Journal of Time Series Analysis

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Adaptive Estimation for Weakly Dependent Functional Times Series

Hassan Maissoro[1,2] | Valentin Patilea[1] | Myriam Vimond[1]

[1]Univ Rennes, Ensai, CNRS, CREST (Center for Research in Economics and Statistics) – UMR 9194, Rennes, France | [2]Datastorm, Palaiseau, France

Correspondence: Myriam Vimond (mvimond@ensai.fr)

Received: 30 September 2024 | **Revised:** 3 June 2025 | **Accepted:** 4 June 2025

Funding: The authors received no specific funding for this work.

Keywords: adaptive estimator | autocovariance function | Hölder exponent | optimal smoothing

ABSTRACT

We propose adaptive mean and autocovariance function estimators for stationary functional time series under $\mathbb{L}^p - m$-approximability assumptions. These estimators are designed to adapt to the regularity of the curves and to accommodate both sparse and dense data designs. The sample paths are observed with error at possibly random design points. Data-driven local bandwidths are selected by minimizing explicit quadratic risk bounds that exploit the local regularity of the process. As a first step, we introduce a local regularity estimator and derive a nonasymptotic concentration bound for it. We also derive the asymptotic normality of the mean estimator, which allows honest inference for irregular mean functions. Simulations and a real data application illustrate the performance of the new estimators.

MSC2020 Classification: Primary 62R10, 62G05, secondary 62M10

1 | Introduction

Functional data analysis (FDA) refers to the case where the observation units are the whole curves (also called trajectories or sample paths). The data set then consists of a collection of N trajectories, modeled by a same stochastic process defined over some domain. Dependent functional data arise in fields such as environment (Aue et al. 2015), energy (Chen et al. 2021), biology (Stoehr et al. 2021), or clinical research (Martínez-Hernández and Genton 2021; Li and Yang 2023). They are often collected sequentially at regular time intervals (e.g., days and weeks) and exhibit a serial dependence. Functional time series (FTS) analysis aims to understand the serial dependence between curves and their dynamics over time. Several types of dependence for functional data have been studied, such as cumulant mixing conditions, strong mixing, physical dependence, $\mathbb{L}^p - m$-approximability. See, for example, Hörmann and Kokoszka (2012); Panaretos and Tavakoli (2013); Chen and Song (2015); Rubín and Panaretos (2020) and their references. We consider FTS that are $\mathbb{L}^p - m$-approximable, that is, satisfying a general moment-based notion of weak dependence involving m-dependence (see Hörmann and Kokoszka 2012).

Most of the textbooks and many FDA articles consider the sample paths observed without error at each point in the domain. In this case, the FDA permits straightforward nonparametric approaches (such as the empirical mean and covariance function estimators), for which an elegant theory is derived based on limit theorems for Hilbert space variables. See, for example, Horváth and Kokoszka (2012). In real data problems, the curves are only observed by a finite number of noisy measurements, at observation design (or domain) points that are not necessarily regular or identical from one curve to another. Two cases are usually studied: the points of the domain where the curves are observed are the same for all the curves (common design) and they are completely different (independent design). The two situations are different in nature and usually lead to different theoretical results. With a common design, there is no information about the stochastic process between the design points.

A common practice in FDA is to first create smoothed curves (usually called functional data objects) from the data points on each curve separately, and then proceed as when the sample paths were observed without error everywhere in the domain. For each curve separately, smoothed curves can be constructed by nonparametric smoothing (splines, kernel smoothing, etc.), or simple linear interpolation. There is no reason however why constructing smoothed curves ignoring the other curves generated by the same stochastic process, should always be an appropriate way to proceed with the FDA. Alternatively, for example, for the mean and covariance function estimation, one can pool all the data points and proceed with nonparametric procedures. See Zhang and Wang (2016) for independent functional data and Rubín and Panaretos (2020) for the dependent sample paths. However, while pooling the data points of all the curves appears to be effective for independent curves, in a time series context it removes the information about the stochastic dependence between the sample paths.

Nonparametric methods with separately smoothed curves, sometimes called "smooth first, then estimate" approaches, are usually recommended for curves observed over a *dense* set of domain points, while pooling is preferred for *sparse* functional data (see Yao et al. 2005; Zhang and Wang 2016). It is worth noting that the definition of sparse and dense regimes depend on the regularity of the sample paths (see Zhang and Wang 2016). Furthermore, the minimax convergence rates for the nonparametric methods are expected to depend on the regularity of the sample paths (see Cai and Yuan 2011, for the mean function estimation). While the regularity of the sample paths, an intrinsic property of the process generating the functional data, has a major impact on nonparametric methods in FDA, it appears that little effort has been devoted so far to estimating this regularity and to constructing adaptive methods. In most cases, the sample paths are assumed to have a certain regularity, for example, twice continuously differentiable. However, many applications produce irregular curves, such as photovoltaic or wind power generation which depend on natural phenomena.

The objective of this paper is to propose adaptive mean and autocovariance function estimators. The local bandwidths are data-driven and chosen by minimizing explicit quadratic risk bounds. As far as we are aware, these risk bounds represent a novel contribution to the field of FTS. They rely on the knowledge of the local regularity of the process generating the weakly dependent sequence of curves. Therefore, we first study a method to estimate the local regularity of the process. In the case of differentiable sample paths, the increments of the highest order derivative of the sample paths are considered instead. Our results on the local regularity estimation extend those of Golovkine et al. (2022), who have studied independent and identically distributed functional data. The definition of local regularity considered below is closely related to the notion of local intrinsic stationarity introduced by (Hsing et al. 2016, 2060). Similar concepts of local regularity are common in continuous-time processes, (see, for example, Bibinger et al. 2017, Section 5, and its references), where the regularity estimation is usually based on a single sample path.

The paper is organized as follows. Section 2 first presents the statistical model associated with the observation of FTS at discrete points in the domain, in the presence of additive heteroscedastic noise. It then introduces the notion of local regularity for processes with nondifferentiable trajectories. The notion involves two parameters, the local Hölder exponent and the local Hölder constant, both of which are allowed to vary over the domain. Several examples of processes and their local regularity parameters are presented. In particular, we show that the local regularity exponent is related to the Hurst function of a multifractional Brownian motion. Given suitable estimates of the local regularity parameters, in Section 2.2 we propose adaptive kernel estimators for the mean function and the autocovariance functions. After the formal definition of the local regularity and the type of $\mathbb{L}^p - m$-approximability we consider as weak dependence, Section 3 presents the theory on the pointwise convergence rates for our kernel estimators. They adapt to the regularity of the sample paths and the design which can be independent or common, sparse or dense. We also prove the asymptotic normality of the adaptive mean function estimator. Our estimators are new in the context of FTS. In Section 4, we present a few results from an extensive simulation study and a real data analysis, and in Section 5 we make some final remarks. The Supporting Information presents the estimators of the local regularity parameters, and also contains the proofs of the main results, additional technical statements, and some further empirical results.

2 | Adaptive Estimation Methodology

A functional time series (FTS) is a sequence of random functions $\{X_n\} = \{X_n(u),\ u \in I, n \in \mathbb{Z}\} \subset \mathcal{H}$, which are temporal dependent, that is, stochastically dependent with respect to the index n. Here, I is a bounded domain over the real line, for instance, $I = (0, 1]$. Moreover, $\mathcal{H} = \mathbb{L}^2(I)$ is the Hilbert space of real-valued, square integrable functions defined over I. In applications, the index n can represent the day, while u can be the daily clock time rescaled to I. We assume that, almost surely, the paths X_n belong to the Banach space $\mathcal{C} = \mathcal{C}(I)$ of continuous functions, equipped with the sup-norm $\|\cdot\|_\infty$.

2.1 | Observation Scheme

For each $1 \le n \le N$, the trajectory (or curve) X_n is observed at the domain points $\{T_{n,i}, 1 \le i \le M_n\} \subset I$, with additive noise. The data points associated with X_n consist of the pairs $(Y_{n,i}, T_{n,i}) \in \mathbb{R} \times I$, where

$$Y_{n,i} = X_n(T_{n,i}) + \sigma(T_{n,i})\varepsilon_{n,i}, \qquad 1 \le n \le N,\ \ 1 \le i \le M_n \qquad (1)$$

The data generating process described in (1) satisfies the following assumptions.

- **H**1. The series $\{X_n\}$ is a (strictly) stationary $\mathcal{H}$-valued series, and $\mathbb{E}(\|X_n\|_\infty^2) < \infty$.
- **H**2. The $M_1, \ldots, M_N$ are random draws of an integer variable $M \ge 2$, with expectation λ.

H3. Either all the $T_{n,i}$ are independent copies of a variable $T \in I$ which admits a strictly positive density g over I (independent design case), or the $T_{n,i}$, $1 \le i \le \lambda = M_n$, are the points of the same equidistant grid of λ points in I (common design case).

H4. The $\varepsilon_{n,i}$ are independent copies of a centered error variable ε with unit variance, and $\sigma^2(\cdot)$ is a Lipschitz continuous function.

H5. The series $\{X_n\}$ and the copies of M, T, and ε are mutually independent.

The domain points $\{T_{n,i}\}$ are either obtained as random copies of $T \in I$ or they are the elements of a grid of length λ, which we consider to be equidistant for simplicity. Assumption (**H**4) allows for heteroscedastic errors. In addition to the stationarity (**H**1), we assume weak dependence, the $\mathbb{L}^p - m$-approximability, see Section 3.2. In the following, X denotes a generic random function having the stationary distribution of $\{X_n\}$. Assumption (**H**1) also implies that the mean function μ and the lag-ℓ autocovariance function Γ_ℓ,

$$\mu(t) = \mathbb{E}[X_n(t)], \qquad t \in I,$$
$$\Gamma_\ell(s,t) = \gamma_\ell(s,t) - \mu(s)\mu(t) := \mathbb{E}\left[X_n(s)X_{n+\ell}(t)\right] - \mu(s)\mu(t),$$
$$s,t \in I, \quad \ell \in \mathbb{Z}$$

are well defined. Our goal is to provide estimators of the mean μ and the lag-ℓ cross-product γ_ℓ designed to adapt to the regularity of the curves.

The *local regularity* of X, and thus that of the stationary distribution of $\{X_n\}$, is first discussed in the case where the sample paths of X_n are not differentiable. Assume that a constant $\beta > 0$ exists and for any $t \in I$, $H_t \in (0,1]$ and $L_t \in (0,\infty)$ exist such that

$$\mathbb{E}\left[\{X(u) - X(v)\}^2\right] = L_t^2 |u-v|^{2H_t}\{1 + O(|u-v|^\beta)\} \tag{2}$$

when $u \le t \le v$ lie in a small neighborhood of t. H_t is then the local Hölder exponent while L_t is the local Hölder constant. They are both allowed to depend on t to allow for curves with general patterns. See Section 3.1 for the formal definition of the local regularity. Examples of processes satisfying (2) include, but are not limited to stationary or stationary increment processes (Golovkine et al. 2022). The class of multifractional Brownian motion processes with domain deformation is another example (Wang et al. 2025). By Kolmogorov's criterion (Revuz and Yor 1999, Theorem 2.1), the local regularity of X is linked to the regularity of the sample paths. We next present examples of FTS and their regularity parameters.

