The Annals of Statistics
2020, Vol. 48, No. 2, 838–862
https://doi.org/10.1214/19-AOS1825

HURST FUNCTION ESTIMATION

BY JINQI SHEN* AND TAILEN HSING**

Department of Statistics, University of Michigan, *jqshen@umich.edu; **thsing@umich.edu

This paper considers a wide range of issues concerning the estimation of the Hurst function of a multifractional Brownian motion when the process is observed on a regular grid. A theoretical lower bound for the minimax risk of this inference problem is established for a wide class of smooth Hurst functions. We also propose a new nonparametric estimator and show that it is rate optimal. Implementation issues of the estimator including how to overcome the presence of a nuisance parameter and choose the tuning parameter from data will be considered. An extensive numerical study is conducted to compare our approach with other approaches.

1. Introduction. Since the introduction by Mandelbrot and Van Ness [13], fractional Brownian motion (fBm) has found many applications, in hydrology, financial mathematics, network analysis, to name a few. Specifically, a fBm with Hurst parameter $H \in (0, 1)$ is a Gaussian process $B_H(t)$, $t \geq 0$, with stationary increments and satisfying $\mathbb{E}[B_H(t)] = 0$ and $\mathbb{E}[B_H^2(t)] = t^{2H}$, $t \geq 0$. In spatial statistics, the latter expression is referred to as a power variogram (cf. [16]). One of the appealing features of fBm is that the Hurst index characterizes the nature of dependence or, equivalently, sample path smoothness of the process globally. The book by Nourdin [14] contains a detailed introduction of the properties of fBm.

However, in many circumstances, a more flexible model is desirable that allows sample path smoothness to vary with time or location while retains some of the other key features of fBm. The multifractional Brownian motion (mBm) is such an example. The mBm was independently introduced in [11] using a moving average type construction and in [2] based on a harmonizable integral representation. Cohen [6] proved that these two definitions are equivalent up to a multiplicative deterministic function. Stoev and Taqqu [17] proposed a generalizations of these two definitions. In this paper, we use the definition of mBm in a general dimension d introduced in [9].

For convenience, let $|\cdot|$ denote both absolute value and the Euclidean norm in $\mathbb{R}^d$. Let

$$D(H) = \left(\int_{\mathbb{R}^d} \frac{1 - \cos x_1}{|\boldsymbol{x}|^{2H+d}} \, d\boldsymbol{x} \right)^{\frac{1}{2}}, \quad H \in (0, 1),$$

where x_1 is the first component of the vector $\boldsymbol{x}$.

DEFINITION 1.1. The multifractional Brownian motion $\{X(\boldsymbol{t}), \boldsymbol{t} \in (0, 1)^d\}$ is a zero-mean Gaussian process with covariance function

$$\begin{aligned} &C(\boldsymbol{t}, \boldsymbol{s}) \\ &\quad = \sigma^2 \mathcal{D}(H(\boldsymbol{t}), H(\boldsymbol{s}))\left(|\boldsymbol{t}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} + |\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} - |\boldsymbol{s} - \boldsymbol{t}|^{H(\boldsymbol{t})+H(\boldsymbol{s})}\right), \end{aligned} \tag{1.1}$$

where $\sigma^2 \in (0, \infty)$, $H(\boldsymbol{t})$ is a Hölder continuous function with range in $(0, 1)$, and

$$\mathcal{D}(H(\boldsymbol{t}), H(\boldsymbol{s})) = \{2D(H(\boldsymbol{t}))D(H(\boldsymbol{s}))\}^{-1} D^2\left(\frac{H(\boldsymbol{t}) + H(\boldsymbol{s})}{2}\right).$$

Received June 2018; revised January 2019.

MSC2010 subject classifications. Primary 62G05, 62G20; secondary 62M30.

Key words and phrases. Infill asymptotics, minimax rate, multifractional Brownian motion, nonparametric estimation.

The function H in this definition will be referred to as the Hurst function of the mBm. Note that the covariance function $C(\boldsymbol{t},\boldsymbol{s})$ in (1.1) is normalized such that $\mathrm{Var}(X(\boldsymbol{t}))=\sigma^2|\boldsymbol{t}|^{2H(\boldsymbol{t})}$. The sample paths of a mBm are still Hölder continuous but the degree of smoothness varies from point to point according to H. Also, properties of self-similarity and stationary increments only hold in a local sense.

In this paper, our primary concern is the estimation of H. We will address both cases where σ is known and unknown. We view σ as a nuisance parameter if it is unknown, in which case we also consider its estimation. While we focus on the case where σ is constant over the entire domain of $\{X(\boldsymbol{t})\}$, we will discuss some extensions where σ is allowed to vary with $\boldsymbol{t}$.

