Learning the smoothness of noisy curves with application to online curve estimation[*]

Steven Golovkine[1], Nicolas Klutchnikoff[2] and Valentin Patilea[3]

[1] *Groupe Renault & CREST - UMR 9194, Rennes, France*
e-mail: steven.golovkine@ensai.fr

[2] *Univ Rennes, IRMAR - UMR 6625, France*
e-mail: nicolas.klutchnikoff@univ-rennes2.fr

[3] *Ensai, CREST - UMR 9194, Rennes, France*
e-mail: valentin.patilea@ensai.fr

Abstract: Combining information both within and across trajectories, we propose a simple estimator for the local regularity of the trajectories of a stochastic process. Independent trajectories are measured with errors at randomly sampled time points. The proposed approach is model-free and applies to a large class of stochastic processes. Non-asymptotic bounds for the concentration of the estimator are derived. Given the estimate of the local regularity, we build a nearly optimal local polynomial smoother from the curves from a new, possibly very large sample of noisy trajectories. We derive non-asymptotic pointwise risk bounds uniformly over the new set of curves. Our estimates perform well in simulations, in both cases of differentiable or non-differentiable trajectories. Real data sets illustrate the effectiveness of the new approaches.

Contents

1. Introduction

More and more phenomena in modern society produce observation entities in the form of a sequence of measurements recorded intermittently at several discrete points in time. Very often the measurements are noisy and the observation points in time are neither regularly distributed nor the same across the entities. Functional data analysis (FDA) considers such data as being values on the trajectories of a stochastic process, recorded with some error, at discrete random times. One of the main purposes of the FDA is to recover the trajectories, also called curves or functions, at any point in time. See, *e.g.*, [29, 21, 38, 41] for some recent references. Whatever the approach for recovering the curve is, in the existing literature it is usually assumed that, for each curve, a certain number of derivatives exist. However, many applications, some of them presented in

the following, indicate that assuming that the curves admit second, third,... order derivatives is not realistic. Assuming that the curves to be reconstructed are smoother than they really are could lead to missing important information carried by the data. In this contribution, we propose a definition of the local regularity of the curves which could be easily estimated from the data and used to estimate the curves.

To formalize the framework, let $I \subset \mathbb{R}$ be a compact interval of time. We consider N functions $X^{(1)}, \ldots, X^{(n)}, \ldots, X^{(N)}$ generated as a random sample of a stochastic process $X = (X_t : t \in I)$ with continuous trajectories. For each $1 \leq n \leq N$, and given a positive integer M_n, let $T_m^{(n)}$, $1 \leq m \leq M_n$, be the random observation times for the curve $X^{(n)}$. These times are obtained as independent copies of a variable T taking values in I. The integers $M_1, \ldots, M_N$ represent an independent sample of an integer-valued random variable M with expectation μ. Thus $M_1, \ldots, M_N$ is the Nth line in a triangular array of integer numbers. We assume that the realizations of X, M and T are mutually independent. The observations associated with a curve, or trajectory, $X^{(n)}$ consist of the pairs $(Y_m^{(n)}, T_m^{(n)}) \in \mathbb{R} \times I$ where $Y_m^{(n)}$ is defined as

$$Y_m^{(n)} = X^{(n)}(T_m^{(n)}) + \varepsilon_m^{(n)}, \qquad 1 \leq n \leq N, \ \ 1 \leq m \leq M_n, \tag{1}$$

and $\varepsilon_m^{(n)}$ are independent copies of a centered error variable ε. For the sake of readability, here and in the following, we use the notation X_t for the value at t of the generic process X and $X^{(n)}(t)$ for the value at t of the realization $X^{(n)}$ of X. The $N-$sample of X is composed of two sub-populations: a *learning* set of N_0 curves and a set of N_1 curves to be recovered that we call the *online* set. Thus, $1 \leq N_0, N_1 < N$ and $N_0 + N_1 = N$. Let $X^{(1)}, \ldots, X^{(N_0)}$ denote the curves corresponding to the *learning* set.

Our first aim is to define a meaningful, model-free concept of local regularity for the process X and to build an estimator for it. The estimator could be computed easily and rapidly from the observations $(Y_m^{(n)}, T_m^{(n)})$ corresponding to the curves in the *learning* set, and does not require a very large number N_0 of curves. Moreover, it could be easily updated if more curves are added to the *learning* set. The problem of estimating the regularity of X is related to the estimation of the Hausdorff, or fractal, dimension of time series. See, for instance, [9, 8, 16] and the references therein. However, herein, we adopt the FDA point of view and use the so-called *replication* and *regularization* features of functional data (see [29], ch.22). More precisely, we combine information both across and within curves. Thus, taking strength from the information contained in the whole set of N_0 available time series, we are able to investigate more general situations: X need not to be a Gaussian, or a transformed Gaussian process, it is not necessarily stationary or with stationary increments, it could have a fractal dimension which changes over time, it is observed with possibly heteroscedastic noise, at random moments in time.

The local regularity we study determines the regularity of the sample paths of X. Sample paths regularity determines, for instance, the minimax optimal rate for the nonparametric estimators of the mean and covariance functions. In par-

ticular, knowing the local regularity serves to distinguish the so-called sparsely and dense sampled curve cases. See [6, 41]. For some widely used examples, the local regularity is also related to the rate of decrease of the eigenvalues of the covariance operator, a property which is commonly used in FDA literature. In almost all existing contributions, the sample paths regularity and the rate of the eigenvalues are supposed given. We here propose a simple method to estimate them.

Based on the regularity estimates, our second objective is to build an adaptive, nearly optimal smoothing for a possibly very large set of N_1 new curves. While several smoothers could be used, we focus on local polynomials. Optimal curve reconstruction is an important step in FDA, for instance for computing the median curve or the depth of a curve, to detect outliers. See, *e.g.,* [25]. It can also serve to compute optimal mean and covariance functions estimator in the dense case. See, *e.g.,* [6, 41]. Let

$$X^{[1]} = X^{(N_0+1)}, \ldots, X^{[N_1]} = X^{(N)},$$

denote the curves from the *online* set to be recovered from the corresponding observations $(Y_m^{(n)}, T_m^{(n)})$. This issue is a nonparametric estimation problem and, if each curve regularity is given, nonparametric estimators of the curves $X^{[1]}, \ldots, X^{[N_1]}$ could be easily built, for instance using the local linear smoother or the series estimator. Nevertheless in applications, there is no reason to suppose that the sample paths of the random process X have a known regularity. When it is not reasonable to assume a given regularity for the trajectories, one could use one of the existing data-driven procedures for determining the optimal smoothing parameter. However, the existing procedures, such as the cross-validation or the Goldenshluger-Lepski method [17], were designed for the case where one observes only one curve. Thus one has to apply them for each curve separately, which could require large amounts of resources.

In Section 2, we define the local regularity and provide concentration bounds for the estimator of the local regularity of the trajectories of X. Our results are new and of non-asymptotic type, in the sense that they hold for any values of the sample sizes N_0 and the mean value of observation times μ, provided these values are sufficiently large. In Section 3, we explain the relationship between the probabilistic concept of local regularity for the trajectory of X and the analytic regularity of the curves which usually determines the optimal risk rate in nonparametric estimation. We also provide insight into the relationship between the local regularity and the rate of decrease for the eigenvalues of the covariance operator. Given the estimate of the local regularity of the trajectories of X, in Section 3 we build adaptive local polynomial estimators and provide a non-asymptotic bound for the pointwise risk of the local polynomial smoother, uniformly over the *online* set. This uniform bound is obtained using an exponential-type moment bound for the pointwise risk for the local polynomial smoother, a new result of interest in itself. The pointwise risk bound is optimal, in the nonparametric regression estimation sense, up to some logarithmic factors induced by our stochastic curves model, the concentration of

the local regularity estimator, and the uniformity over the *online* set. Assuming that X has a constant regularity over the interval I, we also derive a non-asymptotic bound for the risk of the local polynomial smoother uniformly over I, and uniformly over the *online* set. In Section 4, we provide some additional guidance for the implementation of the local polynomial smoother and report results from simulation showing that our estimator of the local regularity and the adaptive local polynomial estimator perform well in both cases, whether or not the trajectories are differentiable. As a further application of our local regularity estimation approach, we consider the median curve estimation problem for samples of noisy curves. Our median curve is obtained from the smoothed curves with the optimal bandwidth given by the local regularity estimate. We compare the accuracy of our median curve with that obtained with the curves smoothed by cross-validation. While the accuracy is comparable, the computation time is far shorter when using our approach, and this makes it suitable for embedded systems or for applications with online data. A real data application on vehicle traffic flow analysis illustrates the effectiveness of our approaches. The proofs of our results are postponed to the Appendix. Additional technical aspects, simulation results, and details on traffic flow application are also relegated to the Appendix. To further illustrate the irregularity of the curves in applications, we also report in the Appendix the local regularity estimates for another three functional data sets often analyzed in the literature.

2. Local regularity estimation

The new local regularity estimator is introduced and studied in this section. After providing some insight into the ideas behind the construction, we provide a concentration result for our estimator under general mild assumptions which do not impose a specific distribution for X. In particular, X could, but need not, be a Gaussian process. The case where the variance of the noise is not constant is also discussed.

2.1. The methodology

Let us present the main ideas behind the construction of the regularity estimate. For this, let us introduce some more notation used throughout the paper. Let K_0 be an integer value which will be defined below, and consider the order statistics of a M-sample $T_1, \ldots, T_M$ distributed as T which admits the density f. Let $t_0 \in I$ such that $f(t_0) > 0$. We extract the subvector of the K_0 closest values to t_0 and denote these values $T_{(1)} \leq \ldots \leq T_{(K_0)}$. If $t_0 = \inf(I)$ then $t_0 \leq T_{(1)}$, while if $t_0 = \sup(I)$, then $T_{(K_0)} \leq t_0$. When t_0 is an interior point of I, t_0 likely lies between $T_{(1)}$ and $T_{(K_0)}$. Next, we define the interval

$$J_\mu(t_0) = \big(t_0 - |I|/\log(\mu), t_0 + |I|/\log(\mu)\big) \cap I,$$

where $|I|$ denotes the length of the interval I and, recall, μ is the expectation of M. In the following, we introduce our conditions using the interval $J_\mu(t_0)$,

which depends on μ. The theoretical results we derive are non-asymptotic, in particular they hold for any fixed μ, provided it is sufficiently large. If one is interested by asymptotic results corresponding to the case where μ increases to infinity, then our $J_\mu(t_0)$ is eventually contained in any fixed neighborhood of t_0. In this case, one can state all the assumptions on X in a more standard way, using a fixed interval instead of our $J_\mu(t_0)$.

We assume that the process X generating the continuous curves $X^{(1)}, \ldots, X^{(N)}$ satisfies

$$\mathbb{E}\left[(X_u - X_v)^2\right] \approx L_{t_0}^2 |v - u|^{2H_{t_0}}, \quad u, v \in J_\mu(t_0), \tag{2}$$

for some $H_{t_0} \in (0, 1]$ and $L_{t_0} > 0$ which could both change with t_0. Here and in the following, $\approx$ means the left-hand side is equal to the right-hand side times a quantity which tends towards 1 when $|v - u| \to 0$. When the trajectories of X are not differentiable, H_{t_0} is what we call the *local regularity of the process X at t_0*. For now, we focus on this case. When, with probability 1, the trajectories of X admits derivatives of order $\mathbf{d} \geq 1$ in a neighborhood of t_0, the property (2) will be used for the derivative of order $\mathbf{d}$ of the smooth trajectories. In this smooth case, the local regularity of the process X at t_0 will be $\mathbf{d} + H_{t_0}$. See the comment following Theorem 1.

Some commonly used processes have the eigenvalues of the covariance operator such that, for some $\nu > 1$, $\lambda_j \sim j^{-\nu}$, $j \geq 1$. Such processes have a constant local regularity. Moreover, $\mathbf{d} + H_{t_0} = (\nu - 1)/2$. As an example, the stationary fractional Ornstein-Uhlenbeck process with index $\rho \in (0, 2)$ has the covariance function

$$\Gamma(s, t) = \exp(-a|s - t|^\rho), \text{ for some } a > 0,$$

which yields $\nu = 1 + \rho$, $H_{t_0} \equiv \rho/2$ and $\mathbf{d} = 0$. Among the nonstationary processes satisfying our condition (2), we can mention the fractional Brownian motion with Hurst exponent $H \in (0, 1)$, for which $H_{t_0} \equiv H$ and $\lambda_j \sim j^{-(1+2H)}$. Examples with $\mathbf{d} > 0$ could be obtained by $\mathbf{d}-$times integration of the processes with $\mathbf{d} = 0$, such as for instance the so-called $\mathbf{d}-$integrated Brownian motion. See, *e.g.*, [24] for more details on these examples.

To construct our estimator of H_{t_0}, we consider the event

$$\mathcal{B} = \{M \geq K_0, T_{(1)} \in J_\mu(t_0), \ldots, T_{(K_0)} \in J_\mu(t_0)\},$$

which is expected to be of high probability. Let $\mathbf{1}_\mathcal{B}$ denote the indicator of $\mathcal{B}$ and let us define the expectation operator

$$\mathbb{E}_\mathcal{B}(\cdot) = \mathbb{E}(\cdot \mathbf{1}_\mathcal{B}).$$

Using (2) and the independence between X and T, for any $1 \leq k < l \leq K_0$,

$$\mathbb{E}_\mathcal{B}\left[(X_{T_{(l)}} - X_{T_{(k)}})^2\right] \approx L_{t_0}^2 \mathbb{E}_\mathcal{B}\left(|T_{(l)} - T_{(k)}|^{2H_{t_0}}\right).$$

From this and the moments of the spacing $T_{(l)} - T_{(k)}$ as given in the Lemma 2, we obtain

$$\mathbb{E}_\mathcal{B}\left[(X_{T_{(l)}} - X_{T_{(k)}})^2\right] \approx L_{t_0}^2 \left(\frac{l - k}{f(t_0)(\mu + 1)}\right)^{2H_{t_0}}.$$

Now, for any $1 \leq k \leq K_0$, let $\varepsilon_{(k)}$ be a generic error term corresponding to the generic realization $X_{T_{(k)}}$, and denote

$$Y_{(k)} = X_{T_{(k)}} + \varepsilon_{(k)}.$$

Moreover, for k such that $2k - 1 \leq K_0$, let

$$\theta_k = \mathbb{E}_{\mathcal{B}}\left[(Y_{(2k-1)} - Y_{(k)})^2\right].$$

Let σ^2 denote the variance of the error term, assumed to be finite. We then obtain

$$\frac{\theta_k - 2\sigma^2}{L_{t_0}^2} \approx \left(\frac{k-1}{f(t_0)(\mu+1)}\right)^{2H_{t_0}}. \tag{3}$$

We distinguish two situations : the case where σ^2 is known and the case where it is unknown. In the former case, we suppose that $4k - 3$ is also less than K_0 and use twice the relationship (3) with k and $2k - 1$, respectively. We deduce

$$\frac{\theta_{2k-1} - 2\sigma^2}{\theta_k - 2\sigma^2} \approx 4^{H_{t_0}}.$$

Taking the logarithm on both sides, we obtain the proxy value

$$H_{t_0}(k, \sigma^2) = \frac{\log(\theta_{2k-1} - 2\sigma^2) - \log(\theta_k - 2\sigma^2)}{2\log 2},$$

of the local regularity parameter H_{t_0}, when σ^2 is given. In the case where σ^2 is unknown, assuming that $8k - 7 \leq K_0$, we use the relationship (3) three times with k, $2k - 1$ and $4k - 3$, respectively, to obtain

$$\frac{\theta_{4k-3} - \theta_{2k-1}}{\theta_{2k-1} - \theta_k} \approx 4^{H_{t_0}}.$$

A natural proxy of H_{t_0} is then given by

$$H_{t_0}(k) = \frac{\log(\theta_{4k-3} - \theta_{2k-1}) - \log(\theta_{2k-1} - \theta_k)}{2\log 2}. \tag{4}$$

Our estimator of the local regularity parameter H_{t_0} is the empirical version of the proxy value $H_{t_0}(k)$, or $H_{t_0}(k, \sigma^2)$, built from a random sample of N_0 trajectories of X, the learning set of curves. Formally, we consider the sequence of events, for $1 \leq n \leq N_0$,

$$\mathcal{B}_n = \mathcal{B}_n(\mu, N_0) = \left\{M_n \geq K_0, T_{(1)}^{(n)} \in J_\mu(t_0), \ldots, T_{(K_0)}^{(n)} \in J_\mu(t_0)\right\}, \tag{5}$$

and we define

$$\hat{\theta}_k = \frac{1}{N_0} \sum_{n=1}^{N_0} \left[Y_{(2k-1)}^{(n)} - Y_{(k)}^{(n)}\right]^2 \mathbf{1}_{\mathcal{B}_n}, \tag{6}$$

where, for any n and k, $Y_{(k)}^{(n)}$ denotes the noisy measurement of $X^{(n)}(T_{(k)}^{(n)})$. If $H_{t_0}(k, \sigma^2)$ is indeed a good approximation of H_{t_0}, a simple estimator of H_{t_0} when σ^2 is known is then

$$\widehat{H}_{t_0}(k, \sigma^2) = \begin{cases} \dfrac{\log(\hat{\theta}_{2k-1} - 2\sigma^2) - \log(\hat{\theta}_k - 2\sigma^2)}{2\log 2} & \text{if } \min(\hat{\theta}_{2k-1}, \hat{\theta}_k) > 2\sigma^2 \\ 1 & \text{otherwise.} \end{cases} \quad (7)$$

The default value 1 is arbitrary and could be replaced by any number between 0 and 1. When σ^2 is unknown the corresponding estimator is

$$\widehat{H}_{t_0}(k) = \begin{cases} \dfrac{\log(\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}) - \log(\hat{\theta}_{2k-1} - \hat{\theta}_k)}{2\log 2} & \text{if } \hat{\theta}_{4k-3} > \hat{\theta}_{2k-1} > \hat{\theta}_k \\ 1 & \text{otherwise,} \end{cases} \quad (8)$$

where $\hat{\theta}_{4k-3}$ is obtained from the formula of $\hat{\theta}_{2k-1}$ after replacing k by $2k - 1$.

It is worth noting that our estimator could be easily updated every time new curves are included in the learning sample, without revisiting the learning set already used. Indeed, one should only add new terms in the sums defining $\hat{\theta}_k$, $\hat{\theta}_{2k-1}$ and $\hat{\theta}_{4k-3}$.

The Associate Editor drew our attention on a large literature related to the estimation of the regularity of nonparametric functions. [15] consider that data are noisy measurements of one sample path from a scaled fractional Brownian motion (fBm) with unknown Hurst parameter and unknown scale. The measurements are sampled on an equidistant grid, and the noise is allowed to be heteroscedastic. The authors derive the optimal rate for estimating the Hurst parameter. Moreover, they provide an estimator achieving the optimal rate. The estimator of [15] is based on the self-similarity property of the fBm process. Our estimator relies on a related property imposed to the second order moment of the increments; see (2) above. Our condition defines a significantly larger class of processes. By construction, our results are not easily comparable to that of Gloter and Hoffmann, which are more refined but derived in a different, more restrictive context. Our estimator is designed to take advantage of the replication feature of functional data, where sample averages provide simple estimates for the moments of the increments. Another related topic extensively studied in the literature, is the construction of confidence or credible sets for a curve of unknown regularity. See, for instance, [14, 3, 4], or [31] for the Bayesian approach. Such quite elaborate methods, which often require the calibration of some tuning parameters, are not specifically designed for the functional data context. Finally, another related problem is testing the regularity of a signal. See, for instance, [7]. However, such methods, designed to be applied to one signal, do not provide a direct estimator of the regularity.

2.2. Concentration bounds for the local regularity estimator

Below, we focus on the more complicated and realistic case with unknown variance. The case with given variance could be treated after obvious adjustments.

The results in this section depend on μ, the mean number of observation times T, and the cardinality N_0 of the learning set of curves. However, they are non-asymptotic in the sense that they hold true for any sufficiently large μ and N_0 satisfying our conditions. For deriving our results, we impose the following mild assumptions.

(H1) The data consist of the pairs $(Y_m^{(n)}, T_m^{(n)}) \in \mathbb{R} \times I$ defined as in (1), with $I \subset \mathbb{R}$ a compact interval, and the realizations of X, M and T are mutually independent.

(H2) The random variable T admits a density $f : I \to \mathbb{R}$ such that $f(t_0) > 0$. Moreover, there exist $L_f > 0$ and $0 < \beta_f \leq 1$ such that

$$|f(u) - f(v)| \leq L_f |u - v|^{\beta_f}, \qquad \forall u, v \in J_\mu(t_0).$$

(H3) There exist a function $\phi_{t_0}(\cdot, \cdot) > 0$, the constants $L_{t_0}, L_\phi > 0$ and $0 < \beta_\phi \leq 1$ such that, for any $u, v \in J_\mu(t_0)$, we have

$$\mathbb{E}\left[(X_u - X_v)^2\right] = L_{t_0}^2 |u - v|^{2H_{t_0}} \left\{1 + \phi_{t_0}(u, v)\right\} \quad \text{and}$$
$$|\phi_{t_0}(u, v)| \leq L_\phi |u - v|^{\beta_\phi}.$$

(H4) Two constants $\mathfrak{a}, \mathfrak{A} > 0$ exist such that

$$\mathbb{E}\left[|X_u - X_v|^{2p}\right] \leq \frac{p!}{2} \mathfrak{a} \mathfrak{A}^{p-2} |u - v|^{2pH_{t_0}}, \qquad \forall p \geq 2, \ \forall u, v \in J_\mu(t_0).$$

(H5) The variables $\varepsilon_m^{(n)}$, $n, m \geq 1$, are independent copies of a centered variable ε, with finite variance σ^2, for which constants $\mathfrak{b} \geq \sigma^2 > 0$ and $\mathfrak{B} > 0$ exist such that

$$\mathbb{E}(|\varepsilon|^{2p}) \leq \frac{p!}{2} \mathfrak{b} \mathfrak{B}^{p-2}, \qquad \forall p \geq 1.$$

(H6) The random variable M is such that $M \geq 9$ and $\gamma_0 > 0$ exists such that, for any $s > 0$, $\mathbb{P}\left(|M - \mu| > s\right) \leq \exp(-\gamma_0 s)$.

Assumption (H2) imposes a mild condition on the distribution of the random observation points which provides convenient moment bounds for their spacings. In particular, it implies that, for a sufficiently large μ, $f(t_0)/2 \leq f(t) \leq 2f(t_0)$, $\forall t \in J_\mu(t_0)$. Assumption (H3) is a version of the so-called local stationarity condition. More precisely, (H3) implies that the trajectories of the process $X = (X_u : u \in I)$ are Hölder continuous in quadratic mean in the neighborhood of t_0, with exact exponent H_{t_0} and local Hölder constant L_{t_0}. Let us call H_{t_0} the *local regularity of the process X at t_0*. Examples include, but are not limited to, stationary or stationary increment processes X. See, *e.g.*, [2] for some examples and references on processes satisfying the mild condition in (H3). Assumptions (H4) and (H5) are needed for deriving exponential bounds for the concentration of our local regularity estimator, and are satisfied by the sub-Gaussian random variables. In particular, (H4) is satisfied by the Gaussian processes. (H6) is a mild condition for controlling the variability of number of observation points on the curves. The lower bound on M guarantees that each curve in the learning

set has a sufficient number of observation times for building our estimator. For a real number a, let $\lfloor a \rfloor$ denote the largest integer not exceeding a.

Theorem 1. *Let Assumptions (H1)–(H6) hold true. Let K_0 be a positive integer such that*

$$(\mu + 1)^{\frac{\beta_f \alpha}{4 + \beta_f \alpha}} \leq K_0 \leq \frac{\mu}{2 \log(\mu)}, \tag{9}$$

with $\alpha = 2H_{t_0} + \beta_\phi \in (0, 3]$. Let

$$\mathfrak{c} \left(\frac{K_0 - 1}{f(t_0)(\mu + 1)} \right)^{\min(\beta_\phi, \beta_f H_{t_0}/2)} < \epsilon < \frac{2}{\log 2}, \tag{10}$$

with $\mathfrak{c}$ a constant depending only on L_f, β_f, β_ϕ, $f(t_0)$ and H_{t_0}. Define $k = \lfloor (K_0 + 7)/8 \rfloor$ and let $\widehat{H}_{t_0} = \widehat{H}_{t_0}(k)$ be defined as in (8). Then, for a sufficiently large μ :

$$\mathbb{P} \left(\left| \widehat{H}_{t_0} - H_{t_0} \right| > \epsilon \right) \leq 12 \exp \left[-\mathfrak{f} N_0 \epsilon^2 \left(\frac{k - 1}{f(t_0)(\mu + 1)} \right)^{4H_{t_0}} \right],$$

where $\mathfrak{f}$ is a positive constant depending on $\mathfrak{a}, \mathfrak{A}, \mathfrak{b}, \mathfrak{B}$ and the length of the interval I.

To obtain a non-trivial estimator of H_{t_0}, we need $k \geq 2$, thus the upper bound in (9) should be larger than 9, and this happens as soon as $\mu \geq 80$. For the estimator in (7), which requires an estimate of σ^2, we would only need $\mu > 35$. The exact expressions of the constants $\mathfrak{c}$ and $\mathfrak{f}$ could be traced in the proof of Theorem 1. The condition imposed on K_0 provides a panel of choices depending on N_0 and μ. As a result, up to some constants, and depending on $L_f, \beta_f, \beta_\phi, f(t_0)$ and H_{t_0}, the concentration rate ϵ, one could expect, could be in a range such that $\epsilon \mu \gg 1$ and $\epsilon \log^{1/2}(\mu) \ll 1$. The best possible concentration of $\widehat{H}_{t_0}$ is guaranteed as soon as N_0 is larger than some power of μ, while for a concentration as fast as some negative power of $\log(\mu)$, one only needs a small number N_0 of curves in the learning set, that is larger than some power of $\log(\mu)$.

For the purpose of building an adaptive optimal kernel estimator for the trajectories of X, we will impose $\epsilon = \log^{-2}(\mu)$ and an exponential bound equal to $\exp(-\mu)$. The following corollary proposes a data-driven choice of K_0 which guarantees these requirements. This choice is guided by the fact that, for any constants $a, b > 0$, we have the relationship $\log^a(\mu) \leq \exp((\log \log(\mu))^2) \leq \mu^b$, provided μ is sufficiently large.

Corollary 1. *Assume the conditions of Theorem 1 hold true. Let*

$$\widehat{\mu} = N_0^{-1} \sum_{n=1}^{N_0} M_n, \qquad \widehat{K}_0 = \lfloor \widehat{\mu} \exp(-(\log \log(\widehat{\mu}))^2) \rfloor,$$

and $\widehat{H}_{t_0} = \widehat{H}_{t_0}(\lfloor (\widehat{K}_0 + 7)/8 \rfloor)$, with $\widehat{H}_{t_0}$ defined in (8). Then, for any constant $C > 0$,

$$\mathbb{P} \left(\left| \widehat{H}_{t_0} - H_{t_0} \right| > C \log^{-2}(\mu) \right) \leq \exp(-\mu),$$

provided $N_0 \geq \mu^{1+b}$ for some $b > 0$ and μ is sufficiently large.

The conditions in Corollary 1 impose μ to be large, but still allow for many cases in the three regimes non-dense, dense and ultra-dense, as defined by [41].

One could also build $\widehat{H}_{t_0}$ with only one trajectory of a process X with stationary increments. If the density of T is uniform and sufficiently many measurements are available, it suffices to split the interval $[0, 1]$ into N_0 intervals of the same length and apply our methodology considering the measuring times and the noisy measured values in each block as belonging to a different curve in the learning set. Theorem 1 and Corollary 1 remain valid.

2.3. The case of conditionally heteroscedastic noise

In some applications, the assumption of constant variance for the error term ε could be unrealistic. Therefore, we consider the following conditional heteroscedastic error extension of model (1):

$$Y_m^{(n)} = X^{(n)}(T_m^{(n)}) + \sigma\left(X^{(n)}(T_m^{(n)}), T_m^{(n)}\right) u_m^{(n)}, \ 1 \leq n \leq N, 1 \leq m \leq M_n, \ (11)$$

where $\sigma(\cdot, \cdot)$ is some unknown function and $u_m^{(n)}$ are independent copies of a centered variable u with unit variance.

Our approach also applies to the model (11) under some additional mild conditions. Indeed, assuming the expectations exist, we have

$$\begin{aligned}
\theta_k = \mathbb{E}_{\mathcal{B}}\left[(Y_{(2k-1)} - Y_{(k)})^2\right] = \mathbb{E}_{\mathcal{B}}\left[(X_{T_{(2k-1)}} - X_{T_{(k)}})^2\right] \\
+ \mathbb{E}_{\mathcal{B}}\left[\sigma^2\left(X_{T_{(2k-1)}}, T_{(2k-1)}\right)\right] \\
+ \mathbb{E}_{\mathcal{B}}\left[\sigma^2\left(X_{T_{(k)}}, T_{(k)}\right)\right].
\end{aligned}$$

From this identity it is clear that the arguments presented in Section 2.1 remain valid as long as the value of the last two expectations on the right-hand side of the last display does not depend on k. Thus, in this case, even if the conditional variance of $\varepsilon_m^{(n)}$ is not given, we could consider the same estimator $\widehat{H}_{t_0}$. This remark leads us to the following additional assumption.

(E1) The variables $u_m^{(n)}$ from model (11) satisfy the Assumption (H5) with unit variance. Moreover, the function $\sigma(\cdot, \cdot)$ is bounded and the map $u \mapsto \mathbb{E}\left[\sigma^2(X_u, u)\right]$, $u \in I$, is constant in a fixed neighborhood of t_0.

Assumption (E1) allows the error term to be conditionally heteroscedastic, but imposes marginal (unconditional) homoscedasticity in a neighborhood of t_0.

Under Assumption (E1), for any k we have

$$\begin{aligned}
\mathbb{E}_{\mathcal{B}}\left[\sigma^2(X_{T_{(k)}}, T_{(k)})\right] &= \mathbb{E}\left[\mathbb{E}\left(\sigma^2(X_{T_{(k)}}, T_{(k)}) \mid M, T_1, T_2, \ldots, T_M\right) \mathbf{1}_{\mathcal{B}}\right] \\
&= \mathbb{E}\left[\sigma^2(X_u, u)\right] \mathbb{P}(\mathcal{B}),
\end{aligned}$$

and thus the terms like $\mathbb{E}_{\mathcal{B}}[\sigma^2(X_{T_{(k)}}, T_{(k)})]$ cancel when considering the differences $\theta_{4k-3} - \theta_{2k-1}$ and $\theta_{2k-1} - \theta_k$.

