_i)}.$$

However, the last inequality in (B.6.1) cannot be guaranteed without further constraints because $\mathfrak{m}_i$ could be close to 1 even when i is very large.

Remark 2. *We here explain the definition of the constant $C_t^{(i)}$ in (18). Let us consider a linear smoother as defined in (20), with kernel-based weights $W_m^{(i)}$, and the average quadratic risk, given the design:*

$$\mathcal{R}_c\left(\widehat{X}_t^{(i)};h\right) = \mathbb{E}_M\left[\widehat{X}_t^{(i)}(h) - {X_t^{(i)}}^2\right].$$

This risk averages over the realizations of X. It can be bounded by

$$\mathcal{W}_1(t;h,M_i)L_t^2h^{2H_t} + \mathcal{W}_2(t;h,M_i),$$

where

$$\mathcal{W}_1(t;h,M) = \sum_{m=1}^{M}|W_m^{(i)}(t;h)| \times \sum_{m=1}^{M}\left\{\left|t - T_m^{(i)}\right|/h\right\}^{2H_t}|W_m^{(i)}(t;h)|,$$

$$\mathcal{W}_2(t;h,M) = \sum_{m=1}^{M}\sigma^2(T_m^{(i)})|W_m^{(i)}(t;h)|^2.$$

Assuming the map $t \mapsto \sigma^2(t)f_T(t)$ is Hölder continuous, we have

$$\mathcal{W}_2(t;h,M) \approx \sigma^2(t)f_T(t) \times \sum_{m=1}^{M}|W_m^{(i)}(t;h)|^2.$$

The optimal bandwidth can be selected by minimization of the conditional risk $\mathcal{R}_c$ given the design, but this in general does not have an explicit solution.

In the functional data framework, due to the replication feature of the data, we can naturally consider the unconditional risk *of $\widehat{X}_t^{(i)}(h)$, that is the risk integrated with respect to the design points distribution:*

$$\mathcal{R}\left(\widehat{X}_t^{(i)};h\right) = \mathbb{E}\left[\left\{\widehat{X}_t^{(i)}(h) - X_t^{(i)}\right\}^2\right].$$

Using the local regularity concept, the unconditional risk can be bounded by an explicit function of h, yielding an explicit optimal bandwidth. More precisely, for the local constant smoother, using the local regularity definition and standard calculations, we get

$$\mathcal{R}\left(\widehat{X}_t^{(i)};h\right) \le L_t^2\int |u|^{2H_t}|K(u)|du \times h^{2H_t} + \frac{\sigma^2(t)\int K^2(u)du}{f_T(t)} \times \frac{1}{M_ih} + \text{ negligible terms.}$$

The bandwidth minimizing the risk bound is then

$$\widehat{h}_t^{(i)} = \left[\frac{\sigma^2(t)\int K^2(u)du}{2H_tL_t^2 \times \int |u|^{2H_t}|K(u)|du \times f_T(t)} \times \frac{1}{M_i}\right]^{\frac{1}{2H_t+1}}.$$

Remark 3. Details on equation (B.5.4). *We here justify the following statement: for $a > 0$, we have*

$$\sum_{j=1}^{i} \log^{-a}(j+1) \le A_1 (i+1) \log^{-a}(i+1) + A_2, \tag{B.6.4}$$

where $A_1, A_2 > 0$ are constants depending only on a.

First, we note that

$$\begin{aligned}\sum_{j=1}^{i} \log^{-a}(j+1) = \log^{-a}(2) + \sum_{j=2}^{i} \log^{-a}(j+1) \le \log^{-a}(2) + \sum_{j=1}^{i-1} \int_{j+1}^{j+2} \log^{-a}(x)dx \\ = \log^{-a}(2) + \int_{2}^{i+1} \log^{-a}(x)dx.\end{aligned}$$

Next, we have the identity

$$\left(\frac{x}{\log^a(x)}\right)' = \frac{1}{\log^a(x)} - \frac{a}{\log^{a+1}(x)},$$

from which we deduce

$$\begin{aligned}\int_2^{i+1} \log^{-a}(x)dx &= a \int_2^3 \log^{-a-1}(x)dx + a \int_3^{i+1} \log^{-a-1}(x)dx + \left[\frac{x}{\log^a(x)}\right]_2^{i+1} \\ &= a \int_2^3 \log^{-a-1}(x)dx + a \int_3^{i+1} \log^{-a-1}(x)dx + \frac{i+1}{\log^a(i+1)} - \frac{2}{\log^a(2)}.\end{aligned}$$