The problem of estimating $H(\boldsymbol{t})$ is challenging, at least nonstandard for time series or spatial statistics due to the nonstationarity of the process mentioned above. Fortunately, self-similarity and stationary increments still hold locally in some sense for the mBm, which ensures that H can be identified with probability one if the entire path of X is observed. Of course, one never observes the whole path in applications. Instead, we will follow the convention for this problem and assume for the most part that the data are observed on a regular grid. However, a brief discussion will be provided to address how this restriction might be relaxed.

A number of papers in the literature address the estimation of H. All of them focus on the case $d=1$ and the estimators are formulated by considering the relationship between H and moments of functions of generalized differences of gridded data. With no intention to provide a complete list, we mention [1, 5] and [3]. To the best of our understanding, [1] contains the most comprehensive results to date that unify the approaches of [5] and [3]; furthermore, it discusses the extension to a class of processes that behave like the mBm in a local sense. More details on these will be given in Sections 2 and 3. None of these papers approach the inference of H in a principled manner so as to thoroughly address issues such as how higher-order smoothness of H should be accounted for in the inference problem and how to formulate rate optimal estimators.

In the context of spatial data, a general framework for nonstationarity termed local intrinsic stationarity is developed in [10]. The mBm falls in that framework. However, they focus on the scenario that H is twice continuously differentiable and the estimator introduced is not tailored to the mBm and consequently leads to suboptimal rates. More importantly, their estimator does not satisfactorily address the subtle but important computational issues of the problem.

The main contributions and organization of the paper are summarized as follows. Our goal is to explore a range of issues concerning the inference of the Hurst function. First, we formulate the nonparametric estimation of H based on gridded data in a general dimension d, taking into account the degree of smoothness of H, for both cases of known σ^2 (Section 2) and unknown σ^2 (Section 4). The existing results focus on $d=1$ and essentially do not consider the smoothness of H beyond Hölder continuity with index 2. Second, we provide thorough asymptotic theories (e.g., Theorem 3.4 and Theorem 4.3) for our estimators under different scenarios. In that vein, we also establish for $d=1$ a lower bound for the minimax risk of estimating H by all possible estimators assuming a broad class of H. This is the first time such a lower bound is developed in the mBm context. With properly tuned parameters, the rate of our estimator matches the lower bound, which makes it rate optimal. We also address the issue of data-driven bandwidth selection (Section 5), which is important for the implementation of the procedures. Some extensions are given in Section 7. Section 7.1 considers a nongridded data scenario under which some of the key results established for gridded data continue to hold. Section 7.2 relaxes the assumption of constant σ by replacing σ^2 in (1.1) with $\sigma(\boldsymbol{t})\sigma(\boldsymbol{s})$ for some smooth, nonconstant function $\sigma(\cdot)$. A numerical study is conducted (Section 6) to illustrate the results and compare with existing approaches. For clarity of presentation and to keep the paper under page limit, all proofs and technical details are given in Section 8 and the Supplementary Material [15].

2. Basic formulation of the estimator. In this section, we consider the estimation of the Hurst function $H(\boldsymbol{t})$, $\boldsymbol{t} \in (0,1)^d$ in the covariance of the mBm in (1.1). We first assume that $\sigma^2 \in (0,\infty)$ is known. In Section 4, we will address the case where σ^2 is unknown and, in Section 7.2, some extensions to nonconstant σ^2.

Assume that we observe $X(\boldsymbol{t})$ for all $\boldsymbol{t}$ belonging to the grid

$$\Omega_n = \{(i_1, i_2, \ldots, i_d)/n, \text{ with } i_s = (j-0.5)/n \text{ for } s = 1,\ldots,d,\ j = 1,\ldots,n\}.$$

For convenience, the generic notion $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{id})$ will be used to denote the grid points. Define the differencing operator in the direction $\boldsymbol{h}$: for a function w,

$$\Delta_{\boldsymbol{h}} w(\boldsymbol{t}) := w(\boldsymbol{t}) - w(\boldsymbol{t}+\boldsymbol{h}) \quad \text{and} \quad \Delta_{\boldsymbol{h}}^j w(\boldsymbol{t}) := \Delta_{\boldsymbol{h}} \Delta_{\boldsymbol{h}}^{j-1} w(\boldsymbol{t}), \quad j \geq 1.$$

It follows that, for any $q \in \mathbb{N}$,

$$\Delta_{\boldsymbol{h}}^q w(\boldsymbol{t}) = \sum_{i=0}^{q} (-1)^i \binom{q}{i} w(\boldsymbol{t} + i\boldsymbol{h}).$$

For the rest of this section let q be fixed and define

$$g(H, \boldsymbol{u}, \boldsymbol{h}) := -\frac{1}{2} \sum_{i=0}^{q} \sum_{j=0}^{q} (-1)^{i+j} \binom{q}{i} \binom{q}{j} |\boldsymbol{u} + (i-j)\boldsymbol{h}|^{2H}. \tag{2.1}$$

The choice of q will be discussed later in condition [A3] and in Section 5.