Corollary 2. *Assume the observations consist of the pairs* $(Y_m^{(n)}, T_m^{(n)}) \in \mathbb{R} \times I$ *where* $Y_m^{(n)}$ *defined as in* (11) *and the realizations of* X, M *and* T *are mutually independent. Assume that Assumptions (H2)–(H4), (H6), (E1) hold. Then Corollary 1 remains valid with the same local regularity estimator* $\widehat{H}_{t_0}$.

The proof of Corollary 2 follows from the proof of Theorem 1 after obvious modifications, and hence will be omitted. It is worthwhile noting that, even if the regularity H_{t_0} is the same at any point t_0, one may not be able to estimate the regularity H_{t_0} using only one observed noisy trajectory with conditionally heteroscedastic noise. This because, intuitively, it might be impossible to identify the oscillations of the signal of interest, that is to separate the increments of the trajectory of X from the differences of the error terms with variable variance. With our approach based on local observed increments averaged over several curves, the effect of the noise vanishes, provided the expectation of the conditional variance is constant. Hence, eventually the identification of the oscillations of X is recovered and there is no difference with respect to the case of homoscedastic errors.

2.4. The case of differentiable sample paths

The definition of the local regularity extends to the case of differentiable curves. When the curve admits derivatives up the order $\mathbf{d} > 0$, condition (2) has to be stated for the $\mathbf{d}$-th order derivative of the curve. To build an estimate of the local regularity of the $\mathbf{d}$-th derivative of the curve, we propose to use a smoothing-based approximation of the $\mathbf{d}$-th derivative. In the Appendix, we derive concentration bounds for the estimator of $\mathbf{d} + H_{t_0}$ when $\mathbf{d} > 0$. In particular, we propose an estimator of $\mathbf{d}$. The implementation of the estimator of $\mathbf{d} + H_{t_0}$ is described in Section 4.1, and the simulation results we report shows that it performs well. However, many real data we analyzed, revealed that in many applications, the sample paths do not seem differentiable. For this reason, and to save space, we leave to the supplement, the details of learning the smoothness in the general case $\mathbf{d} \geq 0$.

3. Adaptive optimal smoothing

With at hand an estimate of the local regularity $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$ obtained from a learning set of N_0 curves, we aim at recovering N_1 new noisy trajectories of X from what we call the online dataset. One of the most popular smoother is the local polynomial estimator, see [11]. This estimator depends on a tuning parameter, the bandwidth, which should ideally be chosen according to the regularity of the target function. Using the local regularity estimate, one can build optimal smoothing using alternative approaches, such as the splines. Here we focus on local polynomials.

One has to connect a definition of local regularity that is meaningful from the theory of stochastic processes to the usual definition of function regularity used

in nonparametric curve estimation. Fortunately, in our framework, the parameter ς_{t_0}, which is understood as the local regularity of the process $(X_t : t \in J_\mu(t_0))$ in *quadratic mean*, see (2), is intrinsically linked with the regularity of the sample paths of the process. Indeed, in many important situations, which are covered by our assumptions, the regularity of the sample paths of a process does not depend on the realization of this process. For example, the regularity of any Brownian path is $1/2$, in the sense that for any $\epsilon > 0$, almost surely the sample path belongs to the Hölder space $\mathcal{C}^{1/2-\epsilon}(I)$ and does not belong to $\mathcal{C}^{1/2+\epsilon}(J)$ whatever $J \subset I$. Here, for any $a > 0$, $\mathcal{C}^a(I)$ denotes the space of uniformly $a-$Hölder continuous functions defined on I, see Theorem 2.2 and Corollary 2.6 of [33] for precise definition. More generally the regularity of the sample paths of a process is linked to integrated regularities through the Kolmogorov's Continuity Theorem [33, Theorem 2.1]. In particular, Assumption (H3) ensures that, with probability 1, the trajectories of the process $(X_t^{(\mathbf{d})} : t \in J_\mu(t_0))$ are Hölder continuous with any exponent parameter $0 < a < H_{t_0}$.

Below, we define the local polynomial estimator and derive its theoretical properties. Since our focus of interest is the *simultaneous* denoising of the additional N_1 curves, we consider the following pointwise risk: for a generic estimator $\widehat{X}_{t_0}^{[n_1]}$ of $X_{t_0}^{[n_1]}$, let

$$\mathcal{R}(\widehat{X}; t_0) = \mathbb{E}\left[\max_{1 \le n_1 \le N_1} \left| \widehat{X}_{t_0}^{[n_1]} - X_{t_0}^{[n_1]} \right|^2 \right]. \tag{12}$$

First, we provide a sharp bound for this risk with $N_1 = 1$, in the case where a suitable estimator of $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$, computed from another independent sample, is given. Such a result, of interest in itself in nonparametric curve estimation, seems to be new. In this case, the expectation defining the risk $\mathcal{R}(\widehat{X}; t_0)$ should be understood as the conditional expectation given the estimator of ς_{t_0}. Next, we provide a sharp bound for $\mathcal{R}(\widehat{X}; t_0)$ in the case where $N_1 \ge 1$ and the estimator of ς_{t_0} is obtained using the approach introduced in Section 2.

3.1. *Local polynomial estimation*

We assume that $\mathbf{d} \ge 0$ is an integer and $H_{t_0} \in (0, 1)$. Let $\hat{\mathbf{d}}$ and $\widehat{H}_{t_0}$ be some generic estimators of $\mathbf{d}$ and H_{t_0}, respectively, and let $\widehat{\varsigma}_{t_0} = \hat{\mathbf{d}} + \widehat{H}_{t_0}$ be the corresponding estimator of $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$. We assume that $\hat{\mathbf{d}}$ and $\widehat{H}_{t_0}$ are independent of the N_1 from the online dataset, generated according to (1).

The estimator of ς_{t_0} could be used to smooth any curve $Y^{[n_1]}$ $(n_1 = 1, \ldots, N_1)$ from the online dataset. For the sake of readability, we omit the superscript $[n_1]$ and we consider a generic curve from the online dataset:

$$Y_m = X(T_m) + \varepsilon_m, \qquad 1 \le m \le M.$$

For any $u \in \mathbb{R}$, we consider the vector $U(u) = (1, u, \ldots, u^{\hat{\mathbf{d}}}/\hat{\mathbf{d}}!)$. Let $K : \mathbb{R} \to \mathbb{R}$

be a positive kernel and define:

$$\vartheta_{M,h} = \arg\min_{\vartheta \in \mathbb{R}^{\mathbf{d}+1}} \sum_{m=1}^{M} \left\{ Y_m - \vartheta^\top U\left(\frac{T_m - t_0}{h}\right) \right\}^2 K\left(\frac{T_m - t_0}{h}\right),$$

where h is the bandwidth. The vector $\vartheta_{M,h}$ satisfies the normal equations $A\vartheta_{M,h} = a$ with

$$A = A_{M,h} = \frac{1}{Mh} \sum_{m=1}^{M} U\left(\frac{T_m - t_0}{h}\right) U^\top\left(\frac{T_m - t_0}{h}\right) K\left(\frac{T_m - t_0}{h}\right) \tag{13}$$

$$a = a_{M,h} = \frac{1}{Mh} \sum_{m=1}^{M} Y_m U\left(\frac{T_m - t_0}{h}\right) K\left(\frac{T_m - t_0}{h}\right). \tag{14}$$

Let λ be the smallest eigenvalue of the matrix A and remark that, whenever $\lambda > 0$, we have $\vartheta_{M,h} = A^{-1}a$.

Taking into account the expression of the bandwidth minimizing the pointwise mean squared risk for a regression function defined on I, with derivative of order $\mathbf{d}$ which is Hölder continuous in a neighborhood of t_0, with exact exponent H_{t_0}, we consider a bandwidth

$$\widehat{h} \sim M^{-1/(2\widehat{\varsigma}_{t_0}+1)}.$$

Our focus of interest is on determining a nearly optimal rate of the bandwidth to be used to recover the trajectories of X. For the applications, one could also be interested in a nearly optimal constant, which in general needs to be estimated. In Section 4.1 we propose a simple way to estimate a suitable constant for the applications.

With at hand the bandwidth $\widehat{h}$, we propose the following definition of the local polynomial estimator of X_{t_0} of order $\mathbf{d}$:

$$\hat{X}_{t_0} = \begin{cases} U^\top(0)\widehat{\vartheta} & \text{if } \lambda > \log^{-1}(M) \text{ and } |U^\top(0)\widehat{\vartheta}| \leq \widehat{\tau}^{5/12}(M) \\ \widehat{\tau}^{5/12}(M) & \text{if } \lambda > \log^{-1}(M) \text{ and } |U^\top(0)\widehat{\vartheta}| > \widehat{\tau}^{5/12}(M) , \\ 0 & \text{otherwise,} \end{cases}$$

where $\widehat{\vartheta} = \vartheta_{M,\widehat{h}}$ and, for any $y > 1$,

$$\widehat{\tau}(y) = \frac{1}{\log^2(y)} \left(\frac{y}{\log(y)}\right)^{2\widehat{\varsigma}_{t_0}/(2\widehat{\varsigma}_{t_0}+1)}.$$

The upper trimming with $\widehat{\tau}^{5/12}(M)$ is a technical device used to control the tails of $\hat{X}_{t_0}$. It has practically no influence in applications. For deriving our results on $\hat{X}_{t_0}$, we impose the following mild assumptions.

(LP1) There exist two positive constants, $\mathfrak{a}$ and $\mathfrak{A}$, such that for any $p \geq 1$:

$$\mathbb{E}\left[|X_t|^{2p}\right] \leq \frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2}, \quad \forall t \in [0,1].$$

Moreover, for $\mathbf{d} \geq 0$ in Assumption (H3), if $\nabla^{\mathbf{d}} X$ denotes the $\mathbf{d}$–th derivative of the process X,

$$\mathbb{E}\left[|\nabla^{\mathbf{d}} X_u - \nabla^{\mathbf{d}} X_v|^{2p}\right] \leq \left(\frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2}\right)|u-v|^{2pH_{t_0}}, \ \forall p \geq 2, \ \forall u, v \in J_\mu(t_0).$$

(LP2) We assume that, almost surely, $\mu/\log(\mu) \leq M \leq \mu\log(\mu)$.

(LP3) The estimator $\widehat{H}_{t_0}$ satisfies the property

$$\mathbb{P}\left(\left|\widehat{H}_{t_0} - H_{t_0}\right| > \log^{-2}(\mu)\right) \leq \mathfrak{K}_1 \exp(-\mu), \qquad \forall \mu > 0,$$

where $\mathfrak{K}_1$ is some positive constant.

(LP4) The estimator $\hat{\mathbf{d}}$ satisfies the property $\mathbb{P}(\hat{\mathbf{d}} \neq \mathbf{d}) \leq \mathfrak{K}_1 \exp(-\mu)$, $\forall \mu > 0$.

The first part of Assumption (LP1) provides a suitably tight control on the moments of X_t, but still allows for unbounded trajectories. The second part of Assumption (LP1), is a technical condition which reinforces Assumption (H4). It allows to control the analytic regularity of the sample paths. More precisely, it is implicitly used in the definition of the variable Λ_β in (33). Assumption (LP2) is a convenient, but mild, technical condition. It could be relaxed at the price of controlling the probability of the complement of the event $\{\mu/\log(\mu) \leq M \leq \mu\log(\mu)\}$, for instance using (H6). Assumptions (LP3) and (LP4) are very mild conditions that the generic estimators of the regularity should satisfy. Since $\mu^{1/\log^2(\mu)} = e^{1/\log(\mu)}$ for any $\mu > 1$, the concentration of $\widehat{H}_{t_0}$ at a suitable negative power of $\log(\mu)$ will suffice for the smoothing purposes. For simplicity, and without loss of generality we consider the same constant $\mathfrak{K}_1$ in Assumptions (LP3) and (LP4).

Theorem 2. *Assume that Assumptions (H1), (H2), (H5) and (H6) and Assumptions (LP1)–(LP4) hold true and let $K(\cdot)$ be a kernel such that, for any $t \in \mathbb{R}$:*

$$\kappa^{-1}\mathbf{1}_{[-\delta,\delta]}(t) \leq K(t) \leq \kappa\mathbf{1}_{[-1,1]}(t), \quad for \ some \ 0 < \delta < 1 \ and \ \kappa \geq 1. \tag{15}$$

There then exists a constant Γ_0 such that for any $\mu \geq 1$,

$$\mathbb{E}\left[\exp\left\{\left(\tau(\mu)\left|\hat{X}_{t_0} - X_{t_0}\right|^2\right)^{1/4}\right\}\right] \leq \Gamma_0$$

where

$$\tau(\mu) = \frac{1}{\log^2(\mu)}\left(\frac{\mu}{\log(\mu)}\right)^{\frac{2\varsigma_{t_0}}{2\varsigma_{t_0}+1}}.$$

The bound on the $\exp(\sqrt{x})$–moment of the $|\hat{X}_{t_0} - X_{t_0}|$ seems a new result for local polynomial estimators. For our purposes, it will entail a sharp bound for $\mathcal{R}(\widehat{X}; t_0)$. More precisely, the price for considering a risk measure uniformly over the whole online dataset is very low, that is a multiplying factor as large as

a power of $\log(N_1)$ in the risk bound we derive below. [13] derived sharp bounds for all the moments of $|\hat{X}_{t_0} - X_{t_0}|$. However, his bounds on the moments would induce a power of N_1 as multiplying factor for our risk bound, instead of the power of $\log(N_1)$.

Theorem 3. *Assume that assumptions of Theorem 2 hold true, and let K be a kernel which satisfies* (15). *There then exists a positive constant Γ_1 such that*

$$\mathcal{R}(\hat{X}; t_0) = \mathbb{E}\left[\max_{1 \leq n_1 \leq N_1} \left| \hat{X}_{t_0}^{[n_1]} - X_{t_0}^{[n_1]} \right|^2 \right]$$

$$\leq \Gamma_1 \log^2(\mu) \log^4(1 + N_1) \{\log(\mu)\}^{\frac{2\varsigma_{t_0}}{2\varsigma_{t_0}+1}} \mu^{-\frac{2\varsigma_{t_0}}{2\varsigma_{t_0}+1}}.$$

If all the trajectories X were in $\mathcal{C}^{\varsigma_{t_0}}(J_\mu(t_0))$, ς_{t_0} were known and $N_1 = 1$, the risk bound for $\mathcal{R}(\hat{X}; t_0)$ would be of the usual nonparametric rate $\mu^{-2\varsigma_{t_0}/(2\varsigma_{t_0}+1)}$. Let us note that the fact that H_{t_0} is not known does not have any consequence on the risk bound in Theorem 3. Indeed, since $\mu^{1/\log^2 \mu} = e^{1/\log \mu}$ for any μ, the order of the risk bound does not change as soon as the probability of the event $\{\hat{\mathbf{d}} = \mathbf{d}\} \cap \{|\hat{H}_{t_0} - H_{t_0}| \leq 1/\log^2 \mu\}$ tends to 1. The $\log(1 + N_1)$ factor is given by the maximum over the N_1 curves in the online dataset. The factor $\{\log(\mu)\}^{2\varsigma_{t_0}/(2\varsigma_{t_0}+1)}$ is due to the concentration properties of M around its mean μ. This factor would not appear if M/μ is almost surely bounded and bounded away from zero. The factor $\log^2(\mu)$ comes from probability theory. The trajectories of a stochastic process X with local regularity H_{t_0} does not necessarily belong to $\mathcal{C}^{\varsigma_{t_0}}(J_\mu(t_0))$ but they are almost surely in any $\mathcal{C}^{\varsigma_{t_0}-\epsilon}(J_\mu(t_0))$ for any $0 < \epsilon < \varsigma_{t_0}$.

Finally, let us notice that Corollary 1 states that the estimator defined by (8) satisfies (LP3) for $\mathbf{d} = 0$ and any $0 < H_{t_0} < 1$. This leads us to the following result.

Corollary 3. *Assume $\mathbf{d} = 0$ and let $\hat{H}_{t_0}$ be the estimator of $0 < H_{t_0} < 1$ defined in Corollary 1. Moreover, Assumptions (H1)–(H6) and Assumptions (LP1)–(LP2) hold true. If $N_0 \geq \mu^{1+b}$ and $N_1 \leq \mu^B$ for some $b, B > 0$, then*

$$\mathcal{R}(\hat{X}; t_0) \leq \Gamma_1 B^4 \log^7(\mu) \mu^{-\frac{2H_{t_0}}{2H_{t_0}+1}}.$$

Finally, we establish the uniform convergence of the recovered trajectories. Below, the uniformity is with respect to $t_0 \in I$ and over all the curves in the online sample.

Theorem 4. *Assume that assumptions of Theorem 2 hold true, uniformly, for any $t_0 \in I$. Let $\varsigma = \varsigma_{t_0}$ be the global regularity of the process. Let also K be a kernel which satisfies* (15). *There then exists a positive constant Γ_1 such that*

$$\mathbb{E}\left[\max_{1 \leq n_1 \leq N_1} \sup_{t_0 \in I} \left| \hat{X}_{t_0}^{[n_1]} - X_{t_0}^{[n_1]} \right|^2 \right] \leq \Gamma_1 \Psi(\mu, N_1)$$

where

$$\Psi(\mu, N_1) = \frac{\log^4(N_1) + \log^8(\mu)}{\log^2(\mu)} \left(\frac{\log(\mu)}{\mu} \right)^{\frac{2\varsigma}{2\varsigma+1}}. \tag{16}$$

4. Empirical analysis

In the usual local polynomial (LP) smoothing framework, for a regression function defined on I, given a sample of size M, a bound of the pointwise, mean squared error risk is first derived and this bound is then minimized with respect to the bandwidth h. See for instance [37]. When the regression function admits a derivative of order $\mathbf{d}$ which is Hölder continuous in a neighborhood of t_0, with exact exponent H_{t_0} and local Hölder constant L_{t_0}, the optimal bandwidth is

$$h_{opt} = \left(\frac{C}{M}\right)^{1/(2\varsigma_{t_0}+1)} \quad \text{with} \quad C = C_{t_0} = \frac{q_2}{2\varsigma_{t_0} q_1^2}, \tag{17}$$

with q_1 and q_2 defined on pages 39-40 of [37]. With the Nadaraya-Watson estimator, we take $q_1 = L_{t_0}\{\lfloor \varsigma_{t_0}\rfloor!\}^{-1} \int K(v)|v|^{\varsigma_{t_0}}\,dv$ and $q_2 = \sigma_{t_0}^2 f(t_0)^{-1} \int K^2(v)dv$, where $\sigma_{t_0}^2$ is the variance of the noise that could depend on t_0. This yields a refined constant C in this case. Details are given in the Appendix. The value $f(t_0)$ can be estimated using all the $T_m^{(n)}$.

Thus our target bandwidth h_{opt} depends on two more unknown quantities, L_{t_0} and $\sigma_{t_0}^2$, for which we now propose estimation procedures. The estimation of L_{t_0} could be based on similar ideas as used for H_{t_0}. For simplicity, we assume $\mathbf{d} = 0$. The extension to the case $\mathbf{d} \geq 1$ could follow the same pattern as for the estimation of the local regularity, using the trajectories of the derivatives. Using twice the relationship (3) with k and $2k-1$, respectively, we deduce

$$L_{t_0}^2 \approx \frac{\theta_{2k-1} - \theta_k}{4^{H_{t_0}} - 1} \left(\frac{f(t_0)(\mu+1)}{k-1}\right)^{2H_{t_0}}.$$

On the other hand, using the approximation of the moments of the spacings, as given in Lemma 2, we have

$$\eta_{2k-1} - \eta_k := \mathbb{E}_{\mathcal{B}}\left[|T_{(4k-3)} - T_{(2k-1)}|^{2H_{t_0}}\right] - \mathbb{E}_{\mathcal{B}}\left[|T_{(2k-1)} - T_{(k)}|^{2H_{t_0}}\right]$$

$$\approx (4^{H_{t_0}} - 1)\left(\frac{k-1}{f(t_0)(\mu+1)}\right)^{2H_{t_0}}.$$

Given an estimator of H_{t_0}, the empirical counterparts of η_k obtained from the learning set of N_0 independent trajectories of X is

$$\widehat{\eta}_k = \frac{1}{N_0} \sum_{n=1}^{N_0} \left|T_{(2k-1)}^{(n)} - T_{(k)}^{(n)}\right|^{2\widehat{H}_{t_0}} \mathbf{1}_{\mathcal{B}_n},$$

where $\mathcal{B}_n$ is the sequence of events defined in (5). An estimate of η_{2k-1} could be obtained similarly. These facts lead us to the following estimator of the local Hölder constant L_{t_0} :

$$\widehat{L}_{t_0}^2 = \widehat{L}_{t_0}^2(\widehat{H}_{t_0}) = \begin{cases} \dfrac{\widehat{\theta}_{2k-1} - \widehat{\theta}_k}{\widehat{\eta}_{2k-1} - \widehat{\eta}_k} & \text{if } \widehat{\eta}_{2k-1} > \widehat{\eta}_k \text{ and } \widehat{\theta}_{2k-1} > \widehat{\theta}_k, \\ 1 & \text{otherwise.} \end{cases} \tag{18}$$

For the implementation, we propose $k = \lfloor (\widehat{K}_0 + 7)/8 \rfloor$ with

$$\widehat{K}_0 = \lfloor \widehat{\mu} \exp\left(-(\log\log\widehat{\mu})^2\right) \rfloor \quad \text{and} \quad \widehat{\mu} = N_0^{-1} \sum_{n=1}^{N_0} M_n.$$

To estimate the variance, we propose

$$\widehat{\sigma}^2 = \widehat{\sigma}_{t_0}^2 = \frac{1}{N_0} \sum_{n=1}^{N_0} \frac{1}{2|\mathcal{S}_n|} \sum_{m \in \mathcal{S}_n} \left[Y_{(m)}^{(n)} - Y_{(m-1)}^{(n)} \right]^2, \tag{19}$$

where $\mathcal{S}_n \subset \{2, 3, \ldots, M_n\}$ is a set of indices for the $n-$th trajectory and $|\mathcal{S}_n|$ is the cardinal of $\mathcal{S}_n$. When the variance of the error ε is considered constant, one could take $\mathcal{S}_n = \{2, 3, \ldots, M_n\}$. When it depends on t_0, one could take

$$\mathcal{S}_n = \left\{ m : T_{(1)}^{(n)} \leq T_{(m)}^{(n)} \leq T_{(\widehat{K}_0)}^{(n)} \right\},$$

with $\widehat{K}_0$ defined above. This is the choice we used in our empirical investigation. When the variance of the errors also depends on the realizations X_u, as described in Section 2.3, in general it is no longer possible to consistently estimate $\sigma^2(X_{t_0}, t_0)$. Our simulation experiments indicate that the estimate (19) remains a reasonable choice.

Finally, the constant involved in the definition of the bandwidth could be estimated by $\widehat{C}$ obtained by plugging the estimates of the unknown quantities into the definition of C in (17). Concerning the kernel, we use $K(t) = (3/4)\left(1 - t^2\right)\mathbf{1}_{[-1,1]}(t)$, that is the Epanechnikov kernel for which $\|K\|^2 = 3/5$ and $\int |K(v)||v|^{\varsigma_{t_0}} dv = 3\{(\varsigma_{t_0} + 1)(\varsigma_{t_0} + 3)\}^{-1}$.

4.1. Simulation experiments

We now illustrate the behavior of our local regularity estimator $\widehat{\varsigma}_{t_0} = \hat{\mathbf{d}} + \widehat{H}_{t_0}$ computed using the *learning* set of noisy curves, and the performance of kernel smoother it induces for estimating the noisy curves from the *online* set. The procedure for calculating $\widehat{\varsigma}_{t_0}$ is summarized in the following algorithm where $LP(d)$ means local polynomial smoother with degree $d \geq 0$. The Nadaraya-Watson smoother corresponds to $LP(0)$.

For the curve estimation, we use $(Y_1^{[n_1]}, T_1^{[n_1]}), \ldots, (Y_{M_{n_1}}^{[n_1]}, T_{M_{n_1}}^{[n_1]})$, $1 \leq n_1 \leq N_1$, and $LP(\hat{\mathbf{d}})$ with $\hat{\mathbf{d}}$ delivered by Algorithm 1. The bandwidth is calculated as $\widehat{h}_{n_1} = (\widehat{C}/M_{n_1})^{1/(2\widehat{\varsigma}_{t_0}+1)}$, $1 \leq n_1 \leq N_1$, with $\widehat{\varsigma}_{t_0}$ obtained from Algorithm 1. The constant estimate $\widehat{C}$ is the same for all curves in the *online* set, that is that obtained with $\widehat{\varsigma}_{t_0}$, $\widehat{L}_{t_0}(\widehat{H}_{t_0})$ and $\widehat{\sigma}_{t_0}^2$. We compare our approach with the classical *cross-validation (CV)* (least-squares leave-one-out) method applied for each curve $X^{[n_1]}$ separately. For CV, we use the **R** package np [19], after rescaling the CV bandwidth to account for their different definition of the Epanechnikov kernel. At this stage, we want to point out that our smoothing method is much

---

Algorithm 1: Estimation of the local regularity $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$

Result: Estimation of ς_{t_0} from the learning set of N_0 noisy curves

Calculate $\widehat{\mu} = N_0{}^{-1} \sum_{n=1}^{N_0} M_n$ and $\widehat{K}_0 = \lfloor \widehat{\mu} \exp(-(\log\log(\widehat{\mu}))^2) \rfloor$;

Calculate $\widehat{H}_{t_0}$ and set $\hat{\mathbf{d}} = 0$;

while $\widehat{H}_{t_0} > 1 - \log^{-2}(\widehat{\mu})$ **do**

    Calculate $\widehat{L}_{t_0}(1)$, as in (18), and $\widehat{\sigma}_{t_0}^2$;

    Calculate $\widehat{C}$ with $\widehat{\varsigma}_{t_0} = \hat{\mathbf{d}} + \widehat{H}_{t_0}$, $\widehat{L}_{t_0}(1)$ and $\widehat{\sigma}_{t_0}^2$;

    Calculate the bandwidth $\widehat{h}_n = (\widehat{C}/M_n)^{1/(2\widehat{\varsigma}_{t_0}+1)}$, $1 \le n \le N_0$;

    Estimate the $(\hat{\mathbf{d}}+1)$−th derivative of the trajectories of X with $LP(\hat{\mathbf{d}}+1)$;

    Calculate $\widehat{H}_{t_0}$ using the estimated trajectories of the $(\hat{\mathbf{d}}+1)$−th derivative;

    Set $\hat{\mathbf{d}} = \hat{\mathbf{d}} + 1$;

end

---

faster than any standard, trajectory-by-trajectory approach, such as CV. We report a time comparison in the Appendix, and as expected, the ratio between the times needed for CV and for our approach is at least of the same order as N_1. It is worth noting that one cannot follow an *ad-hoc* approach and transfer one CV bandwidth from a curve $X^{[n_1]}$ to another because H_{t_0} is not known, and could even vary with t_0.

The data are generated from the model (1) using different settings for X, the distribution of T and the variance of the noise, as well as for N_0 and N_1. For X, we consider three types of Gaussian processes: fractional Brownian motion (fBm) with constant Hurst parameter $H \in (0,1)$, fBm with piecewise constant Hurst parameter, and integrated fBm. In the later case, $X_t = \int_0^t W_H(s)ds$, where W_H denotes a fBm with constant Hurst parameter H. The local regularity is constant for the first and the third type, and variable for the second. The third type is an example of X with smooth trajectories. We identify the setting for X by $s \in \{1,2,3\}$. A more detailed description of these processes, as well as plots of their trajectories, are provided in the Appendix. The number M of measuring times of a curve is a Poisson random variable with expectation μ, while for the measuring times T, we considered either a uniform distribution (identified by `unif`), or a deterministic equispaced grid (`equi`) on the range $[0,1]$. For the noise, we considered the Gaussian distribution with both constant and variable variance. The cases are identified by σ^2 which could be a number or a list, respectively. The values of σ^2 are chosen in such a way that the variance ratio signal-to-noise remains almost unchanged. Thus, one simulation setting is defined by the 7-tuple $(s, N_0, N_1, \mu, f, H, \sigma^2)$, with $f \in \{\texttt{unif}, \texttt{equi}\}$ and H, the Hurst parameter, is a list in the case of fBm with piecewise constant local regularity. Below, we present the results for a few settings, complementary results are reported in the Appendix. For each type of experiment, the reported results are obtained from 500 replications of the experiment. Our methodology is implemented in the **R** package `denoisr` available at https://github.com/StevenGolovkine/denoisr.

Figure 1 presents the results for the local regularity estimation for piecewise fbM with homoscedastic noise. The local estimations of H_{t_0} are performed at
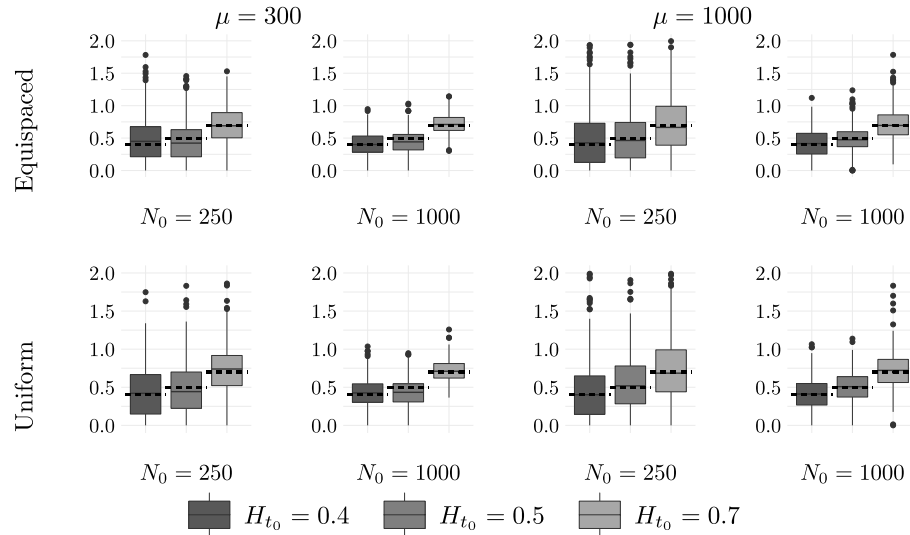
FIG 1. *Estimation of the local regularity for piecewise fBm, with constant noise variance* $\sigma^2 = 0.05$, *at* $t_0 = 1/6, 1/2$ *and* 5/6. *True values:* $\varsigma_{t_0} = H_{t_0}$ *equal to* 0.4, 0.5 *and* 0.7, *respectively.*
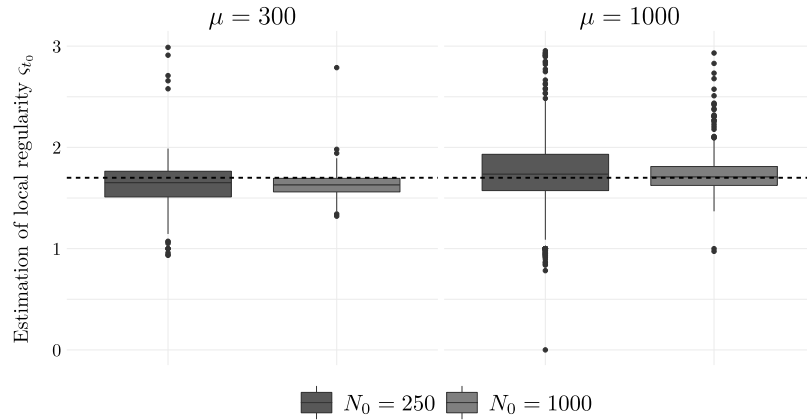

FIG 2. *Estimation of the local regularity for integrated fBm, with constant noise variance* $\sigma^2 = 0.005$, *at* $t_0 = 0.5$. *True value:* $\varsigma_{t_0} = 1.7$.