Since

$$\int_3^{i+1} \log^{-a-1}(x)dx \le \frac{1}{\log^a(3)} \int_3^{i+1} \log^{-a}(x)dx,$$

and $\log(3) > 1$, we deduce that positive constants C_1, C_2 exist such that

$$\int_2^{i+1} \log^{-a}(x)dx \le C_1 \frac{i+1}{\log^a(i+1)} + C_2,$$

and this implies the statement (B.6.4)*. Finally,* (B.5.4) *follows from* (B.6.4) *because we have*

$$\frac{1}{i} \sum_{j=1}^{i} \log^{-a}(j+1) \le A_1 \frac{i+1}{i} \log^{-a}(i+1) + \frac{A_2}{i} \le A_3 \log^{-a}(i+1),$$

for some $A_3 > A_1$.

Lemma 5. *Let $\beta, \delta > 0$, $\beta \ne 1$, and $1 \le i_0 < i$ integers. Then*

$$\frac{1}{1-\beta}\left[(i+1)^{1-\beta} - i_0^{1-\beta}\right] \le \sum_{j=i_0}^{i} j^{-\beta} \le \frac{1}{1-\beta}\left[i^{1-\beta} - i_0^{1-\beta}\right] + i_0^{1-\beta},$$

and

$$\frac{1}{1+\delta}\left[i^{1+\delta} - i_0^{1+\delta}\right] + i_0^{1+\delta} \le \sum_{j=i_0}^{i} j^{\delta} \le \frac{1}{1+\delta}\left[(i+1)^{1+\delta} - i_0^{1+\delta}\right],$$

Proof of Lemma 5. For any $j \geq 1$ and $x \in [j, j+1]$, we have

$$(j+1)^{-\beta} \leq x^{-\beta} \leq j^{-\beta}.$$

We then deduce

$$\sum_{j=i_0+1}^{i+1} j^{-\beta} = \sum_{j=i_0}^{i} (j+1)^{-\beta} \leq \int_{i_0}^{i+1} x^{-\beta} dx = \frac{1}{1-\beta}\left[(i+1)^{1-\beta} - i_0^{1-\beta}\right] \leq \sum_{j=i_0}^{i} j^{-\beta},$$

from which the first statement of the result follows.

Since $\delta > 0$, for any $j \geq 1$ and $x \in [j, j+1]$, we have $j^{\delta} \leq x^{\delta} \leq (j+1)^{\delta}$, and thus

$$\sum_{j=i_0}^{i} j^{\delta} \leq \int_{i_0}^{i+1} x^{\delta} dx = \frac{1}{1+\delta}\left[(i+1)^{1+\delta} - i_0^{1+\delta}\right] \leq \sum_{j=i_0}^{i} (j+1)^{\delta} = \sum_{j=i_0+1}^{i+1} j^{\delta}.$$

□

B.7 Batch estimation and initial values for recursions

To initiate the online estimation procedures, we use a batch sample generated from the same data generating process as the online observations. Let $X^{(k)}$, $k = 1, \ldots, n$ be a batch sample of independent curves. The batch sample size is n, which can be as small as 10. To distinguish these realizations from the online ones, we index them by the superscript (k) instead of (i). The data points associated with the batch sample path $X^{(k)}$ consists of the pairs $(Y_m^{(k)}, T_m^{(k)}) \in \mathbb{R} \times \mathcal{T}$, generated according to

$$Y_m^{(k)} = X^{(k)}(T_m^{(k)}) + \varepsilon_m^{(k)}, \qquad \varepsilon_m^{(k)} = \sigma(T_m^{(k)}) e_m^{(k)}, \qquad 1 \leq m \leq M_{b,k}.$$

The $X_m^{(k)}$, $M_{b,k}$, $T_m^{(k)}$, $e_m^{(k)}$ are generated in the same way as $X_m^{(i)}$, M_i, $T_m^{(i)}$, $e_m^{(i)}$, respectively.

Let $t \in \mathcal{T}$. Let $\widehat{f}_T(t)$ be the histogram density estimator of the density of T obtained in the batch. Let

$$m_t^{(k)} = \arg\min_{2 \leq m \leq M_{b,k}} \left\{ \left| T_m^{(k)} - t \right| + \left| T_{m-1}^{(k)} - t \right| \right\}.$$

Also, let the batch variance estimator be

$$\widehat{\sigma}^2(t) = \frac{1}{n} \sum_{k=1}^{n} \left\{ Y_{m_t^{(k)}}^{(k)} - Y_{m_t^{(k)}-1}^{(k)} \right\}^2.$$

This is then the initial value $\widehat{\sigma}^{(0)}(t)^2$ we use for the online error variance estimator (19).