Let us consider the properties of X around a fixed $\boldsymbol{t}$. It is well known (cf. [7]) that, as $n \to \infty$,

$$U_n(\boldsymbol{h}) := n^{H(\boldsymbol{t})} \Delta_{\boldsymbol{h}/n} X(\boldsymbol{t}) \xrightarrow{d} \sigma B_{H(\boldsymbol{t})}(\boldsymbol{h}), \tag{2.2}$$

where $B_{H(\boldsymbol{t})}$ is fBm with index $H(\boldsymbol{t})$, and $\xrightarrow{d}$ stands for convergence in distribution for the process $U_n(\boldsymbol{h})$ in the space of continuous functions endowed with the uniform metric on any compact set. Consequently,

$$\begin{aligned} n^{H(\boldsymbol{t})} \Delta_{\boldsymbol{h}/n}^q X(\boldsymbol{t}) &= n^{H(\boldsymbol{t})} \sum_{i=0}^{q} (-1)^i \binom{q}{i} \big(X(\boldsymbol{t} + i\boldsymbol{h}/n) - X(\boldsymbol{t})\big) \\ &\xrightarrow{d} \sigma \sum_{i=0}^{q} (-1)^i \binom{q}{i} B_{H(\boldsymbol{t})}(i\boldsymbol{h}). \end{aligned}$$

Recall that

$$\operatorname{Cov}\big(B_{H(\boldsymbol{t})}(\boldsymbol{s}_1), B_{H(\boldsymbol{t})}(\boldsymbol{s}_2)\big) = \frac{1}{2}\big(|\boldsymbol{s}_1|^{2H(\boldsymbol{t})} + |\boldsymbol{s}_2|^{2H(\boldsymbol{t})} - |\boldsymbol{s}_1 - \boldsymbol{s}_2|^{2H(\boldsymbol{t})}\big).$$

Thus, for any direction $\boldsymbol{h}$,

$$\mathbb{E}\big(\Delta_{\boldsymbol{h}/n}^q X(\boldsymbol{t})\big)^2 \sim n^{-2H(\boldsymbol{t})} \sigma^2 g\big(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}\big). \tag{2.3}$$

Note that the right-hand side is a one-to-one function in $H(\boldsymbol{t})$. Thus, a plausible approach might be that, for large n, if we could estimate $\mathbb{E}(\Delta_{\boldsymbol{h}/n}^q X(\boldsymbol{t}))^2$ well using a nonparametric approach based on differenced data $\Delta_{\boldsymbol{h}/n}^q X(\boldsymbol{t}_i)$ for $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$, then in principle we could also estimate $H(\boldsymbol{t})$ by inverting g. The choice of the direction $\boldsymbol{h}$ is relevant in two ways in multiple dimensions. Most importantly, since we consider gridded data, we must make sure that the $\boldsymbol{h}$ picked will lead to the full utilization of the data as differences are formed. Given that this is fulfilled, the choice of $\boldsymbol{h}$ may affect the asymptotics in a minor way but not the rate of convergence. An alternative approach is to consider a more general notion of differencing as in [5] and [1] but that is unlikely to improve the rate either. In our setting, we conjecture that the optimal choice of $\boldsymbol{h}$ is a unit vector that parallels any of the d axes. We shall fix $\boldsymbol{h}$ to be such a vector in the remainder of this paper.

Approaches similar in spirit to what was proposed above have been considered in the literature. For instance, in the case $d = 1$, [5] considers a kernel approach to estimate $\mathbb{E}(\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}))^2$ by averaging those $(\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i))^2$ for which $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$. Unfortunately, the paper contains some technical issues which were later corrected by [1]. On the other hand, [10] considers a larger class of model for a general d and adopts local linear estimation. As explained in Section 1, none of the existing works satisfactorily address the wide range of statistical issues explored in this paper.