Example 1. Let $\{\xi_n\}$ be a sequence of i.i.d. multifractional Brownian motion (MfBm) of Hurst exponent function $H_\xi : \mathbb{R}_+ \to (0,1)$. That means ξ_n are independent copies of ξ, a centered Gaussian process with covariance function

$$\mathbb{E}[\xi(u)\xi(v)] = D(H_{\xi,u}, H_{\xi,v})\left[u^{H_{\xi,u}+H_{\xi,v}} + v^{H_{\xi,u}+H_{\xi,v}} - |v-u|^{H_{\xi,u}+H_{\xi,v}}\right]$$

where $u, v \ge 0$, and

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

See, for example, Balança (2015) for the formal definition of the MfBm. The fractional Brownian motion is an MfBm with constant Hurst index function. For any bounded interval $I \subset \mathbb{R}_+$, it can be shown that ξ satisfies (2) provided H_ξ is twice continuously differentiable (Wang et al. 2025). Note that defining $\eta(t) = \int_a^t \xi(u)du$, $t \ge 0$, for some $a \ge 0$, the trajectories of η are differentiable and their derivatives satisfy (2). Repeated application of the integral operator gives examples of processes with any regularity exponent greater than 1. See the Supporting Information for the formal definition of the local regularity for processes with smooth paths.

Example 2 (FAR(1) model). Let $\{X_n\}$ be the zero-mean, stationary Functional AutoRegressive (FAR) time series that is the stationary solution of the equation

$$X_n(t) = \Psi(X_{n-1})(t) + \xi_n(t), \qquad t \in I \subset \mathbb{R}_+, \quad n \in \mathbb{Z} \tag{3}$$

where $\{\xi_n\}$ is an MfBm functional white noise as in Example 1, with the twice continuously differentiable Hurst index function $H_\xi \in (0,1)$. We next assume that Ψ is the integral operator

$$\forall x \in C, \quad \Psi(x)(t) = \int_I \psi(s,t)x(s)ds, \qquad \text{with} \quad \iint_{I\times I} \psi^2(s,t)dsdt < 1$$

The stationary solution for (3) then exists (see, for instance, Kokoszka and Reimherr 2017, section 8.8). Assume further that constants $C > 0$, $H_\psi \in (0,1]$ exist such that

$$\sup_{u\in I} H_{\xi,u} < H_\psi \le 1$$
$$\text{and} \quad |\psi(s,u) - \psi(s,v)|^2 \le C|u-v|^{2H_\psi}, \quad \forall s,u,v \in I$$

Then, the local regularity parameters (H_t, L_t) of $\{X_n\}$ are $(H_{\xi,t}, 1)$ (see Supporting Information).

2.2 | Adaptive Mean and Autocovariance Functions Estimators

We propose nonparametric estimates of $\mu(t)$ and $\gamma_\ell(s,t)$. Our estimates adapt to the regularity of X and, to the best of our knowledge, are the first of their kind in the context of weakly dependent FTS. For simplicity, we restrict the study to the case where the trajectories X_n are not differentiable. Our estimators depend on (H_t, L_t^2), and for now we assume that suitable estimators $(\widehat{H}_t, \widehat{L}_t^2)$ of the local regularity parameters are given. Such local regularity estimators are constructed and studied in Section A in the Appendix.

Let $\widehat{\sigma}^2(t)$ be a consistent estimator of errors' variance $\sigma^2(t)$ at t. A simple choice is

$$\widehat{\sigma}^2(t) := \frac{1}{N}\sum_{n=1}^N \frac{1}{2}\left(Y_{n,i(t)} - Y_{n,i(t)+1}\right)^2$$

where, for each n, $i(t), i(t)+1$ are the indices of the two closest domain points $T_{n,i}$ to t. For each $1 \leq n \leq N$, we consider a local linear smoother of the curve X_n at $t \in I$,

$$\widehat{X}_n(t;h) = \sum_{i=1}^{M_n} W_{n,i}(t;h) Y_{n,i} \tag{4}$$

where $\{W_{n,i}(t;h)\}_{i=1\dots M_n}$ are local weights, and h is a local smoothing parameter which we let depending on t. For simplicity, we consider the Nadaraya–Watson (NW) estimator with a non-negative kernel supported on $[-1,1]$, see (**H**8). Given the $\widehat{X}_n(t;h)$, we follow the *smooth first, then estimate* approach and define the estimators for the mean and the lag-ℓ cross-products as the empirical estimators with the true curves replaced by the smoothed ones.

The rule $0/0 = 0$ applies in (4), meaning that the estimator $\widehat{X}_n(t;h)$ is set to 0 if there are no domain points $T_{n,i}$ in $[t-h, t+h]$. Similar situations occur with any smoothing-based method, and is more likely when the curves are sparsely sampled. When a curve is not observed in the neighborhood of t, it does not carry any useful information about $X_n(t)$, and should therefore be removed from the data set when estimating $\mu(t)$ or $\gamma_\ell(s,t)$. The neighborhood of t is defined by h. A trade-off must be found between the large bias induced by large h and the large variance resulting from dropping curves when h is small. Let

$$\pi_n(t;h) = 1 \quad \text{if} \sum_{i=1}^{M_n} \mathbb{1}\{|T_{n,i} - t| \leq h\} \geq 1, \quad \text{and}$$

$$\pi_n(t;h) = 0 \text{ otherwise}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. The number of curves X_n with at least one observation in the interval $[t-h, t+h]$ is then

$$P_N(t;h) = \sum_{n=1}^{N} \pi_n(t;h)$$

We finally denote the conditional expectation given the $\{M_n\}$ and the realizations of T by:

$$\mathbb{E}_{M,T}(\cdot) = \mathbb{E}\big(\cdot \mid M_n, \{T_{n,i}, 1 \leq i \leq M_n\}, 1 \leq n \leq N\big)$$

2.2.1 | Adaptive Mean Function Estimator

There are several contributions to the problem of estimating the mean function in the context of stationary FTS. First, if the curves are fully observed without error, under the $L^2_{\mathcal{H}} - m$-approximation assumption, the empirical mean is a $\sqrt{N}$-consistent estimator as per Hörmann and Kokoszka (2012). A central limit theorem for the empirical mean is also established under cumulant mixing dependence. Sabzikar and Kokoszka (2023) have defined a broad class of models for FTS that can be used to quantify near long-range dependence. They established rates of consistency for the empirical mean function assuming error-free, fully observed sample paths. Recently, authors have been focusing on the case where FTS are discretely sampled with an additive noise. Chen and Song (2015) and Li and Yang (2023) propose a method, based on B-splines, for constructing simultaneous confidence bands for the mean function under physical-dependence and infinite average FTS models, respectively. Their procedures assume an equidistant common design and the mean function is at least continuously differentiable. Rubín and Panaretos (2020) propose a local linear estimator of the mean function when the design is random and when the curves are sparsely observed. They derive asymptotic results assuming a mean function that is twice differentiable and two types of dependence conditions, namely cumulant mixing and strong mixing conditions. We here propose an adaptive nonparametric mean function estimator for irregular mean functions, and derive its asymptotic normality.

Let $t \in I$ be fixed. Our adaptive mean function pointwise estimator is

$$\widehat{\mu}^*_N(t) = \widehat{\mu}_N(t; h^*_\mu) \quad \text{with}$$

$$\widehat{\mu}_N(t;h) = \frac{1}{P_N(t;h)} \sum_{n=1}^{N} \pi_n(t;h) \widehat{X}_n(t;h) \tag{5}$$

where h^*_μ is an adaptive, optimal bandwidth. To define the selection rule for h, we derive a bound $2R_\mu(t;h)$ of the local quadratic risk $\mathbb{E}_{M,T}\left[(\widehat{\mu}_N(t;h) - \mu(t))^2\right]$. More precisely, let

$$\begin{aligned} &R_\mu(t; h, H_t, L_t^2, \sigma^2(t)) \\ &\quad = L_t^2 h^{2H_t} \mathbb{B}(t; h, 2H_t) + \sigma^2(t) \mathbb{V}_\mu(t;h) + \mathbb{D}_\mu(t;h)/P_N(t;h) \end{aligned} \tag{6}$$

and set $R_\mu(t;h) := R_\mu(t; h, H_t, L_t^2, \sigma^2(t))$, where

$$\mathbb{V}_\mu(t;h) = \frac{1}{P_N^2(t;h)} \sum_{n=1}^{N} \pi_n(t;h) \max_{1 \leq i \leq M_n} |W_{n,i}(t;h)|,$$

$$\mathbb{B}(t;h,\alpha) = \frac{1}{P_N(t;h)} \sum_{n=1}^{N} \pi_n(t;h) b_n(t;h,\alpha),$$

$$\begin{aligned} \mathbb{D}_\mu(t;h) &= \mathbb{E}\left[\{X_0(t) - \mu(t)\}^2\right] \\ &\quad + 2\sum_{\ell=1}^{N-1} p_\ell(t;h) \mathbb{E}\big(\{X_0(t) - \mu(t)\}\{X_\ell(t) - \mu(t)\}\big) \end{aligned}$$

with

$$b_n(t;h,\alpha) = \sum_{i=1}^{M_n} \left| \frac{T_{n,i} - t}{h} \right|^\alpha |W_{n,i}(t;h)|,$$

$$p_\ell(t;h) = \sum_{i=1}^{N-\ell} \frac{\pi_i(t;h)\pi_{i+\ell}(t;h)}{P_N(t;h)}$$

The three terms on the right-hand side of (6) can be interpreted as a bias, a stochastic and a penalty term, respectively. Regarding the latter, which is specific to the FDA framework, $\mathbb{D}_\mu(t;h)/P_N(t;h)$ increases as h decreases because more curves are excluded from the mean estimation. The weak dependence of $\{X_n\}$ will be defined such that the autocovariances of the time series $\{X_n(t), n \geq 1\}$ are absolutely summable. This means that, without using any additional information on the FTS model, we can simply take absolute values and bound $\mathbb{D}_\mu(t;h)$ by a constant equal to the limit of the series of the absolute values of the autocovariances. For now, we suppose that $\mathbb{D}_\mu(t;h)$ is given.

The local bandwidth h^*_μ is selected to minimize an estimate of $R_\mu(t;h)$. More precisely,

$$h^*_\mu \in \arg\min_{h\in\mathcal{H}_N} \widehat{R}_\mu(t;h) \quad \text{with}$$
$$\widehat{R}_\mu(t;h) = R_\mu(t;h,\widehat{H}_t,\widehat{L}_t^2,\widehat{\sigma}^2(t)) \tag{7}$$

where $\widehat{H}_t, \widehat{L}_t^2$ are local regularity estimators, and $\widehat{\sigma}^2(t)$ is an estimate of $\sigma^2(t)$. We will show that, under mild conditions, $\widehat{R}_\mu(t;h)/R_\mu(t;h) = 1 + o_{\mathbb{P}}(1)$, uniformly with respect to $h \in \mathcal{H}_N$. As a consequence, the rate of h^*_μ will coincide with that of the minimizer of $R_\mu(t;h)$.

2.2.2 | Adaptive Autocovariance Function Estimator

The nonparametric estimation of the lag-ℓ autocovariance function Γ_ℓ, $\ell \geq 1$, seems less explored in the literature. Kokoszka et al. (2017) consider the case of fully observed, error-free sample paths and derive asymptotic results for the empirical autocovariance functions. Zhong and Yang (2023) consider $MA(\infty)$ FTS observed with error over a fixed grid of design points, and use splines to estimate the sample paths. Their grid size corresponds to a dense regime and allows then to show that the empirical lag-ℓ autocovariance function constructed from the smoothed curves is asymptotically equivalent to the infeasible one obtained from the true sample paths. We propose here a nonparametric estimator of the lag-ℓ cross-product function $\gamma_\ell(s,t) = \mathbb{E}\left[X_n(s)X_{n+\ell}(t)\right]$, $s,t \in I$, with independent or common design, in a sparse or dense regime, and which adapts to the regularity of the process generating the FTS. The estimator of Γ_ℓ is constructed from the estimators of γ_ℓ and μ.