$t_0 = 1/6, 1/2$ and 5/6 which correspond to the middle of the interval for each regularity. The true values of H_{t_0} are 0.4, 0.5 and 0.7, respectively. The results show a quite accurate estimator $\widehat{H}_{t_0}$ and confirm the theoretical result on its concentration. Increasing either μ or N_0 improves the concentration. The results for unif and equi are quite similar. Figure 2 presents the estimation of
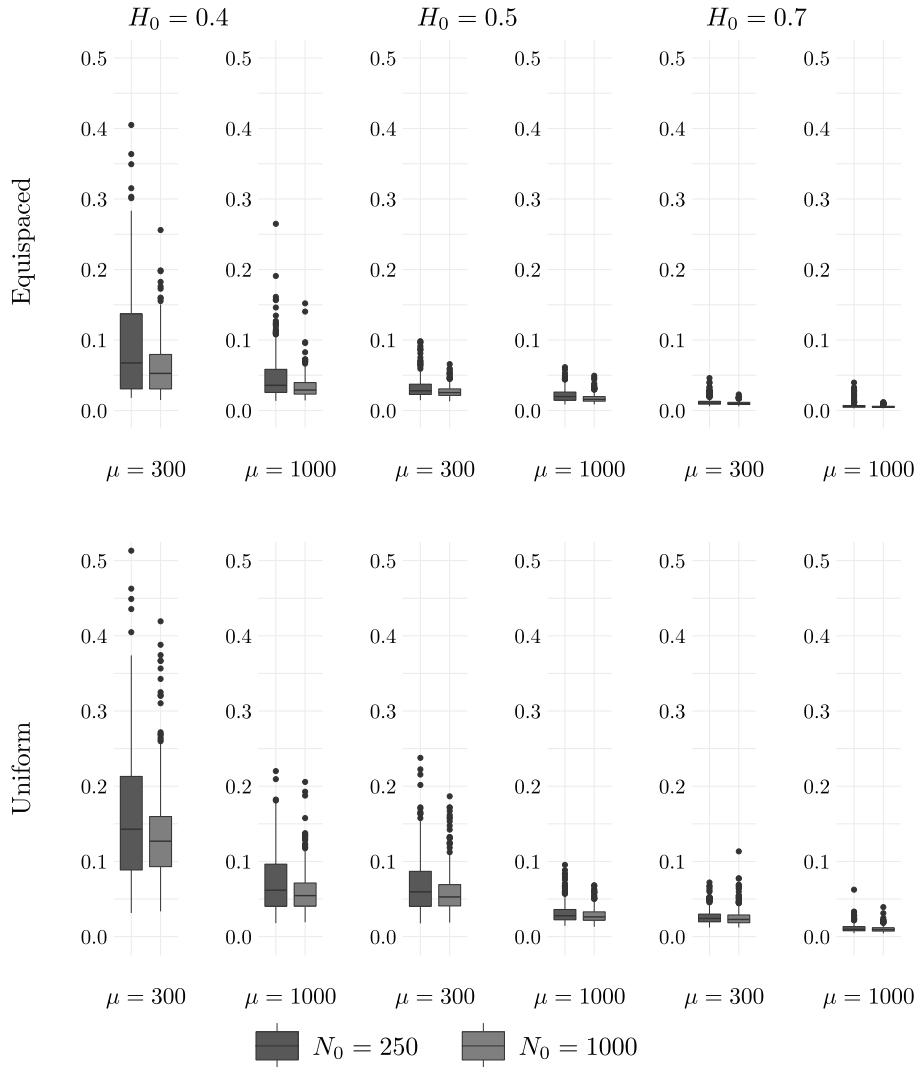
FIG 3. *Estimation of the risks* $\mathcal{R}(\widehat{X}; 1/6)$, $\mathcal{R}(\widehat{X}; 0.5)$ *and* $\mathcal{R}(\widehat{X}; 5/6)$ *for piecewise fBm, with constant noise variance* $\sigma^2 = 0.05$.

ς_{t_0} for different settings $(3, N_0, 500, \mu, \mathtt{equi}, 1.7, 0.005)$. As expected, our local regularity estimation approach also performs well for smooth trajectories.

Next, we present the results on the risk $\mathcal{R}(\widehat{X}; t_0)$. Figure 3 presents the box-plots of the risk $\mathcal{R}(\widehat{X}; t_0)$ defined in (12) in the case of piecewise constant local regularity, with three values of t_0, each one in the middle of the interval of the changes of regularity are defined. The results are quite good. Part of the curves with lower regularity are harder to estimate and thus results in higher risks than the more regular parts. It appears that N_0 and μ do not have the same influence
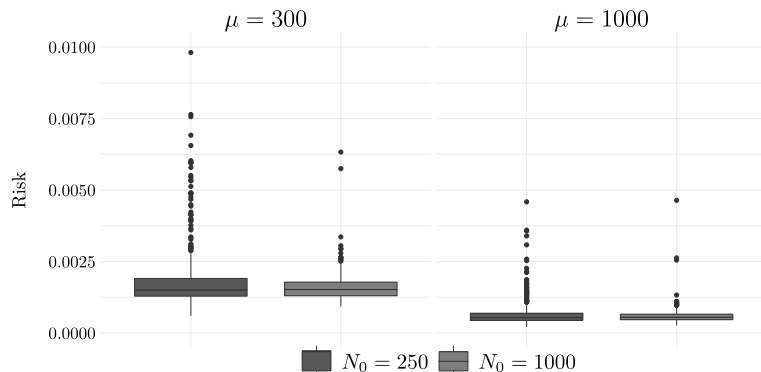
FIG 4. *Estimation of the risk* $\mathcal{R}(\widehat{X}; 0.5)$ *for smoothing the noisy trajectories of an integrated fBm, with constant noise variance* $\sigma^2 = 0.005$.
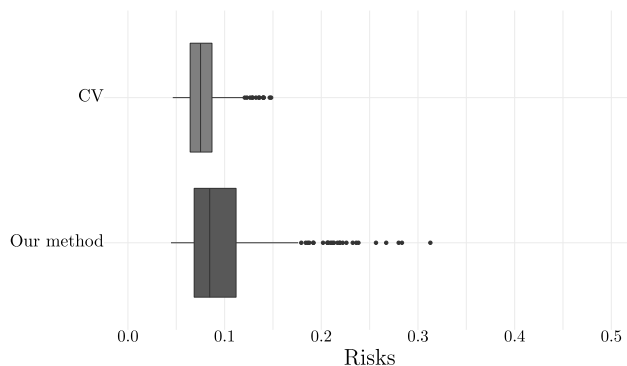

FIG 5. *CV versus our method: comparing the pointwise risk* $\mathcal{R}(\widehat{X}; 0.5)$ *for smoothing the noisy trajectories of a fBm; simulation* $(1, 1000, 500, 300, equi, 0.5, 0.05)$.

on the risk as the estimation of the local regularity, and this is in line with the risk bound in Theorem 3. Thus, going from 300 to 1000 sampling points leads to large improvement in terms of risk whereas going from 250 to 1000 curves in the *learning* dataset only results in little or no improvement. Finally, it seems that the method achieves better results for equispaced sampling points.

The same conclusions could be drawn from the results presented in Figure 4, obtained for the simulation experiment defined by the 7-tuple $(3, 1000, 500, 1000, \texttt{equi}, 1.7, 0.005)$.

Finally, we present a comparison with the CV. Because of the large amount of computing resources required by CV, we only considered a few cases. Figure 5 presents the results in terms of the risk calculated at $t_0 = 0.5$ for the setting $(1, 1000, 500, 300, \texttt{equi}, 0.5, 0.05)$. We make the remark that our method and CV perform similarly despite the fact that CV uses a specifically tailored bandwidth for each curve in the *online* set. The homoscedastic setting is favorable to CV
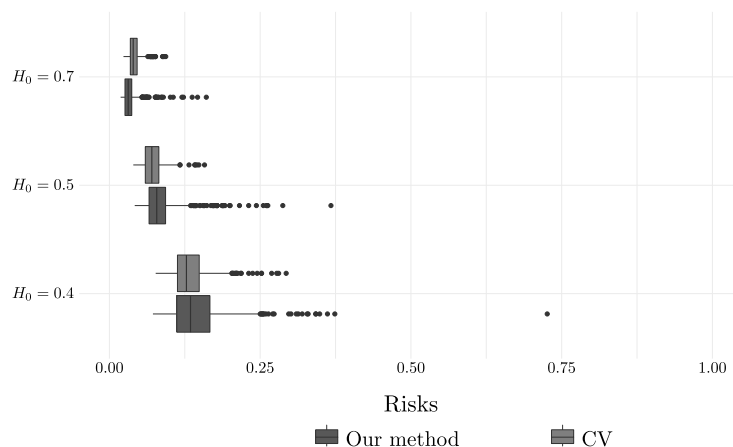
FIG 6. *CV versus our method: comparing* $\mathcal{R}(\widehat{X};1/6)$, $\mathcal{R}(\widehat{X};0.5)$ *and* $\mathcal{R}(\widehat{X};5/6)$ *for smoothing the noisy trajectories of a piecewise fBm; simulation* $(2,1000,500,1000, equi, (0.4,0.5,0.7),0.05)$.
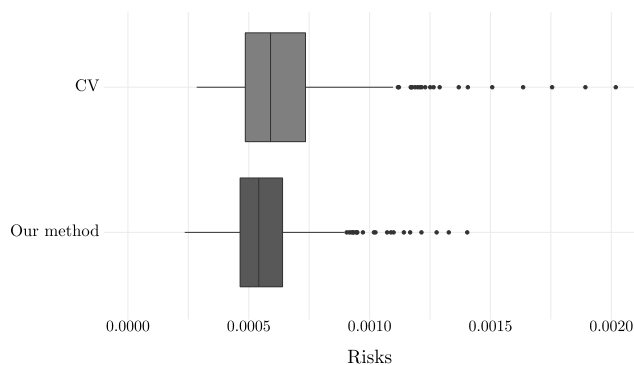

FIG 7. *CV versus our method: comparing the pointwise risk* $\mathcal{R}(\widehat{X};0.5)$ *for smoothing the noisy trajectories of an integrated fBm; simulation* $(3,1000,500,1000, equi, 1.7, 0.005)$.

which, for a given curve, uses a global bandwidth at any t_0. Figure 6 presents the the heteroscedastic setting $(2,1000,500,1000, equi, (0.4,0.5,0.7),0.05)$. CV preserves good performances when the local regularity varies moderately. Our method shows close performance in this case, slightly better when $H_{t_0} = 0.7$. Figure 7 presents the results in the setting $(3,1000,500,1000, \texttt{equi}, 1.7, 0.005)$. Again, CV and our method perform quite similarly.

Let us end our simulation experiments presentation with the results on the median curve estimation from noisy curves. We consider a Gaussian simulation design given by Model 4 in [25], page 726. The mean function is $g(t) = 3\sin(3\pi(t+0.5)) + 2t^3$ and the covariance function is $\gamma(s,t) = \exp(-|t-s|^{0.8})$. The curves are measured on a common equidistant grid with $M \in \{300, 1000\}$
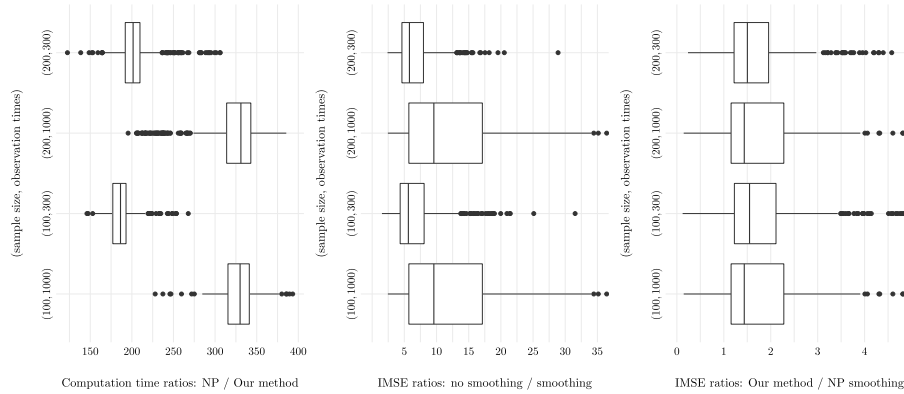
FIG 8. *Median curve estimation from samples of $N \in \{100, 200\}$ noisy curves generated from Model 4 of [25], observed at $M \in \{300, 1000\}$ points. The noise variance is $\sigma^2 = 1$. Plotted ratios: computation times with CV and our method (left); IMSE for the median curve without smoothing and with smoothing using method (middle); IMSE with our method and CV (right).*

points and the samples of size $N \in \{100, 200\}$ have a proportion of 25% contaminated observations. We used the contamination by peaks with the parameter values used by [25] in their Figure 6. To the simulation design used in [25], we add a Gaussian measurement error with constant variance σ^2. We compute the median curve using the mode depth with the noiseless sample (the benchmark median) and with the noisy samples, without smoothing. Next we estimate the median curve using the smoothed curves. For smoothing, we use, on the one hand, $LP(\hat{\mathbf{d}})$ with $\hat{\mathbf{d}}$ and the bandwidth delivered by Algorithm 1, and, on the other hand, the **R** package np with the bandwidth selected by CV for each curve separately. We compare the mean integrated squared error (IMSE) of the smoothing-based median curves with respect to the benchmark. We also compare the IMSE obtained by our approach with that obtained with the median curves computed from the noisy samples directly. Finally, we compare the computing times for our method and that using CV. The results obtained from 500 replications, with $\sigma^2 = 1$, are reported in Figure 8. The accuracy of the estimates, obtained in a small fraction of a second using our smoothing method, is slightly lower than that obtained by CV, which requires far more computation time. Ignoring the noise in the sample leads to very large errors for the median curve estimation.

4.2. Real data analysis: the NGSIM Study

In this section, our method is applied to data from the Next Generation Simulation (NGSIM) study, which aims to "describe the interactions of multimodal travelers, vehicles and highway systems", see [18]. This study is known to be one

of the largest publicly available source of naturalistic driving data. This dataset is widely used in traffic flow studies from the interpretation of traffic phenomena such as congestion to the validation of models for trajectories prediction (see *e.g.* [10, 22, 40, 26, 20] for some recent references). However, such data have been proved to be subject to measurement errors revealed by physical inconsistency between the space traveled, velocity and acceleration of the vehicles, *cf.* [28]. Montanino and Punzo [27] developed a trajectory-by-trajectory four-steps method to recover the signals from the noisy curves, and their methodology is now considered as a benchmark in the traffic flow engineering community for analyzing NGSIM data. The steps, finely tuned for the NGSIM data, are : 1. removing the outliers; 2. cutting off the high- and medium-frequency responses in the speed profile; 3. removing the residual nonphysical acceleration values, preserving the consistency requirements; 4. cutting off the high- and medium-frequency responses generated from step 3. The detailed description of these steps is provided in the Appendix.

To compare our smoothed curves to those of [27], we consider the following ratio:

$$r(\widehat{X}, \widetilde{X}) = \frac{\sum_{m=1}^{M_n} \left[Y_m^{(n)} - \widehat{X}(T_m^{(n)}) \right]^2}{\sum_{m=1}^{M_n} \left[Y_m^{(n)} - \widetilde{X}(T_m^{(n)}) \right]^2}, \quad 1 \leq n \leq 1714,$$

where $\widehat{X}$ denotes our curve estimation while $\widetilde{X}$ is that obtained by [27]. A value of the ratio $r(\widehat{X}, \widetilde{X})$ less than 1 indicates smoothed values closer to the observations.

For our illustration, we consider a subset of the NGSIM dataset, known as the I-80 dataset. It contains 45 minutes of trajectories for vehicles on the Interstate 80 Freeway in Emeryville, California, segmented into three 15-minute periods (from 4:00 p.m. to 4:15 p.m.; from 5:00 p.m. to 5:15 p.m. and from 5:15 p.m. to 5:30 p.m.) on April 13, 2005 and corresponds to different traffic conditions (congested, transition between uncongested and congested and fully congested). In total, the dataset contains trajectories, velocities and accelerations for $N_0 = 1714$ individual vehicles that passed through this highway during this period, recorded every 0.1s. The number M_n of measurements for each curve varies from 165 to 946. We focus on the velocity variable and rescale the measurement times for each of the 1714 velocity curves such that the first velocity measurement corresponds to $t = 0$ and the last one to $t = 1$. Figure 9 presents a sample of five curves from this data. It can easily be noticed that the velocities are quite erratic and their variation is not physically realistic, indicating the presence of a noise. Moreover, the data have been recorded at a moment of the day when traffic is evolving, it goes from fluid to dense traffic. Therefore, we consider that there are three groups in the data: a first group corresponding to a fluid (high-speed) traffic, a second one for in-between fluid and dense traffic, and a third groups corresponding to the dense (low-speed) traffic. To determine the three clusters, we fit a finite Gaussian mixture model to the vector of number of sampling points. The model is estimated by an EM algorithm initialized by hierarchical model-based agglomerative clustering as proposed by Fraley and
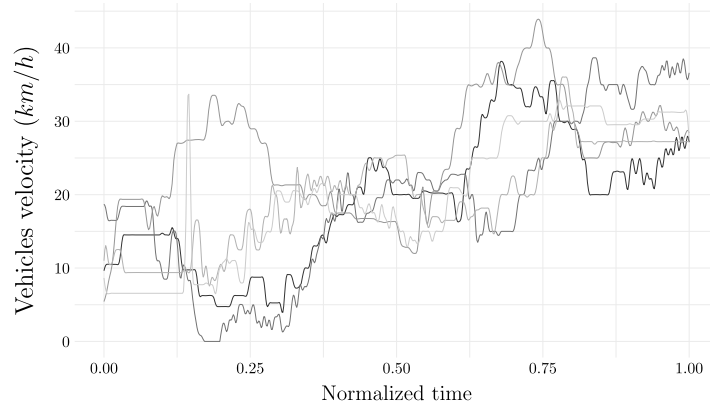
FIG 9. *I-80 dataset illustration: a sample of five velocity curves.*
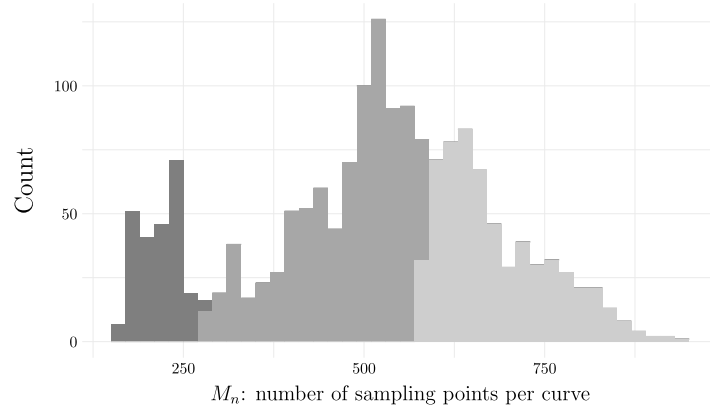
FIG 10. *I-80 dataset clusters: density of sampling points for fluid (darkest gray), in-between fluid and dense, and dense traffic (lightest gray).*

Raftery [12] and implemented in the **R** package mclust [34]. The optimal model is then selected according to BIC. The three resulting classes have 239, 869 and 606 velocity trajectories, respectively. Plots of randomly selected subsamples of trajectories from each groups are provided in the Appendix. The respective numbers of measures M_n are plotted in Figure 10. The mean estimates $\widehat{\mu}$ obtained in the three groups are $218, 474$ and 684, respectively, and the corresponding values $\widehat{K}_0$, as defined as in Corollary 1, are $13, 17$ and 20.

Figure 11 presents the results of the estimation of ς_{t_0} for values of t_0 from 0.2 to 0.8, for each group. The evolution of ς_{t_0} is quite smooth, except for Group 1 (Figure 11b). A possible explanation could be the small number of curves and the average of M_n in this group, which correspond to low values of N_0 and $\widehat{\mu}$. We also provide the estimation of the regularity using the whole sample of
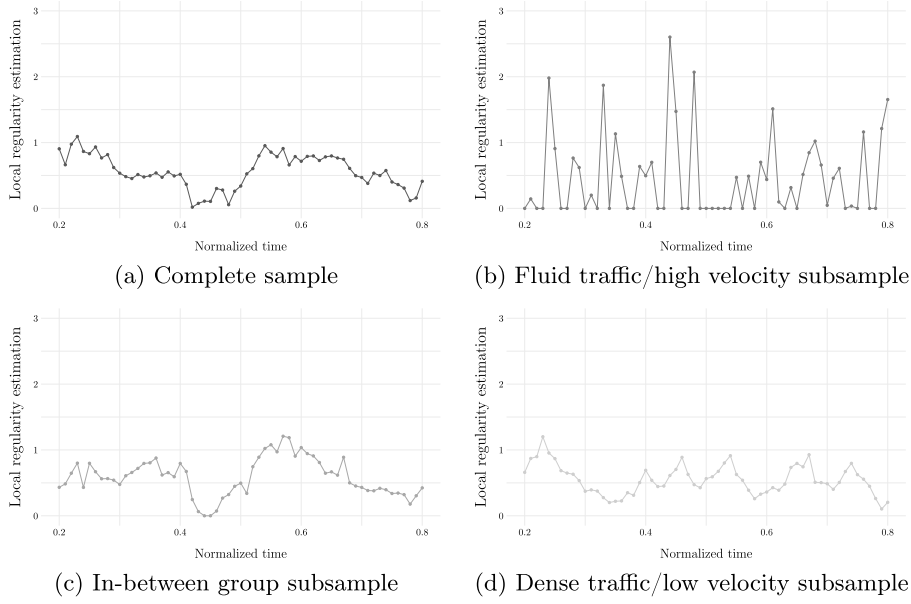
FIG 11. *Estimation of the local regularity of the velocity curves for different t_0.*

size 1714. The differences we notice between the estimates of ς_{t_0} from different groups support our preliminary clustering step.

To compute the curve estimate we adopt a leave-one-curve-out procedure: each curve is smoothed using the local regularity estimates computed from the other curves in the group (or the other 1713 curves when the data is not split into groups). The densities of the resulting ratios $r(\widehat{X}, \widetilde{X})$ are plotted in Figure 12. When the traffic is fluid and the speed is high (group 1), our method perform much better than that of Montanino and Punzo. When the traffic is dense with low speed (group 3), the smoothed values obtained with the two methods are more similar, though our method still exhibits better performance for the majority of the curves.

Appendix A: Proof of Theorem 1

The proof of Theorem 1 is based on several lemmas that we present in the following. For these lemmas, we implicitly assume that the conditions of Theorem 1 are satisfied.

Lemma 1. *Let r be an integer such that*

$$(\mu + 1)^{\frac{\beta_f (2H_{t_0} + \beta_\phi)}{4 + \beta_f (2H_{t_0} + \beta_\phi)}} \leq 8r \leq K_0.$$

Let $\mathfrak{s} \in \{1, 2, 4, 8\}$ and let $1 \leq k, l \leq K_0$ be such that $l - k = \mathfrak{s}r$. Then, for

(a) Complete sample

(b) Fluid traffic/high velocity subsample

(c) In-between group subsample

(d) Dense traffic/low velocity subsample
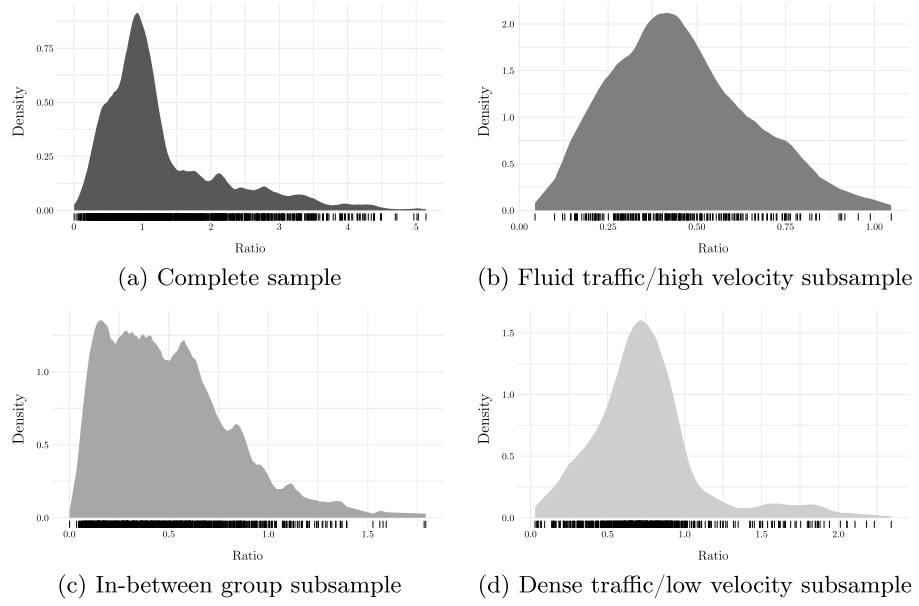
FIG 12. *Densities of the ratio $r(\widehat{X}, \widetilde{X})$ within different groups.*

sufficiently large μ, we have

$$
\left| \mathbb{E}_{\mathcal{B}}\left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \right] - L_{t_0}^2 \left(\frac{l - k}{f(t_0)(\mu + 1)} \right)^{2H_{t_0}} \right|
$$
$$
\leq \mathfrak{c}\left(\frac{l - k}{f(t_0)(\mu + 1)} \right)^{2H_{t_0} + \min(\beta_\phi, \beta_f H_{t_0}/2)},
$$

where $\mathfrak{c} = \max(2L_\phi, \mathfrak{c}_1)$ and $\mathfrak{c}_1$ is a constant depending on H_{t_0}, β_ϕ, L_f, β_f and $f(t_0)$.

Proof of Lemma 1. Note that, by the definition of $\mathbb{E}_{\mathcal{B}}$, elementary properties of the conditional expectation, and Assumption (H3),

$$
\begin{aligned}
\mathbb{E}_{\mathcal{B}}\left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \right] &= \mathbb{E}\left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \mathbf{1}_{\mathcal{B}} \right] \\
&= \mathbb{E}\left\{ \mathbb{E}\left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \mathbf{1}_{\mathcal{B}} \middle| M, T_1, \ldots, T_M \right] \right\} \\
&= \mathbb{E}\left\{ \mathbb{E}\left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \middle| M, T_1, \ldots, T_M \right] \mathbf{1}_{\mathcal{B}} \right\} \\
&= \mathbb{E}_{\mathcal{B}}\left\{ L_{t_0}^2 |T_{(l)} - T_{(k)}|^{2H_{t_0}} \left[1 + \phi_{t_0}(T_{(k)}, T_{(l)}) \right] \right\} \\
&=: (I) + (II),
\end{aligned}
$$

where $(I) = L_{t_0}^2 \mathbb{E}_{\mathcal{B}}\left\{ |T_{(l)} - T_{(k)}|^{2H_{t_0}} \right\}$.

By Lemma 2 applied with $\alpha = 2H_{t_0} \leq 2$,

$$(I) = L_{t_0}^2 \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{2H_{t_0}} (1 + R_1) \tag{20}$$

with

$$|R_1| \leq \mathfrak{c}_1(2H_{t_0}) \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{\beta_f H_{t_0}/2}.$$

On the other hand, Assumption (H3) implies that

$$|(II)| \leq L_{t_0}^2 L_\phi \mathbb{E}_\mathcal{B} \left(|T_{(l)} - T_{(k)}|^{2H_{t_0}+\beta_\phi} \right),$$

and using again Lemma 2 with $\alpha = 2H_{t_0} + \beta_\phi \leq 3$ we obtain

$$|(II)| \leq 2L_{t_0}^2 L_\phi \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{2H_{t_0}+\beta_\phi}, \tag{21}$$

for μ large enough such that $\mathfrak{c}_1(2H_{t_0} + \beta_\phi) \leq 1$. Then, from (20) and (21) we obtain

$$\mathbb{E}_\mathcal{B} \left[\left| X_{T_{(l)}} - X_{T_{(k)}} \right|^2 \right] = L_{t_0}^2 \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{2H_{t_0}} (1 + R),$$

where R is a remainder term such that, for sufficiently large μ,

$$|R| \leq \max \left\{ 2L_\phi \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{\beta_\phi}, \mathfrak{c}_1(2H_{t_0}) \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{\beta_f H_{t_0}/2} \right\}$$

$$\leq \mathfrak{c} \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{\min(\beta_\phi, \beta_f H_{t_0}/2)},$$

with $\mathfrak{c} = \max(2L_\phi, \mathfrak{c}_1(2H_{t_0}))$ with $\mathfrak{c}_1(\cdot)$ defined in Lemma 2. $\qquad \square$

For the sake of readability, we state below a technical lemma on the moments of the spacings $T_{(l)} - T_{(k)}$, for which the proof is given in the Appendix E. In Lemma 2, we consider that μ is sufficiently large to ensure $(\mu+1)^{\beta_f \alpha/(4+\beta_f \alpha)} + 1 \leq \mu/\{2\log(\mu)\}$.

Lemma 2. *Let $0 < \alpha \leq 3$ be a fixed parameter and let r be an integer such that*

$$(\mu+1)^{\frac{\beta_f \alpha}{4+\beta_f \alpha}} \leq 8r \leq K_0 \quad with \quad K_0 \leq \frac{\mu}{2\log(\mu)}.$$

Let $\mathfrak{s} \in \{1, 2, 4, 8\}$ and let $1 \leq k, l \leq K_0$ be such that $l - k = \mathfrak{s}r$. Then, for sufficiently large μ,

$$\left| \mathbb{E}_\mathcal{B} \left[|T_{(l)} - T_{(k)}|^\alpha \right] - \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^\alpha \right| \leq \mathfrak{c}_1 \left(\frac{l-k}{f(t_0)(\mu+1)} \right)^{\alpha(1+\beta_f/4)},$$

with $\mathfrak{c}_1 = \mathfrak{c}_1(\alpha) = 8\mathfrak{c}_0\{2f(t_0)\}^{\beta_f \alpha/4}$ and $\mathfrak{c}_0$ a constant depending on α, L_f, β_f and $f(t_0)$.