A major problem with the approach motivated by (2.3) described earlier is that since the quantity in (2.3) is small for large n, a smoothing approach such as local polynomial regression that takes into account of higher-order smoothness could yield negative values. When that happens, there is no sensible way to define an estimate for $H(\boldsymbol{t})$. Note that the issue is nonexistent if ones uses the Nadaraya–Watson estimator with a nonnegative kernel, but the approach would not account for higher-order smoothness and the usual associated boundary issues would be exaggerated in higher dimensions. A natural remedy is to consider instead the estimation of $\mathbb{E}\log(\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}))^2 = 2\mathbb{E}\log|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t})|$. For $H \in (0, 1)$, define

$$G(H; n, \boldsymbol{h}) := -2H \log n + \log \sigma^2 + \log g(H, \mathbf{0}, \boldsymbol{h}) + \mathbb{E}\log \chi^2_1, \tag{2.4}$$

with χ^2_1 denoting a χ^2 random variable with one degree of freedom. Note that $g(H, \mathbf{0}, \boldsymbol{h}) > 0$ for any Hurst index $H \in (0, 1)$ and direction $\boldsymbol{h}$ by Lemma S.2.1 in the Supplementary Material [15]. We can also define $G(0; n, \boldsymbol{h})$ as $\lim_{H\downarrow 0} G(H; n, \boldsymbol{h})$. It is easy to see (cf. Lemma 8.3) that

$$2\mathbb{E}\log\left|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t})\right| \approx G\left(H(\boldsymbol{t}); n, \boldsymbol{h}\right). \tag{2.5}$$

Thus, our basic strategy is to estimate $\mathbb{E}\log|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t})|$, and hence the quantity $G(H(\boldsymbol{t}); n, \boldsymbol{h})$ nonparametrically based on $\log|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i)|$ for $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$.

The nonparametric approach of our choice will be local polynomial regression (cf. [8]). The advantages of the approach have been extensively documented in the literature. For vectors $\boldsymbol{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ and $\boldsymbol{i} = (i_1, \ldots, i_d)^T \in \{0, 1, 2, \ldots\}^d$, let

$$\boldsymbol{x}^{\boldsymbol{i}} = \prod_{l=1}^{d} x_l^{i_l} \quad \text{and} \quad \boldsymbol{i}! = \prod_{l=1}^{d} i_l!. \tag{2.6}$$

Fix $p \in [1, \infty)$ and let $\lceil p \rceil$ be the smallest integer no smaller than p. Sort the finite set $\{(j_1, \ldots, j_d)^T : j_l \in \{0, 1, 2, \ldots\}, \sum_{l=1}^{d} j_l \le \lceil p \rceil - 1\}$ in any manner and denote the sorted set as $\{\boldsymbol{j}_m, m = 1, \ldots, S\}$. However, we set $\boldsymbol{j}_1 = \mathbf{0}$ for convenience. Then solve

$$\begin{aligned}
&(\widehat{\beta}_{\boldsymbol{j}_1}, \ldots, \widehat{\beta}_{\boldsymbol{j}_S}) \\
&\quad = \operatorname*{argmin}_{\beta_{\boldsymbol{j}_1}, \ldots, \beta_{\boldsymbol{j}_S}} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left\{2\log\left|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i)\right| - \sum_{m=1}^{S} \beta_{\boldsymbol{j}_m}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^{\boldsymbol{j}_m}\right\}^2,
\end{aligned}$$

where K is a kernel function and $b > 0$ is bandwidth. The local polynomial regression estimator of $G(H(\boldsymbol{t}); n, \boldsymbol{h})$ is $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) := \widehat{\beta}_{\boldsymbol{j}_1}$. Define

$$\mathbf{A}(\boldsymbol{x}) = (\boldsymbol{x}^{\boldsymbol{j}_1}, \ldots, \boldsymbol{x}^{\boldsymbol{j}_S})^T \in \mathbb{R}^S,$$

and let $s_{\boldsymbol{t},p,b}(\boldsymbol{s})$ be the first element of the vector

$$K\left(\frac{\boldsymbol{s} - \boldsymbol{t}}{b}\right)\left(\sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^T\right)^{-1}\mathbf{A}\left(\frac{\boldsymbol{s} - \boldsymbol{t}}{b}\right).$$

It follows that

$$G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big) = \sum_i 2s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log\big|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i)\big|. \tag{2.7}$$

Finally, the estimator of $H(\boldsymbol{t})$ that we will focus on is

$$\widehat{H}(\boldsymbol{t}) = G^{-1}\big(G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big)\big); n, \boldsymbol{h}\big). \tag{2.8}$$

When n is sufficiently large, the upper bound for $G(H; n, \boldsymbol{h})$ is $G(0; n, \boldsymbol{h}) < \infty$ and the lower bound is $-\infty$. As it is possible to have

$$G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big) > G(0; n, \boldsymbol{h}) = \log\sigma^2 + \log g(0, \boldsymbol{0}, \boldsymbol{h}) + \mathbb{E}\log\chi^2_1,$$

we define $G^{-1}(x; n, \boldsymbol{h}) = 0$ for all the $x \ge G(0; n, \boldsymbol{h})$. The asymptotic properties of $\widehat{H}(\boldsymbol{t})$ will be considered in Section 3.