Let $s,t \in I$, and ℓ be an integer greater than or equal to 1. Let

$$P_{N,\ell}(s,t;h) = \sum_{n=1}^{N-\ell} \pi_n(s;h)\pi_{n+\ell}(t;h)$$

be the number of pairs $(X_n, X_{n+\ell})$ with at least one pair $(T_{n,i}, T_{n+\ell,k})$ in the rectangle $[s-h,s+h]\times[t-h,t+h]$. The adaptive estimator of $\gamma_\ell(s,t)$ is

$$\widehat{\gamma}^*_{N,\ell}(s,t) = \widehat{\gamma}_{N,\ell}(s,t;h^*_\gamma) \quad \text{with}$$
$$\widehat{\gamma}_{N,\ell}(s,t;h) = \sum_{n=1}^{N-\ell} \frac{\pi_n(s;h)\pi_{n+\ell}(t;h)}{P_{N,\ell}(s,t;h)} \widehat{X}_n(s;h)\widehat{X}_{n+\ell}(t;h) \tag{8}$$

with $\widehat{X}_n(s;h)$ and $\widehat{X}_{n+\ell}(t;h)$ as in (4). The local bandwidth h^*_γ is data-driven, defined as

$$h^*_\gamma \in \arg\min_{h\in\mathcal{H}_N} \widehat{R}_\gamma(s,t;h) \tag{9}$$

where $\widehat{R}_\gamma(s,t;h)$ is the estimate of

$$\begin{aligned}
R_\gamma(s,t;h) = {} & 3v_2^2\left(X_{1+\ell}(t)\right)L_s^2 h^{2H_s}\mathbb{B}(s|t;h,2H_s,0) \\
& + 3v_2^2\left(X_1(s)\right)L_t^2 h^{2H_t}\mathbb{B}(t|s;h,2H_t,\ell) \\
& + 3\left\{\sigma^2(s)v_2^2(X_{1+\ell}(t))\mathbb{V}_{\gamma,0}(s,t;h)\right. \\
& \left. + \ \sigma^2(t)v_2^2(X_1(s))\mathbb{V}_{\gamma,\ell}(s,t;h)\right\} \\
& + 3\sigma^2(s)\sigma^2(t)\mathbb{V}_\gamma(s,t;h) + \mathbb{D}(s,t;h)/P_{N,\ell}(s,t;h)
\end{aligned}$$

where for any $h > 0$, $\alpha > 0$, and any integer $\ell' \geq 0$,

$$\begin{aligned}
\mathbb{B}(t|s;h,\alpha,\ell') &= \sum_{n=1}^{N-\ell} \frac{\pi_n(s;h)\pi_{n+\ell}(t;h)}{P_{N,\ell}(s,t;h)} b_{n+\ell'}(t;h,\alpha), \\
b_n(t;h,\alpha) &= \sum_{i=1}^{M_n} \left|\frac{T_{n,i}-t}{h}\right|^\alpha W_{n,i}(t;h), \\
\mathbb{V}_{\gamma,\ell'}(s,t;h) &= \frac{1}{P_{N,\ell}(s,t;h)} \sum_{n=1}^{N-\ell} \frac{\pi_n(s;h)\pi_{n+\ell}(t;h)}{P_{N,\ell}(s,t;h)} \\
&\quad \times \max_{1\leq i\leq M_{n+\ell'}} W_{n+\ell',i}(t;h), \\
\mathbb{V}_\gamma(s,t;h) &= \frac{1}{P_{N,\ell}(s,t;h)} \sum_{n=1}^{N-\ell} \frac{\pi_n(s;h)\pi_{n+\ell}(t;h)}{P_{N,\ell}(s,t;h)} \\
&\quad \times \max_{1\leq i\leq M_n} W_{n,i}(s;h) \max_{1\leq k\leq M_{n+\ell}} W_{n+\ell,k}(t;h), \\
\mathbb{D}(s,t;h) &= \mathbb{E}(X_0 \otimes X_\ell - \gamma_\ell)^2(s,t) \\
&\quad + 2\sum_{k=1}^{N-\ell-1} p_k(s,t;h)\mathbb{E}(X_0 \otimes X_\ell - \gamma_\ell)(X_k \otimes X_{k+\ell} - \gamma_\ell)(s,t)
\end{aligned}$$

where

$$p_k(s,t;h) = \sum_{i=1}^{N-k-\ell} \frac{\pi_i(s;h)\pi_{i+k}(s;h)\pi_{i+\ell}(t;h)\pi_{i+\ell+k}(t;h)}{P_{N,\ell}(s,t;h)}$$

Here, for any f and g real-valued functions, $(f\otimes g)(s,t) = f(s)g(t)$.

We show that $2R_\gamma(s,t;h)$ is a bound of the quadratic risk $\mathbb{E}_{M,T}\{\widehat{\gamma}_\ell(s,t;h) - \gamma_\ell(s,t)\}^2$. Like for the mean function estimation, $\widehat{R}_\gamma(s,t;h)$ is obtained by replacing the values of H, L^2 and σ^2 by the estimates introduced above. Moreover, the moments $v_2^2\left(X_1(s)\right)$ and $v_2^2\left(X_{1+\ell}(t)\right)$ are simply obtained as empirical variances of the presmoothing estimator $\{\widetilde{X}_n\}$, see Section A in the Appendix. Assuming a stronger weak dependence assumption on $\{X_n\}$, the autocovariances of the series $\{X_n \otimes X_{n+\ell}(s,t), n \geq 1\}$ are absolutely summable. Similar to mean estimation, this means that we can simply take absolute values and bound $\mathbb{D}(s,t;h)$ by a constant. The details are given in Lemma E.8 in the Appendix. For now, we assume that $\mathbb{D}(s,t;h)$ is given.

3 | Asymptotic Results

We develop asymptotic theory as the number of curves N is growing to infinity. The expectation λ of the variable M depends on N, namely it has to increase to infinity with N. Thus, the domain points $T_{n,i}$, $1 \leq i \leq M_n$, $1 \leq n \leq N$ form a triangular array. Before presenting the results, we formally present two important hypotheses: the local regularity of X and the weak dependence of $\{X_n\}$. Then we prove the asymptotic behavior of our estimators under the independent design. The common design case will be discussed in Section 3.5.

3.1 | The Local Regularity Assumption

For simplicity, we focus on the case where the sample paths of X are almost surely nondifferentiable. Functional data in fields such as energy, environmental science, chemistry, physics, medicine, and meteorology are often highly irregular, making the consideration of nondifferentiable curves X_n realistic. The notion of local regularity is extended to cases where the sample paths of X have derivatives, with detailed proofs and results provided in the Supporting Information. Let $J \subset I$ be an open interval.

H6. The stationary distribution of $\{X_n\}$ satisfies the following conditions: Lipschitz functions $H : J \to (0,1]$, $L : J \to (0,\infty)$, and constants $\beta_0, S_0 > 0$ exist such that:

a. the paths of X belongs a.s. to the Banach space $C(I)$ and

$$0 < \underline{a}_0 := \inf_{u\in J} \mathbb{E}\left[X^2(u)\right] \leq \sup_{u\in J} \mathbb{E}\left[X^2(u)\right] =: \overline{a}_0 < \infty$$

b. there exists $\Delta_0 > 0$ such that $\forall t,u,v \in J$ with $t - \Delta_0/2 \leq u \leq t \leq v \leq t + \Delta_0/2$,

$$\left|\mathbb{E}\left[\{X(u) - X(v)\}^2\right] - L_t^2|u-v|^{2H_t}\right| \leq S_0^2|u-v|^{2H_t+2\beta_0}$$

Definition 3.1. Let $\mathcal{X}(H,L;J)$ denote the class of stochastic processes X with continuous paths satisfying (**H**6) and

$$0 < \inf_{u\in J} H_u \leq \max_{u\in J} H_u < 1 \quad \text{and} \quad 0 < \inf_{u\in J} L_u \leq \sup_{u\in J} L_u < \infty$$

The local regularity of X at t, an interior point of I, is defined by the parameters H_t and L_t^2.

3.2 | Weak Dependence Assumption

We consider a general notion of weak dependence which allows for a refined study of the local regularity of FTS. More precisely, we reconsider the concept of $\mathbb{L}^p_{\mathcal{H}} - m$*-approximability* in the context of $(C, \|\cdot\|_\infty)$-valued (instead of $\mathcal{H}$-valued) random processes. This is slightly more restrictive than the setup considered by Hörmann and Kokoszka (2010), but in this way the type of weak dependence between the curves X_n is inherited by the sequences $\{X_n(t)\}$, for all $t \in I$. The general idea with the weak dependence type considered by Hörmann and Kokoszka is to approximate $\{X_n\}$ by an m-dependent sequence $\{X_n^{(m)}, m \geq 1\}$ such that, for every $n \in \mathbb{Z}$, the sequence $\{X_n^{(m)}, m \geq 1\}$ converges in some sense to X_n as $m \to \infty$. The limiting behavior of the original process can then be obtained from that of its coupled m-dependent sequences provided they are sufficiently close to the original process.

Some more notations are needed: $\langle\cdot,\cdot\rangle_{\mathcal{H}}$ and $\|\cdot\|_{\mathcal{H}}$ are the inner product of the Hilbert space $\mathcal{H}$ and the associated norm, respectively. For $p \geq 1$, $\mathbb{L}^p$ is the space of real-valued variables Z with $\nu_p(Z) = (\mathbb{E}[|Z|^p])^{1/p} < \infty$. Moreover, $\mathbb{L}^p_{\mathcal{H}}$ and $\mathbb{L}^p_C$ are the spaces of $\mathcal{H}$-valued and C-valued random functions X with $\nu_p(\|X\|_{\mathcal{H}}) < \infty$ and $\nu_p(\|X\|_\infty) < \infty$, respectively.

Definition 3.2. The stationary FTS $\{X_n\}$ is $\mathbb{L}^p_C - m$-approximable with $p \geq 1$ if:

1. $\{X_n\} \subset \mathbb{L}^p_C$ admits a moving average (MA) representation, that is,

$$X_n = f(\xi_n, \xi_{n-1}, \ldots)$$

with $\{\xi_n\}$ independent copies of $\xi \in S$, S a measurable space and f a measurable function from S^∞ to C.

2. For every $n \in \mathbb{Z}$, let $\{\xi_k^{(n)}, k \in \mathbb{Z}\}$ be a sequence of independent copies of ξ defined over the same probability space. The coupled version of X_n is defined by

$$X_n^{(m)} = f(\xi_n, \xi_{n-1}, \ldots, \xi_{n-m+1}, \xi_{n-m}^{(n)}, \xi_{n-m-1}^{(n)}, \ldots)$$

3. The sequence $\{X_n^{(m)}, m \geq 1\}$ converges to X_n as $m \to \infty$ in the sense that

$$\sum_{m\geq 0} \nu_p\left(\|X_m - X_m^{(m)}\|_\infty\right) < \infty$$

As in Hörmann and Kokoszka (2010), having $p \geq 4$ will be convenient for our applications. Moreover, in Assumption (**H**7) below we impose stronger restrictions on the rate of convergence of the coupled sequences, see also Rice and Shum (2019).

H7. The stationary FTS $\{X_n\}$ is $\mathbb{L}^p_C - m$-approximable with some $p \geq 4$ such that constants $C > 0$ and $\alpha > 3/2$ exist and $\nu_p(\|X_m - X_m^{(m)}\|_\infty) \leq Cm^{-\alpha}$, $m \geq 1$.