Lemma 2 is a generic result that will be applied with two values $\alpha \in (0,3]$, that are $\alpha = 2H_{t_0}$ and $\alpha = 2H_{t_0} + \beta_\phi$.

Lemma 3. *Let k be a positive integer such that $2k - 1 \leq K_0$. Then for any $\eta > 0$,*

$$q_k(\eta) := \max\left\{\mathbb{P}(\hat{\theta}_k - \theta_k \geq \eta), \ \mathbb{P}(\hat{\theta}_k - \theta_k \leq -\eta)\right\} \leq \exp\left(-\mathfrak{e}N_0\eta^2\right),$$

where, using the notations introduced in Assumptions (H4) and (H5),

$$\mathfrak{e} = 1/(2\mathfrak{d} + 2\mathfrak{D}) \quad with \quad \mathfrak{d} = 27\left(\mathfrak{a} + 2\mathfrak{b}\right) \quad and \quad \mathfrak{D} = 9\max(\mathfrak{A}, \mathfrak{B}).$$

Proof of Lemma 3 . By the definition in (5) and (6),

$$\hat{\theta}_k = \frac{1}{N_0}\sum_{n=1}^{N_0} Z_n \qquad \text{where} \qquad Z_n = \left[Y_{(2k-1)}^{(n)} - Y_{(k)}^{(n)}\right]^2 \mathbf{1}_{\mathcal{B}_n},$$

and $\mathcal{B}_n = \left\{M_n \geq K_0, T_{(1)}^{(n)} \in J_\mu(t_0), \ldots, T_{(K_0)}^{(n)} \in J_\mu(t_0)\right\}$. Note that $\mathbb{E}(\hat{\theta}_k) = \theta_k$. Moreover, for any $p \geq 2$, using Assumptions (H4) and (H5), we have

$$\begin{aligned}
\mathbb{E}\left(|Z_n|^p\right) &= \mathbb{E}_{\mathcal{B}}\left(|Y_{(2k-1)} - Y_{(k)}|^{2p}\right) \\
&\leq 3^{2p-1}\mathbb{E}_{\mathcal{B}}\left(|X_{T_{(2k-1)}} - X_{T_{(k)}}|^{2p} + |\varepsilon_{(2k-1)}|^{2p} + |\varepsilon_{(k)}|^{2p}\right) \\
&\leq 3^{2p-1}\frac{p!}{2}\left(\mathfrak{a}\mathfrak{A}^{p-2} + 2\mathfrak{b}\mathfrak{B}^{p-2}\right) \\
&\leq \frac{p!}{2}\mathfrak{d}\mathfrak{D}^{p-2},
\end{aligned}$$

where $\mathfrak{d}$ and $\mathfrak{D}$ are defined in the statement of this lemma. Bernstein's inequality implies

$$\mathbb{P}(\hat{\theta}_k - \theta_k \geq \eta) \leq \exp\left(-\frac{N_0\eta^2}{2\mathfrak{d} + 2\mathfrak{D}\eta}\right) \leq \exp\left(-\mathfrak{e}N_0\eta^2\right),$$

and the same bound is valid for $\mathbb{P}(\hat{\theta}_k - \theta_k \leq -\eta)$. The bound for $q_k(\eta)$ follows. $\quad\square$

Lemma 4. *Let $2 \leq k < l \leq (K_0 + 1)/2$ be two positive integers. For any $\eta > 0$, define*

$$p_{k,l}^+(\eta) = \mathbb{P}(\hat{\theta}_l - \hat{\theta}_k \geq (1+\eta)(\theta_l - \theta_k)) \quad and \quad p_{k,l}^-(\eta) = \mathbb{P}(\hat{\theta}_l - \hat{\theta}_k \leq (1-\eta)(\theta_l - \theta_k)).$$

Then, for sufficiently large μ,

$$\max\left\{p_{k,l}^+(\eta), p_{k,l}^-(\eta)\right\} \leq 2\exp\left[-\frac{\mathfrak{e}}{16}N_0\eta^2\left(\frac{l-k}{\mu+1}\right)^{4H_{t_0}}\right],$$

with $\mathfrak{e}$ defined in Lemma 3.

Proof of Lemma 4. Assume that k and l satisfy the assumptions stated in the Lemma and assume moreover that μ is large enough so that $\eta(\theta_l - \theta_k)/2 < 1$. Then

$$
\begin{aligned}
p_{k,l}^+(\eta) &= \mathbb{P}\big[(\hat{\theta}_l - \theta_l) - (\hat{\theta}_k - \theta_k) \geq \eta(\theta_l - \theta_k)\big] \\
&\leq \mathbb{P}\big[\hat{\theta}_l - \theta_l \geq \eta(\theta_l - \theta_k)/2\big] + \mathbb{P}\big[\hat{\theta}_k - \theta_k \leq -\eta(\theta_l - \theta_k)/2\big] \\
&\leq q_l\left(\eta(\theta_l - \theta_k)/2\right) + q_k\left(\eta(\theta_l - \theta_k)/2\right),
\end{aligned}
$$

and the same bound is valid for $p_{k,l}^-(\eta)$. By (22), we have

$$
\theta_l - \theta_k \geq \{(l - k)/(\mu + 1)\}^{2H_{t_0}}/2,
$$

provided μ is sufficiently large. We obtain the bound for $\max(p_{k,l}^+(\eta), p_{k,l}^-(\eta))$ after applying Lemma 3. $\qquad\square$

Proof of Theorem 1. Let $\epsilon > 0$. With the notation from (4) and (8), we can write

$$
\mathbb{P}\left(\big|\widehat{H}_{t_0}(k) - H_{t_0}\big| > \varepsilon\right) \leq \mathbf{1}_{\left\{|H_{t_0}(k) - H_{t_0}| > \epsilon/2\right\}} + \mathbb{P}\left(\big|\widehat{H}(k) - H_{t_0}(k)\big| > \varepsilon/2\right)
$$
$$
=: B + V,
$$

and thus it suffices to bound the terms B and V.

The term B. The study of the set in the indicator function boils down to the study of the convergence of $H_{t_0}(k)$ to H_{t_0}. Using Lemma 1 with $l - k = k - 1 = r$ we have

$$
\theta_k - 2\sigma^2 = \mathbb{E}_{\mathcal{B}}\left[\left(X_{T_{(2k-1)}} - X_{T_{(k)}}\right)^2\right] = L_{t_0}^2\left(\frac{k - 1}{f(t_0)(\mu + 1)}\right)^{2H_{t_0}}(1 + \rho_k),
$$

and

$$
|\rho_k| \leq \mathfrak{c}\left(\frac{k - 1}{f(t_0)(\mu + 1)}\right)^{\min(\beta_\phi, \beta_f H_{t_0}/2)} =: \rho_k^*,
$$

with $\mathfrak{c}$ a constant defined in Lemma 1. Using again Lemma 1 with $k = 2k - 1$, $l = 4k - 3$ and $\mathfrak{a} = 2$ and taking the difference, we deduce that there exists R_k such that

$$
\begin{aligned}
\theta_{2k-1} - \theta_k &= L_{t_0}^2\left(\frac{2(k - 1)}{f(t_0)(\mu + 1)}\right)^{2H_{t_0}}(1 + \rho_{2k-1}) \\
&\quad - L_{t_0}^2\left(\frac{k - 1}{f(t_0)(\mu + 1)}\right)^{2H_{t_0}}(1 + \rho_k) \\
&= (4^{H_{t_0}} - 1)L_{t_0}^2\left(\frac{k - 1}{f(t_0)(\mu + 1)}\right)^{2H_{t_0}}(1 + R_k), \qquad (22)
\end{aligned}
$$

where

$$
|R_k| = \left|\frac{4^{H_{t_0}}\rho_{2k-1} - \rho_k}{4^{H_{t_0}} - 1}\right| =\leq \frac{4^{H_{t_0}} + 1}{4^{H_{t_0}} - 1}\rho_{2k-1}^* \leq \frac{4^{H_{t_0}} + 1}{4^{H_{t_0}} - 1}\rho_{K_0}^*.
$$

Similarly, we obtain :

$$\theta_{4k-3} - \theta_{2k-1} = (4^{H_{t_0}} - 1)L_{t_0}^2 \left(\frac{2(k-1)}{f(t_0)(\mu+1)} \right)^{2H_{t_0}} (1 + R_{2k-1}). \qquad (23)$$

Combining (22) and (23), we obtain

$$\log(\theta_{4k-3} - \theta_{2k-1}) - \log(\theta_{2k-1} - \theta_k) = H_{t_0} \log 4 + \log(1 + R_{2k-1}) - \log(1 + R_k),$$

which leads, using the definition of $H_{t_0}(k)$ given by (4), to:

$$H_{t_0}(k) = H_{t_0} + \eta_k \qquad \text{where} \qquad \eta_k = \frac{\log(1 + R_{2k-1}) - \log(1 + R_k)}{2\log 2}.$$

Note that, for sufficiently large μ, both R_k and R_{2k-1} are greater that $-1/2$. This implies

$$|\eta_k| \leq \frac{|R_{2k-1} - R_k|}{\log 2} \leq \left(\frac{2}{\log 2} \frac{4^{H_{t_0}} + 1}{4^{H_{t_0}} - 1} \right) \rho_{4k-3}^*.$$

Thus, since $\rho_{4k-3}^* \leq \rho_{K_0}^*$, the condition $\left| H_{t_0}(k) - H_{t_0} \right| > \varepsilon/2$ fails and $B = 0$ as soon as

$$\epsilon > \left(\frac{4}{\log 2} \frac{4^{H_{t_0}} + 1}{4^{H_{t_0}} - 1} \right) \rho_{K_0}^*,$$

that is as soon as condition (10) is satisfied, provided μ is sufficiently large.

The term V. Defining the event $\mathcal{D} = \{\hat{\theta}_{4k-3} > \hat{\theta}_{2k-1} > \hat{\theta}_k\}$, we can write

$$\mathbb{P}\left(\left| \widehat{H}(k) - H_{t_0}(k) \right| > \epsilon/2 \right) \leq \mathbb{P}\left(\left| \widehat{H}(k) - H_{t_0}(k) \right| > \epsilon/2, \mathcal{D} \right) + \mathbb{P}(\overline{\mathcal{D}}). \qquad (24)$$

First note that using Lemma 4 we have, for sufficiently large μ :

$$\mathbb{P}(\overline{\mathcal{D}}) \leq \mathbb{P}(\hat{\theta}_k \geq \hat{\theta}_{2k-1}) + \mathbb{P}(\hat{\theta}_{2k-1} \geq \hat{\theta}_{4k-3})$$

$$\leq p_{2k-1,k}^-(1) + p_{4k-3,2k-1}^-(1) \leq 4\exp\left[-\frac{\mathfrak{e}}{16}N_0 \left(\frac{k-1}{\mu+1} \right)^{4H_{t_0}} \right]. \qquad (25)$$

Now, it remains to bound the quantity

$$\wp = \mathbb{P}\left(\left| \widehat{H}(k) - H_{t_0}(k) \right| > \epsilon/2, \mathcal{D} \right)$$

$$= \mathbb{P}\left[\left| \log\left(\frac{\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}}{\theta_{4k-3} - \theta_{2k-1}} \times \frac{\theta_{2k-1} - \theta_k}{\hat{\theta}_{2k-1} - \hat{\theta}_k} \right) \right| > \epsilon \log 2, \mathcal{D} \right].$$

Since both $\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}$ and $\hat{\theta}_{2k-1} - \hat{\theta}_k$ are positive under $\mathcal{D}$, we have

$$\wp \leq \mathbb{P}\left[\frac{\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}}{\theta_{4k-3} - \theta_{2k-1}} \times \frac{\theta_{2k-1} - \theta_k}{\hat{\theta}_{2k-1} - \hat{\theta}_k} > 2^\epsilon, \mathcal{D} \right]$$

$$+ \mathbb{P}\left[\frac{\hat{\theta}_{4k-3} - \hat{\theta}_{2k-1}}{\theta_{4k-3} - \theta_{2k-1}} \times \frac{\theta_{2k-1} - \theta_k}{\hat{\theta}_{2k-1} - \hat{\theta}_k} < 2^{-\epsilon}, \mathcal{D} \right]$$

$$\leq \mathbb{P}\left[\frac{\hat\theta_{4k-3}-\hat\theta_{2k-1}}{\theta_{4k-3}-\theta_{2k-1}}>2^{\frac{\epsilon}{2}}\right]+\mathbb{P}\left[\frac{\hat\theta_{2k-1}-\hat\theta_k}{\theta_{2k-1}-\theta_k}<2^{-\frac{\epsilon}{2}}\right]$$

$$+\mathbb{P}\left[\frac{\hat\theta_{4k-3}-\hat\theta_{2k-1}}{\theta_{4k-3}-\theta_{2k-1}}<2^{-\frac{\epsilon}{2}}\right]+\mathbb{P}\left[\frac{\hat\theta_{2k-1}-\hat\theta_k}{\theta_{2k-1}-\theta_k}>2^{\frac{\epsilon}{2}}\right].$$

Applying Lemma 4, we obtain :

$$\wp \leq p^+_{4k-3,2k-1}(2^{\frac{\epsilon}{2}}-1)+p^-_{4k-3,2k-1}(1-2^{-\frac{\epsilon}{2}})+p^+_{2k-1,k}(2^{\frac{\epsilon}{2}}-1)$$
$$+p^-_{2k-1,k}(1-2^{-\frac{\epsilon}{2}}).$$

Now remark that

$$p^+_{2k-1,k}(2^{\frac{\epsilon}{2}}-1)\leq 2\exp\left[-\frac{\mathfrak{e}}{16}N_0\left(2^{\frac{\epsilon}{2}}-1\right)^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right]$$

$$\leq 2\exp\left[-\frac{\mathfrak{e}\log^2(2)}{64}N_0\epsilon^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right],$$

and, as soon as $\epsilon<2/\log 2$, we have $1-2^{-\epsilon/2}\leq\epsilon/4$, which implies:

$$p^-_{2k-1,k}(1-2^{-\frac{\epsilon}{2}})\leq 2\exp\left[-\frac{\mathfrak{e}}{16}N_0\left(1-2^{-\frac{\epsilon}{2}}\right)^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right]$$

$$\leq 2\exp\left[-\frac{\mathfrak{e}\log^2(2)}{256}N_0\epsilon^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right].$$

Using similar derivations for the others terms, we obtain:

$$\wp\leq 8\exp\left[-\frac{\mathfrak{e}\log^2(2)}{256}N_0\epsilon^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right]. \tag{26}$$

Combining (24) with (25) and (26), we obtain, for sufficiently large μ and $\epsilon<2/\log 2$:

$$\mathbb{P}\left(\left|\widehat{H}(k)-H_{t_0}(k)\right|>\epsilon/2\right)\leq 12\exp\left[-\mathfrak{f}N_0\epsilon^2\left(\frac{k-1}{\mu+1}\right)^{4H_{t_0}}\right],$$

where $\mathfrak{f}=\mathfrak{e}\log^2(2)/256$. $\qquad\square$

Appendix B: Proofs of Theorems 2 and 3

The proofs of Theorems 2 and 3 are based on the following lemmas for which the proofs are provided in the Appendix E. For the first lemma we consider the

matrix A defined in (13) with the bandwidth $\widehat{h} = M^{-1/(2\widehat{\varsigma}_{t_0}+1)}$. Let λ be the smallest eigenvalue of this matrix. Let

$$\mathbf{A} = f(t_0) \int_{\mathbb{R}} U(u)U^{\top}(u)K(u)du,$$

and let λ_0 denote its smallest eigenvalue. In the following, we assume that $K(\cdot)$ satisfies (15). Then $\mathbf{A}$ is positive definite [see 37, for details] and thus $\lambda_0 > 0$.

Lemma 5. *Let $K(\cdot)$ be a kernel such that, for any $t \in \mathbb{R}$:*

$$\kappa^{-1}\mathbf{1}_{[-\delta,\delta]}(t) \leq K(t) \leq \kappa\mathbf{1}_{[-1,1]}(t), \quad \text{for some } 0 < \delta < 1 \text{ and } \kappa \geq 1.$$

Under Assumptions (LP2), (LP3) and (LP4), the matrix A defined as in (13), with $h = \widehat{h}$, is positive semidefinite. Moreover, there exists a positive constant $\mathfrak{g}$ that depends only on K, $\mathbf{d}$, $f(t_0)$ and λ_0 such that, for M sufficiently large,

$$\mathbb{P}(\lambda \leq \beta|M) \leq 2\exp(-\mathfrak{g}M\widehat{h}), \quad \forall 0 < \beta \leq \lambda_0/2, \tag{27}$$

and, for sufficiently large μ,

$$\sup_{0<\beta\leq\lambda_0/2} \mathbb{P}(\lambda \leq \beta) \leq \mathfrak{K}_2 \exp\left[-\frac{\mathfrak{g}}{2}\tau(\mu)\log^2(\mu)\right] \tag{28}$$

where

$$\tau(\mu) = \frac{1}{\log^2(\mu)}\left(\frac{\mu}{\log(\mu)}\right)^{\frac{2\varsigma_{t_0}}{2\varsigma_{t_0}+1}},$$

with $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$. Here, $\mathfrak{K}_2$ is a universal constant.

Since the dimension of A and $\mathbf{A}$ are given by $\hat{\mathbf{d}}$, the probability $\mathbb{P}(\cdot)$ in Lemma 5 should be understood as the conditional probability given the estimator $\hat{\mathbf{d}}$.

Lemma 6. *Let ξ be a positive random variable such that $c_1 := \mathbb{E}\left[\exp\left(\eta_0\xi^4\right)\right] < \infty$, for some positive constant η_0. Then, for any $\tau \geq 1$:*

$$\mathbb{E}\left[\exp\left(\tau\xi\right)\right] \leq c_1\exp\left(c_2\tau^{4/3}\right) \qquad \text{where} \qquad c_2 = (5/16\eta_0)^{1/3}.$$

Proof of Theorem 2. Without loss of generality, we could suppose that $f(t_0)/2 \leq f(t) \leq 2f(t_0)$, $\forall t \in J_\mu(t_0)$. We define the events

$$\mathcal{E} = \{\lambda > \lambda_0/2\}, \quad \mathcal{F} = \left\{|\widehat{H}_{t_0} - H_{t_0}| \leq \log^{-2}(\mu)\right\} \cap \left\{\hat{\mathbf{d}} = \mathbf{d}\right\},$$

and $\mathcal{G} = \{|X_{t_0}| \leq \tau^{\tilde{\alpha}}(\mu)\}$, with $1/3 < \tilde{\alpha} < \alpha = 5/12$. Next, let $Z = \left|\widehat{X}_{t_0} - X_{t_0}\right|$. Assume that μ is such that $\log^{-1}(\mu/\log(\mu)) < \lambda_0/2$, then using Assumption (H2), we have:

$$\mathbb{E}[\varphi(\tau(\mu)Z^2)] = (A) + (B) + (C) + (D),$$

where $\varphi(x) = \exp(x^{1/4})$ and

$$(A) = \mathbb{E}\left[\varphi\big(\tau(\mu)Z^2\big)\mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\mathbf{1}_{\mathcal{G}}\right]$$
$$(B) = \mathbb{E}\left[\varphi\big(\tau(\mu)Z^2\big)\mathbf{1}_{\overline{\mathcal{E}}}\right] \leq \mathbb{E}^{1/2}\left[\varphi^2\big(\tau(\mu)Z^2\big)\right]\mathbb{P}^{1/2}(\overline{\mathcal{E}})$$
$$(C) = \mathbb{E}\left[\varphi\big(\tau(\mu)Z^2\big)\mathbf{1}_{\overline{\mathcal{F}}}\right] \leq \mathbb{E}^{1/2}\left[\varphi^2\big(\tau(\mu)Z^2\big)\right]\mathbb{P}^{1/2}(\overline{\mathcal{F}})$$
$$(D) = \mathbb{E}\left[\varphi\big(\tau(\mu)Z^2\big)\mathbf{1}_{\overline{\mathcal{G}}}\right] \leq \mathbb{E}^{1/2}\left[\varphi^2\big(\tau(\mu)Z^2\big)\right]\mathbb{P}^{1/2}(\overline{\mathcal{G}}).$$

We show that (A) is the main term, and it is bounded by a constant.

By construction, $\left|\hat{X}_{t_0}\right| \leq \tau^\alpha(M)$ and, by (LP2),

$$\tau^\alpha(M) \geq \tau^\alpha(\mu/\log(\mu)) \geq \tau^{\tilde{\alpha}}(\mu),$$

provided that μ is sufficiently large. Thus, for sufficiently large μ,

$$\left|\hat{X}_{t_0} - X_{t_0}\right|\mathbf{1}_{\mathcal{G}} \leq \left|U^\top(0)\hat{\vartheta} - X_{t_0}\right|\mathbf{1}_{\mathcal{G}} \leq \left|U^\top(0)\hat{\vartheta} - X_{t_0}\right|,$$

with $\hat{\vartheta} = A^{-1}a$ and A and a defined in (13) and (14), respectively. Therefore,

$$\sqrt{\tau(\mu)}Z\mathbf{1}_{\mathcal{G}} \leq \sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}(X_{T_m} - X_{t_0})W_m\right| + \sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}\varepsilon_m W_m\right|,$$

where
$$W_m = \frac{1}{Mh}U^\top(0)A^{-1}U\left(\frac{T_m - t_0}{h}\right)K\left(\frac{T_m - t_0}{h}\right).$$

This leads to

$$\left(\sqrt{\tau(\mu)}Z\right)^{1/2}\mathbf{1}_{\mathcal{G}} \leq \left(\sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}(X_{T_m} - X_{t_0})W_m\right|\right)^{1/2} +$$
$$\left(\sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}\varepsilon_m W_m\right|\right)^{1/2}.$$

Then, to show that (A) is finite, it suffices to show that

$$A_1 = \mathbb{E}\left[\exp\left\{\left(4\sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}(X_{T_m} - X_{t_0})W_m\right|\right)^{1/2}\right\}\mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\right],$$

and

$$A_2 = \mathbb{E}\left[\exp\left\{\left(4\sqrt{\tau(\mu)}\left|\sum_{m=1}^{M}\varepsilon_m W_m\right|\right)^{1/2}\right\}\mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\right],$$

are finite and to apply Cauchy-Schwarz inequality. To control the stochastic term A_2, remark that:

$$A_2 = 1 + \sum_{p \geq 1}\frac{2^p[\tau(\mu)]^{p/4}}{p!}B_p, \tag{29}$$

where

$$B_p = \mathbb{E}\left[\left|\sum_{m=1}^{M} \varepsilon_m W_m\right|^{p/2} \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\mathcal{F}}\right].$$

By Jensen's inequality, $B_1 \leq B_2^{1/2} \leq B_3^{1/3} \leq B_4^{1/4}$. Thus, it remains to control B_p for any $p \geq 4$. For such values of p, we use Marcinkiewicz-Zygmund's inequality and obtain:

$$B_p \leq \left(\frac{p}{2} - 1\right)^{p/2} \mathbb{E}\left[\left(\sum_{m=1}^{M} \varepsilon_m^2 W_m^2\right)^{p/4} \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\mathcal{F}}\right].$$

By a version of Lemma 1.3 of [37], for κ defined in (15),

$$\sup_{1 \leq m \leq M} |W_m| \mathbf{1}_{\mathcal{E}_{\lambda_0/2}} \leq \frac{4\kappa}{\lambda_0 M h},$$

and

$$\sum_{m=1}^{M} |W_m| \mathbf{1}_{\mathcal{E}_{\lambda_0/2}} \leq \frac{4\kappa}{\lambda_0} \frac{1}{Mh} \sum_{j=1}^{M} \mathbf{1}_{\{t_0 - h \leq T_m \leq t_0 + h\}}. \tag{30}$$

Let $\chi_m = h^{-1} \mathbf{1}_{\{t_0 - h \leq T_m \leq t_0 + h\}}$. By the Rosenthal inequality, there exists a universal constant C, such that, for any $q \geq 1$,

$$\tilde{\mathbb{E}}\left[\left(\sum_{m=1}^{M} \chi_m\right)^q\right] \leq \frac{C^q q^q}{\log^q q} \max\left\{\sum_{m=1}^{M} \tilde{\mathbb{E}} \chi_m^q, \left(\sum_{m=1}^{M} \tilde{\mathbb{E}} \chi_m\right)^q\right\} \tag{31}$$
$$\leq \left(\frac{4qCf(t_0)M}{\log q}\right)^q.$$

See [23]. Since $W_m^2 \leq |W_m| \sup_{1 \leq j \leq M} |W_j|$, we deduce

$$\tilde{\mathbb{E}}\left[\left(\sum_{m=1}^{M} W_m^2\right)^q \mathbf{1}_{\mathcal{E}}\right] \leq \left(\frac{4\kappa}{\lambda_0 Mh}\right)^q \left(\frac{16q\kappa Cf(T)}{\lambda_0 \log q}\right)^q =: \left(\frac{1}{Mh}\right)^q \left(\frac{c_0 q}{\log q}\right)^q.$$

Using Assumption (LP2), we deduce:

$$\mathbb{E}\left[\left(\sum_{m=1}^{M} W_m^2\right)^q \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\mathcal{F}}\right] \leq \left(\frac{c_0 q}{\log q}\right)^q \mathbb{E}\left[\left(\frac{\log \mu}{\mu}\right)^{q\frac{2\widehat{H}_{t_0}}{2\widehat{H}_{t_0}+1}} \mathbf{1}_{\mathcal{F}}\right]$$
$$= 2\left(\frac{c_0 q}{\log q}\right)^q \frac{1}{\{\tau(\mu) \log^2 \mu\}^q}.$$

The last line can be deduced using similar arguments to those used to obtain (45) in the proof of Lemma 5. Next, let $\widetilde{W}_m = W_m^2 / \sum_{j=1}^{M} W_j^2$. Since the error terms

are independent on the T_m's, and using (H5), by Jensen's inequality

$$B_p \left(\frac{p}{2} - 1\right)^{-p/2} \le \mathbb{E}\left[\left(\sum_{m=1}^{M} |\varepsilon_m|^{p/2}\widetilde{W}_m\right)\left(\sum_{j=1}^{M} W_j^2\right)^{p/4} \mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\right]$$

$$= \mathbb{E}\left\{\mathbb{E}\left[\left(\sum_{m=1}^{M} |\varepsilon_m|^{p/2}\widetilde{W}_m\right) \mid M, W_1, \dots, W_M\right]\left(\sum_{j=1}^{M} W_j^2\right)^{p/4} \mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\right\}$$

$$\le \left(\mathbb{E}|\varepsilon|^{2p}\right)^{1/4} \mathbb{E}\left[\left(\sum_{m=1}^{M} W_m^2\right)^{p/4} \mathbf{1}_{\mathcal{E}}\mathbf{1}_{\mathcal{F}}\right]$$

$$\le \left(\frac{p!}{2}\mathfrak{b}\mathfrak{B}^{p-2}\right)^{1/4} \left(\frac{c_0 p}{4\log(p/4)}\right)^{p/4} \left(\frac{1}{\tau(\mu)\log^2 \mu}\right)^{p/4}.$$

Thus we have

$$B_p \le \left(\frac{p}{2} - 1\right)^{p/2} \left(\frac{p!}{2}\mathfrak{b}\mathfrak{B}^{p-2}\right)^{1/4} \left(\frac{c_0 p}{4\log(p/4)}\right)^{p/4} \left(\frac{1}{\tau(\mu)\log^2 \mu}\right)^{p/4}$$

$$= \left(\frac{\mathfrak{b}}{2\mathfrak{B}^2}\right)^{1/4} \left(\frac{1}{\tau(\mu)\log^2 \mu}\right)^{p/4} D_p,$$

where

$$D_p = \left(\frac{p}{2} - 1\right)^{p/2} (p!)^{1/4} \left(\frac{c_0 p \mathfrak{B}}{4\log(p/4)}\right)^{p/4} \le p! \left(\frac{c_1}{\log p}\right)^{p/4},$$

for some constant c_1. For the last inequality, we use Stirling's formula. This implies that there exists a universal constant c_2 such that

$$\frac{B_p}{p!} \le \left(\frac{c_2}{\log p}\right)^{p/4} \left(\frac{1}{\tau(\mu)\log^2 \mu}\right)^{p/4}. \tag{32}$$

Combining (29) with (32) we obtain:

$$A_2 = 1 + \left\{2B_1\tau^{1/4}(\mu) + 2B_2\tau^{1/2}(\mu) + \frac{4B_3}{3}\tau^{3/4}(\mu)\right\} + \sum_{p\ge 4} \frac{2^p\tau(\mu)^{p/4}}{p!} B_p$$

$$\le 1 + \left\{2\big(B_4\tau(\mu)\big)^{1/4} + 2\big(B_4\tau(\mu)\big)^{1/2} + \frac{4\big(B_4\tau(\mu)\big)^{3/4}}{3}\right\}$$

$$+ \sum_{p\ge 4} \left(\frac{16c_2}{\log p}\right)^{p/4} \left(\frac{1}{\log^2 \mu}\right)^{p/4}$$

$$< \infty.$$

The inequality on the last line comes from the fact that $B_4\tau(\mu)\log^2(\mu)$ is bounded.