3. Asymptotic properties. This section contains two major asymptotic results as $n \to \infty$. Since we consider gridded data on a fixed bounded set, these results belong to the realm of the so-called fixed-domain or infill asymptotics (cf. [16]). Our first result is a uniform lower bound for the risk of estimating the Hurst function H that belongs to a class of smooth functions. The second result addresses the properties of the estimator $\widehat{H}$ defined in (2.8).

We begin by presenting a minimax bound for the risk of estimating H for the case $d = 1$. For functions f on (0, 1), define

$$\|f\|_s = \left[\int_0^1 \big|f(t)\big|^s\,dt\right]^{1/s}, \quad s \in [1, \infty), \quad \text{and} \quad \|f\|_\infty = \sup_{t\in(0,1)} \big|f(t)\big|.$$

Let $\lfloor x\rfloor$ be the largest integer no larger than x. For open set $B \in \mathbb{R}^d$ and constants $p \ge 0$ and $M \in (0, \infty)$, define $\mathcal{H}_p(B, M)$ as the space of $\lfloor p\rfloor$-times differentiable functions $f : B \mapsto \mathbb{R}$ such that $f^{(\lfloor p\rfloor)}$ is Hölder continuous on B with $|f^{(\lfloor p\rfloor)}(\boldsymbol{x}) - f^{(\lfloor p\rfloor)}(\boldsymbol{y})| \le M|\boldsymbol{x} - \boldsymbol{y}|^{p-\lfloor p\rfloor}$ for all $\boldsymbol{x}, \boldsymbol{y} \in B$.

THEOREM 3.1. *Consider all estimator $\widetilde{H}_n$ of H based on data $\{X((i - 1/2)/n), i = 1, \dots, n\}$. Then for any $\gamma \in (0, 1)$, $s \in [1, \infty)$, $p > 1$ and $M \in (0, \infty)$, there exists a $\delta \in (0, \infty)$ that only depends on s, p, M, γ such that*

$$\liminf_{n\to\infty} \inf_{\widetilde{H}_n} \sup_{H\in\mathcal{H}_p((0,1),M)} \mathbb{P}_H\big(\|\widetilde{H}_n - H\|_s > \delta\big(n\log^2 n\big)^{-\frac{p}{2p+1}}\big) > \gamma, \tag{3.1}$$

where $\mathbb{P}_H$ denotes the probability measure under H.

To the best of our knowledge, (3.1) is the first bound of the kind for the inference H. The proof borrows a familiar strategy from the development of minimax bounds in density and regression function estimation (cf. Chapter 2 of [19]). The core of the proof is to compute a tight bound for the Kullback–Leibler (KL) divergence between two mBms with Hurst functions that are close. It is interesting to note that the minimax bound established in Theorem 3.1 is slightly faster, due to presence of the $\log n$ term, than the corresponding bounds in classical problems such as density and regression function estimation. While the lower bound is only established for $d = 1$, we conjecture the corresponding lower bound for $d = 2$ has the rate $(n^2\log^2 n)^{-\frac{p}{2p+2}}$. Unfortunately, we have not been able to establish it so far. Below we will see that the lower bound for $d = 1$ can be attained by the estimator $\widehat{H}$ defined in (2.8) in Section 2.

We next proceed to consider the asymptotic properties of the estimator $\widehat{H}(\boldsymbol{t})$. For clarity, we list below the assumptions that will be frequently referred to in this and future sections. Let $\mathcal{H}_p(B) := \bigcup_{M=1}^\infty \mathcal{H}_p(B, M)$.

[K] K is a nonnegative kernel function with support $B_1(\mathbf{0})$ and has continuous second-order partial derivatives, where $B_r(\boldsymbol{t}_0) = \{\boldsymbol{t} : |\boldsymbol{t} - \boldsymbol{t}_0| < r\}$.

[A1] $H \in \mathcal{H}_p((0,1)^d)$ and is bounded away from 0 and 1.

[A2] $b = b_n$ is a bandwidth parameter varying with n, such that $nb \to \infty$ and $b \log^k n \to 0$ for all $k > 0$ as $n \to \infty$.

[A3] $p \geq q \geq 1$, where q is the order of differencing in defining $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$.

We also collect some common notations here for easy reference. Define

$$\psi(\boldsymbol{t}) := 2q - 2H(\boldsymbol{t}), \qquad \bar{\psi} = \inf_{\boldsymbol{t}} \psi(\boldsymbol{t}) \quad \text{and} \quad \rho_n(\boldsymbol{t}) = \log(n)/n + n^{-\psi(\boldsymbol{t})}.$$

Also, let

$$\Omega_\delta = (0,1)^d - B_\delta(\mathbf{0}),$$

where $B_\delta(\mathbf{0})$ is the d-dimensional ball centered at $\mathbf{0}$ with radius $\delta > 0$. We will focus on δ that are close to 0 (see next paragraph).