The basic properties of $\mathbb{L}^p_{\mathcal{H}} - m$-approximability established by Hörmann and Kokoszka (2010, Lemma 2.1) remain true with Definition 3.2 (see our Lemma B.1 in the Appendix). Moreover, Lemma B.2 in the Appendix shows that $\mathbb{L}^p_C - m$-approximability entails the pointwise $\mathbb{L}^p - m$-approximability of $\{X_n(t)\}$, for all $t \in I$. The condition $p \geq 4$ guarantees that the long-run variances $\mathbb{D}(t;h)$ and $\mathbb{D}(s,t;h)$, of $\{X_n\}$ and $\{X_nX_{n+\ell}\}$, are summable. Note also that $\mathbb{L}^p_C - m$-approximability implies $\mathbb{L}^p_{\mathcal{H}} - m$-approximability, because the $\|\cdot\|_{\mathcal{H}}$ is bounded by the sup-norm. Definition 3.2 can also be considered with other Banach spaces C than $(C(I), \|\cdot\|_\infty)$. For instance, when C is the real line, our definition of $\mathbb{L}^p_C - m$-approximability becomes the $\mathbb{L}^p - m$-approximability for scalar times series, see Wu (2005).

Common FTS models, such as the FAR(1) model, functional linear process, product model, and Functional ARCH(1), are $\mathbb{L}^p_C - m$-approximable in the sense of Definition 3.2. We consider these examples under slightly stronger assumptions than those in Hörmann and Kokoszka (2010), corresponding to our choice of the space C in Definition 3.2. The full justification for these examples is provided in the Supporting Information. Let $\mathcal{L} = \mathcal{L}(C,C)$ denote the space of bounded linear operators on $(C, \|\cdot\|_\infty)$. For any operator $\Psi \in \mathcal{L}$, let $|||A|||_\infty = \sup\{\|Ax\|_\infty : \|x\|_\infty \leq 1\}$.

Example 3 (FAR(1) model revisited). Consider the model in Example 2, with $\Psi \in \mathcal{L}$ such that $|||\Psi|||_\infty < 1$, and a zero-mean i.i.d. sequence $\{\varepsilon_n\} \subset \mathbb{L}^p_C$. By Bosq (2000, Theorem 3.1), there exists then, a unique mean zero, stationary solution $\{X_n\} \subset C$ of the FAR(1) equation (3), provided $p \geq 2$. Then, $\{X_n\}$ is $\mathbb{L}^p_C - m$-approximable.

3.3 | Asymptotic Results on the Adaptive Mean Function Estimator

To establish the consistency and the convergence rate of the estimators in the independent design case, we make use of the following further assumptions.

H8. The estimator $\widehat{X}_n(t;h)$ is the NW estimator with

$$W_{n,i}(t;h) = K\left(\frac{T_{n,i}-t}{h}\right)\left[\sum_{k=1}^{M_n} K\left(\frac{T_{n,k}-t}{h}\right)\right]^{-1}, \quad i = 1\ldots M_n$$

where the convention $0/0 = 0$ applies. The kernel K is a nonnegative, symmetric and bounded kernel K, supported in $[-1,1]$ such that $\inf_{|u|<1} K(u) > 0$.

H9. The bandwidth set $\mathcal{H}_N$ is a grid with at most $(N\lambda)^c$ points, for some $c > 0$, such that $\max \mathcal{H}_N \to 0$ and $N\lambda \min \mathcal{H}_N / \log(N\lambda) \to \infty$. Moreover, $\log(N)/\log^2(\lambda) \to 0$.

H10. Constants $\underline{c}, \overline{c} > 0$ exist such that, for any N, $\underline{c} \le M/\lambda = M/\mathbb{E}(M) \le \overline{c}$.

H11. The density g of the observation points $T_{n,i}$ is Hölder continuous, and constants $\underline{c}_g, \overline{c}_g$ exist such that $0 < \underline{c}_g \le g(t) \le \overline{c}_g, \forall t \in I$.

H12. The estimators of $(H_t, L_t^2) \in (0,1)\times(0,\infty)$ admit concentration bounds as in Theorem A.1 and Theorem A.2 in the Appendix with $\varphi = \left(\log\lambda\right)^{-2}$ and $\psi = \left(\log\lambda\right)^{-1}$, respectively.

The condition on a kernel bounded from below (e.g., the uniform kernel) in (**H**8), and the condition (**H**10) can be relaxed at the cost of more involved technical arguments. Regarding (**H**11), if the design density g vanishes at t, the pointwise convergence rate at t of any nonparametric estimator would be degraded, and our assumption prevents this. We conjecture that by construction our risk bound (6) adapts to low design, but we leave the study of this aspect for future work. Finally, since we necessarily have $h > (N\lambda)^{-1}$ for every $h \in \mathcal{H}_N$, and, by the last part of (**H**9), $(N\lambda)^{-1/\log^2(\lambda)} \to 1$, the assumption (**H**12) guarantees that replacing the exponent H_t by its estimate does not change the rate of the risk bound. For $\widehat{L}_t^2$, which appears as a factor in the risk bound, a slower concentration rate is sufficient.

Theorem 3.3. *Let $t \in I$ and assume that (**H**1) to (**H**12) hold true. Then, we have $h_\mu^* = \mathcal{O}_{\mathbb{P}}\{(N\lambda)^{-1/(1+2H_t)}\}$, and the estimator $\widehat{\mu}_N^*(t) = \widehat{\mu}_N(t;h_\mu^*)$ defined in (5) and (7) satisfies*

$$\widehat{\mu}_N^*(t) - \mu(t) = \mathcal{O}_{\mathbb{P}}\left\{(N\lambda)^{-\frac{H_t}{1+2H_t}} + N^{-1/2}\right\}$$

The rate of the optimal bandwidth h_μ^* and the rate of convergence of $\widehat{\mu}_N^*(t)$ coincide with those obtained by Golovkine et al. (2025) in the i.i.d. case. Our mean function estimator achieves the minimax rate derived by Cai and Yuan (2011) for the mean function estimation. This convergence rate is slower than the parametric rate $\mathcal{O}_{\mathbb{P}}(N^{-1/2})$ in the *sparse regime* ($\lambda^{2H_t} \ll N$), and achieves the parametric rate in the *dense regime* ($\lambda^{2H_t} \gg N$). See Zhang and Wang (2016) for the terminology.

We next derive the pointwise asymptotic distribution of our adaptive mean function estimation. Usually, the rate of convergence in distribution for a nonparametric curve estimator is given by the power $-1/2$ of the effective sample, and the limit has the mean corrected by a bias term. In our context, the effective sample size is expected to be given by $N\lambda$ times the bandwidth. Meanwhile, the rate of convergence of the mean function estimator cannot be faster than the parametric rate $N^{-1/2}$ which corresponds to the ideal situation where all N curves are observed without error at t. In the following, we show that the effective sample size is given by $P_N(t;h_N)$ which, by construction, adaptively accounts for the two aspects.

In a functional data context, the regularity of the mean function is necessarily equal to or larger than that of the sample paths. As a consequence, the minimax optimal rates for the mean function estimation are given by the sample path regularity, see Cai and Yuan (2011). Hence, from the minimax optimality perspective, the rate of convergence in distribution for a mean function estimator has to depend on H_t. Assuming a higher regularity than the true one makes the convergence in distribution break down, as the bias term will tend to infinity.

Theorem 3.4. *Let $t \in I$ and assume that (**H**1) to (**H**12) hold true, and the error variable ε has a finite moment of order $(2+\delta)$, for some $\delta > 0$. Let $h_N \in \mathcal{H}_N$, $N \ge 1$, such that*

$$(N\lambda)^{1/(2H_t+1)} h_N \to 0 \qquad (10)$$

Moreover, assume that

$$\frac{\sigma^2(t)}{P_N(t;h_N)} \sum_{n=1}^{N} \pi_n(t;h_N)\left\{\sum_{i=1}^{M_n} W_{n,i}^2(t;h_N)\right\} \xrightarrow{\mathbb{P}} \Sigma(t) \in [0,\infty) \quad (11)$$

and

$$\mathrm{Var}_{M,T}\left(\frac{1}{\sqrt{P_N(t;h_N)}} \sum_{n=1}^{N} \pi_n(t;h_N)\{X_n(t) - \mu(t)\}\right) \xrightarrow{\mathbb{P}} \mathbb{S}_\mu(t) \in (0,\infty) \qquad (12)$$

Then $\sqrt{P_N(t;h_N)}\left\{\widehat{\mu}_N(t;h_N) - \mu(t)\right\} \xrightarrow{d} \mathcal{N}\left(0, \mathbb{S}_\mu(t) + \Sigma(t)\right)$.

To prove the pointwise asymptotic normality of $\widehat{\mu}_N(t;h_N)$, we first write the sequence $\sqrt{P_N(t;h_N)}\left\{\widehat{\mu}_N(t;h_N) - \mu(t)\right\}$ as the sum of the sequences $\sqrt{P_N(t;h)}\left\{\widehat{\mu}_N(t;h) - \widetilde{\mu}_N(t;h)\right\}$ and $\sqrt{P_N(t;h)}\left\{\widetilde{\mu}_N(t;h) - \mu(t)\right\}$, that are conditionally independent given $\{M_n, T_{n,i}, 1 \le i \le M_n, 1 \le n \le N\}$. Here, $\widetilde{\mu}_N(t;h)$ denotes the infeasible mean estimator obtained after replacing $\widehat{X}_n(t;h)$ by $X_n(t)$ in the definition of $\widehat{\mu}_N(t;h)$. Then, it suffices to study separately the convergence in distribution of these two sequences conditionally given the M_n's and the domain points $T_{n,i}$. For the former, after centering it, we use the Lyapounov's Central Limit Theorem (CLT) for independent variables, and the condition (11) requires that its conditional variance has a limit. Condition (10) makes the bias term negligible and thus avoids the usual mean correction used in the nonparametric regression. For the second, we use a suitable CLT for dependent variables, and the condition (12) requires that its conditional variance also has a limit. In the dense regime case, if in addition $\lambda h_N \to \infty$, then $\Sigma(t) = 0$ and $\mathbb{S}_\mu(t)$ is the long-run variance from the CLT for the infeasible empirical mean estimator (Horváth and Kokoszka 2012, Theorem 16.3). As expected, in the sparse regime case, the rate of convergence in distribution, given by

$\mathbb{E}[P_N(t;h_N)]^{-1/2}$, is slower than $N^{-1/2}$. Moreover, in this case $\Sigma(t) = \sigma^2(t)$ and $\mathbb{S}_\mu(t) = \mathrm{Var}(X(t))$. Finally, let us notice that Theorem 3.4 can be extended to a vector of values of the mean function. For brevity, we omit this extension.

3.4 | Asymptotic Results on the Adaptive Autocovariance Function Estimator

To derive the asymptotic rate of $\hat{\gamma}^*_{N,\ell}(s,t)$, we add an assumption on the bandwidth range.

H13. $N(\lambda \min \mathcal{H}_N)^2/\log(N\lambda) \to \infty$.