To control the bias term A_1, let us define, for any $0 < \beta < H_{t_0}$:

$$\Lambda_\beta = \sup_{\substack{u,v \in J_\mu(t_0) \\ u \neq v}} \frac{|X_u^{(\mathbf{d})} - X_v^{(\mathbf{d})}|}{|u - v|^\beta}, \tag{33}$$

where here $X_u^{(\mathbf{d})}$ denotes the $\mathbf{d}$-th derivative of the trajectory X_u. Applying Taylor's formula and using the basic properties satisfied by the weights W_m, we obtain:

$$\left| \sum_{m=1}^M X(T_m)W_m - X_{t_0} \right| \leq \left| \sum_{m=1}^M \sum_{k=1}^{\mathbf{d}} \frac{X^{(k)}(t_0)}{k!}(T_m - t_0)^k W_m \right|$$

$$+ \sum_{m=1}^M \frac{|X^{(\mathbf{d})}(t_0) - X^{(\mathbf{d})}(\zeta_m)|}{\mathbf{d}!}|T_m - t_0|^{\mathbf{d}}|W_m|$$

$$\leq \frac{\Lambda_\beta}{\mathbf{d}!} \sum_{m=1}^M |T_m - t_0|^{\mathbf{d}+\beta}|W_m|,$$

where $|\zeta_m - t_0| \leq |T_m - t_0|$. Note that this result is obtained using :

$$\sum_{m=1}^M (T_m - t_0)^k W_m = 0.$$

Since, under $\mathcal{E}$ we have, $W_m = 0$ as soon as $|T_m - t_0| > h$,:

$$\left| \sum_{m=1}^M X(T_m)W_m - X_{t_0} \right| \mathbf{1}_\mathcal{E} \leq \frac{\Lambda_\beta h^{\mathbf{d}+\beta}}{\mathbf{d}!} \sum_{m=1}^M |W_m|\mathbf{1}_\mathcal{E}$$

$$\leq \frac{\Lambda_\beta}{\mathbf{d}!} \frac{4\kappa}{\lambda_0} \frac{h^{\mathbf{d}+\beta}}{Mh} \sum_{m=1}^M \mathbf{1}_{\{t_0 - h \leq T_m \leq t_0 + h\}}.$$

The last line follows from (30). Moreover, combining the result obtained by [33, p. 27], with (H4), for any $0 < H_{t_0} - \beta < \beta_0$ where β_0 is some sufficiently small fixed value, we have:

$$\mathbb{E}\Lambda_\beta^{p/2} \leq 2^{\frac{1}{4} + \frac{p}{2}(H_{t_0}+1)} \left(\frac{1}{1 - 2^{\beta - H_{t_0}}} \right)^{p/2} \left(\frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2} \right)^{1/4}$$

$$\leq \frac{\mathfrak{a}^{1/4}}{\mathfrak{A}^{1/2}}(p!)^{1/4} \left(\frac{8\log 2\sqrt{\mathfrak{A}}}{H_{t_0} - \beta} \right)^{p/2}.$$

Since, by definition, the random variable Λ_β is independent of $\widehat{H}_{t_0}$, M and the T_m's, by the last inequality above and inequality (31), we have:

$$\mathbb{E}\left(\left|\sum_{m=1}^{M} X(T_m)W_m - X_{t_0}\right|^{p/2} \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\mathcal{F}}\right)$$
$$\leq \left(\frac{2pCf(t_0)}{\log(p/2)}\right)^{p/2} \mathbb{E}\left[\left(h^{\mathbf{d}+\beta}\right)^{p/2}\right] \mathbb{E}\left[(\Lambda_\beta)^{p/2}\right].$$

We thus obtain:

$$A_1 \leq \sum_{p \geq 0} \frac{(16\tau(\mu))^{p/4}}{p!} \mathbb{E}\left(\left|\sum_{m=1}^{M} X(T_m)W_m - X_{t_0}\right|^{p/2} \mathbf{1}_{\mathcal{E}} \mathbf{1}_{\mathcal{F}}\right)$$
$$\leq \sum_{p \geq 0} \frac{(16\tau(\mu))^{p/4}}{p!} \left(\frac{2pCf(t_0)}{\log(p/2)}\right)^{p/2} \mathbb{E}\left[\left(h^{\mathbf{d}+\beta}\right)^{p/2} \mathbf{1}_{\mathcal{F}}\right] \mathbb{E}\left[(\Lambda_\beta)^{p/2}\right].$$

Note that, on the event $\mathcal{F}$,

$$\left(h^{\mathbf{d}+\beta}\right)^{p/2} \leq C^{p/2} \left(\frac{\log \mu}{\mu}\right)^{\frac{p}{2}\frac{2(\mathbf{d}+\beta)}{2(\mathbf{d}+H_{t_0})+1}}.$$

Taking $\beta = H_{t_0} - \log^{-1} \mu$, since $(\mu/\log \mu)^{1/\log \mu}$ is bounded, we deduce that, for some constant $C > 0$,

$$A_1 \leq \sum_{p \geq 0} \frac{C^{p/2}}{\log^{p/2}(p)} < \infty.$$

It remains to control (B), (C) and (D). For this purpose, let us first note that, by the Assumption (LP1), $c_1 := \mathbb{E}[\exp(\eta_0 X_{t_0}^2)] < \infty$, for $\eta_0 = 1/(2\mathfrak{A})$. We deduce that

$$\mathbb{E}\left[\varphi^2\left(\tau(\mu)Z^2\right)\right] \leq \mathbb{E}\left[\exp\left(2\tau^{1/4}(\mu)\left|\hat{X}_{t_0} - X_{t_0}\right|^{1/2}\right)\right]$$
$$\leq \mathbb{E}\left[\exp\left(2\tau^{1/4}(\mu)\{|\hat{X}_{t_0}|^{1/2} + |X_{t_0}|^{1/2}\}\right)\right]$$
$$\leq \exp\left[2\tau^{1/4}(\mu)\tau^{\alpha/1}(\mu\log(\mu))\right] \mathbb{E}\left[\exp\left(2\tau^{1/4}(\mu)|X_{t_0}|^{1/2}\right)\right]$$
$$\leq c_1 \exp\left[2\tau^{1/4}(\mu)\tau^{\alpha/2}(\mu\log(\mu)) + 2^{4/3}c_2\tau^{1/3}(\mu)\right],$$

where for the last inequality, we apply Lemma 6 with and $\xi = |X_{t_0}|^{1/2}$, and thus $c_2 = (5\mathfrak{A}/8)^{1/3}$. Now notice that, using Markov's inequality

$$\mathbb{P}(\overline{\mathcal{G}}) = \mathbb{P}(|X_{t_0}| > \tau^{\tilde{\alpha}}(\mu))$$
$$\leq c_1 \exp\left(-\eta_0\tau^{2\tilde{\alpha}}(\mu)\right).$$

Since $\tilde{\alpha} < 1/2$, Assumptions (LP3) and (LP4) imply that for sufficiently large μ:

$$\mathbb{P}(\overline{\mathcal{F}}) \leq 2\mathfrak{K}_1 \exp(-\mu) \leq \exp\left(-\eta_0\tau^{2\tilde{\alpha}}(\mu)\right).$$

Moreover, Lemma 5 also implies that, for sufficiently large μ:

$$\mathbb{P}(\overline{\mathcal{E}}) \leq \mathfrak{K}_2 \exp\left(-\frac{\mathfrak{g}}{2}\tau(\mu)\log^2(\mu)\right) \leq \exp\left(-\eta_0\tau^{2\tilde{\alpha}}(\mu)\right).$$

Finally, if $\mathcal{H}$ denotes either $\mathcal{E}$, $\mathcal{F}$ or $\mathcal{G}$, we have

$$\mathbb{E}\left[\varphi^2\left(\tau(\mu)Z^2\right)\right]\mathbb{P}(\overline{\mathcal{H}})$$
$$\leq C\exp\left[2\tau^{1/4}(\mu)\tau^{\alpha/2}(\mu\log(\mu)) + 2^{4/3}c_2\tau^{1/3}(\mu) - \eta_0\tau^{2\tilde{\alpha}}(\mu)\right],$$

where C denotes a positive constant. The choice $\alpha = 5/12$ and $\tilde{\alpha} = 9/24$ allows us to deduce that $\mathbb{E}\left[\varphi^2\left(\tau(\mu)Z^2\right)\right]\mathbb{P}(\overline{\mathcal{H}})$ is bounded. This concludes the proof of Theorem 2. □

Proof of Theorem 3. By Theorem 2,

$$\max_{1 \leq n_1 \leq N_1} \mathbb{E}\left[\varphi\left\{\tau(\mu)\left|\widehat{X}_{t_0}^{[n_1]} - X_{t_0}^{[n_1]}\right|^2\right\}\right] \leq \Gamma_0$$

where

$$\tau(\mu) = \frac{1}{\log^2(\mu)}\left(\frac{\mu}{\log(\mu)}\right)^{\frac{2s_{t_0}}{2s_{t_0}+1}},$$

and $\varphi(x) = \exp(x^{1/4})$. Now, let $x_0 = 256$ and consider $\widetilde{\varphi} \leq \varphi$ defined by

$$\widetilde{\varphi}(x) = \begin{cases} \varphi'(x_0)(x - x_0) + \varphi(x_0) & \text{if } x \leq x_0 \\ \varphi(x) & \text{if } x \geq x_0 \end{cases},$$

and note that $\widetilde{\varphi}$ is nondecreasing and convex. Then, by Lemma 1.6 in [37],

$$\mathbb{E}\left(\max_{1 \leq n_1 \leq N_1}\left|\widehat{X}_{t_0}^{[n_1]} - X_{t_0}^{[n_1]}\right|^2\right) \leq \tau^{-1}(\mu)\widetilde{\varphi}^{\leftarrow}(\Gamma_0 N_1),$$

where $\widetilde{\varphi}^{\leftarrow}$ denotes the inverse function of $\widetilde{\varphi}$. Moreover, for N_1 sufficiently large, we have $\widetilde{\varphi}^{\leftarrow}(\Gamma_0 N_1) = \log^4(\Gamma_0 N_1)$. □

Appendix C: Proof of Theorem 4

Proof of Theorem 4. Assume, without loss of generality that

$$K = \log^4(N_1)\exp(\log^2(\mu)),$$

is an integer and let $\min(I) = s_0 < s_1 < \ldots < s_K < s_{K+1} = \max(I)$ be a regular grid of the interval I. For any $t \in I$, let $k_t \in \{1, \ldots, K\}$ be such that $|t - s_{k_t}| \leq 1/(2K - 2) =: \epsilon$. We have

$$\mathbb{E}\left(\max_{1 \leq n_1 \leq N_1}\sup_{t \in I}\left|\widehat{X}_t^{[n_1]} - X_t^{[n_1]}\right|^2\right) \leq 3(A + B + C),$$

where

$$A = \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \max_{k=1,\ldots,K} \left| \widehat{X}_{s_k}^{[n_1]} - X_{s_k}^{[n_1]} \right|^2 \right),$$

$$B = \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \sup_{t \in I} \left| \widehat{X}_t^{[n_1]} - \widehat{X}_{s_{k_t}}^{[n_1]} \right|^2 \right)$$

$$C = \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \sup_{t \in I} \left| X_t^{[n_1]} - X_{s_{k_t}}^{[n_1]} \right|^2 \right).$$

Bound for A. Using arguments similar to those of the proof of Theorem 3, we obtain:

$$A \leq c\tau^{-1}(\mu) \log^4(N_1 K) = c\Psi(\mu, N_1),$$

where c denotes an absolute positive constant and Ψ is defined by (16).

Bound for B. Note that using [37, p. 45] we obtain that there exists a positive constant ℓ such that, for any $1 \leq n_1 \leq N_1$, we have almost surely:

$$\sup_{|t-s|<\epsilon} |\widehat{X}_t^{[n_1]} - \widehat{X}_s^{[n_1]}| \leq \frac{\epsilon\ell}{\lambda_0 M_{n_1} h^2} \sum_{i=1}^{M_{n_1}} |Y_i^{[n_1]}|.$$

Using (H5) and (LP1), this implies that there exists a positive constant $\mathfrak{Y}$ such that:

$$B \leq \frac{\ell\mathfrak{Y}}{\lambda_0} \frac{\epsilon}{h^2} \leq \frac{\ell\mathfrak{Y}}{\lambda_0} \epsilon\mu^2 \ll \Psi(\mu, N_1).$$

Bound for C. Set $0 < \eta < H/2$ and, for any $1 \leq n_1 \leq N_1$, define the random variable:

$$\Lambda_{n_1} = \sup_{s \neq t \in I} \frac{|X_t^{[n_1]} - X_s^{[n_1]}|}{|t-s|^\eta}.$$

We have:

$$C \leq \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \sup_{|t-s| \leq \epsilon} \left| X_t^{[n_1]} - X_s^{[n_1]} \right|^2 \right) \leq \epsilon^{2\eta} \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \Lambda_{n_1}^2 \right).$$

It remains to bound the last expectation. By (H4), we have, for any $1 \leq n_1 \leq N_1$ and any $p \geq 2$:

$$\mathbb{E}\left(\Lambda_{n_1}^{2p} \right) \leq \left(\frac{p!}{2} \mathfrak{a}\mathfrak{A}^{p-2} \right) |I|^{2p(H-\eta)} = p!\mathfrak{M}^p,$$

where $|I|$ denotes the length of I and

$$\mathfrak{M} = \max \left[\frac{\mathfrak{a}}{2} |I|^{4(H-\eta)}, \mathfrak{A}|I|^{2(H-\eta)}, 1 \right].$$

Thus, we have

$$\mathbb{E}\left[\exp\left(\frac{\Lambda_{n_1}^2}{2\mathfrak{M}} \right) \right] \leq 2 \qquad \text{which implies} \qquad \mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \Lambda_{n_1}^2 \right) \leq 2\mathfrak{M} \log(2N_1).$$

We finally obtain that:

$$C \leq 2\mathfrak{M} \log(2N_1)\epsilon^{2\eta} \ll \Psi(\mu, N_1).$$

Gathering the bounds, we deduce that:

$$\mathbb{E}\left(\max_{1 \leq n_1 \leq N_1} \sup_{t \in I} \left|\widehat{X}_t^{[n_1]} - X_t^{[n_1]}\right|^2\right) \leq 3c\Psi(\mu, N_1)\{1 + o(1)\}. \qquad \square$$

Appendix D: Alternative local regularity estimator

D.1. Main assumptions

In this section we propose an alternative approach to estimate the regularity $\varsigma_{t_0} = \mathbf{d} + H_{t_0}$ of the process X without any restriction on $\mathbf{d} \in \mathbb{N}$. The main idea is to replace the noisy observations in (6) by smoothed versions of the sample paths of the process. To construct this estimator of ς_{t_0} and to derive its theoretical properties we need a set of assumptions that slightly differ from the one presented in the main manuscript. In what follows we fix an open subinterval $\mathcal{O}$ of $J_\mu(t_0)$ with length $0 < \Delta \leq 1$ and, for the sake of homogeneity, we denote $H_{\mathbf{d}} = H_{t_0}$.

Definition 1. *For any $d \in \mathbb{N}$, $0 < H_d \leq 1$ and $L_d > 0$, the class $\mathcal{X}(d + H_d, L_d; \mathcal{O})$ is the set of stochastic processes indexed by $t \in \mathcal{O}$ for which the following conditions hold true.*

(G1) With probability 1, for any $\ell \in \{0, \ldots, d\}$ the ℓ-th order derivative $\nabla^\ell X_t$ of X_t exists for all $t \in \mathcal{O}$, and satisfies:

$$0 < \underline{a}_\ell = \inf_{u \in \mathcal{O}} \mathbb{E}\left[(\nabla^\ell X_u)^2\right] \leq \sup_{u \in \mathcal{O}} \mathbb{E}\left[(\nabla^\ell X_u)^2\right] = \overline{a}_\ell < \infty.$$

(G2) Two positive constants S_d and β_d exist such that:

$$\left|\mathbb{E}\left[(\nabla^d X_t - \nabla^d X_s)^2\right] - L_d^2|t - s|^{2H_d}\right| \leq S_d^2|t - s|^{2H_d}\Delta^{2\beta_d}, \qquad s, t \in \mathcal{O}.$$

(G3) $\mathfrak{a} > 0$ and $\mathfrak{A} > 0$ exist such that, for any $\ell \in \{0, \ldots, d\}$ and any $p \geq 1$:

$$\mathbb{E}\left[|\nabla^\ell X_t - \nabla^\ell X_s|^{2p}\right] \leq \frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2}, \qquad s, t \in \mathcal{O}.$$

The quantity $d + H_d$ is the local regularity of the process on $\mathcal{O}$, while L_d is the Hölder constant of the d−th derivative of the trajectories.

These classes of processes satisfy embedding properties that will be useful to construct an estimator of $\mathbf{d}$.

Lemma 7. *Assume that X restricted to $\mathcal{O}$ belongs to $\mathcal{X}(\mathbf{d} + H_{\mathbf{d}}, L_{\mathbf{d}}; \mathcal{O})$ for some $0 < H_{\mathbf{d}} \leq 1$ and $L_{\mathbf{d}} > 0$. Then, for any $d \in \{0, \ldots, \mathbf{d} - 1\}$, two positive real numbers L_d and S_d exist such that*

$$\left|\mathbb{E}\left[(\nabla^d X_t - \nabla^d X_s)^2\right] - L_d^2|t - s|^2\right| \leq S_d^2|t - s|^2\Delta^{H_{d+1}}, \quad s, t \in \mathcal{O},$$

with $H_{d+1} = \mathbf{1}_{\{d \neq \mathbf{d}-1\}} + H_{\mathbf{d}} \mathbf{1}_{\{d = \mathbf{d}-1\}}$. *This implies that, for any* $0 \leq d \leq \mathbf{d}$, *the process X restricted to $\mathcal{O}$ belongs to the class* $\mathcal{X}(d + H_d, L_d; \mathcal{O})$.

The three parameters $\mathbf{d} \in \mathbb{N}$, $0 < H_{\mathbf{d}} < 1$ and $L_{\mathbf{d}} > 0$ are fixed for the rest of the Section. We also assume that X restricted to $\mathcal{O}$ belongs to $\mathcal{X}(\mathbf{d}+H_{\mathbf{d}}, L_{\mathbf{d}}; \mathcal{O})$.

D.2. Heuristics on the definition of the alternative estimator

Using Lemma 7, we remark that

$$\mathbf{d} = \min\{d \in \mathbb{N} : H_d < 1\}.$$

A natural idea to construct an estimator of ς_{t_0}, is thus to find an estimator $\hat{H}_d$ of H_d, for any $d \in \mathbb{N}$, and to define:

$$\widehat{\mathbf{d}} = \min\{d \in \mathbb{N} : \hat{H}_d < 1 - \varphi(\mu)\} \quad \text{and} \quad \widehat{\varsigma_{t_0}} = \widehat{\mathbf{d}} + \hat{H}_{\widehat{\mathbf{d}}},$$

for some decreasing function $\varphi(\cdot)$ which will be defined later.

Thus, our problem reduces to the construction of accurate estimators of H_d for all $d \in \mathbb{N}$. For simplicity, let us denote, for any $s, t \in \mathcal{O}$:

$$\theta_d(s, t) = \mathbb{E}\left[(\nabla^d X_t - \nabla^d X_s)^2\right] \approx L_d^2 |t - s|^{2H_d} \quad \text{if } \Delta \text{ is small.}$$

Now, let t_1 and t_3 be such that $[t_1, t_3] \subset \mathcal{O}$ and $t_3 - t_1 = \Delta/2$. Denote by t_2 the middle point of $[t_1, t_3]$. It is easily seen that

$$H_d \approx \tilde{H}_d = \frac{\log(\theta_d(t_1, t_3)) - \log(\theta_d(t_1, t_2))}{2\log(2)} \quad \text{if } \Delta \text{ is small.} \qquad (34)$$

This suggests to define

$$\widehat{H}_d = \frac{\log(\hat{\theta}_d(t_1, t_3)) - \log(\hat{\theta}_d(t_1, t_2))}{2\log(2)},$$

where, for any $s, t \in \mathcal{O}$

$$\hat{\theta}_d(s, t) = \frac{1}{N_0} \sum_{n=1}^{N_0} \left(\widetilde{\nabla^d X}_t^{(n)} - \widetilde{\nabla^d X}_s^{(n)}\right)^2.$$

Here, $\widetilde{\nabla^d X}^{(n)}$ denotes a pilot estimator of the curve $\nabla^d X^{(n)}$ that can be obtained by a presmoothing procedure.

D.3. Concentration properties

The quality of the estimator $\widehat{\varsigma_{t_0}}$ depends on the quality of the generic nonparametric estimators $\widetilde{\nabla^d X}$ of $\nabla^d X$. To quantify their behavior, we consider the local $\mathbb{L}^p$-risk

$$R_p(d) = R_p(d; \mathcal{O}) = \sup_{t \in \mathcal{O}} \mathbb{E}\left(|\xi_d(t)|^p\right), \quad \text{where} \quad \xi_d(t) = \widetilde{\nabla^d X}_t - \nabla^d X_t.$$

Our method applies with any type of nonparametric estimator $\widetilde{\nabla^d X}$ (local polynomials, splines,...) as soon as, for any $p \in \mathbb{N}$, its $\mathbb{L}^p$-risk is suitably bounded. The following mild condition is satisfied by common estimators, see for instance Theorem 1 in [13] for the case of local polynomials.

(LP5) There exist two positive constants $\mathfrak{c}$ and $\mathfrak{C}$ such that

$$R_{2p}(d) \leq \frac{p!}{2}\mathfrak{c}\mathfrak{C}^{p-2}, \qquad \forall p \geq 1, \; d \in \{0, \ldots, \delta\}.$$

We can now derive an exponential bound for the concentration of all the estimators $\hat{H}_d$, $d \in \{0, \ldots, \mathbf{d}\}$. To make this exponential bound useful for deriving optimal rates for our estimators of the mean and covariance functions, we will require the largest risk among $R_2(0), \ldots, R_2(\mathbf{d})$ to tend to zero as μ increases to infinity.

Theorem 5. *Assume that X restricted to $\mathcal{O}$ belongs to $\mathcal{X}(\mathbf{d} + H_{\mathbf{d}}, L_{\mathbf{d}}; \mathcal{O})$, for some integer $\mathbf{d} \geq 0$ and $0 < H_{\mathbf{d}} < 1$, and that (LP5) holds. Assume also that there exists $\tau > 0$ and $B > 0$ such that:*

$$\rho(\mu) = \max_{d \in \{0, \ldots, \mathbf{d}\}} R_2(d) \leq B\mu^{-\tau}.$$

Let $0 < \gamma < 1$ and $\Gamma > 0$, and consider

$$\Delta(\mu) = 2\exp\left(-\log^{\gamma}(\mu)\right) \quad and \quad \varphi(\mu) = \log^{-\Gamma}(\mu).$$

Then, for any μ larger than some constant μ_0 depending on B, τ, γ, Γ, $H_{\mathbf{d}}$, $\beta_{\mathbf{d}}$ and for some positive constant $\mathfrak{f}$, we have

$$\mathbb{P}\left(|\widehat{\varsigma_{t_0}} - \varsigma_{t_0}| > \varphi(\mu)\right) \leq 8(1 + \mathbf{d})\exp\left(-\mathfrak{f}N_0\varphi^2(\mu)\left[\Delta(\mu)\right]^{4H_{\mathbf{d}}}\right).$$

The three quantities $\rho(\mu)$, $\Delta(\mu)$ and $\varphi(\mu)$ are required to decrease to zero, as μ tends to infinity, in such a way that $\rho(\mu)/\Delta(\mu) + \Delta(\mu)/\varphi(\mu) \to 0$. We propose $\Gamma = 2$ and $\gamma = 1/2$. The choices of the rates for $\rho(\mu)$, $\Delta(\mu)$ and $\varphi(\mu)$ satisfy some additional requirements. First, it will be shown below that, in order to achieve optimal rates of convergence for the mean and covariance estimators, the local regularity has to be estimated with a concentration rate $\varphi(\mu)$ faster than $\log^{-1}(\mu)$. This is a consequence of the identity $\mu^{1/\log(\mu)} = e$ for any $\mu > 1$, and of a mild condition on N and μ, such as

$$\limsup_{N,\mu \to \infty}\{\log(N)/\log(\mu)\} < \infty. \tag{35}$$

The technical condition (35) matches general situations found in applications. Second, we want to allow for reasonable rates of increase for N_0, the size of the learning set. In Theorem 5, N_0 can increase as fast as an arbitrary positive power of μ. Third, since $\tau > 0$ could be arbitrarily small, the rate imposed on the nonparametric estimators $\widetilde{\nabla^d X}$ of $\nabla^d X$ is a very mild requirement which could be achieved by the common estimators, with random or fixed design, under mild conditions, in particular on the distribution of the M_i and the smoothing parameter. See, for instance, [37] and [1]. In particular, the required rate for the $\widetilde{\nabla^d X}$ can be obtained under general forms of heteroscedasticity.

D.4. Proofs for the alternative local regularity estimator

To prove Theorem 5, we state and prove some auxiliary lemmas where the following notations will be used:

$$\rho_* = \max_{d \in \{0,...,\mathbf{d}\}} \left\{ (R_2(d))^{\frac{1}{4H_d}} \right\}, \quad \text{and} \quad \Delta_* = \Delta(\mu).$$

Remark that, for μ large enough, we have $\mathfrak{M} \max(\Delta_*^{H_\mathbf{d}}, \Delta_*^{\beta_\mathbf{d}}) < \varphi(\mu) < 2$ and $A_1 \rho_* \leq \Delta_* \leq A_2$, where:

$$A_1 = \max_{d \in \{0,...,\mathbf{d}\}} \left[\frac{2^{4H_d+3}}{L_d^2} \left(\sqrt{\frac{\mathfrak{a}}{\mathfrak{A}}} + 1 \right) \right]^{\frac{1}{2H_d}}, \quad A_2 = \min_{d \in \{0,...,\mathbf{d}\}} \left[\frac{1}{2} \left(\frac{L_d}{S_d} \right)^2 \right]^{\frac{1}{H_{d+1}}} \wedge 1,$$

and $\mathfrak{M} = 4 \max_{d \in \{0,...,\mathbf{d}\}} (S_d/L_d)^2$.

Proof of Lemma 7. Using Taylor's formula, there exists $\xi \in (s \wedge t, s \vee t)$ such that:

$$\mathbb{E}\left[|\nabla^d X_t - \nabla^d X_s|^2 \right] = (t-s)^2 \mathbb{E}\left[\left(\nabla^{d+1} X_\xi \right)^2 \right]$$
$$= (t-s)^2 \left\{ L_d^2 + 2E_1(d) + E_2(d) \right\},$$

where

$$L_d^2 = \mathbb{E}\left[\left(\nabla^{d+1} X_{t_1} \right)^2 \right]$$
$$E_1(d) = \mathbb{E}\left[\nabla^{d+1} X_{t_1} \left(\nabla^{d+1} X_\xi - \nabla^{d+1} X_{t_1} \right) \right]$$
$$E_2(d) = \mathbb{E}\left[\left(\nabla^{d+1} X_\xi - \nabla^{d+1} X_{t_1} \right)^2 \right].$$

Remark that (G1) implies that $\underline{a}_{d+1} < L_d^2 < \overline{a}_{d+1}$. Using the Cauchy-Schwartz inequality,

$$\left| \mathbb{E}\left[(\nabla^d X_t - \nabla^d X_s)^2 \right] - L_d^2 (t-s)^2 \right| \leq |2E_1(d) + E_2(d)|(t-s)^2$$
$$\leq \left(2L_d \sqrt{E_2(d)} + E_2(d) \right) (t-s)^2.$$

Thus, it remains to bound $E_2(d)$. First, consider $d = \delta - 1$. Then using (G2) combined with the fact that $|\xi - t_1| \leq \Delta(\mu) \leq 1$, we have:

$$E_2(\delta) = \mathbb{E}\left[\left(\nabla^\delta X_\xi - \nabla^\delta X_{t_1} \right)^2 \right] \leq (L_\delta^2 + S_\delta^2) \left[\Delta(\mu) \right]^{2H_\delta}.$$

This implies that

$$|2E_1(\delta-1) + E_2(\delta-1)| \leq S_{\delta-1} \left[\Delta(\mu) \right]^{H_\delta} \quad \text{with} \quad S_{\delta-1} = 2L_d \sqrt{L_\delta^2 + S_\delta^2} + L_\delta^2 + S_\delta^2.$$

Next, consider the case of $d < \delta - 1$. Using Taylor's formula and (G1), we have

$$E_2(d) = \mathbb{E}\left[\left(\nabla^{d+1} X_\xi - \nabla^{d+1} X_{t_1} \right)^2 \right] \leq \overline{a}_{d+2} (\xi - t_1)^2 \leq \overline{a}_{d+2} \Delta^2(\mu),$$

which implies $|2E_1(d) + E_2(d)| \leq S_d \Delta(\mu)$ with $S_d = 2L_d \sqrt{\overline{a}_{d+2}} + \overline{a}_{d+2}$. Lemma 7 is now proved. □

Lemma 8. *Assume that the condition of Theorem 5 hold true. Assume also that $\Delta_*^{2\beta_d} \leq (L_{\mathbf{d}}/S_{\mathbf{d}})^2/2$. For $d \in \{0, \ldots, \mathbf{d}\}$, define*

$$\eta_*(d) = 4 \left(\sqrt{\frac{\mathfrak{a}}{\mathfrak{A}}} + 1 \right) \sqrt{R_2(d)}.$$

Let $s, t \in \mathcal{O}$ such that $\theta_d(s,t) > \eta_*(d)$. For any $\kappa > 0$, define

$$p_d^+(s,t;\kappa) = \mathbb{P}\left[\hat{\theta}_d(s,t) > (1+\kappa)\theta_d(s,t) \right]$$

and

$$p_d^-(s,t;\kappa) = \mathbb{P}\left[\hat{\theta}_d(s,t) < (1-\kappa)\theta_d(s,t) \right].$$

There exists a constant $\mathfrak{c} > 0$ such that, for any κ such that $\eta_*(d) < \kappa\theta_d(s,t) < 1$, we have

$$\max\left[p_d^+(s,t;\kappa), p_d^-(s,t;\kappa) \right] \leq \exp\left(-\mathfrak{c} N_0 \kappa^2 \underline{a}_{d+1}^2 |t-s|^{4H_d} \right),$$

where $\underline{a}_{d+1}$ is defined in (G1), for $d < \mathbf{d}+1$, and $\underline{a}_{\mathbf{d}+1} = L_{\mathbf{d}}^2/2$.

Proof of Lemma 8. First, let us point out that, by the definition of the space $\mathcal{X}(\mathbf{d} + H_{\mathbf{d}}, L_{\mathbf{d}}; \mathcal{O})$, the quantity $\theta_d(s,t)$ could not be equal to zero for any s, t in an open interval. Thus, the points $s, t \in \mathcal{O}$ in the statement of Lemma 8 are well-defined.