To investigate the asymptotic properties of $\widehat{H}(t)$, we consider the decomposition

$$\begin{aligned} &G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h}) \\ &\quad = \{G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) - \mathbb{E}(G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}))\} \\ &\qquad + \{\mathbb{E}(G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\}. \end{aligned}$$

The first term on the right-hand side is a centered random variable, which determines the variance of the estimator. The second term corresponds to the bias. We will address the asymptotic behavior of these terms in the following two results. In doing so, our approach is to establish uniform bounds with respect to $\boldsymbol{t}$, which requires us to focus on $\boldsymbol{t} \in \Omega_\delta$ for an arbitrarily small but fixed δ. This is due to the fact that, since $\operatorname{Var}(X(\boldsymbol{t})) = \sigma^2 |\boldsymbol{t}|^{H(\boldsymbol{t})}$, the information contained in $X(\boldsymbol{t})$ becomes increasingly scarce as $\boldsymbol{t}$ approaches zero, and consequently, the asymptotic theory for $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ and $\widehat{H}(\boldsymbol{t})$ with $\boldsymbol{t}$ close to $\mathbf{0}$ has to be dealt with differently. This is completely unrelated to the usual boundary-effect issues in nonparametric estimation.

Our first result considers the rates of bias and variance of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$. For convenience of presentation, define

$$T_1(n, b, \boldsymbol{t}) = \log(n)(b^p + (nb)^{-2 \wedge p}) + \rho_n(\boldsymbol{t}) \tag{3.2}$$

and

$$T_2(n, b, \boldsymbol{t}) = \begin{cases} (nb)^{-d} & \text{if } 2\psi(\boldsymbol{t}) > d, \\ (nb)^{-d} \log(nb) & \text{if } 2\psi(\boldsymbol{t}) = d, \\ (nb)^{-2\psi(\boldsymbol{t})} & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases} \tag{3.3}$$

In the remaining part of the paper, the statement $f(n, b, \boldsymbol{t}) = O(g(n, b, \boldsymbol{t}))$ uniformly for $\boldsymbol{t} \in \Omega_\delta$ means $\sup_{\boldsymbol{t} \in \Omega_\delta} |\frac{f(n,b,\boldsymbol{t})}{g(n,b,\boldsymbol{t})}| \leq C_\delta$ for some finite constant C_δ.

THEOREM 3.2. *Suppose that the conditions* [K], [A1]–[A3] *hold. Then for any* $\delta \in (0,1)$, *we have uniformly for any* $\boldsymbol{t} \in \Omega_\delta$,

$$\mathbb{E}(G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})) - G(H(\boldsymbol{t}); n, \boldsymbol{h}) = O(T_1(n, b, \boldsymbol{t})), \tag{3.4}$$

and

$$\operatorname{Var}(G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})) = O(T_2(n, b, \boldsymbol{t})).$$

REMARKS.

(i) In the result for bias, (3.4), the term $\rho_n(\boldsymbol{t})$ stands out as one that is absent in a classical nonparametric regression context. In some situations, for example, when the dimension d exceeds 3, $\rho_n(\boldsymbol{t})$ could play a major role in deciding the estimation rate (cf. Remark (i) following Theorem 3.4 below).

(ii) Consider the case where $p \in \mathbb{N}$ and the term $\log(n)b^p$ dominates in $T_1(n, b, \boldsymbol{t})$, that is,

$$\frac{1}{n^{2\wedge p} b^{p+2\wedge p}} \to 0 \quad \text{and} \quad \frac{\rho_n(\boldsymbol{t})}{\log(n)b^p} \to 0. \tag{3.5}$$

The proof of Theorem 3.2 shows that if

$$\begin{aligned} R(\boldsymbol{t}) &:= \mathbf{e}_1^T \left(\int_{B_1(\mathbf{0})} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})\,d\mathbf{z} \right)^{-1} \\ &\quad \times \int_{B_1(\mathbf{0})} K(\mathbf{z})\mathbf{A}(\mathbf{z}) \left(\sum_{|\boldsymbol{\alpha}|=p} \frac{D^{\boldsymbol{\alpha}} H(\boldsymbol{t})}{\boldsymbol{\alpha}!} \mathbf{z}^{\boldsymbol{\alpha}} \right) d\mathbf{z} \end{aligned} \tag{3.6}$$

is well-defined in $(0, \infty)$, where $\mathbf{e}_1$ stands for $(1, 0, \dots, 0)^T$, then

$$\operatorname{Bias}\big(G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big)\big) \sim -2\log(n) b^p R(\boldsymbol{t}). \tag{3.7}$$

This bias expression is useful for deriving the asymptotic distribution of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ and $\widehat{H}(\boldsymbol{t})$ (cf. Corollary 3.5).