Theorem 3.5. *Assume the conditions* (**H**1) *to* (**H**6), (**H**7) *for* $p \geq 8$, (**H**8) *to* (**H**11), (**H**12) *for* $s,t \in I$, *and* (**H**13) *hold true. Moreover, assume that a constant* $\mathfrak{C} > 0$ *exists such that*

$$\mathbb{E}(X(u) - X(v))^4 \leq \mathfrak{C}\left[\mathbb{E}(X(u) - X(v))^2\right]^2, \qquad \forall u, v \in I$$

Let $H(s,t) = \min\{H_s, H_t\}$. *Then,*

$$h^*_\gamma = O_{\mathbb{P}}\left(\max\left\{(N\lambda^2)^{-\frac{1}{2\{H(s,t)+1\}}}, (N\lambda)^{-\frac{1}{2H(s,t)+1}}\right\}\right)$$

and

$$\hat{\gamma}^*_{N,\ell}(s,t) - \gamma_\ell(s,t) = \mathcal{O}_{\mathbb{P}}\left((N\lambda^2)^{-\frac{H(s,t)}{2\{H(s,t)+1\}}} + (N\lambda)^{-\frac{H(s,t)}{2H(s,t)+1}} + N^{-1/2}\right)$$

Let us note that

$$(N\lambda^2)^{-\frac{1}{2\{H(s,t)+1\}}} \gg (N\lambda)^{-\frac{1}{2H(s,t)+1}} \quad \Longleftrightarrow \quad \lambda^{2H(s,t)} \ll N$$

and

$$(N\lambda)^{-\frac{H(s,t)}{2H(s,t)+1}} \ll N^{-1/2} \quad \Longleftrightarrow \quad \lambda^{2H(s,t)} \gg N$$

As a consequence, in the "sparse" regime (i.e., $\lambda^{2H(s,t)} \ll N$),

$$\max\left\{\left|\hat{\mu}^*_N(s) - \mu(s)\right|, \left|\hat{\mu}^*_N(t) - \mu(t)\right|\right\} = o_{\mathbb{P}}\left(\left|\hat{\gamma}^*_{N,\ell}(s,t) - \gamma_\ell(s,t)\right|\right)$$

Moreover, the convergence rates of $\hat{\gamma}^*_{N,\ell}(s,t)$, $\hat{\mu}^*_N(s)$ and $\hat{\mu}^*_N(t)$ are all slower than the parametric rate $O_{\mathbb{P}}(N^{-1/2})$. In the "dense" regime (i.e., $\lambda^{2H(s,t)} \gg N$), the three estimators attain the parametric rate. As a consequence, the estimator $\hat{\Gamma}^*_{N,\ell}(s,t) = \hat{\gamma}^*_{N,\ell}(s,t) - \hat{\mu}^*_N(s)\hat{\mu}^*_N(t)$ of the autocovariance function estimator $\Gamma_\ell(s,t)$ has the same convergence rate as $\hat{\gamma}^*_{N,\ell}(s,t)$.

The pointwise convergence rate for the lag $-\ell$ autocovariance function, obtained in Theorem 3.5, coincides with that for the estimation of the covariance function in the i.i.d. case, see Golovkine et al. (2025). This rate is given by the lowest regularity exponent H at s and t.

3.5 | Asymptotic Results for the Common Design Case

As noted by Golovkine et al. (2025), the local bandwidth selection rules defined in (7) and (9) can be used for the mean and covariance function estimation, with both independent and common design. In the case of common design, where $T_{n,i} \equiv T_i$, $1 \leq i \leq \lambda$, the indicators $\pi_n(t;h)$ no longer depend on n, and they are all equal either to 0 or 1. That means that h^*_μ and h^*_γ are automatically chosen in the set of admissible bandwidths where the $\pi_n(t;h)$ are all equal to 1. That also means that the penalty terms $\mathbb{D}_\mu(t;h)/P_N(t;h)$ and $\mathbb{D}(s,t;h)/P_{N,\ell}(s,t;h)$ can be removed from the risk bounds, because they are constant on the range of admissible bandwidth values, and the risk bounds minimization is constrained to the admissible set. If the common design is equidistant, h cannot be smaller than $1/\lambda$. For both mean and autocovariance functions, two cases can occur: the minimum of the risk bound without the penalty term is attained in the interior of the admissible set of h (dense regime case), or on the left boundary where the bias term will be larger than the variance term (sparse regime case). Thus, our kernel smoothing automatically selects between linear interpolation and smoothing by choosing the optimal bandwidth in a data-driven manner. This is illustrated in our real data analysis for the mean function estimation. As a consequence of these facts, we can deduce the following result for which the justification is obvious and is thus omitted.

Theorem 3.6. *Assume that* $T_{n,i}$ *belong to a common design as in condition* (**H**3).

1. *Assume that the conditions of Theorem 3.3 are satisfied and* $\hat{\mu}^*_N(t) = \hat{\mu}_N(t;h^*_\mu)$ *with* h^*_μ *defined as in* (7). *Then,* $\hat{\mu}^*_N(t) - \mu(t) = \mathcal{O}_{\mathbb{P}}(\lambda^{-H_t} + N^{-1/2})$.
2. *Assume that the conditions of Theorem 3.5 are satisfied and* $\hat{\gamma}^*_{N,\ell}(s,t) = \hat{\gamma}_{N,\ell}(s,t;h^*_\gamma)$ *with* h^*_γ *defined as in* (9). *Then,* $\hat{\gamma}^*_{N,\ell}(s,t) - \gamma_\ell(s,t) = \mathcal{O}_{\mathbb{P}}(\lambda^{-H(s,t)} + N^{-1/2})$.

4 | Numerical Study

This section presents a Monte-Carlo study and an application to daily voltage curves from the Individual Household Electricity Consumption dataset (Hebrail and Berard 2012). The results were obtained using an R package which is publicly available at https://github.com/hmaissoro. The Epanechnikov kernel ($K(u) = (3/4)(1-u^2)$ for $|u| \leq 1$, and 0 otherwise) was used in all experiments.

4.1 | Simulation Setting

We consider three types of FTS $\{X_n\}$ and investigate the effectiveness of our methods in the case of nondifferentiable sample paths with different local regularity exponents. The three types of FTS that are considered are versions of a FAR(1) process with an associated innovation process $\{\xi_n\}$,

$$X_n(u) = \mu(u) + \int_0^1 \psi(u,s)(X_{n-1}(s) - \mu(s))ds + L_t\xi_n(u) \qquad (13)$$

where $\mu(t) = 4\sin(3\pi t/2)$, $\psi(u,s) = \kappa\exp(-(u+2s)^2)$ and the constant κ is chosen so that the operator norm $|||\cdot|||_\infty$ of the integral operator defined by $\psi(u,s)$ is equal to 0.5, while $L_t = 2$. A series of 100 burn-in steps are used to initialize (13).

The simulation results we present here are obtained with the series $\{X_n\}$ generated in what we call **FTS Model 2**: $\{X_n\}$ is

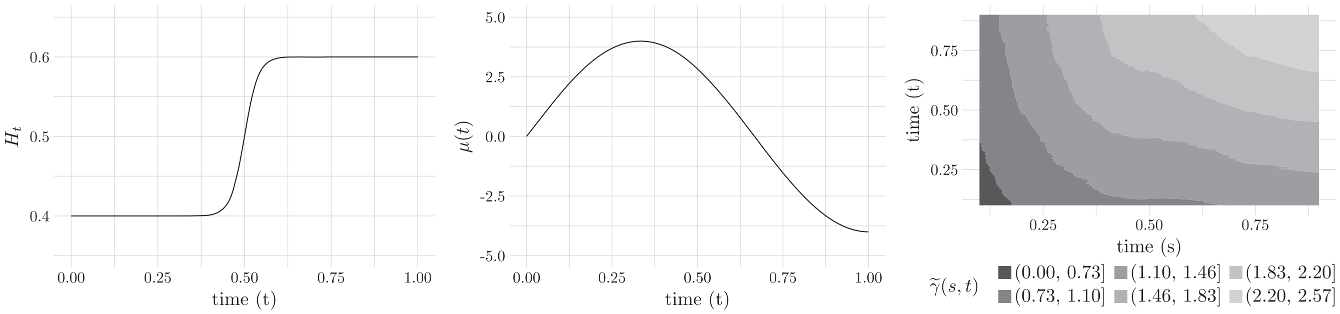

FIGURE 1 | Simulation parameters. **Left:** logistic local exponent function H_t used in FTS Model 2 and 3. **Middle:** the mean function μ used in FTS Model 1 and 2. **Right:** the empirical approximation of the lag $-$ 1 autocovariance function $\gamma_1(s,t)$ obtained from a large sample in FTS Model 2 when $\mu \equiv 0$.

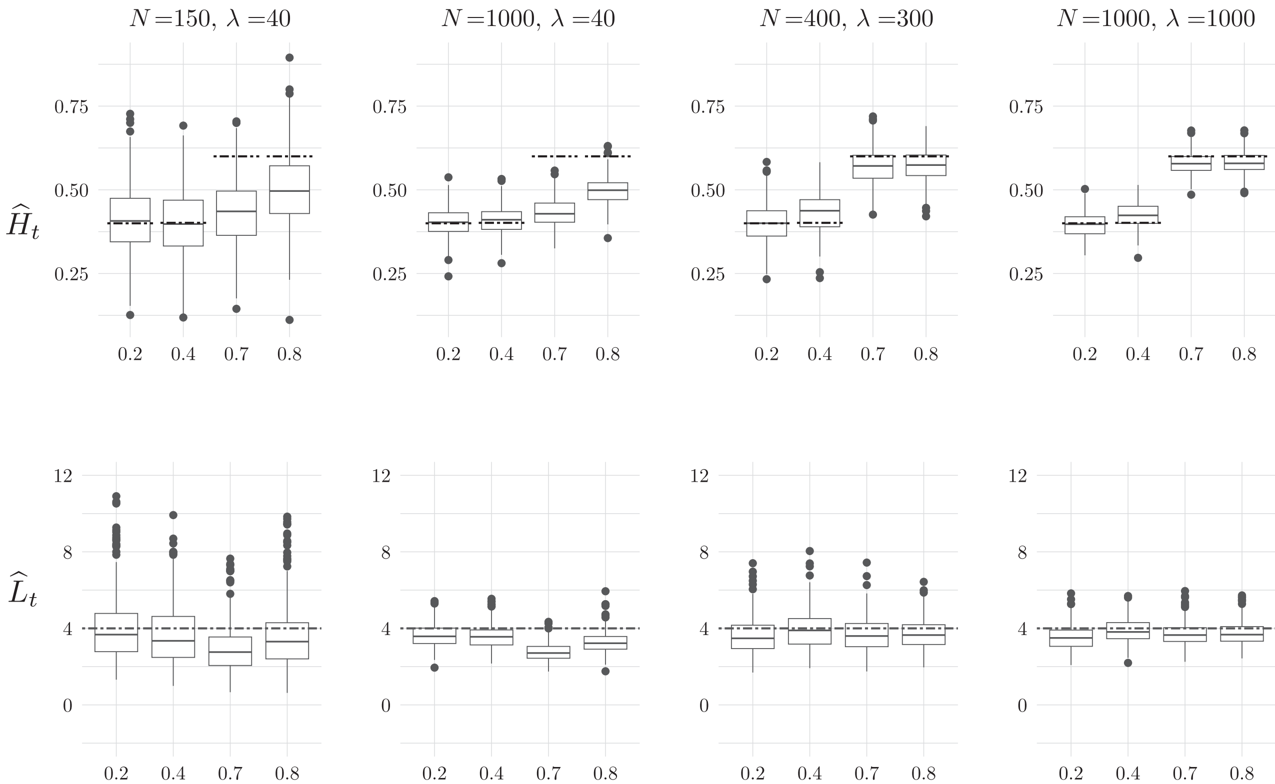

FIGURE 2 | Boxplots of $R = 400$ pointwise estimates of $\widehat{H}_t$ and $\widehat{L}_t^2$, for $t \in \{0.2, 0.4, 0.7, 0.8\}$ and four pairs (N, λ), in FTS Model 2. The dashed horizontal lines indicate the true values of H_t and L_t^2.

a FAR(1) as in (13), with $\{\xi_n\}$ independently generated from a MfBm with a logistic Hurst index function (see Figure 1). **FTS Model 1** is a version of **FTS Model 2** with a constant Hurst index H_t instead of the logistic one, while **FTS Model 3** is another version of **FTS Model 2** with the mean function $\mu(t)$ and the function $\psi(u, s)$ learned from the daily voltage of the Individual Household Electricity Consumption dataset (Hebrail and Berard 2012). The results obtained with **FTS Model 1** and **3**, as well as details of the setups of these models, are presented in Supporting Information.