Set $d \in \{0, \ldots, \mathbf{d}\}$ and $s, t \in \mathcal{O}$. Let us decompose

$$\hat{\theta}_d(s,t) - \theta_d(s,t) = \frac{1}{N_0} \sum_{n=1}^{N_0} \bar{Z}_n + \mathbb{E}\left(\hat{\theta}_d(s,t) \right) - \theta_d(s,t),$$

where, for any $n = 1, \ldots, N_0$:

$$\bar{Z}_n = Z_n - \mathbb{E}(Z_n) \quad \text{with} \quad Z_n = \left(\widetilde{\nabla^d X}_t^{(n)} - \widetilde{\nabla^d X}_s^{(n)} \right)^2.$$

Bounding the bias term. Note that

$$\mathbb{E}\left(\hat{\theta}_d(s,t) \right) - \theta_d(s,t) = 2\mathbb{E}\left[(\xi_d(t) - \xi_d(s)) \left(\nabla^d X_t - \nabla^d X_s \right) \right]$$
$$+ \mathbb{E}\left[(\xi_d(t) - \xi_d(s))^2 \right].$$

Since $\mathbb{E}(\xi_d(t) - \xi_d(s))^2 \leq 2R_2(d)$, using the Cauchy-Schwartz inequality and (G3), we obtain:

$$\left| \mathbb{E}\left(\hat{\theta}_d(s,t) \right) - \theta_d(s,t) \right| \leq 2\sqrt{\frac{\mathfrak{a}}{\mathfrak{A}} R_2(d)} + 2R_2(d),$$

Considering without loss of generality that $R_2(d) \leq 1$, we obtain

$$\left| \mathbb{E}\left(\hat{\theta}_d(s,t) \right) - \theta_d(s,t) \right| \leq \eta_*(d)/2.$$

Moments of the stochastic term. Let us note that, for any $p \geq 1$,

$$\mathbb{E}\left(\left|\bar{Z}_n\right|^p\right) = \mathbb{E}\left(\left|Z_n - \mathbb{E}(Z_n)\right|^p\right) \leq 2^p \mathbb{E}\left(\left|Z_n\right|^p\right)$$

Moreover,

$$
\begin{aligned}
|Z_n|^p &= \left|\left(\nabla^d X_t^{(n)} - \nabla^d X_s^{(n)}\right) - \left(\xi_d^{(n)}(t) - \xi_d^{(n)}(s)\right)\right|^{2p} \\
&\leq \left(\left|\nabla^d X_t^{(n)} - \nabla^d X_s^{(n)}\right| + \left|\xi_d^{(n)}(t)\right| + \left|\xi_d^{(n)}(s)\right|\right)^{2p} \\
&\leq 3^{2p-1}\left\{\left|\nabla^d X_t^{(n)} - \nabla^d X_s^{(n)}\right|^{2p} + \left|\xi_d^{(n)}(t)\right|^{2p} + \left|\xi_d^{(n)}(s)\right|^{2p}\right\}.
\end{aligned}
$$

This implies that

$$
\begin{aligned}
\mathbb{E}\left(\left|\bar{Z}_n\right|^p\right) &\leq \frac{18^p}{3}\left\{\mathbb{E}\left(\left|\nabla^d X_t^{(n)} - \nabla^d X_s^{(n)}\right|^{2p}\right) + 2R_{2p}(d)\right\} \\
&\leq \frac{18^p}{3}\frac{p!}{2}\left\{\mathfrak{a}\mathfrak{A}^{p-2} + 4\mathfrak{c}\mathfrak{C}^{p-2}\right\} \\
&\leq \frac{p!}{2}\mathfrak{d}\mathfrak{D}^{p-2},
\end{aligned}
$$

where $\mathfrak{d} = 108(\mathfrak{a} + 4\mathfrak{c})$ and $\mathfrak{D} = 18\max(\mathfrak{A}, \mathfrak{C})$. The second line in the last display comes from (G3) and (LP5).

Exponential bounds. Since $\eta_*(d)$ has the rate of $\sqrt{R_2(d)}$, we could consider $\eta_*(d) < \eta < 1$ and, using Bernstein's inequality, we obtain:

$$
\begin{aligned}
\mathbb{P}\left(\hat{\theta}_d(s,t) - \theta_d(s,t) > \eta\right) &\leq \mathbb{P}\left(\hat{\theta}_d(s,t) - \mathbb{E}\left(\hat{\theta}_d(s,t)\right) > \eta/2\right) \\
&\leq \exp\left(-\frac{N_0\eta^2}{8\mathfrak{d} + 4\mathfrak{D}\eta}\right) \\
&\leq \exp\left(-\mathfrak{e}N_0\eta^2\right),
\end{aligned}
$$

where $\mathfrak{e} = 1/(8\mathfrak{d} + 4\mathfrak{D})$. Since $\kappa\theta_d(s,t) > \eta_*(d)$, this quantity could replace η in the above inequality. Hence:

$$\mathbb{P}\left(\hat{\theta}_d(s,t) > (1 + \kappa)\theta_d(s,t)\right) \leq \exp\left(-\mathfrak{e}N_0\kappa^2\theta_d^2(s,t)\right).$$

Assume first that $d < \mathbf{d} - 1$. Applying Talor's formula, there exists $\xi \in \mathcal{O}$ such that $|\xi - s| \leq |t - s|$ and

$$\theta_d(s,t) = (t - s)^2\mathbb{E}\left[\left(\nabla^{d+1}X_\xi\right)^2\right] \geq (t - s)^2\underline{a}_{d+1} = (t - s)^{2H_d}\underline{a}_{d+1}.$$

The last inequality is a consequence of (G3). Assume now that $d = \mathbf{d}$. Using (G2), we have:

$$\theta_{\mathbf{d}}(s,t) \geq L_{\mathbf{d}}^2|t - s|^{2H_{\mathbf{d}}} - S_{\mathbf{d}}^2|t - s|^{2H_{\mathbf{d}}}\Delta_*^{2\beta_d}$$

$$= |t - s|^{2H_{\mathbf{d}}} \left(L_{\mathbf{d}}^2 - S_{\mathbf{d}}^2 \Delta_*^{2\beta_d} \right)$$

$$\geq \frac{L_{\mathbf{d}}^2}{2} |t - s|^{2H_{\mathbf{d}}} = \underline{a}_{\mathbf{d}+1} |t - s|^{2H_{\mathbf{d}}}.$$

This implies that, for any $d \in \{0, \ldots, \mathbf{d}\}$:

$$p_d^+(s, t; \kappa) \leq \exp\left(-\mathfrak{c} N_0 \kappa^2 \underline{a}_{d+1}^2 |t - s|^{4H_d} \right).$$

The same reasoning can be applied to bound the term $p_d^-(s, t; \kappa)$.   □

Lemma 9. *Assume that the condition of Theorem 5 hold true. Let d be an element of $\{0, \ldots, \mathbf{d}\}$. There then exists a positive constant $\mathfrak{f}_d$, depending on $\mathfrak{a}$, $\mathfrak{A}$, $\mathfrak{c}$, $\mathfrak{C}$, $\underline{a}_{d+1}$ and H_d such that, for any ϵ which satisfies*

$$4 \left(\frac{S_d}{L_d} \right)^2 \Delta_*^{H_{d+1}} < \epsilon \log(2) < 2,$$

the following inequality holds:

$$\mathbb{P}\left(|\hat{H}_d - H_d| > \epsilon \right) \leq 4 \exp\left(-\mathfrak{f}_d N_0 \epsilon^2 \Delta_*^{4H_d} \right).$$

Proof of Lemma 9. We first have to control the distance between H_d and the proxy value $\tilde{H}_d$ defined in (34). To do so, note that, for $k = 2, 3$, we have

$$\theta_d(t_1, t_k) = L_{d+1}^2 |t_k - t_1|^{2H_d} (1 + \rho_d(k)),$$

where, using (G2) and Lemma 7

$$|\rho_d(k)| \leq \left(\frac{S_d}{L_d} \right)^2 \Delta_*^{H_{d+1}} \leq \frac{1}{2}.$$

This implies

$$
\begin{aligned}
|\tilde{H}_d - H_d| &= \left| \frac{\log(1 + \rho_d(3) - \log(1 + \rho_d(2)))}{2\log(2)} \right| \\
&\leq \frac{|\rho_d(3) - \rho_d(2)|}{\log(2)} \\
&\leq \frac{2}{\log(2)} \left(\frac{S_d}{L_d} \right)^2 \Delta_*^{H_{d+1}} \\
&\leq \frac{\epsilon}{2}.
\end{aligned}
$$

We deduce that

$$
\begin{aligned}
\mathbb{P}(|\hat{H}_d - H_d| > \epsilon) &\leq \mathbb{P}(|\hat{H}_d - \tilde{H}_d| > \epsilon - |\tilde{H}_d - H_d|) \\
&\leq \mathbb{P}(|\hat{H}_d - \tilde{H}_d| > \epsilon/2)
\end{aligned}
$$

$$\leq \mathbb{P}\left(\frac{\hat{\theta}_d(t_2, t_3)}{\theta_d(t_2, t_3)}\frac{\theta_d(t_1, t_2)}{\hat{\theta}_d(t_1, t_2)} > 2^\epsilon\right)$$

$$+ \mathbb{P}\left(\frac{\hat{\theta}_d(t_2, t_3)}{\theta_d(t_2, t_3)}\frac{\theta_d(t_1, t_2)}{\hat{\theta}_d(t_1, t_2)} < 2^{-\epsilon}\right).$$

By simple algebra and using the definition of the functions p_d^+ and p_d^- introduced in Lemma 8, we obtain

$$\mathbb{P}(|\hat{H}_d - H_d| > \epsilon) \leq p_d^+(t_1, t_3; 2^{\epsilon/2} - 1) + p_d^-(t_1, t_3; 1 - 2^{-\epsilon/2})$$
$$+ p_d^+(t_1, t_2; 2^{\epsilon/2} - 1) + p_d^-(t_1, t_2; 1 - 2^{-\epsilon/2}).$$

Note that, using Lemma 7 and (G2), we have for $k = 2, 3$:

$$\theta_d(t_1, t_k) \geq |t_3 - t_1|^{2H_d}\left(L_d^2 - S_d^2 \Delta_*^{H_{d+1}}\right)$$

$$= \left(\frac{\Delta_*}{2^{k-1}}\right)^{2H_d}\left(L_d^2 - S_d^2 \Delta_*^{H_{d+1}}\right)$$

$$\geq \frac{L_d^2}{2}\left(\frac{\Delta_*}{2^{k-1}}\right)^{2H_d}$$

$$\geq 4\left(\sqrt{\frac{\mathfrak{a}}{\mathfrak{A}}} + 1\right)\sqrt{R_2(d)}.$$

Thus, Lemma 8 can be used to write

$$p_d^+(t_1, t_k; 2^{\epsilon/2} - 1) \leq \exp\left(-\mathfrak{c}N_0(2^{\epsilon/2} - 1)^2 \underline{a}_{d+1}^2 |t_k - t_1|^{4H_d}\right)$$

$$\leq \exp\left(-\frac{\mathfrak{c}\log^2(2)}{4}N_0\epsilon^2 \underline{a}_{d+1}^2\left(\frac{\Delta_*}{2^{k-1}}\right)^{4H_d}\right)$$

$$\leq \exp\left(-\frac{\underline{a}_{d+1}^2 \mathfrak{c}\log^2(2)}{2^{2+8H_d}}N_0\epsilon^2 \Delta_*^{4H_d}\right).$$

The same reasoning can be applied to bound $p_d^-(t_1, t_k; 1 - 2^{-\epsilon/2})$. However, note that in this case $1 - 2^{-\epsilon/2} \leq \epsilon/4$. This implies that:

$$p_d^-(t_1, t_k; 1 - 2^{-\epsilon/2}) \leq \exp\left(-\frac{\underline{a}_{d+1}^2 \mathfrak{c}\log^2(2)}{2^{4+8H_d}}N_0\epsilon^2 \Delta_*^{4H_d}\right).$$

To complete the proof of Lemma 9, if suffices to take $\mathfrak{f}_d = \underline{a}_{d+1}^2 \mathfrak{c}\log^2(2)/2^{4+8H_d}$.
□

Proof of Theorem 5. Note that:

$$\mathbb{P}(|\hat{\varsigma}_{t_0} - \varsigma_{t_0}| > \varphi(\mu)) \leq \mathbb{P}(|\hat{\varsigma}_{t_0} - \varsigma_{t_0}| > \varphi(\mu), \hat{\mathbf{d}} = \mathbf{d}) + \mathbb{P}(\hat{\mathbf{d}} \neq \mathbf{d})$$
$$\leq \mathbb{P}(|\hat{H}_\mathbf{d} - H_\mathbf{d}| > \varphi(\mu)) + \mathbb{P}(\hat{\mathbf{d}} < \mathbf{d}) + \mathbb{P}(\hat{\mathbf{d}} > \mathbf{d})$$

$$\leq \mathbb{P}(|\hat{H}_{\mathbf{d}} - H_{\mathbf{d}}| > \varphi(\mu)) + \sum_{d=0}^{\mathbf{d}-1} \mathbb{P}(\hat{H}_d < 1 - \varphi(\mu))$$
$$+ \mathbb{P}(\hat{H}_{\mathbf{d}} > 1 - \varphi(\mu)).$$

Now, recall that, for $d < \mathbf{d}$ we have $H_d = 1$. Note also that $H_{\mathbf{d}} < 1$. This implies that:

$$\mathbb{P}(|\hat{\varsigma}_{t_0} - \varsigma_{t_0}| > \varphi(\mu)) \leq \mathbb{P}(|\hat{H}_{\mathbf{d}} - H_{\mathbf{d}}| > \varphi(\mu))$$
$$+ \sum_{d=0}^{\mathbf{d}-1} \mathbb{P}(|\hat{H}_d - H_d| > \varphi(\mu)) + \mathbb{P}(|\hat{H}_{\mathbf{d}} - H_{\mathbf{d}}| > 1 - H_{\mathbf{d}}).$$

Since $1 - H_{\mathbf{d}} > \varphi(\mu)$ for μ sufficiently large, such that $\varphi(\mu)$ could replace ϵ in Lemma 9, we have:

$$\mathbb{P}(|\hat{\varsigma}_{t_0} - \varsigma_{t_0}| > \varphi(\mu)) \leq 8(1 + \mathbf{d}) \exp\left(-\mathfrak{f} N_0 \varphi^2(\mu) \Delta_*^{4H_{\mathbf{d}}}\right).$$

The Theorem 5 is proved. $\qquad\square$

Appendix E: Technical lemmas

Proof of Lemma 2. Let $\mathcal{C} = \{M \geq K_0\} \setminus \mathcal{B}$. We have

$$\mathbb{E}_{\mathcal{B}}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha}\right] = \mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha} \mathbf{1}_{\mathcal{B}}\right] = (I) - (II),$$

where

$$(I) = \mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha} \mathbf{1}_{M \geq K_0}\right] \quad \text{and} \quad (II) = \mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha} \mathbf{1}_{\mathcal{C}}\right].$$

We study separately the two terms of the right hand side of the above equation.

Study of (II)**.** Note that

$$\mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha} \mathbf{1}_{\mathcal{C}}\right] \leq |I|^{\alpha} \mathbb{P}(\mathcal{C}).$$

The event $\mathcal{C}$ happens if, less than K_0 random times among $T_1, \ldots, T_M$ fall into the interval $J_\mu(t_0)$. This implies that

$$\mathbb{P}(\mathcal{C}) \leq \mathbb{E}\left[\mathbb{P}(B_M < K_0 \mid M) \mathbf{1}_{\{M \geq K_0\}}\right]$$

where, for any integer $m \geq 1$, B_m denotes a Binomial random variable defined as $\mathcal{B}(m, |J_\mu(t_0)|/|I|)$, independent of M. Using the Bernstein inequality, we obtain

$$\mathbb{P}(B_M < K_0 \mid M) \leq \exp\left(-\frac{2|J_\mu(t_0)|}{|I|} M + 2K_0\right).$$

Since $|J_\mu(t_0)|/|I| \leq (\log(\mu))^{-1}$ and $K_0 \leq (2\log(\mu))^{-1}\mu$, we obtain

$$\mathbb{P}(\mathcal{C}) \leq \exp\left(-\frac{2|J_\mu(t_0)|}{|I|}\mu + 2K_0\right) \mathbb{E}\left[\exp\left(-\frac{2|J_\mu(t_0)|}{|I|}(M - \mu)\right) \mathbf{1}_{\{M \geq K_0\}}\right]$$

$$\leq \exp\left(-\frac{\mu}{\log(\mu)}\right) \mathbb{E}\left[\exp\left(-\frac{2|J_\mu(t_0)|}{|I|}(M-\mu)\right)\mathbf{1}_{\{M \geq K_0\}}\right].$$

To bound the last expectation, let $0 < \epsilon < 1$ be some real number. Then

$$\exp\left(-\frac{\mu}{\log(\mu)}\right)\mathbb{E}\left[\exp\left(-\frac{2|J_\mu(t_0)|}{|I|}(M-\mu)\right)\mathbf{1}_{\{M \geq K_0\}}\right]$$

$$\leq \exp\left(-\frac{\mu}{\log(\mu)}\right)\left\{\exp\left(\frac{2|J_\mu(t_0)|}{|I|}\mu\epsilon\right)\right.$$

$$\left. +\mathbb{E}\left[\exp\left(-\frac{2|J_\mu(t_0)|}{|I|}(M-\mu)\right)\mathbf{1}_{\{K_0 \leq M \leq \mu - \mu\epsilon\}}\right]\right\}$$

$$\leq \exp\left(-\frac{\mu}{\log(\mu)}\right)\left\{\exp\left(\frac{4\mu\epsilon}{\log(\mu)}\right) + \exp\left(\frac{4\mu}{\log(\mu)}\right)\mathbb{P}\left[|M-\mu| > \mu\epsilon\right]\right\}$$

$$\leq \exp\left(-\frac{\mu}{\log(\mu)}\right)\left\{\exp\left(\frac{4\mu\epsilon}{\log(\mu)}\right) + \exp\left(\frac{2\mu}{\log(\mu)}\right)\exp(-\gamma_0\mu\epsilon)\right\}.$$

This implies that:

$$\mathbb{P}(\mathcal{C}) \leq \exp\left[-\frac{\mu}{\log(\mu)}(1-4\epsilon)\right] + \exp\left[-\mu\epsilon\left(\gamma_0 - \frac{1}{\epsilon\log(\mu)}\right)\right].$$

Taking $\epsilon = 1/8$, we obtain, for sufficiently large μ :

$$\mathbb{P}(\mathcal{C}) \leq 2\exp\left[-\frac{\mu}{2\log(\mu)}\right].$$

We finally obtain, for sufficiently large μ,

$$(II) = \mathbb{E}\left[|T_{(l)} - T_{(k)}|^\alpha \mathbf{1}_\mathcal{C}\right] \leq 2|I|^\alpha \exp\left[-\frac{\mu}{2\log(\mu)}\right]. \tag{36}$$

Study of (I)**.** We define the random variable ρ by the equation:

$$\mathbb{E}\left[|T_{(l)} - T_{(k)}|^\alpha \mid M\right] = \left(\frac{l-k}{f(t_0)(M+1)}\right)^\alpha(1+\rho)$$

Using Lemma 12, we have, almost surely:

$$|\rho| \leq \mathfrak{c}_0\left\{\frac{1}{M} + \frac{1}{\mathfrak{s}r} + \frac{1}{M\mathfrak{s}r} + \left(\frac{\mathfrak{s}r}{M+1}\right)^{\beta_f\alpha/4}\right\}.$$

Whenever $\mathfrak{s}r \geq (\mu+1)^{\beta_f\alpha/(4+\beta_f\alpha)}$, by bounding smaller terms by the dominant ones and balancing the dominant terms on the right hand side of the last inequality, we have for μ large enough:

$$|\rho| \leq 3\mathfrak{c}_0\left(\frac{\mathfrak{s}r}{M+1}\right)^{\beta_f\alpha/4} + \mathfrak{c}_0\left(\frac{\mathfrak{s}r}{\mu+1}\right)^{\beta_f\alpha/4}. \tag{37}$$

On the other hand we have

$$\mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha}\mathbf{1}_{M \geq K_0}\right] = \mathbb{E}\left(\mathbb{E}\left[\left|T_{(l)} - T_{(k)}\right|^{\alpha} \mid M\right]\mathbf{1}_{M \geq K_0}\right)$$

$$= \mathbb{E}\left[\left(\frac{l-k}{f(t_0)(M+1)}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\right]$$

$$= \left(\frac{l-k}{f(t_0)(\mu+1)}\right)^{\alpha}\mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\right].$$

$$(38)$$

Now, define:

$$t = \frac{(\log(\mu+1))^2}{2} \leq (\mu+1)/2,$$

and consider the following decomposition:

$$\mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\right] = \mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| \leq t}\right]$$

$$+ \mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| > t}\right].$$

Using Assumption ([H6](#)), combined with the fact that $r \leq \mu$, the term of the right hand side can be roughly bounded as follows:

$$\mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| > t}\right]$$

$$\leq 4\mathfrak{c}_0(\mu+1)^{\alpha(1+\alpha\beta_f/4)}\mathbb{P}\left(|M-\mu| > t\right)$$

$$\leq 4\mathfrak{c}_0(\mu+1)^{\alpha(1+\alpha\beta_f/4)}\exp\left(-\frac{\gamma_0}{2}\left(\log(\mu+1)\right)^2\right).$$

Thus, for sufficiently large μ, we have:

$$\mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| > t}\right] \leq \frac{1}{\mu+1}. \tag{39}$$

It remains to study the term

$$\mathbb{E}\left[\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| \leq t}\right].$$

To do so, let us define

$$\tilde{\rho}_{\alpha} = \left(\frac{\mu+1}{M+1}\right)^{\alpha} - 1.$$

Since $K_0 < (\mu+1)/2$, we have

$$\left(\frac{\mu+1}{M+1}\right)^{\alpha}(1+\rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M-\mu| \leq t} = \left(1 + \tilde{\rho}_{\alpha}\right)\left(1+\rho\right)\mathbf{1}_{|M-\mu| \leq t}$$

$$= 1 + \left(\tilde{\rho}_{\alpha} + \rho + \tilde{\rho}_{\alpha}\rho\right)\mathbf{1}_{|M-\mu| \leq t} - \mathbf{1}_{|M-\mu| > t}. \tag{40}$$

Under the event $\{|M - \mu \leq t|\}$, since $t < (\mu + 1)/2$, we have:

$$1 - \frac{\alpha t}{\mu + 1} \leq \left(\frac{\mu + 1}{M + 1}\right)^{\alpha} \leq 1 + \frac{2(2^{\alpha} - 1)t}{\mu + 1},$$

which leads to

$$|\tilde{\rho}_{\alpha}|\,\mathbf{1}_{\{|M - \mu \leq t|\}} \leq \frac{2(2^{\alpha} - 1)t}{\mu + 1} = (2^{\alpha} - 1)\frac{\log^2(\mu + 1)}{\mu + 1}. \tag{41}$$

Note also that by (37):

$$|\rho|\mathbf{1}_{|M - \mu| \leq t} \leq 4\mathfrak{c}_0 \left(\frac{2\mathfrak{s}r}{\mu + 1}\right)^{\beta_f \alpha/4}. \tag{42}$$

Gathering (40), (41) and (42) we obtain, for sufficiently large μ :

$$\left| \mathbb{E}\left[\left(\frac{\mu + 1}{M + 1}\right)^{\alpha}(1 + \rho)\mathbf{1}_{M \geq K_0}\mathbf{1}_{|M - \mu| \leq t} \right] - 1 \right|$$

$$\leq 5\mathfrak{c}_0 \left(\frac{2\mathfrak{s}r}{\mu + 1}\right)^{\beta_f \alpha/4} + \mathbb{P}(|M - \mu| > t)$$

$$\leq 6\mathfrak{c}_0 \left(\frac{2\mathfrak{s}r}{\mu + 1}\right)^{\beta_f \alpha/4}. \tag{43}$$

Combining (38) with (39) and (43), we obtain, for sufficiently large μ :

$$\mathbb{E}\left[|T_{(l)} - T_{(k)}|^{\alpha}\mathbf{1}_{M \geq K_0}\right] = \left(\frac{l - k}{f(t_0)(\mu + 1)}\right)^{\alpha}\left(1 + \tilde{R}\right), \tag{44}$$

where

$$|\tilde{R}| \leq 7\mathfrak{c}_0 \left(\frac{2\mathfrak{s}r}{\mu + 1}\right)^{\beta_f \alpha/4}.$$

From (36) and (44), we obtain, for μ large enough:

$$\mathbb{E}_{\mathcal{B}}\left[|T_{(l)} - T_{(k)}|^{\alpha}\right] = \left(\frac{l - k}{f(t_0)(\mu + 1)}\right)^{\alpha}(1 + R),$$

where

$$|R| \leq 8\mathfrak{c}_0 \left(\frac{2\mathfrak{s}r}{\mu + 1}\right)^{\beta_f \alpha/4} = 8\mathfrak{c}_0(2f(t_0))^{\beta_f \alpha/4}\left(\frac{l - k}{f(t_0)(\mu + 1)}\right)^{\beta_f \alpha/4}.$$

This ends the proof. $\qquad\square$

Let us recall the definitions

$$A = A_{M,h} = \frac{1}{Mh}\sum_{m=1}^{M} U\left(\frac{T_m - t_0}{h}\right)U^{\top}\left(\frac{T_m - t_0}{h}\right)K\left(\frac{T_m - t_0}{h}\right),$$

and

$$\mathbf{A} = f(t_0) \int_{\mathbb{R}} U(u)U^\top(u)K(u)du,$$

with $U(u) = (1, u, \dots, u^{\mathbf{d}}/\mathbf{d}!)$. Moreover, λ and λ_0 are the smallest eigenvalues of A and $\mathbf{A}$, respectively. The matrix $\mathbf{A}$ is positive definite and thus $\lambda_0 > 0$. See [37]. The following result shows that, with high probability, λ stays away from zero. Let us recall that in our context, $\hat{\mathbf{d}}$ is a generic estimator of $\mathbf{d}$, independent of the online set of curves. Since dimension of the matrices A and $\mathbf{A}$ are given by this estimator, the probability $\mathbb{P}(\cdot)$ in Lemma 5 should be understood as the conditional probability given the estimator $\hat{\mathbf{d}}$. Finally, recall that

$$\widehat{h} = \left(\frac{1}{M}\right)^{1/(2\widehat{\varsigma}_{t_0}+1)}.$$

Proof of Lemma 5. Without loss of generality, we could work on the set $\{\hat{\mathbf{d}} = \mathbf{d}\}$. Moreover, for simplicity, we write h instead of $\widehat{h}$ below in this proof.

Note that, using Assumption (LP2), for any $1 \leq i \leq j \leq \mathbf{d}$, the element $A_{i,j}$ tends almost surely to the element $\mathbf{A}_{i,j}$ as μ goes to infinity. This implies that the matrix A tends to the matrix $\mathbf{A}$. This also implies that, for sufficiently large μ, we have $\lambda > 0$. More precisely, we have:

$$|\lambda - \lambda_0| \leq \|A - \mathbf{A}\|_2 \leq (\mathbf{d}+1)\|A - \mathbf{A}\|_\infty,$$

where $\|\cdot\|_2$ denotes the norm induced by the Euclidean norm whereas $\|\cdot\|_\infty$ denotes the entrywise sup-norm. Let $\tilde{\mathbb{P}}(\cdot)$ and $\tilde{\mathbb{E}}(\cdot)$ denote the conditional probability $\mathbb{P}(\cdot|M)$ and conditional expectation $\mathbb{E}(\cdot|M)$, respectively. Then:

$$\tilde{\mathbb{P}}(\lambda \leq \beta) \leq \tilde{\mathbb{P}}(|\lambda - \lambda_0| \geq \lambda_0/2) \leq \sum_{0 \leq i,j \leq \mathbf{d}} \tilde{\mathbb{P}}\left(|(A_n)_{i,j} - \mathbf{A}_{i,j}| \geq \lambda_0/\{2(\mathbf{d}+1)\}\right).$$

Next, we decompose

$$A_{i,j} - \mathbf{A}_{i,j} = A_{i,j} - \tilde{\mathbb{E}}(A_{i,j}) + \tilde{\mathbb{E}}(A_{i,j}) - \mathbf{A}_{i,j}.$$

Using Assumption (H2) and the fact that $K(\cdot)$ has the support $[-1, 1]$, we have:

$$\left|\tilde{\mathbb{E}}(A_{i,j}) - \mathbf{A}_{i,j}\right| \leq \left|\int_{\mathbb{R}} \left[U(u)U^\top(u)\right]_{i,j} K(u)\{f(t_0 + hu) - f(t_0)\}du\right|$$

$$= L_f h^{\beta_f} \int_{\mathbb{R}} \left|\left[U(u)U^\top(u)\right]_{i,j} uK(u)\right| du$$

$$\leq L_f h^{\beta_f} \int_{\mathbb{R}} |u| K(u)du$$

$$=: L_f \|K\|_1 h^{\beta_f}.$$

This implies that, for h sufficiently small, that is for M sufficiently large,

$$\tilde{\mathbb{P}}(\lambda \leq \beta) \leq \sum_{0 \leq i,j \leq \mathbf{d}} \tilde{\mathbb{P}}(|A_{i,j} - \tilde{\mathbb{E}}(A_{i,j})| \geq \lambda_0/\{2(\mathbf{d}+1)\} - L_f\|K\|_1 h^{\beta_f})$$

$$\leq \sum_{0 \leq i,j \leq \mathbf{d}} \tilde{\mathbb{P}}(|A_{i,j} - \tilde{\mathbb{E}}(A_{i,j})| > \lambda_0/\{4(\mathbf{d}+1)\}).$$

Let us define

$$\xi_{m,i,j} = \left[U\left(\frac{T_m - t_0}{h}\right) U^\top \left(\frac{T_m - t_0}{h}\right) \right]_{i,j} K\left(\frac{T_m - t_0}{h}\right)$$

$$= \frac{1}{i!j!} \left(\frac{T_m - t_0}{h}\right)^{i+j} K\left(\frac{T_m - t_0}{h}\right).$$

By property (15), we have

$$\left| \xi_{m,i,j} - \tilde{\mathbb{E}}(\xi_{m,i,j}) \right| \leq 2\kappa.$$

Moreover, for h sufficiently small, that is for M sufficiently large, $f(t) \leq 2f(t_0)$, $\forall |t - t_0| \leq h$, and thus

$$\sum_{m=1}^{M} \widetilde{\mathrm{Var}}(\xi_{i,j}^{(m)}) \leq \sum_{m=1}^{M} \tilde{\mathbb{E}}\left[\{\xi_{i,j}^{(m)}\}^2 \right]$$

$$\leq 2f(t_0)Mh \int_{\mathbb{R}} \left| \left[U(u)U^\top(u) \right]_{i,j} \right| K^2(u)du$$

$$\leq 2f(t_0)\|K\|_2^2 Mh.$$

Applying the Bernstein inequality [see 35, p. 95], we obtain, for any $x > 0$:

$$\tilde{\mathbb{P}}\left(\frac{1}{Mh} \sum_{m=1}^{M} \left| \xi_{m,i,j} - \tilde{\mathbb{E}}(\xi_{m,i,j}) \right| > x \right) \leq 2\exp\left(-\frac{M^2 x^2}{\frac{2\|f\|_\infty \|K\|_2^2 M}{h} + \frac{4\kappa x M}{3h}} \right).$$

Then equation (27) follows if we define:

$$\mathfrak{g} = \psi\left(\frac{\lambda_0}{4(\mathbf{d}+1)}\right) \quad \text{with} \quad \psi(x) = \frac{x^2}{2\|f\|_\infty \|K\|_2^2 + \frac{4\kappa x}{3}}.$$

It remains to prove (28). Let us define the events

$$\mathcal{E}_\beta = \{\lambda > \beta\} \quad \text{and} \quad \mathcal{F} = \left\{ |\widehat{H}_{t_0} - H_{t_0}| \leq \log^{-2}(\mu) \right\} \cap \left\{ \widehat{\mathbf{d}} = \mathbf{d} \right\}.$$