(iii) The variance in Theorem 3.2 was derived by analyzing the dependence of the process $n^{H(\boldsymbol{t})}\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t})$ in $\boldsymbol{t}$. The cases $2\psi(\boldsymbol{t}) > d$ and $2\psi(\boldsymbol{t}) < d$ can be referred to as the short and long memory cases while $2\psi(\boldsymbol{t}) = d$ is the borderline of the two. Limit theorems in those cases are related to classical limit theorems in, for instance, [4] and [18]. This remark also applies to Theorem 3.3 below.

Next, we turn attention to the asymptotic distribution of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$. Let the function $F(x) := \log(x^2) - \mathbb{E}\log\chi_1^2$ have the Hermite polynomial decomposition (cf. [18])

$$F(x) = \sum_{l=2}^{\infty} c_l H_l(x), \tag{3.8}$$

where H_l denotes the Hermite polynomial of order l. Define

$$\sigma_H^2 = \sum_{l=2}^{\infty} c_l^2 l! \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \left(\frac{g(H, \boldsymbol{j}, \boldsymbol{h})}{g(H, \mathbf{0}, \boldsymbol{h})} \right)^l. \tag{3.9}$$

THEOREM 3.3. *Assume that the conditions* [K], [A1]–[A3] *hold. Then, for all* $\boldsymbol{t} \in (0, 1)^d$,

$$a_{n,b,\boldsymbol{t}}\big(G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big) - \mathbb{E}G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big)\big) \xrightarrow{d} Z(\boldsymbol{t}),$$

where the normalizing constant $a_{n,b,\boldsymbol{t}}$ *and the limit* $Z(\boldsymbol{t})$ *depend on* $\psi(\boldsymbol{t})$, *as follows*:

(i) *If* $2\psi(\boldsymbol{t}) > d$, *then* $a_{n,b,\boldsymbol{t}} = (nb)^{d/2}$ *and* $Z(\boldsymbol{t}) \sim N(0, \xi^2(\boldsymbol{t}))$ *where*

$$\xi^2(\boldsymbol{t}) = \sigma^2_{H(\boldsymbol{t})} \int_{B_1(\mathbf{0})} \omega^2(\mathbf{z})\,d\mathbf{z}, \tag{3.10}$$

with

$$\omega(\mathbf{z}) = \mathbf{e}_1^T \left(\int_{B_1(\mathbf{0})} \mathbf{A}(\boldsymbol{v})\mathbf{A}(\boldsymbol{v})^T K(\boldsymbol{v})\,d\boldsymbol{v} \right)^{-1} \mathbf{A}(\mathbf{z}) K(\mathbf{z});$$

(ii) *if* $2\psi(\boldsymbol{t}) = d$, *then* $a_{n,b,\boldsymbol{t}} = (nb)^{d/2}\log^{-1/2}(nb)$ *and* $Z(\boldsymbol{t}) \sim N(\mathbf{0}, \xi^2(\boldsymbol{t}))$, *where*

$$\xi^2(\boldsymbol{t}) = \frac{V_d\omega^2(\mathbf{0})}{g^2(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}))}\int_{B_1(\mathbf{0})}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{x})}{|\boldsymbol{x}|^{d-1}}\,d\boldsymbol{x},$$

V_d *is the volume of a* d*-dimensional unit ball and* $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ *is the coefficient of* $|\boldsymbol{u}|^{-\psi(\boldsymbol{t})}$ *in*

$$-\frac{\partial^{2q}}{\partial\phi^q\partial\eta^q}\frac{1}{2}|\boldsymbol{u} + (\phi - \eta)\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{t}+\eta\boldsymbol{h})}\Big|_{\phi=0,\eta=0};$$

(iii) *if* $2\psi(\boldsymbol{t}) < d$, *then* $a_{n,b,\boldsymbol{t}} = (nb)^{\psi(\boldsymbol{t})}$ *and the characteristic function of* $Z(\boldsymbol{t})$ *is*

$$\phi(u) = \exp\left\{\frac{1}{2}\sum_{k=2}^{\infty}\frac{(2iu)^k}{k}S_k\right\},$$

where

$$S_k = \int_{(B_1(\mathbf{0}))^{\otimes k}}\frac{\prod_{i=1}^k c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i - \mathbf{z}_{i+1})|\mathbf{z}_i - \mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})}\omega(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})}\,d\mathbf{z}_1\cdots d\mathbf{z}_k,$$

in which $\mathbf{z}_{k+1}$ *denotes* $\mathbf{z}_1$ *for notational convenience.*

REMARKS. Recall that we let $\boldsymbol{h}$ be any unit vector parallel to one of the axes. It is worth pointing out that, by Lemma S.2.2 in the Supplementary Material [15], $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ is actually a polynomial of $H(\boldsymbol{t})$ and $\langle\boldsymbol{u}, \boldsymbol{h}\rangle/|\boldsymbol{u}|$. Thus, $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ depends on $\boldsymbol{u}$ only through its direction $\boldsymbol{u}/|\boldsymbol{u}|$. For instance, for $d = 1$,

$$c_h(t, u) = (-1)^{q-1}\prod_{l=0}^{2q-1}\big(2H(t) - l\big).$$

As a consequence of this and symmetry in integration, taking into account that $g(H, \mathbf{0}, \boldsymbol{h})$ does not depend on the direction of $\boldsymbol{h}$, the limits in (i)–(iii) do not depend on the direction of $\boldsymbol{h}$ either.