To obtain the data points according to (1), the integers M_n are randomly generated uniformly between 0.8λ and 1.2λ, while the $\{T_{n,i}\}$ are uniformly distributed over $(0, 1]$. The errors $\varepsilon_{n,i}$ are Gaussian with constant variance $\sigma^2 = 0.25^2$. We consider $(N, \lambda) \in \{(150, 40), (1000, 40), (400, 300), (1000, 1000)\}$. For each setup, we generate $R = 400$ independent series.

4.2 | Local Regularity Estimation

Our approach for the estimation of H_t and L_t^2, detailed in Section A in the Appendix, depends on the tuning parameter Δ in (A.1), and the presmoothing used in (A.2). The presmoothing bandwidth is selected by a standard cross-validation procedure described in the Supporting Information. Our local regularity estimation approach is sensitive to the choice of Δ. The Theorems A.1 and A.2 propose the choice $\Delta = \exp\big(-(\log \lambda)^\gamma\big)$ for some $\gamma \in (0, 1)$. On the basis of an extensive simulation study detailed in the Supporting Information, we set $\gamma = 1/3$. Figure 2 shows the

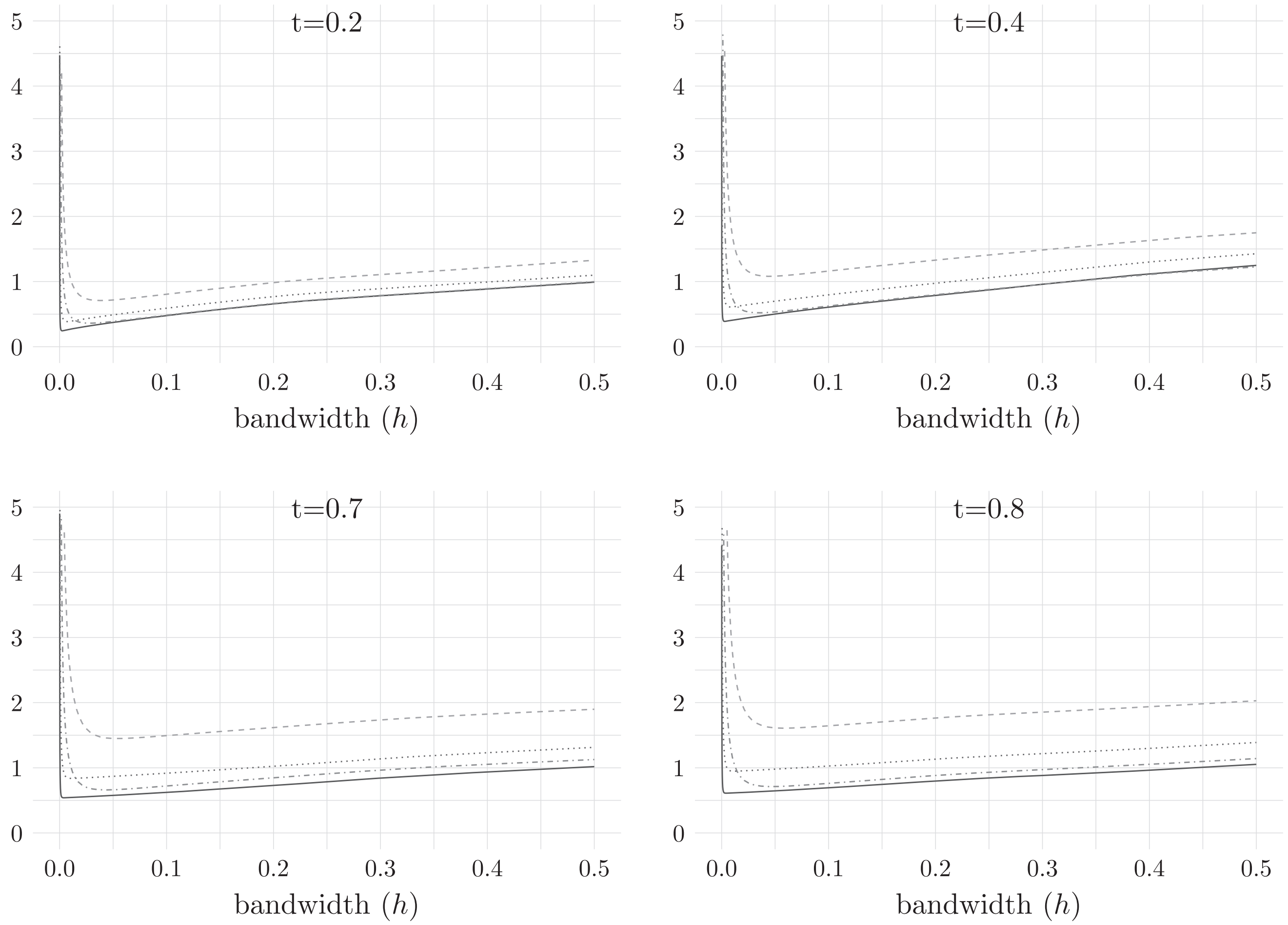

FIGURE 3 | Empirical average of the risk function $\widehat{R}_\mu(t;h)$ at $t \in \{0.2, 0.4, 0.7, 0.8\}$ over 400 independent functional series generated according to FTS Model 2, with four setups (N, λ): ______ $(1000, 1000)$, _ . _ . $(1000, 40)$, _ _ _ _ $(150, 40)$, ... $(400, 300)$.

boxplots of $\widehat{H}_t$ and $\widehat{L}_t^2$ defined in (A.4) for the four pairs (N, λ) at four points $t \in I = (0, 1]$. The bias of the regularity parameters estimates decreases as λ increases, and the boxplot are more concentrated as N increases. Overall, the local regularity estimators show good finite sample performance.

4.3 | Mean Function Estimation

Our adaptive "smooth first, then estimate" estimator of the mean function is constructed with the bandwidth h_μ^* defined in (7), obtained by minimizing the estimated bound $2\widehat{R}_\mu(t;h)$ of the pointwise quadratic risk. Instead of the dependence coefficient $\mathbb{D}_\mu(t;h)$, we simply consider

$$\overline{\mathbb{D}}_\mu(t;h) = \frac{1}{N}\sum_{n=1}^{N}\left\{\widetilde{X}_n(t) - \widehat{\nu}_1(X(t))\right\}^2 + 2\sum_{\ell=1}^{N-1}\frac{1}{N-\ell}\left|\sum_{n=1}^{N-\ell-1}\left\{\widetilde{X}_n(t) - \widehat{\nu}_1(X(t))\right\} \times \left\{\widetilde{X}_{n+\ell}(t) - \widehat{\nu}_1(X(t))\right\}\right|$$

with $\{\widetilde{X}_n\}$ the presmoothed curves as defined in (A.2) and $\widehat{\nu}_1(X(t))$ their empirical mean at t. Figure 3 presents the average of the risk function $\widehat{R}_\mu(t;h)$ over 400 independent time series generated according to FTS Model 2, with four setups (N, λ). The plots provide evidence that $h \to R_\mu(t;h)$ is a convex function which converges to zero as N and λ become larger.

Table 1 shows the bias and standard deviation of the estimates of $\widehat{\mu}_N^*(t) = \widehat{\mu}_N(t; h_\mu^*)$ obtained for functional time series generated according to the FTS Model 2. As expected the bias and the variance decrease as $N, \lambda \to \infty$. The estimated standard deviations increase as t increases, which may be surprising given that the sample paths become smoother to the right of the domain I. However, larger t also means larger $\mathrm{Var}(X_t)$ (see Supporting Information for the variance plot), and the consequence is less precise estimates of the mean. Finally, we study the asymptotic distribution of $\widehat{\mu}_N^*(t)$. The $Q - Q$ plots in Figure 4 show that the Gaussian limit, as stated in Theorem 3.4, is an accurate approximation. Indeed, we notice that the distribution of $P_N(t;h_N)^{1/2}\{\widehat{\Sigma}(t) + \widehat{\mathbb{S}}_\mu(t)\}^{-1/2}\left\{\widehat{\mu}_N(t;h_N) - \mu(t)\right\}$ is close to the standard normal distribution for all (N, λ) considered. The estimates $\widehat{\Sigma}(t)$ and $\widehat{\mathbb{S}}_\mu(t)$ are defined in the Supporting Information.

We conclude this section with a comparison with the procedure of Rubín and Panaretos (2020), procedure referred to as `RP20`, in the context of the FTS Model 2. A similar comparison in the context of the FTS Model 3 can be found in Supporting Information.

TABLE 1 | Bias and standard deviation (Sd) of the mean function estimates obtained from 400 independent time series generated in the FTS Model 2.

		$t = 0.2$		$t = 0.4$		$t = 0.7$		$t = 0.8$	
N	λ	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd
150	40	0.0056	0.2079	0.0112	0.2692	0.0329	0.3259	0.0497	0.3417
1000	40	0.0005	0.0883	−0.0062	0.1139	0.0119	0.1353	0.0213	0.1425
400	300	0.0074	0.1283	0.0049	0.1626	0.0119	0.1944	0.0150	0.2044
1000	1000	−0.0020	0.0849	0.0004	0.1094	−0.0003	0.1301	0.0003	0.1369

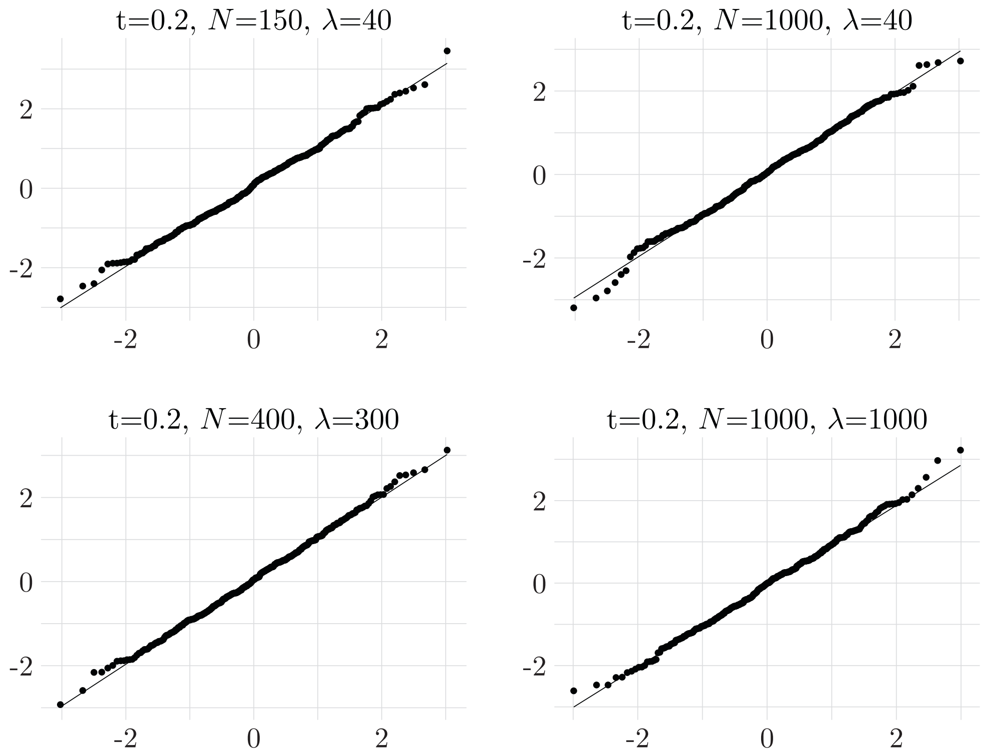

FIGURE 4 | Normal $Q - Q$ plots of $\sqrt{P_N(t;h_N)}\left(\widehat{\mu}_N(t;h_N) - \mu(t)\right)/\sqrt{\widehat{\mathbb{S}}_\mu(t) + \widehat{\Sigma}(t)}$ at $t = 0.2$, with $h_N = \{h^*_\mu\}^{1.1}$. The results obtained with 400 independent series generated in the FTS Model 2.