Using (27), we have

$$\mathbb{P}(\overline{\mathcal{E}}_\beta) = 2\mathbb{E}[\exp(-\mathfrak{g}Mh)\mathbf{1}_\mathcal{F}] + \mathbb{P}(\overline{\mathcal{F}})$$

$$\leq 2\mathbb{E}[\exp(-\mathfrak{g}Mh)\mathbf{1}_\mathcal{F}] + 2\mathfrak{K}_1 \exp(-\mu).$$

The last line comes from Assumption (LP3). Note that under $\mathcal{F}$

$$Mh = M^{\frac{2\{\widehat{\mathbf{d}} + \widehat{H}_{t_0}\}}{2\{\widehat{\mathbf{d}} + \widehat{H}_{t_0}\}+1}} = M^{\frac{2\{\mathbf{d} + \widehat{H}_{t_0}\}}{2\{\mathbf{d} + \widehat{H}_{t_0}\}+1}} = M^{\frac{2\{\mathbf{d} + H_{t_0}\}}{2\{\mathbf{d} + H_{t_0}\}+1} + \eta},$$

with

$$|\eta| = \left| \frac{2(\widehat{H}_{t_0} - H_{t_0})}{(2\{\hat{\mathbf{d}} + \widehat{H}_{t_0}\} + 1)(2\{\mathbf{d} + H_{t_0}\} + 1)} \right| \leq 2|\widehat{H}_{t_0} - H_{t_0}|.$$

Assumption (LP2) implies that, under $\mathcal{F}$ and, for sufficiently large μ,

$$Mh \geq \left(\frac{\mu}{\log(\mu)} \right)^{\frac{2\{\mathbf{d}+H_{t_0}\}}{2\{\mathbf{d}+H_{t_0}\}+1} - \frac{2}{\log^2(\mu)}} \geq \frac{1}{2} \left(\frac{\mu}{\log(\mu)} \right)^{\frac{2\{\mathbf{d}+H_{t_0}\}}{2\{\mathbf{d}+H_{t_0}\}+1}}. \tag{45}$$

Thus, we have

$$\mathbb{P}(\overline{\mathcal{E}}_\beta) \leq 2 \exp\left(-\frac{\mathbf{g}}{2}\tau(\mu)\log^2(\mu) \right) + 2\mathfrak{K}_1 \exp(-\mu)$$
$$\leq \mathfrak{K}_2 \exp\left[-\frac{\mathbf{g}}{2}\tau(\mu)\log^2(\mu) \right],$$

for some positive constant $\mathfrak{K}_2$, that does not depend on $0 < \beta \leq \lambda_0/2$. $\qquad\square$

Lemma 10. *Let ξ be a positive random variable such that*

$$c_1 := \mathbb{E}\left[\exp\left(\eta_0 \xi^4 \right) \right] < \infty,$$

for some positive constant η_0. Then, for any $\tau \geq 1$:

$$\mathbb{E}\left[\exp\left(\tau\xi \right) \right] \leq c_1 \exp\left(c_2 \tau^{4/3} \right) \qquad where \qquad c_2 = \left(\frac{5}{16\eta_0} \right)^{1/3}.$$

Proof of Lemma 10. Defining $\zeta = (16\eta_0/5)^{1/4}\xi$, we can assume, without loss of generality that $\eta_0 = 5/16$. Let $\gamma \geq \tau$. Remark that, since

$$1 - \frac{\tau\xi}{\gamma} = \left(1 - \frac{\tau\xi}{4\gamma} \right)^4 - 6\left(\frac{\tau\xi}{4\gamma} \right)^2 + 4\left(\frac{\tau\xi}{4\gamma} \right)^3 - \left(\frac{\tau\xi}{4\gamma} \right)^4,$$

we obtain:

$$\mathbb{E}\left[\exp\left(\tau\xi \right) \right] = \exp(\gamma)\mathbb{E}\left[\exp\left(-\gamma\left(1 - \frac{\tau\xi}{\gamma} \right) \right) \right]$$
$$\leq \exp(\gamma)\mathbb{E}\left[\exp\left(\frac{3\gamma}{8}\left(\frac{\tau\xi}{\gamma} \right)^2 \right) \exp\left(\frac{\gamma}{256}\left(\frac{\tau\xi}{\gamma} \right)^4 \right) \right]$$
$$\leq \exp(\gamma + \eta)\mathbb{E}\left[\exp\left(-\eta\left(1 - \frac{3\gamma}{8\eta}\left(\frac{\tau\xi}{\gamma} \right)^2 \right) \right) \exp\left(\frac{\gamma}{256}\left(\frac{\tau\xi}{\gamma} \right)^4 \right) \right].$$

Using the fact that

$$1 - \frac{3\gamma}{8\eta}\left(\frac{\tau\xi}{\gamma} \right)^2 = \left[1 - \frac{3\gamma}{16\eta}\left(\frac{\tau\xi}{\gamma} \right)^2 \right]^2 - \left[\frac{3\gamma}{16\eta}\left(\frac{\tau\xi}{\gamma} \right)^2 \right]^2,$$

we obtain:

$$\mathbb{E}\left[\exp\left(\tau\xi\right)\right] \le \exp(\gamma + \eta)\mathbb{E}\left[\exp\left(\frac{9}{256}\frac{\tau^4\xi^4}{\eta\gamma^2} + \frac{1}{256}\frac{\tau^4\xi^4}{\gamma^3}\right)\right].$$

Taking $\gamma = \eta = \tau^{4/3}/2$, we obtain:

$$\mathbb{E}\left[\exp\left(\tau\xi\right)\right] \le \exp(\tau^{4/3})\mathbb{E}\left[\exp\left(\frac{5}{16}\xi^4\right)\right].$$

This completes the proof. □

Appendix F: Moment bounds for spacings

We need to find an accurate approximation for moments like

$$\mathbb{E}[(T_{(k)} - T_{(l)})^\alpha \mid M = m],$$

where $1 \le l < k \le K_0 \le m$, $\alpha > 0$. Here, $T_{(1)} \le \ldots \le T_{(K_0)}$ are defined as in Section 2, that is the subvector of the K_0 closest values to t_0. We assume that T admits a density f. Such moments will be considered with k and l such that, for some fixed value $t_0 \in [0, 1]$ such that $f(t_0) > 0$,

$$\frac{\max(|\lfloor t_0 m \rfloor - k|, |\lfloor t_0 m \rfloor - l|)}{m + 1} \le 8 \frac{k - l}{m + 1} \tag{46}$$

and

$$\frac{k - l}{m + 1} \quad \text{is small,} \tag{47}$$

and converges to zero when $m \to \infty$. Herein, for any real number a, $\lfloor a \rfloor$ denotes the largest integer smaller than or equal to a. These conditions on k and l allows for $(k - l)$ increasing slower than m.

Let us point out that $T_{(1)} \le \ldots \le T_{(K_0)}$ defined in section 2 is not the order statistics from a random sample of T. In fact, $T_{(k)}$, with $1 \le k \le K_0$, is the $(G + k)$−th order statistics of the sample $T_1, \ldots, T_m$. Here G is a random variable and its value is determined by the way the subvector of K_0 closest values to t_0 is built. It is important to notice that G depends of the smallest and the largest values in this subvector, but is independent of the other components of the subvector. In particular, this means that in the case where T has a uniform distribution, the law of the spacings between $T_{(1)} \le \ldots \le T_{(K_0)}$ coincides with the law of the same type of spacings between the order statistics of a uniform sample of size m on $[0, 1]$. In particular, in the uniform case, the law of $T_{(k)} - T_{(l)}$ depends only on m and $k - l$. For this reason, first we consider the case of T with uniform law. In the general case, we use the transformation by the distribution function in order to get back to the uniform case.

F.1. The uniform case

Consider U a uniform random variable on $[0, 1]$. Let $U_1, \ldots, U_m$ be an independent sample of U and let $U_{(1)}, \ldots, U_{(m)}$ be the order statistics. In the case of a uniform sample, $U_{(k)} - U_{(l)}$ and $U_{(k-l)}$ have the same distribution, that is a beta distribution $\text{Beta}(k - l, m - (k - l) + 1)$. Hence in this case, it is equivalent to study the moments of $U_{(r)}$ with $1 \leq r = k - l \leq m - 1$. The variable $U_{(r)}$ has a $\text{Beta}(r, m - r + 1)$ distribution. It also worthwhile to notice that $U_{(k)} - U_{(l)}$ and $U_{(l)}$ are independent, and the same is true for $U_{(k)} - U_{(l)}$ and $U_{(k)}$.

By elementary calculations, we have

$$\mathbb{E}\left[U_{(r)}^\alpha\right] = \frac{B(\alpha + r, m - r + 1)}{B(r, m - r + 1)} = \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{\Gamma(m + 1)}{\Gamma(m + \alpha + 1)},$$

where $B(\cdot, \cdot)$ denotes the beta function and $\Gamma(\cdot)$ the gamma function. To derive the bounds for the moments of interest, we use some existing results on the approximation of the gamma functions and the ratios of the gamma functions. The results are recalled in Section F.3 below.

Let M be a random variable taking positive integer values. In the following proposition we assume that, given the realization of $M \geq K_0$, $T_1, \ldots, T_M$ be an independent sample with uniform distribution on $[0, 1]$.

Lemma 11. *Consider $0 < \alpha \leq 3$ and $1 \leq l < k \leq m$, and let $r = k - l$. Then, for any $m \geq K_0$ in the support of M,*

$$\left| \mathbb{E}\left[(T_{(k)} - T_{(l)})^\alpha \mid M = m\right] - \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{1}{(m + 1)^\alpha} \right| \leq \frac{3}{m} \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{1}{(m + 1)^\alpha},$$

and

$$\left| \mathbb{E}\left[(T_{(k)} - T_{(l)})^\alpha \mid M = m\right] - \left(\frac{r}{m + 1}\right)^\alpha \right| \leq \left(\frac{r}{m + 1}\right)^\alpha \left[\frac{3}{m} + \frac{4}{r} + \frac{12}{mr}\right].$$

Proof of Lemma 11. Given that $M = m$, $T_{(k)} - T_{(l)}$ is distributed as $U_{(r)}$, the r−th order statistic, with $1 \leq r = k - l \leq m - 1$, of an independent sample of size m from the uniform law on $[0, 1]$. Using inequality (52) with $x = m + 1$ and $s = \alpha$, we can write

$$\left| \mathbb{E}\left[U_{(r)}^\alpha \mid M = m\right] - \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{1}{(m + 1)^\alpha} \right|$$
$$= \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{1}{(m + 1)^\alpha} \left| \frac{(m + 1)^\alpha \Gamma(m + 1)}{\Gamma(m + \alpha + 1)} - 1 \right|$$
$$\leq \frac{3}{m} \frac{\Gamma(\alpha + r)}{\Gamma(r)} \frac{1}{(m + 1)^\alpha}.$$

Next, using inequality (51) twice, with $x = r$ and $s = \alpha$, and triangle inequality

$$\left| \mathbb{E}\left[U_{(r)}^{\alpha} \mid M = m \right] - \left(\frac{r}{m+1} \right)^{\alpha} \right|$$

$$\leq \left| \mathbb{E}\left[U_{(r)}^{\alpha} \mid M = m \right] - \frac{\Gamma(\alpha+r)}{\Gamma(r)} \frac{1}{(m+1)^{\alpha}} \right| + \left(\frac{r}{m+1} \right)^{\alpha} \left| \frac{\Gamma(\alpha+r)}{r^{\alpha}\Gamma(r)} - 1 \right|$$

$$\leq \left(\frac{r}{m+1} \right)^{\alpha} \left[\frac{3}{m} \frac{\Gamma(\alpha+r)}{r^{\alpha}\Gamma(r)} + \frac{4}{r} \right]$$

$$\leq \left(\frac{r}{m+1} \right)^{\alpha} \left[\frac{3}{m} \left(1 + \frac{4}{r} \right) + \frac{4}{r} \right]. \quad \square$$

F.2. The general case

Given the realization of M, let $T_1, T_2, \ldots$ be an independent sample from T, a random variable independent of M, with an absolute continuous distribution on $[0,1]$. Let f (resp. F) (resp. Q) denote the density (resp. distribution function) (resp. quantile function) of T. We assume that F is strictly increasing on $[0,1]$ and thus Q is the inverse function for F, and Q is differentiable with $Q' = 1/f$. Then, given $M = m$, for any $1 \leq l < k \leq m$, the joint distribution of the order statistics $(T_{(k)}, T_{(l)})$ is the same as the joint distribution of $(Q(U_{(k)}), Q(U_{(l)}))$, where $U_{(1)}, \ldots, U_{(m)}$ is the order statistics of an independent uniform sample on $[0,1]$.

Assume $\inf_{t \in [0,1]} f(t) > 0$ and f is Hölder continuous around t_0, i.e. there exists $L_f > 0$, $0 < \beta_f \leq 1$, and a neighborhood of t_0 in $[0,1]$ such that for any u, v in this neighborhood, $|f(u) - f(v)| \leq L_f |u - v|^{\beta_f}$.

Lemma 12. *Let m be an integer value in the support of M. Let $t_0 \in [0,1]$, assume that k and l are satisfying the conditions* (46)-(47), *and let $r = k - l$. The for any $0 < \alpha \leq 3$,*

$$\left| \mathbb{E}\left[(T_{(k)} - T_{(l)})^{\alpha} \mid M = m \right] - \frac{\Gamma(\alpha+r)}{\Gamma(r)} \left(\frac{1}{f(t_0)(m+1)} \right)^{\alpha} \right|$$

$$\leq \frac{\Gamma(\alpha+r)}{\Gamma(r)} \left(\frac{1}{f(t_0)(m+1)} \right)^{\alpha} \left[\frac{3}{m} + C \left(\frac{r}{m+1} \right)^{\alpha\beta_f/4} \right],$$

and

$$\left| \mathbb{E}\left[(T_{(k)} - T_{(l)})^{\alpha} \mid M = m \right] - \left(\frac{r}{f(t_0)(m+1)} \right)^{\alpha} \right|$$

$$\leq \left(\frac{r}{f(t_0)(m+1)} \right)^{\alpha} \left[\frac{3}{m} + \frac{4}{r} + \frac{12}{mr} + C \left(\frac{r}{m+1} \right)^{\alpha\beta_f/4} \right]$$

$$\leq \mathfrak{c}_0 \left(\frac{r}{f(t_0)(m+1)} \right)^{\alpha} \left[\frac{1}{m} + \frac{1}{r} + \frac{1}{mr} + \left(\frac{r}{m+1} \right)^{\alpha\beta_f/4} \right],$$

with C and $\mathfrak{c}_0$ are two constants depending only on α and L_f, β_f and $f(t_0)$.

Proof of Lemma 12. In the following, we use several times the following property: for any $a, b, \alpha \geq 0$,

$$(a + b)^\alpha \leq \max(1, 2^{\alpha-1}) (a^\alpha + b^\alpha).$$

Next, given $M = m$,

$$\mathbb{E}\left[(T_{(k)} - T_{(l)})^\alpha \mid M = m\right] = \mathbb{E}\left[\{Q(U_{(k)}) - Q(U_{(l)})\}^\alpha \mid M = m\right].$$

By a first order Taylor expansion of $Q(U_{(k)})$ around the point $U_{(l)}$, we get

$$Q(U_{(k)}) - Q(U_{(l)}) = \frac{1}{f(t_0)} \left[U_{(k)} - U_{(l)}\right] \left[1 + r(m, k, l)\right], \tag{48}$$

with

$$r(m, k, l) = \int_0^1 \frac{f(t_0) - f(U_{(l)} + t[U_{(k)} - U_{(l)}])}{f(U_{(l)} + t[U_{(k)} - U_{(l)}])} dt.$$

Note that due to the fact the Q is increasing and almost surely $U_{(k)} > U_{(l)}$, the identity (48) implies that $1 + r(m, k, l) > 0$ almost surely. Using the triangle inequality and the properties of f,

$$|r(m, k, l)| \leq \frac{L_f}{f(t_0)/2} \left(|U_{(l)} - t_0|^{\beta_f} + |U_{(k)} - U_{(l)}|^{\beta_f}\right).$$

Let

$$t_m = \frac{\lfloor t_0(m - 1) \rfloor + 1}{m + 1}.$$

Note that $1/(m + 1) \leq t_m \leq m/(m + 1)$ and

$$t_m = \mathbb{E}[U_{(t_m(m+1))}].$$

Next, we can bound

$$|U_{(l)} - t_0| \leq |U_{(l)} - t_m| + |t_m - t_0| \leq |U_{(l)} - \mathbb{E}[U_{(t_m(m+1))}]| + \frac{2}{m + 1}$$

$$\leq |U_{(l)} - U_{(t_m(m+1))}| + |U_{(t_m(m+1))} - \mathbb{E}[U_{(t_m(m+1))}]| + \frac{2}{m + 1}.$$

Thus, with the convention $U_{(0)} = 0$,

$$\mathbb{E}\left[|U_{(l)} - t_0|^{\beta_f} \mid M = m\right] \leq \mathbb{E}\left[U_{(|l-t_m(m+1)|)}^{\beta_f} \mid M = m\right]$$

$$+ \mathbb{E}\left[\left|U_{(t_m(m+1))} - \mathbb{E}[U_{(t_m(m+1))}]\right|^{\beta_f} \mid M = m\right] + \left(\frac{2}{m + 1}\right)^{\beta_f}.$$

By the facts presented in the uniform case, when $l \neq t_m(m + 1)$,

$$\mathbb{E}\left[U_{(|l-t_m(m+1)|)}^{\beta_f} \mid M = m\right] = \frac{\Gamma(\beta_f + |l - t_m(m + 1)|)}{\Gamma(|l - t_m(m + 1)|)} \frac{\Gamma(m + 1)}{\Gamma(m + \beta_f + 1)},$$

and using Wendel's double inequality (50) with $s = \beta_f$, and (46), the product of the ratios of the gamma functions is bounded from above by

$$\left(\frac{|l - t_m(m+1)|}{m+1+\beta_f} \right)^{\beta_f} \left(1 + \frac{\beta_f}{m+1} \right) \le 9 \left(\frac{r}{m+1} \right)^{\beta_f}.$$

On the other hand, using Jensen's inequality and the variance of a beta distribution with parameters $t_m(m+1)$ and $(1 - t_m)(m+1)$,

$$\mathbb{E}\left[\left| U_{(t_m(m+1))} - \mathbb{E}[U_{(t_m(m+1))}] \right|^{\beta_f} \mid M = m \right]$$

$$\le \mathbb{E}^{\beta_f/2} \left[\left| U_{(t_m(m+1))} - \mathbb{E}[U_{(t_m(m+1))}] \right|^2 \mid M = m \right] = \left(\frac{t_m(1-t_m)}{m+2} \right)^{\beta_f/2}.$$

Gathering facts and using Lemma 11, there exists a constant c such that

$$\mathbb{E}\left[|U_{(l)} - t_0|^{\beta_f} \mid M = m \right] \le c \left(\frac{r}{m+1} \right)^{\beta_f/2}.$$

On the other hand, since $U_{(k)} - U_{(l)}$ is independent of $U_{(l)}$, from above and Lemma 11 we deduce that for any $0 < \alpha' \le \alpha \le 3$,

$$\mathbb{E}\left[\{U_{(k)} - U_{(l)}\}^\alpha |r(m,k,l)|^{\alpha'} \mid M = m \right] \le C \left(\frac{r}{m+1} \right)^{\alpha + \alpha'\beta_f/2} \tag{49}$$

for some constant C depending on L_f, β_f and $f(t_0)$.

Coming back to relationship (48), taking power α on both sides of the identity, we can write

$$\mathbb{E}\left[\{Q(U_{(k)}) - Q(U_{(l)})\}^\alpha \mid M = m \right] = \frac{1}{f^\alpha(t_0)} \mathbb{E}\left[U_{(r)}^\alpha \mid M = m \right] + R(m,k,l)$$

with

$$R(m,k,l) = \mathbb{E}\left[\{U_{(k)} - U_{(l)}\}^\alpha \{[1 + r(m,k,l)]^\alpha - 1\} \mid M = m \right].$$

Since for any $a > -1$ and $0 < \alpha \le 3$,

$$|(1+a)^\alpha - 1| = |(1+a)^{\alpha/2} - 1||(1+a)^{\alpha/2} + 1| \le 2|a|^{\alpha/2}(|a|^{\alpha/2} + 2),$$

using the bound (49) with $\alpha' = \alpha$ and $\alpha' = \alpha/2$,

$$|R(m,k,l)| \le c_R \left(\frac{r}{m+1} \right)^{\alpha(1+\beta_f/4)},$$

for some constant c_R depending on L_f, β_f and $f(t_0)$. It remains to apply Lemma 11 to complete the proof. $\square$

F.3. Wendel's type inequalities for gamma function ratios

Since in our case, we only need to consider $\alpha \in (0,3]$, we could use the sharp bounds for the ratio of two gamma functions, as deduced by [39]. For any $x > 0$ and $s \geq 0$, let

$$R(x,s) = \frac{\Gamma(x+s)}{\Gamma(x)}.$$

[39] proved that when $0 \leq s \leq 1$,

$$\left(\frac{1}{1+s/x}\right)^{1-s} \leq \frac{R(x,s)}{x^s} \leq 1. \tag{50}$$

Since

$$1 - \frac{s}{x} \leq \left(\frac{1}{1+s/x}\right)^{1-s}, \qquad \forall x \geq 1, 0 \leq s \leq 1,$$

we can deduce that, when $0 \leq s \leq 1$,

$$1 - \frac{1}{x} \leq 1 - \frac{s}{x} \leq \frac{R(x,s)}{x^s} \leq 1, \qquad \forall x \geq 1.$$

Next, using the recurrence formula for the gamma function, when $1 \leq s \leq 2$ we can write

$$\frac{R(x,s)}{x^s} = \left(1 + \frac{s-1}{x}\right)\frac{R(x,s-1)}{x^{s-1}}$$

and deduce

$$1 - \frac{1}{x} \leq \left(1 + \frac{s-1}{x}\right)\left(1 - \frac{s-1}{x}\right) \leq \frac{R(x,s)}{x^s} \leq 1 + \frac{s-1}{x} \leq 1 + \frac{1}{x}, \quad \forall x \geq 1.$$

For our purpose, we could deduce the following bounds: for any $0 \leq s \leq 2$,

$$1 - \frac{1}{x} \leq \frac{R(x,s)}{x^s} \leq 1 + \frac{1}{x}, \qquad \forall x \geq 1,$$

and

$$1 - \frac{1}{x-1} \leq \frac{x^s}{R(x,s)} \leq 1 + \frac{1}{x-1}, \qquad \forall x \geq 2.$$

Finally, using again the recurrence formula for the gamma function, when $2 \leq s \leq 3$, we can write

$$\frac{R(x,s)}{x^s} = \left(1 + \frac{s-1}{x}\right)\left(1 + \frac{s-2}{x}\right)\frac{R(x,s-2)}{x^{s-2}}$$

and deduce, for $2 \leq s \leq 3$, and $x \geq 2$,

$$\frac{R(x,s)}{x^s} \leq \left(1 + \frac{s-1}{x}\right)\left(1 + \frac{s-2}{x}\right) = 1 + \frac{3}{x} + \frac{2}{x^2} \leq 1 + \frac{4}{x}, \quad \forall x \geq 2,$$

and

$$\frac{x^s}{R(x,s)} \geq 1 - \frac{3x+2}{(x+2)(x+1)} \geq 1 - \frac{3}{x+2} \geq 1 - \frac{3}{x}, \quad \forall x \geq 1.$$

On the other hand,

$$\frac{R(x,s)}{x^s} \geq \left(1 + \frac{1}{x}\right)\frac{R(x,s-2)}{x^{s-2}} \geq \left(1 + \frac{1}{x}\right)\left(1 - \frac{s-2}{x}\right) \geq 1 - \frac{1}{x}$$

and

$$\frac{x^s}{R(x,s)} \leq \frac{x}{x+1}\frac{x^{s-2}}{R(x,s-2)} \leq \frac{x}{x+1}\ \frac{x}{x-(s-2)} \leq \frac{x^2}{x^2-1} \leq 1 + \frac{1}{x-1}.$$

Gathering facts, for $0 \leq s \leq 3$

$$\left|\frac{R(x,s)}{x^s} - 1\right| \leq \frac{4}{x}, \qquad \forall x \geq 2, \tag{51}$$

and

$$\left|\frac{x^s}{R(x,s)} - 1\right| \leq \frac{3}{x-1}, \qquad \forall x \geq 2. \tag{52}$$

Appendix G: Additional simulation results

G.1. The settings

In our simulations, we use three types of stochastic processes to generate the trajectories of X that we recall in the following.

- **Setting 1**: *Fractional Brownian motion.* The curves are generated using a classical fractional Brownian motion with constant Hurst parameter $H \in (0,1)$. In this case, the local regularity of the process is the same at every point. Figure 13a illustrates one realization of this setting.
- **Setting 2**: *Piecewise fractional Brownian motion.* The curves are generated as a concatenation of multiple fractional Brownian motions with different regularities, that is with different Hurst parameters for different time periods. In this case, the local regularity is no longer constant. Figure 13b illustrates one realization of this setting.
- **Setting 3**: *Integrated fractional Brownian motion.* The curves X_t are obtained as integrals $\int_0^t W_H(s)ds$, $t \in [0,1]$, of the paths of a fractional Brownian motion process W_H with constant Hurst parameter H. Here, the local regularity of the process is the same at each point but will be greater than 1, thus this setting corresponds to the case of smooth trajectories. Figure 13c illustrates one realization of this setting.

(a) Brownian motion
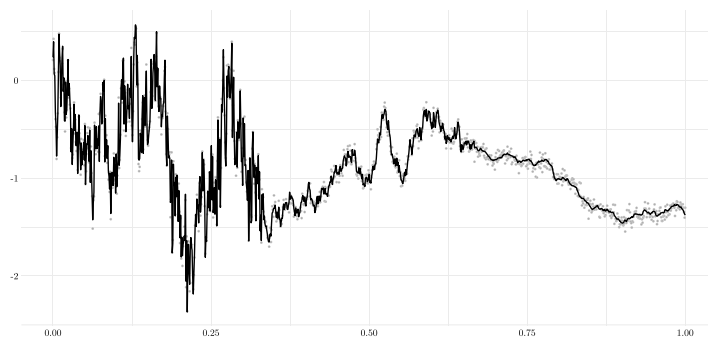
(b) Piecewise Brownian motion
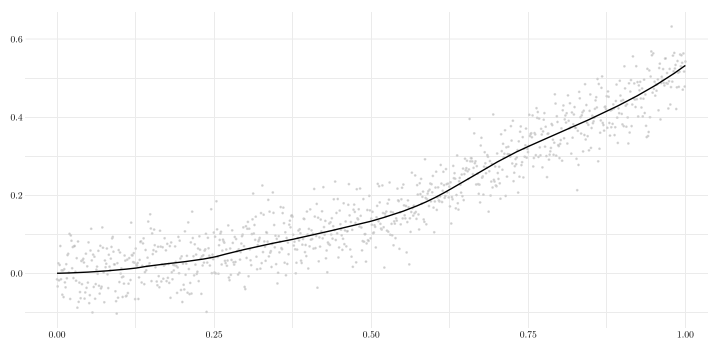
(c) Integrated Brownian motion

Fig 13. *Illustrations of simulated data generated according to the different settings. The curves correspond to the generated trajectories without noise that we aim to recover, and the grey points correspond to the noisy measurements.*

G.2. On the computation time

Figure 14a presents the violin plots of the needed time to smooth $N_1 = 1000$ curves. The results are obtained with the parameters of the simulation $(1, 1000, 1000, 300, \texttt{equi}, 0.5, 0.05)$. They correspond to the total CPU time (system time and user time) to estimate the bandwidth h_n and then estimate the curves at their sampling points. We perform these computations on a personal computer equipped with a processor Intel Core i7-6600U, CPU: 2.60GHz, RAM: 24Go and rerun the estimation 10 times. We observe that our smoothing device outperforms cross-validation and *plug-in* in terms of computation time: about 1000 times faster than the cross-validation. Let $\mathcal{H}_n$ be a set of bandwiths. For the cross-validation, we may explain these differences because of the computation of the estimator for each bandwidth in $\mathcal{H}_n$ and each curve $X^{(n)}$ of the sample $(N_1 \times \mathrm{Card}(\mathcal{H}_n)$ calls to the estimation function) while our estimator requires only one estimation of the regularity of the functions and one evaluation of the estimator per curve (N_1 calls to the estimation function). In a similar way, figure 14b presents the violin plots of the time necessary to smooth $N_1 = 1000$ curves with the parameters of the simulation $(3, 1000, 1000, 1000, \texttt{equi}, 1.7, 0.005)$. The same personal computer is used and the simulation is also run 10 times. For setting 3, our procedure is slower than for setting 1, which can be easily explained by the computation of the derivatives of each curve $X^{(n)}$. However, the computation time for the cross-validation is still not comparable with ours.
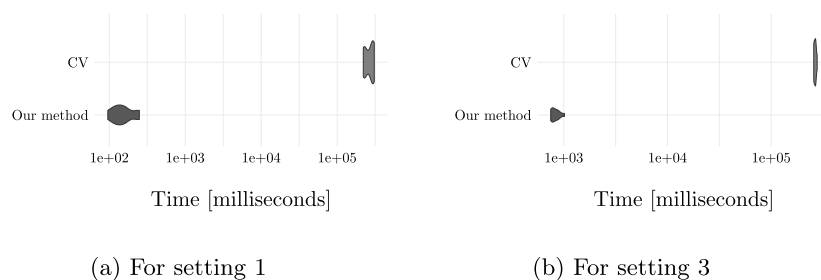

(a) For setting 1   (b) For setting 3

FIG 14. *Computational times (log scale)*

G.3. On the estimation of the local regularity

Figure 15 presents the results for the local regularity estimation for fBm with homoscedastic noise. The local esitmation of H_{t_0} is performed at $t_0 = 1/2$ which correspond to the middle of the interval. The true value of H_{t_0} is 0.5. The results show an accurate estimator $\widehat{H}_{t_0}$, except, maybe, for the simulation $(1, 250, 500, 1000, \texttt{equi}, 0.5, 0.05)$ where there is not enough curves compared to the number of sampling points.