We next build the pilot local constant estimators $\widetilde{X}_u^{(k)}$ and $\widetilde{X}_v^{(k)}$ on the batch curves, and compute

$$\widehat{\theta}(u, v) = \frac{1}{n} \sum_{k=1}^{n} \left(\widetilde{X}_u^{(k)} - \widetilde{X}_v^{(k)} \right)^2. \tag{B.7.1}$$

It may happen that, in the batch, all of the weights of the pilot local constant smoother are equal to zero, which implies there are no $T_m^{(k)}$ close to t. In such cases, we use the $1-$NN estimates for constructing $\widetilde{X}_u^{(k)}$ and/or $\widetilde{X}_v^{(k)}$. The rationale here is simply that our batch contains a small number of curves hence discarding curves when computing initial values $\widehat{\theta}(u, v)$ may create unnecessary instability. Next, we define

$$\widehat{H}_t = \frac{\log\left(\widehat{\theta}(t_1, t_3)\right) - \log\left(\widehat{\theta}(t_1, t_2)\right)}{2\log(2)} \qquad \text{and} \qquad \widehat{L}_t = \frac{\sqrt{\widehat{\theta}(t_1, t_3)}}{|t_1 - t_3|^{\widehat{H}_t}}, \tag{B.7.2}$$

with t_1, t_2 and t_3 close to t, defined as in Section 3.2.

The constructions (B.7.1) and (B.7.2) require a local bandwidth. For $1 \leq k \leq n$, let

$$h_t(H, L) = \left[\frac{C_t(H, L)}{M_{b,k}}\right]^{\alpha_t(H)}, \tag{B.7.3}$$

where

$$C_t(H, L) = \frac{\widehat{\sigma}^2(t) \int K^2(u) du}{2HL^2 \times \int |u|^{2H} |K(u)| du \times \widehat{f}_T(t)} \qquad \text{and} \qquad \alpha_t(H) = \frac{1}{2H+1}. \tag{B.7.4}$$

The local bandwidth rule we propose is based on simple iterations using $C_t(H, L)$ and $\alpha_t(H)$. The algorithm is presented in the following.

Batch iterative local regularity estimation

Input: Estimates $\widehat{f}_T(t)$ and $\widehat{\sigma}^2(t)$; $H \in [\eta, 1-\eta]$ for some small $\eta > 0$ and $L > 0$ (say, $L = 1$)
Initialization: Compute $C_t(H, L)$, $\alpha_t(H)$ in (B.7.4) and the corresponding $h_t(H, L)$ in (B.7.3);

1. Use $h_t(H, L)$ to compute (B.7.1) and update H and L in (B.7.2);
2. Use H and L to update $C_t(H, L)$, $\alpha_t(H)$ in (B.7.4) and $h_t(H, L)$ in (B.7.3);
3. Repeat Steps 1 and 2 until convergence

Let $h_t^{(0)}$ be the last iteration value for $h_t(H, L)$ in iterative procedure. We use $h_t^{(0)}$ to compute the pilot local constant estimators for $\widehat{\theta}^{(0)}(u, v)$ in (6). Meanwhile, the last iteration values for H and L are used as initial values $\widehat{H}_t^{(0)}$ and $\widehat{L}_t^{(0)}$ in (18). For starting the recursion (23) for online mean estimation, we simply set $\mathcal{W}^{(0)}(t) = n$. An initial mean function estimator $\widehat{\mu}^{(0)}(t)$ is simply defined as the average of the batch pilot local constant estimators $\widetilde{X}_t^{(k)}$ built with the bandwidth equal to the last iteration value for $h_t(H, L)$ in the iterative estimation procedure above. For starting the recursion (32) for online covariance estimation, we simply set $\mathcal{W}^{(0)}(s, t) = n$ for any pair $s \neq t$. The initial $\widehat{\gamma}^{(0)}(s, t)$ estimator is simply defined as the average of the products $\widetilde{X}_s^{(k)} \widetilde{X}_t^{(k)}$, $1 \leq k \leq n$, of batch pilot local constant estimators built with the bandwidth equal to $\max\left\{\widehat{h}_s^{(0)}, \widehat{h}_t^{(0)}\right\}$.