Rubín and Panaretos (2020) proposed a locally linear estimator of the mean function, which we denote by $\widehat{\mu}_{\mathrm{RP}}$, in sparsely observed settings. Their bandwidth is selected by K-fold cross-validation using the Bayesian optimization algorithm implemented in MATLAB. The implemented procedure is such that the observations times $\{T_{ij}\}$ are randomly sampled over a regular discrete grid of 241 points. In addition, since the implementation of K-fold cross-validation is time consuming, a projection on a B-spline basis is proposed for dimension reduction in the Bayesian optimization. In Figure 5, we present the boxplots of the selected bandwidths according to `RP20`'s global approach and to our local approach. The selected bandwidths have comparable sizes in almost all setups (N, λ). As expected from the increasing shape of the function H, our local bandwidths are smaller for t in the first half of I and increase as t gets closer to 1. Table 2 shows the ratio of the Monte-Carlo estimates of the mean square error (MSE) of our mean function estimator and the `RP20`'s local linear estimator. Although the ratio is close to 1, our estimator shows slightly better performance (ratio less than 1) in almost all setups.

4.4 | Autocovariance Function Estimation

To focus on the specific aspects related to the estimation of the lag-ℓ autocovariance function, we consider series generated as in FTS Model 2 but with the mean function set equal to zero. We set $\ell = 1$. An accurate approximation of $\gamma_1(s,t) = \mathbb{E}[X_n(s)X_{n+1}(t)]$ is shown in Figure 1. In this case, $\widehat{\Gamma}^*_{N,1}(s,t) = \widehat{\gamma}_{N,1}(s,t;h^*_\gamma)$, with $\widehat{\gamma}_{N,1}(s,t;h)$ defined in (8) and the bandwidth h^*_γ obtained from (9). Further details on the optimization in (9) are given in Supporting Information. The results of the estimation of the lag $-$ 1 autocovariance function for two setups (N, λ) are presented in Table 3. Larger Sd values occur for smaller N and/or for points (s,t) with larger values of $\gamma_1(s,t)$.

4.5 | Real Data Analysis

Predicting electrical energy consumption is essential for planning electricity production and significantly reduces the problems of storage and overproduction. A key step in this objective

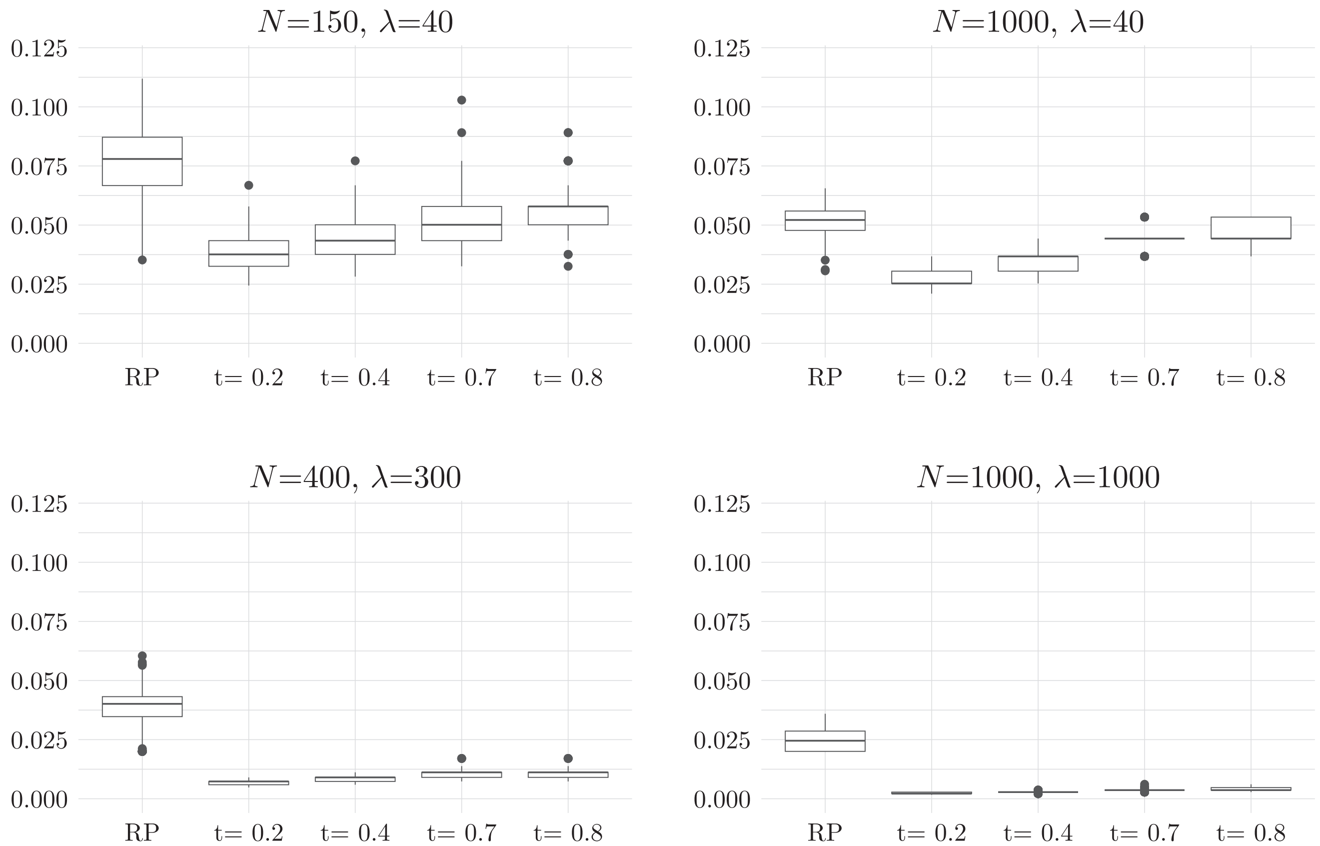

FIGURE 5 | Bandwidths selected by `RP20` (left boxplot) and by our local approach for the mean estimation at $t \in \{0.2, 0.4, 0.7, 0.8\}$; results from 400 independent series generated in the FTS Model 2.

TABLE 2 | MSE ratio for our mean estimator and `RP20`; results from 400 series generated in FTS Model 2.

N	λ	$t = 0.2$	$t = 0.4$	$t = 0.7$	$t = 0.8$
150	40	0.9689	0.9321	0.9520	0.9537
1000	40	0.9710	0.9414	0.9228	0.9208
400	300	1.0131	0.9716	0.9959	0.9867
1000	1000	0.9914	1.0015	0.9917	0.9949

TABLE 3 | Bias and standard deviation (Sd) of the lag-1 cross-product function $\gamma_1(s,t)$ estimation in FTS Model 2 when $\mu \equiv 0$; results obtained from 400 independent series.

		$(s,t) = (0.2, 0.4)$		$(s,t) = (0.4, 0.7)$		$(s,t) = (0.7, 0.8)$		$(s,t) = (0.8, 0.2)$	
N	λ	Bias	Sd	Bias	Sd	Bias	Sd	Bias	Sd
150	40	0.0019	0.3359	0.0307	0.5193	0.0371	0.6675	0.0102	0.4058
1000	40	0.0052	0.1303	−0.0004	0.1893	0.0026	0.2398	0.0126	0.1568

is to be able to accurately estimate the evolution of the electricity production parameters (such as the voltage), and functional time series are an effective approach for this purpose. To illustrate, we consider the data provided by the Individual Household Electricity Consumption dataset from the UC Irvine Machine Learning Repository (Hebrail and Berard 2012). It contains various measurements of electricity consumption in a household near Paris, with a sampling rate of one minute from December 2006 to November 2010. The data of interest here are 1358 voltage curves with a common design of 1440 points (corresponding to minute-by-minute observations), normalized so that $I = (0, 1]$. There are about 5.8% daily curves missing from the dataset, but we decided to neglect the missingness effect and consider the series as complete.

Our model (1), assumes that the measurement errors $\varepsilon_{n,i}$ are i.i.d. To check their serial correlation within the real voltage curves, we compute residuals from the smoothed curves and test

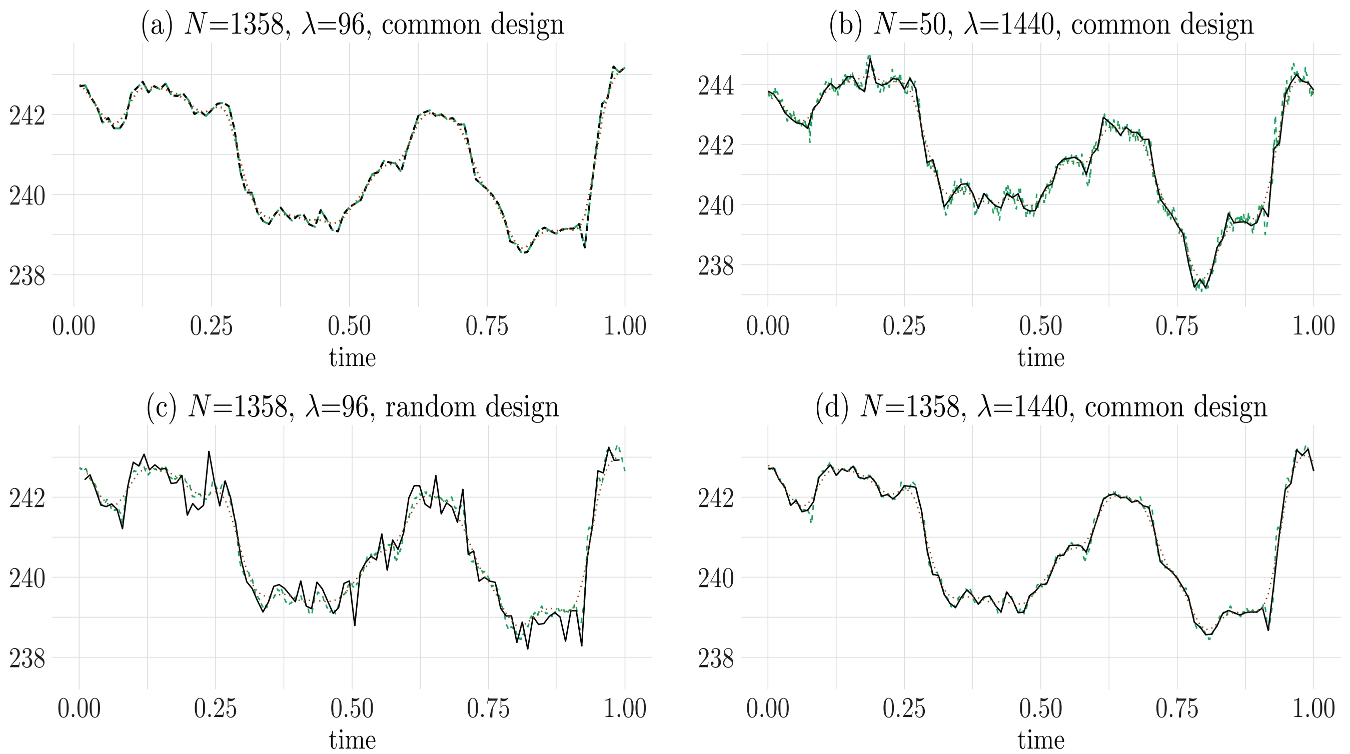

FIGURE 6 | Estimation of the daily mean function using our procedure (——— $\hat{\mu}_N$), the empirical mean (– – – – $\overline{X}_N$), and RP20 (……… $\hat{\mu}_{\text{RP}}$) for the daily voltage curves. The settings are: (a) all the curves with common design each 15 min; (b) the last 50 curve with complete common design (1440 points); (c) all the curves with independent design, the integers M_n are randomly generated between 77 and 115, the design points are selected randomly among the complete design; and (d) all the curves with the complete common design.

their serial dependence using the Durbin–Watson test (Durbin and Watson 1971), as implemented in the dwtest function from the R package lmtest (Hothorn et al. 2020). Using the Benjamini–Yekutieli multiple testing procedure (Benjamini and Yekutieli 2001), we reject the null hypothesis at the level 0.01 for only 5.6% of the curves. The details are provided in the Supporting Information.