Figure 16 presents the results for the local regularity estimation for piecewise fBm with heteroscedastic noise. The local estimations of H_{t_0} are performed at
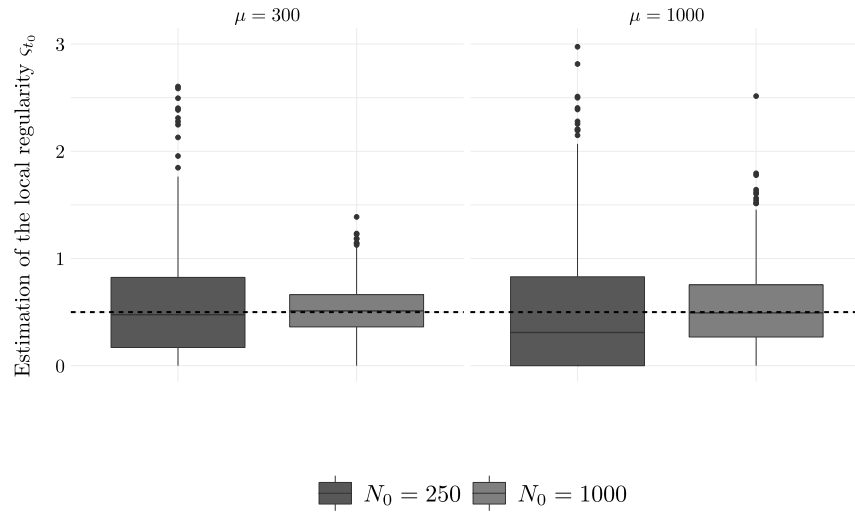
FIG 15. *Estimation of the local regularity for fBm, with constant noise variance $\sigma^2 = 0.05$, at $t_0 = 1/2$. True value: $\varsigma_{t_0} = 0.5$.*
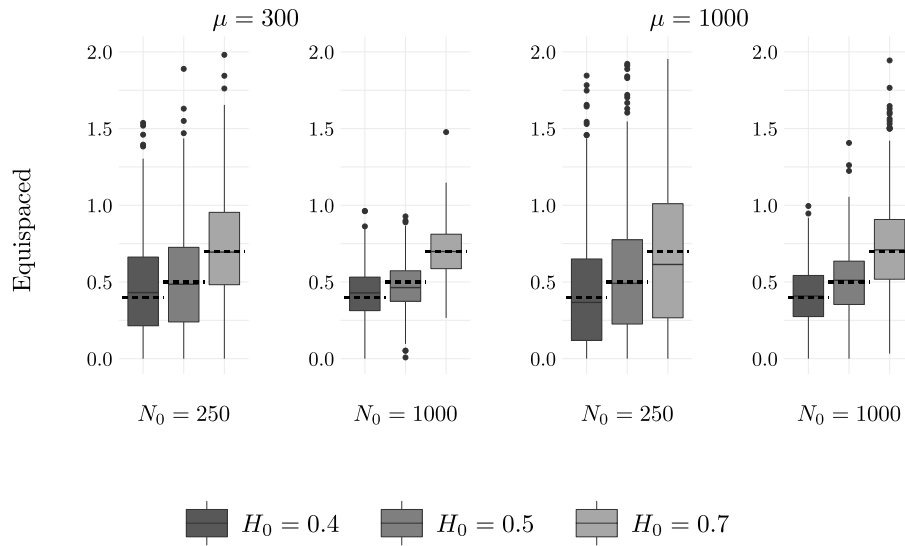

FIG 16. *Estimation of the local regularity for piecewise fBm, with non-constant noise variance $\sigma^2 = 0.04, 0.05$ and 0.07, at $t_0 = 1/6, 1/2$ and $5/6$, respectively. True values: $\varsigma_{t_0} = H_{t_0}$ equal to $0.4, 0.5$ and 0.7, respectively.*

$t_0 = 1/6, 1/2$ and $5/6$ which correspond to the middle of the interval for each regularity. The true values of H_{t_0} are $0.4, 0.5$ and 0.7, respectively. The true values of σ^2 are $0.04, 0.05$ and 0.07, respectively. The results show an accurate estimator $\widehat{H}_{t_0}$.

G.4. On the pointwise risk

For technical convenience, in our theoretical study, we only considered the case where the regularity estimator ς_{t_0} is applied with an independent sample. If one wants to smooth the curves in the learning set, one can use a leave-one-out method. That is, for each curve, one can estimate the local regularity without that curve, and smooth the curve with the estimate obtained. Our method for calculating $\widehat{H}_{t_0}$ is very fast, and such a leave-one-curve-out procedure is feasible. This idea was used to analyze the NGSIM data. However, one could also simply smooth the *learning* set curves using the same local regularity estimates obtained from this dataset. Figure 17 presents the estimation of the risks $\mathcal{R}(\widehat{X}; 1/6)$, $\mathcal{R}(\widehat{X}; 0.5)$ and $\mathcal{R}(\widehat{X}; 5/6)$ for piecewise fBm, with constant noise
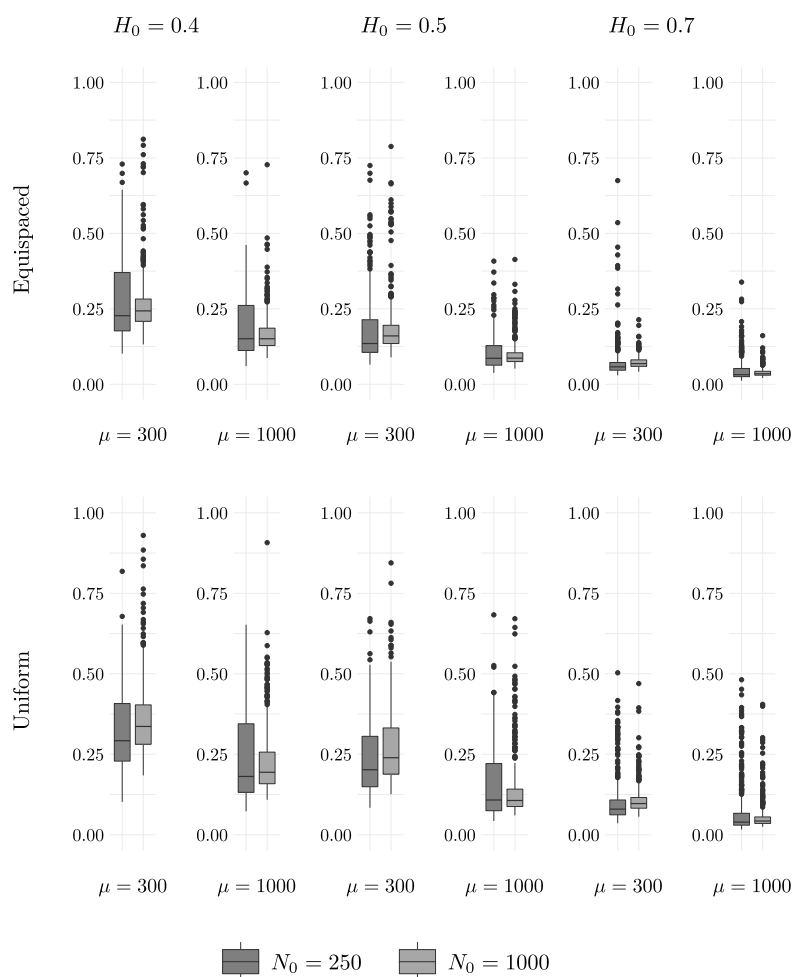
FIG 17. *Estimation of the risks $\mathcal{R}(\widehat{X}; 1/6)$, $\mathcal{R}(\widehat{X}; 0.5)$ and $\mathcal{R}(\widehat{X}; 5/6)$ for piecewise fBm, with constant noise variance $\sigma^2 = 0.05$, when the training and the test set are the same.*
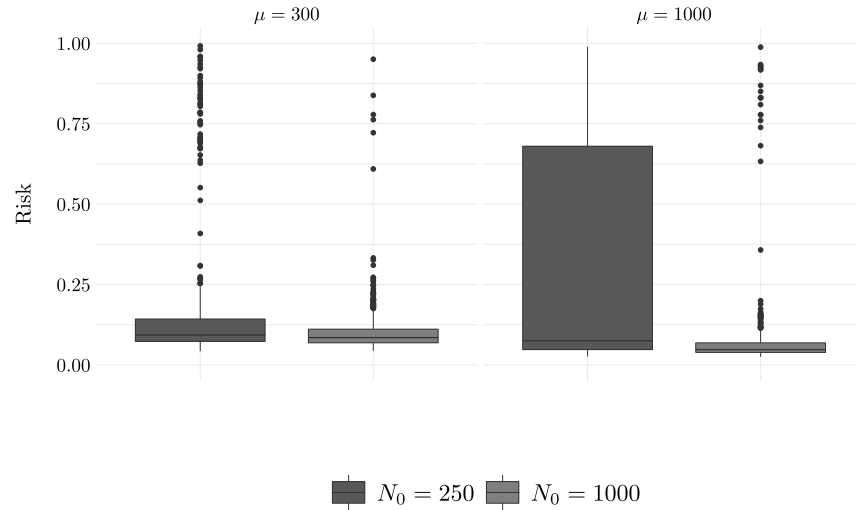
FIG 18. *Estimation of the risk* $\mathcal{R}(\widehat{X}; 0.5)$ *for smoothing the noisy trajectories of a fBm, with constant noise variance* $\sigma^2 = 0.05$.
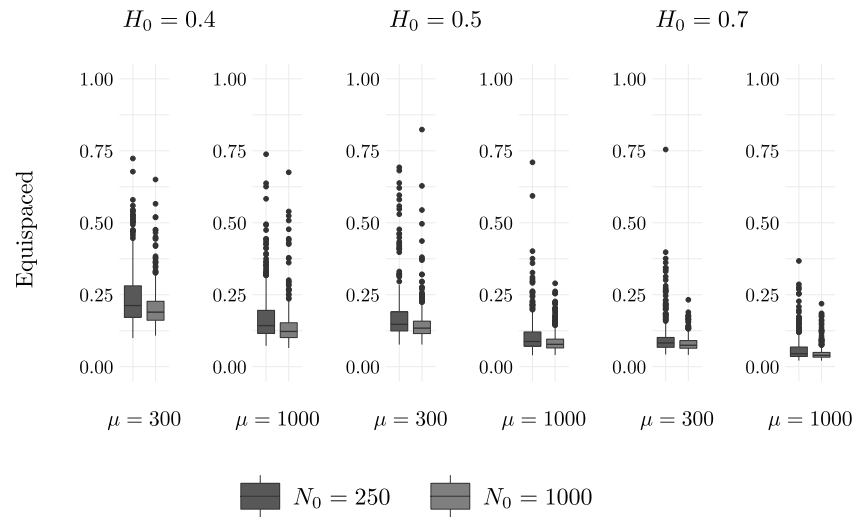

FIG 19. *Estimation of the risks* $\mathcal{R}(\widehat{X}; 1/6)$, $\mathcal{R}(\widehat{X}; 0.5)$ *and* $\mathcal{R}(\widehat{X}; 5/6)$ *for piecewise fBm, with non-constant noise variance* $\sigma^2 = 0.04, 0.05$ *and* 0.07.

variance $\sigma^2 = 0.05$, when the training and the test set are the same. The simulation results indicate that our theoretical results could be extended to the case where the *online* set is taken equal to the *learning* set, though the concentration deteriorates. The theoretical investigation of this issue is left for future work. Figure 18 presents the estimation of the risks $\mathcal{R}(\widehat{X}; 0.5)$ for fBm, with constant noise variance $\sigma^2 = 0.05$. Figure 19 presents the estimation of the risks

$\mathcal{R}(\widehat{X}; 1/6)$, $\mathcal{R}(\widehat{X}; 0.5)$ and $\mathcal{R}(\widehat{X}; 5/6)$ for piecewise fBm, with heteroscedastic noise. The conclusion are the same than the homoscedastic case.

G.5. Details on the constant of the bandwidth h_{opt}

When the regression function admits a derivative of order $\mathbf{d}$ which is Hölder continuous in a neighborhood of t_0, with exact exponent H_{t_0} and local Hölder constant L_{t_0}, the optimal bandwidth for local polynomial smoothing proposed by [37] is

$$h_{opt} = \left(\frac{C}{M}\right)^{1/(2\varsigma_{t_0}+1)} \quad \text{with} \quad C = C_{t_0} = \frac{q_2}{2\varsigma_{t_0}q_1^2}$$

with

$$q_1 = C_* L_{t_0} / \lfloor \varsigma_{t_0} \rfloor! \quad \text{and} \quad q_2 = \sigma_{t_0}^2 C_*^2.$$

Here, C_* is the constant defined on page 39 of [37]. Let us recall the notation used by [37] for the local polynomial estimator of a regression function $r(\cdot)$, at the point t, using a sample $(Y_1, T_1), \ldots, (Y_M, T_M)$:

$$\hat{r}(t_0) = \sum_{m=1}^{M} Y_m W_{Mm}(t_0).$$

In the case of the Nadaraya-Watson (NW) estimator,

$$W_{Mm}(t_0) = \frac{1}{Mh} \frac{K((T_m - t_0)/h)}{\hat{f}(t)},$$

where

$$\hat{f}(t_0) = \frac{1}{Mh} \sum_{j=1}^{M} K((T_j - t_0)/h) \approx f(t_0).$$

A closer look at the proof of Proposition 1.13 of [37] reveals that the absolute value of the bias is bounded by

$$h^{\varsigma_{t_0}} \frac{L_{t_0}}{\lfloor \varsigma_{t_0} \rfloor!} \frac{1}{\hat{f}(t_0)} \sum_{m=1}^{M} |(T_m - t_0)/h|^{\varsigma_{t_0}} K((T_m - t_0)/h) \approx \frac{h^{\varsigma_{t_0}} L_{t_0}}{\lfloor \varsigma_{t_0} \rfloor!} \int K(v)|v|^{\varsigma_{t_0}} dv.$$

Meanwhile, the conditional variance of the NW estimator given the T_m can be bounded by

$$\sigma_{t_0}^2 \frac{1}{Mh\hat{f}^2(t_0)} \frac{1}{Mh} \sum_{m=1}^{M} K^2((T_m - t_0)/h) \approx \sigma_{t_0}^2 \frac{1}{Mhf(t_0)} \int K^2(v)dv.$$

Given that $f(t_0)$ can be estimated using the data points from all the curves, the density of the $T_m^{(n)}$ can be estimated with high accuracy. We therefore use the true value $f(t_0)$ in our simulations, which in the case of a uniform design is equal to 1.

Appendix H: Traffic flow: Montanino and Punzo [27] methodology

Montanino and Punzo [27] presents a four steps methodology to make the NGSIM data usable. For a complete description of the steps, we let the reader refer to their article [27]. We briefly summarize their method here. The four steps below are applied for each trajectory separately.

Step 1. *Removing the outliers*

They remove the measurements that lead to unreliable values of the acceleration by cutting all the records above a deterministic threshold of 30 m/s^2. The missing points are interpolated using a natural cubic spline with 10 reference points before and after the outliers.

Step 2. *Cutting off the high- and medium-frequency responses in the speed profile*

They remove the noise from the signal by linear smoothing of the signal with low-pass filter. The considered one is a first-order Butterworth filter [5] with cutoff frequency of 1.25 Hz.

Step 3. *Removing the residual unphysical acceleration values, keeping the consistency requirements*

They remove residual peaks that exceed defined thresholds (varying with speed levels). For that, they move the position of the vehicle when the peak in acceleration appears in order to fulfill the thresholds. In order to prevent inconsistency, a 5th-degree polynomial interpolation with constraint on the space traveled plus minor conditions was applied on a 1s window around the peak points.

Step 4. Cutting off the high- and medium-frequency reponses generated from step 3

This step is the same as the step 2 but using the results of the step 3.

The methodology of [27] seems very specific to the NGSIM dataset, or at least some trajectory dataset, and by extension can not be easily applied to others. For using the algorithm on other trajectory datasets, their method requires some fine-tuning of the parameters.

As explained in the main text, the 1714 observation units from the I-80 dataset, available in the NGSIM study, have been recorded at moments of the day when traffic is evolving, it goes from fluid to dense traffic. Therefore, we consider that there are three groups in the data: a first group corresponding to a fluid (high-speed) traffic, a second one for in-between fluid and dense traffic, and a third groups corresponding to the dense (low-speed) traffic. Our local regularity approach, and the kernel smoothing induced, are applied for each group separately. The three group clustering was performed using a Gaussian mixture model estimated by an EM algorithm initialized by hierarchical model-based agglomerative clustering as proposed by Fraley and Raftery [12] and implemented in the **R** package mclust [34]. The optimal model is then selected according to BIC. The three resulting classes have 239, 869 and 606 velocity trajectories, respectively. Plots of randomly selected subsamples of trajectories from each groups are provided in Figure 20.
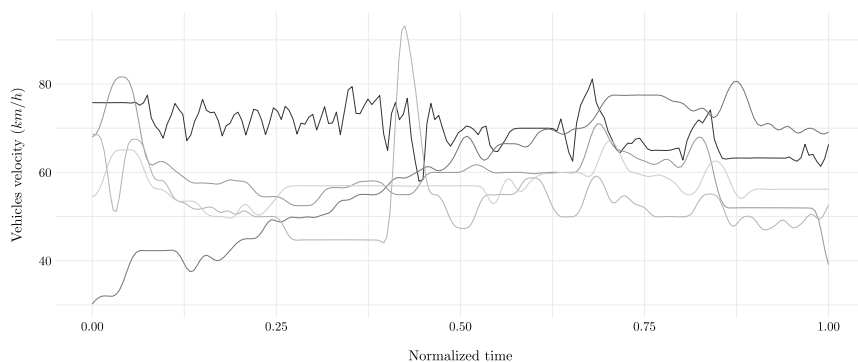
(a) Fluid/high-speed traffic
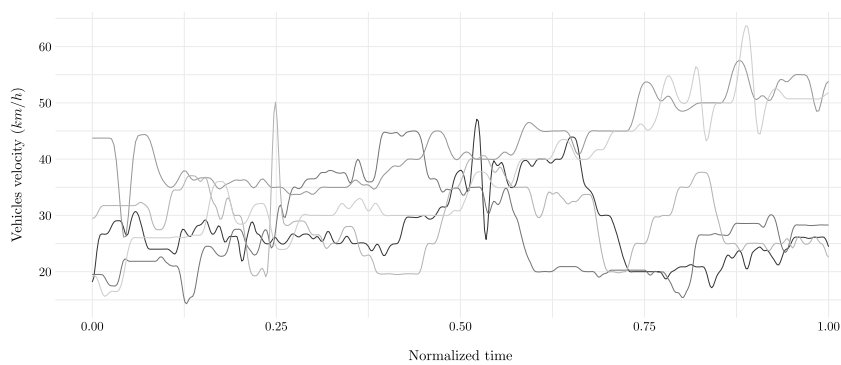
(b) In-between traffic
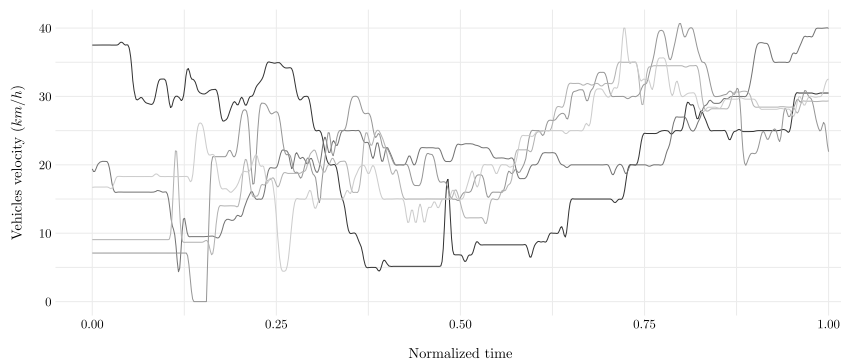
(c) Dense/low-speed traffic

FIG 20. *I-80 dataset illustration of the clusters: a sample of five velocity curves from each of the three groups of curves*
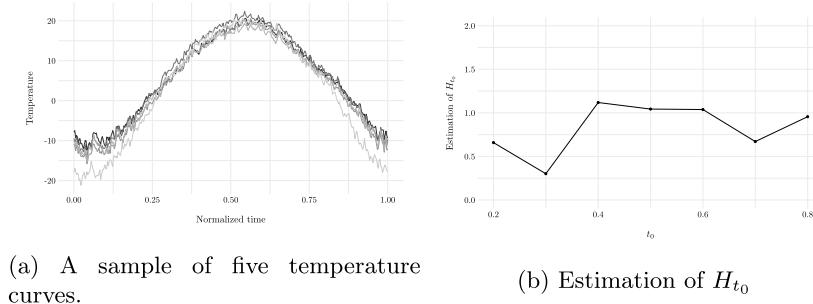
(a) A sample of five temperature curves.

(b) Estimation of H_{t_0}

Fig 21. *Canadian weather dataset illustration.*

Appendix I: Complements on the real-data applications

In this section, we point out the fact that our situation is not specific only to the traffic flow data, but can be applied to other real datasets.

I.1. Canadian weather

The Canadian Weather dataset [30, 29] records the daily temperature and precipitations in Canada averaged over the period from 1960 to 1994. Here, we are interested in the average daily temperature for each day of the year. It contains the measurements of 35 canadian stations. Here, we have $N_0 = 35$ and $\mu = 365$. A sample of five temperature curves has been plotted in the Figure 21a. Figure 21b presents the estimation of H_{t_0} for different t_0. We see that the estimation varies around 1 with $\widehat{K}_0 = 25$.

I.2. Household Active Power Consumption

The Household Active Power Consumption dataset is part of the Monash University, UEA, UCR time series regression archive [36] and was sourced from the UCI repository[1]. The data measures diverse energy related features of a house located in Sceaux, near Paris every minute between December 2006 and November 2010. In total, its represents around 2 million data points. These data are used to predict the daily power consumption of a house. Here, we are only interested in the daily voltage. The dataset contains $N_0 = 746$ time series of $\mu = 1440$ measurements. Figure 22a presents a sample of five curves from this dataset. The estimation of the local regularity H_{t_0}, plotted in Figure 22b, is around 0.5 with $\widehat{K}_0 = 73$.

---

[1]https://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power +consumption
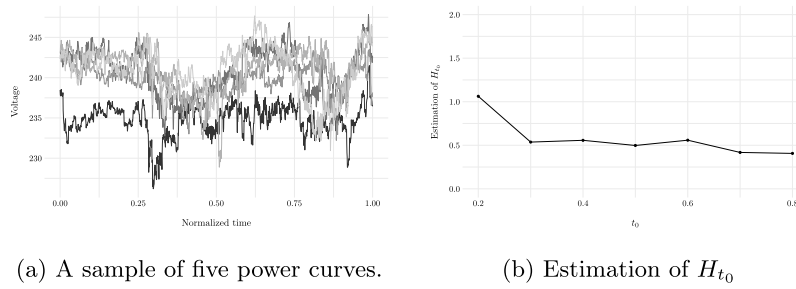
(a) A sample of five power curves.　　　(b) Estimation of H_{t_0}

Fɪɢ 22. *Household active power consumption dataset illustration.*

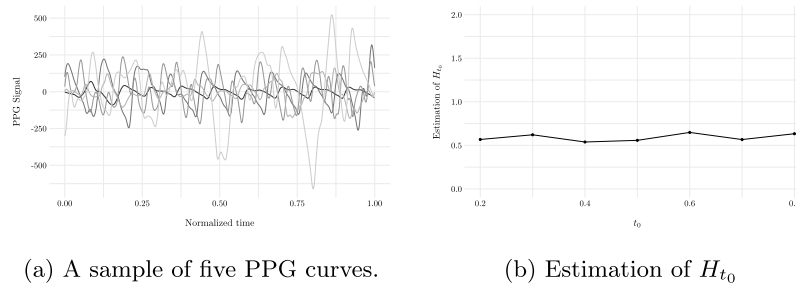(a) A sample of five PPG curves.　　　(b) Estimation of H_{t_0}

Fɪɢ 23. *PPG-Dalia dataset illustration.*

I.3. PPG-Dalia

The PPG-Dalia dataset is also part of the Monash University, UEA, UCR time series regression archive [36] and was also sourced from the UCI repository[2]. PPG sensors are widely used in smart wearable devices to measure heart rate [32]. They contain a single channel PPG and 3D accelerometer motion data recorded from 15 subjects performing various real-life activities. Measurements from each subject are segmented into 8 second windows with 6 second overlaps, resulting in $N_0 = 65000$ time series of $\mu = 512$ features. Here, we are interested in the PPG channel. A sample of five curves is plotted in Figure 23a. The estimation of the local regularity H_{t_0} is also around 0.5 (see Figure 23b) with $\widehat{K}_0 = 25$.

Acknowledgments

---

[2]https://archive.ics.uci.edu/ml/datasets/PPG-DaLiA

References

[1] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Some new asymptotic theory for least squares series: Pointwise and uniform results. *Journal of Econometrics* **186** 345 – 366. MR3343791

[2] BLANKE, D. and VIAL, C. (2014). Global smoothness estimation of a Gaussian process from general sequence designs. *Electronic Journal of Statistics* **8** 1152–1187. MR3263116

[3] BULL, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electron. J. Stat.* **6** 1490–1516. MR2988456

[4] BULL, A. D. and NICKL, R. (2013). Adaptive confidence sets in L^2. *Probab. Theory Relat. Fields* **156** 889–919. MR3078289

[5] BUTTERWORTH, S. (1930). On the theory of filter amplifiers. *Wireless Engineer* **7** 536–541.

[6] CAI, T. T. and YUAN, M. (2011). Optimal estimation of the mean function based on discretely sampled functional data: Phase transition. *Ann. Statist.* **39** 2330–2355. MR2906870

[7] CARPENTIER, A. (2015). Testing the regularity of a smooth signal. *Bernoulli* **21** 465–488. MR3322327

[8] CHAN, G. and WOOD, A. T. A. (2004). Estimation of fractal dimension for a class of non-Gaussian stationary processes and fields. *Annals of Statistics* **32** 1222–1260. MR2065204

[9] CONSTANTINE, A. G. and HALL, P. (1994). Characterizing Surface Smoothness via Estimation of Effective Fractal Dimension. *Journal of the Royal Statistical Society. Series B (Methodological)* **56** 97–113. MR1257799

[10] DONG, C., DOLAN, J. M. and LITKOUHI, B. (2017). Intention estimation for ramp merging control in autonomous driving. In *2017 IEEE Intelligent Vehicles Symposium (IV)* 1584-1589.

[11] FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications. Monographs on statistics and applied probability* **66**. Chapman & Hall, London. MR1383587

[12] FRALEY, C. and RAFTERY, A. E. (2002). Model-Based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association* **97** 611–631. MR1951635

[13] GAÏFFAS, S. (2007). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM: Probability and Statistics* **11** 344–364. MR2339297

[14] GINÉ, E. and NICKL, R. (2010). Confidence bands in density estimation. *Ann. Stat.* **38** 1122–1170. MR2604707

[15] GLOTER, A. and HOFFMANN, M. (2007). Estimation of the Hurst parameter from discrete noisy data. *Ann. Stat.* **35** 1947–1974. MR2363959

[16] GNEITING, T., ŠEVČÍKOVÁ, H. and PERCIVAL, D. B. (2012). Estimators of Fractal Dimension: Assessing the Roughness of Time Series and Spatial Data. *Statistical Science* **27** 247–277. MR2963995

[17] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimal-

ity. *The Annals of Statistics* **39** 1608–1632. MR2850214

[18] Halkias, J. and Colyar, J. (2006). NGSIM interstate 80 freeway dataset Technical Report, US Federal Highway Administration, Washington, DC, USA.

[19] Hayfield, T. and Racine, J. S. (2008). Nonparametric Econometrics: The np Package. *Journal of Statistical Software* **27** 1–32. MR2283034

[20] Henaff, M., Canziani, A. and LeCun, Y. (2019). Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic. *arXiv:1901.02705 [cs, stat]*.

[21] Horvàth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications. Springer Series in Statistics.* Springer-Verlag, New York. MR2920735

[22] Hu, Y., Zhan, W. and Tomizuka, M. (2018). A Framework for Probabilistic Generic Traffic Scene Prediction. *arXiv:1810.12506 [cs, stat]*.

[23] Johnson, W. B., Schechtman, G. and Zinn, J. (1985). Best Constants in Moment Inequalities for Linear Combinations of Independent and Exchangeable Random Variables. *The Annals of Probability* **13** 234–253. MR770640

[24] Luschgy, H. and Pagès, G. (2004). Sharp asymptotics of the functional quantization problem for Gaussian processes. *Ann. Probab.* **32** 1574–1599. MR2060310

[25] López-Pintado, S. and Romo, J. (2009). On the Concept of Depth for Functional Data. *Journal of the American Statistical Association* **104** 718-734. MR2541590

[26] Mercat, J., Zoghby, N. E., Sandou, G., Beauvois, D. and Gil, G. P. (2019). Inertial Single Vehicle Trajectory Prediction Baselines and Applications with the NGSIM Dataset. *arXiv:1908.11472 [cs]*.

[27] Montanino, M. and Punzo, V. (2013). Making NGSIM Data Usable for Studies on Traffic Flow Theory. *Transportation Research Record: Journal of the Transportation Research Board* **2390** 99–111.

[28] Punzo, V., Borzacchiello, M. T. and Ciuffo, B. F. (2009). Estimation of Vehicle Trajectories from Observed Discrete Positions and Next-Generation Simulation Program (NGSIM) Data.

[29] Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis*, 2 ed. *Springer Series in Statistics.* Springer-Verlag, New York. MR2168993

[30] Ramsay, J. O. and Silverman, B. W. (2002). *Applied Functional Data Analysis: Methods and Case Studies. Springer Series in Statistics.* Springer-Verlag, New York. MR1910407

[31] Ray, K. (2017). Adaptive Bernstein-von Mises theorems in Gaussian white noise. *Ann. Stat.* **45** 2511–2536. MR3737900

[32] Reiss, A., Indlekofer, I., Schmidt, P. and Van Laerhoven, K. (2019). Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors (Basel, Switzerland)* **19**.

[33] Revuz, D. and Yor, M. (2013). *Continuous Martingales and Brownian Motion.* Springer Science & Business Media. MR1725357

[34] Scrucca, L., Fop, M., Murphy, T. B. and Raftery, A. E. (2016).

mclust 5: clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal* **8** 289–317.

[35] SERFLING, R. J. (2009). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc., New York Wiley Series in Probability and Mathematical Statistics. MR595165

[36] TAN, C. W., BERGMEIR, C., PETITJEAN, F. and WEBB, G. I. (2020). Monash University, UEA, UCR Time Series Regression Archive. *arXiv:2006.10996 [cs, stat].* MR4250758

[37] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics.* Springer New York. MR2724359

[38] WANG, J.-L., CHIOU, J.-M. and MUELLER, H.-G. (2015). Review of Functional Data Analysis. *arXiv:1507.05135 [stat].*

[39] WENDEL, J. G. (1948). Note on the Gamma Function. *The American Mathematical Monthly* **55** 563–564. MR29448

[40] WULFE, B., CHINTAKINDI, S., CHOI, S.-C. T., HARTONG-REDDEN, R., KODALI, A. and KOCHENDERFER, M. J. (2018). Real-time Prediction of Intermediate-Horizon Automotive Collision Risk. *arXiv:1802.01532 [cs].*

[41] ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44** 2281–2321. MR3546451