We then compare the estimates of the daily mean voltage curve obtained by our procedure, the RP20 procedure and the empirical mean. In our procedure, we first estimate the local regularity parameters, as described in Section A in the Appendix. For the presmoothing and the choice of the tuning parameter Δ, we follow the guidelines given in Section 4.2, which proved to be satisfactory in our simulation study. Given the local regularity estimates, we minimize the risk function in (6) over a grid of bandwidths.

Figure 6 shows the estimates of the daily mean voltage curve with our procedure, the procedure RP20, and the empirical mean $\overline{X}_N$ constructed with the interpolated curves with common design. The plot of $\overline{X}_N$ in Figure 6c is the same as in Figure 6d. The infeasible empirical mean of all the curves X_n is the ideal estimator for the mean function μ. Given the large number of curves and design points in the real data set, one can reasonably consider that the estimator $\overline{X}_N$ in Figure 6d is close to the infeasible empirical mean and thus to the theoretical mean μ. In our application, the regularity exponent H_t is mostly between 0.1 and 0.5 (see Supporting Information). Therefore, the case in Figure 6a, more likely falls in the "sparse" regime ($\lambda^{2H_t} \ll N$). In this case, our approach automatically selects the smallest admissible bandwidth, leading to the interpolation. This explains why our estimator and the empirical mean coincides in Figure 6a. Let us point out that assuming twice differentiable curves, as it is usually considered in the literature, would mean to consider this case as belonging to the "dense" regime (because $96^4 \gg 1358$). When only a sub-series of $N = 50$ curves are selected, we are more likely in the "dense" regime. This explains that our estimator in Figure 6b is smoother than the empirical mean. The plot in Figure 6c compared with that in Figure 6a illustrates the difference between the common and independent designs. The estimator $\overline{X}_N$ exhibits irregular behavior in each plot. Our estimates appear to capture this irregularity more effectively than the procedure proposed by Rubín and Panaretos (2020), which produces smoother results, as can be seen in 6c,d. The case in Figure 6d seems to fall at the frontier between the "sparse" and "dense" regimes, and therefore our estimator balances between local averaging and interpolation. To summarize, our adaptive procedure automatically chooses between the interpolation and smoothing in the common design setting. It is worth recalling that in the common design with sparsely sampled curves (i.e., when $\lambda^{2H_t} \ll N$) interpolation is minimax rate optimal (see Cai and Yuan 2011). In practice, however, the phase transition from sparse to dense is unknown and an adaptive procedure is helpful. In the case of independent design, our adaptive estimator accurately approximates the ideal infeasible mean estimator and is able to capture its irregular patterns, which are likely to be inherited from the theoretical mean.

5 | Conclusions

We have studied a notion of local regularity for the process generating the sample paths of a stationary, weakly dependent functional time series (FTS). The paths are observed with heteroscedastic errors on discrete sets of design points, which may be fixed or random. The weak dependence condition we consider is satisfied by a large panel of FTS. Using a Nagaev-type inequality, we derive bounds on the concentration of the regularity estimators. Using the regularity estimators, adaptive mean and autocovariance function nonparametric estimators are proposed. The estimators adapt to the regularity of the process and to the nature of the design (sparse versus dense, independent versus common). They are simple "smooth first, then estimate" procedures where the kernel estimates of the sample paths are constructed with optimal plug-in local bandwidths. The bandwidths realize the minima of explicit pointwise risk bounds for the mean and autocovariance functions estimators, respectively. We also prove the pointwise asymptotic normality of the mean estimator, a result which permits the construction of adaptive confidence intervals for nondifferentiable mean functions. The study can be extended to other types of dependence, can allow for some serial correlation of the $\varepsilon_{n,i}$ within the curves or for an informative design (see Weaver et al. 2023). The uniform convergence for the mean and (auto)covariance functions can also be investigated. Such extensions are left for future work.

Acknowledgments

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-24-CE40-2439 (FUNMathStat project).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The real data supporting the study in Section 4 are publicly available (see Hebrail and Berard 2012). The simulated data for the Monte-Carlo study are generated with the R code provided at https://github.com/hmaissoro/Adaptive_Estimation_for_Weakly_Dependent_FTS.

References

Aue, A., D. D. Norinho, and S. Hörmann. 2015. "On the Prediction of Stationary Functional Time Series." *Journal of the American Statistical Association* 110, no. 509: 378–392.

Balança, P. 2015. "Some Sample Path Properties of Multifractional Brownian Motion." *Stochastic Processes and Their Applications* 125, no. 10: 3823–3850.

Benjamini, Y., and D. Yekutieli. 2001. "The Control of the False Discovery Rate in Multiple Testing Under Dependency." *Annals of Statistics* 29, no. 4: 1165–1188.

Bibinger, M., M. Jirak, and M. Vetter. 2017. "Nonparametric Change-Point Analysis of Volatility." *Annals of Statistics* 45, no. 4: 1542–1578.

Bosq, D. 2000. "Linear Processes in Function Spaces." In *Theory and Applications*, Lecture Notes in Statistics, vol. 149. Springer-Verlag.

Cai, T. T., and M. Yuan. 2011. "Optimal Estimation of the Mean Function Based on Discretely Sampled Functional Data: Phase Transition." *Annals of Statistics* 39, no. 5: 2330–2355.

Chen, M., and Q. Song. 2015. "Simultaneous Inference of the Mean of Functional Time Series." *Electronic Journal of Statistics* 9, no. 2: 1779–1798.

Chen, Y., T. Koch, K. G. Lim, X. Xu, and N. Zakiyeva. 2021. "A Review Study of Functional Autoregressive Models With Application to Energy Forecasting." *Wiley Interdisciplinary Reviews: Computational Statistics* 13, no. 3: e1525.

Durbin, J., and G. S. Watson. 1971. "Testing for Serial Correlation in Least Squares Regession. III." *Biometrika* 58: 1–19.

Golovkine, S., N. Klutchnikoff, and V. Patilea. 2022. "Learning the Smoothness of Noisy Curves With Application to Online Curve Estimation." *Electronic Journal of Statistics* 16, no. 1: 1485–1560.

Golovkine, S., N. Klutchnikoff, and V. Patilea. 2025. "Adaptive Estimation of Irregular Mean and Covariance Functions." *Bernoulli* 31, no. 2: 1032–1057.

Hebrail, G., and A. Berard. 2012. Individual Household Electric Power Consumption. UCI Machine Learning Repository.

Hörmann, S., and P. Kokoszka. 2010. "Weakly Dependent Functional Data." *Annals of Statistics* 38, no. 3: 1845–1884.

Hörmann, S., and P. Kokoszka. 2012. "Functional Time Series." In *Handbook of Statistics*, vol. 30, 157–186. Elsevier.

Horváth, L., and P. Kokoszka. 2012. "Inference for Functional Data With Applications." In *Springer Series in Statistics*. Springer.

Hothorn, T., A. Zeileis, R. W. Farebrother, et al. 2020. lmtest: Testing Linear Regression Models. CRAN. R Package Version 0.9–40.

Hsing, T., T. Brown, and B. Thelen. 2016. "Local Intrinsic Stationarity and Its Inference." *Annals of Statistics* 44, no. 5: 2058–2088.

Kokoszka, P., and M. Reimherr. 2017. *Introduction to Functional Data Analysis*, Texts in Statistical Science Series. CRC Press.

Kokoszka, P., G. Rice, and H. L. Shang. 2017. "Inference for the Autocovariance of a Functional Time Series Under Conditional Heteroscedasticity." *Journal of Multivariate Analysis* 162: 32–50.

Li, J., and L. Yang. 2023. "Statistical Inference for Functional Time Series." *Statistica Sinica* 33, no. 1: 519–549.

Martínez-Hernández, I., and M. G. Genton. 2021. "Nonparametric Trend Estimation in Functional Time Series With Application to Annual Mortality Rates." *Biometrics* 77, no. 3: 866–878.

Panaretos, V. M., and S. Tavakoli. 2013. "Fourier Analysis of Stationary Time Series in Function Space." *Annals of Statistics* 41, no. 2: 568–603.

Revuz, D., and M. Yor. 1999. *Continuous Martingales and Brownian Motion*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Vol. 293. 3rd ed. Springer-Verlag.

Rice, G., and M. Shum. 2019. "Inference for the Lagged Cross-Covariance Operator Between Functional Time Series." *Journal of Time Series Analysis* 40, no. 5: 665–692.

Rubín, T., and V. M. Panaretos. 2020. "Sparsely Observed Functional Time Series: Estimation and Prediction." *Electronic Journal of Statistics* 14, no. 1: 1137–1210.

Sabzikar, F., and P. Kokoszka. 2023. "Tempered Functional Time Series." *Journal of Time Series Analysis* 44, no. 3: 280–293.

Stoehr, C., J. A. D. Aston, and C. Kirch. 2021. "Detecting Changes in the Covariance Structure of Functional Time Series With Application to fMRI Data." *Econometrics and Statistics* 18: 44–62.

Wang, S. W., V. Patilea, and N. Klutchnikoff. 2025. "Adaptive Functional Principal Components Analysis." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*.

Weaver, C., L. Xiao, and W. Lu. 2023. "Functional Data Analysis for Longitudinal Data With Informative Observation Times." *Biometrics* 79, no. 2: 722–733.

Wu, W. B. 2005. "Nonlinear System Theory: Another Look at Dependence." *Proceedings of the National Academy of Sciences USA* 102, no. 40: 14150–14154.

Yao, F., H.-G. Müller, and J.-L. Wang. 2005. "Functional Data Analysis for Sparse Longitudinal Data." *Journal of the American Statistical Association* 100, no. 470: 577–590.

Zhang, X., and J.-L. Wang. 2016. "From Sparse to Dense Functional Data and Beyond." *Annals of Statistics* 44, no. 5: 2281–2321.

Zhong, C., and L. Yang. 2023. "Statistical Inference for Functional Time Series: Autocovariance Function." *Statistica Sinica* 33, no. 4: 2519–2543.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1.** Supporting Information.