The combination of Theorems 3.2 and 3.3 leads to the following result for $\widehat{H}(\boldsymbol{t})$.

THEOREM 3.4. *Suppose that the conditions* [K], [A1]–[A3] *hold. Then, uniformly for* $\boldsymbol{t} \in \Omega_\delta$, *where* δ *is an arbitrary constant in* $(0, 1)$, *we have*

$$\begin{aligned}\mathbb{E}\big(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\big) = {}&-(2\log n)^{-1}\mathbb{E}\big(G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big) - G\big(H(\boldsymbol{t}); n, \boldsymbol{h}\big)\big)\\ &+ O\big(\big(T_1(n, b, \boldsymbol{t}) + T_2(n, b, \boldsymbol{t})\big)/\log^2 n\big)\end{aligned}$$

and

$$\mathbb{E}\big(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\big)^2 = O\big(\big(T_1^2(n, b, \boldsymbol{t}) + T_2(n, b, \boldsymbol{t})\big)/\log^2 n\big),$$

where T_1, T_2 *are as defined in* (3.2) *and* (3.3). *Moreover, for any* $\boldsymbol{t} \in (0, 1)^d$, $2a_{n,b,\boldsymbol{t}} \times \log(n)(\widehat{H}(\boldsymbol{t}) - G^{-1}(\mathbb{E}G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})))$ *has the same asymptotic distribution as* $a_{n,b,\boldsymbol{t}} \times (G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) - \mathbb{E}G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}))$, *where* $a_{n,b,\boldsymbol{t}}$ *and the corresponding limits are the same as given in Theorem* 3.3.

REMARKS.

(i) To compute the rate of $\widehat{H}(\boldsymbol{t})$ in various situations using Theorems 3.2 and 3.4, we focus on the case where $2\psi(\boldsymbol{t}) > d$ and (3.5) holds. Letting $b \sim (n^d\log^2 n)^{-\frac{1}{2p+d}}$ then leads to the rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$. The following are some key scenarios for which this rate holds:

- $d = 1$, $p \geq q = 1$ and $\sup_t H(t) < 0.75$,
- $d = 2$, $p \geq q = 1$ and $\sup_{\boldsymbol{t}} H(\boldsymbol{t}) < 0.5$,
- $d \leq 2$, $p \geq q \geq 2$ and any H,
- $d = 3$, $p \in [1.5, 3)$, $q = 1$ and $\sup_{\boldsymbol{t}} H(\boldsymbol{t}) < 0.25$,
- $d = 3$, $3 > p \geq q = 2$ and any H.

If $d > 3$, the term $\rho_n(\boldsymbol{t})$ dominates the bias when picking the bandwidth b above and the rate calculations above does not apply (i.e., (3.5) fails). In future results, we will return to these 5 scenarios for comparisons.

(ii) For $d = 1$, $p \geq q \geq 2$ and any H or $d = 1$, $p \geq q = 1$ and $\|H\|_\infty < 0.75$, the rate $(n \log^2 n)^{-\frac{p}{2p+1}}$ matches the rate in the lower bound of Theorem 3.1 and, therefore, $\widehat{H}(t)$ is minimax optimal.

The following corollary describes the asymptotic distributions under the scenarios in Remark (i) above.

COROLLARY 3.5. *Suppose that the conditions* [K], [A1] *and* [A3] *hold. For* $\boldsymbol{t} \in (0,1)^d$, *assume additionally that* $2\psi(\boldsymbol{t}) > d$, $p \in \mathbb{N}$, (3.5) *holds for* $b \sim (n^d \log^2 n)^{-\frac{1}{2p+d}}$ *and that the quantity* $R(\boldsymbol{t})$ *in* (3.6) *is well-defined in* $[0, \infty)$. *Then*

$$(n^p \log^{-1} n)^{\frac{d}{2p+d}} \big(G\big(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}\big) - G\big(H(\boldsymbol{t}); n, \boldsymbol{h}\big)\big) \xrightarrow{d} N\big(R(\boldsymbol{t}), \xi^2(\boldsymbol{t})\big)$$

and

$$2\big(n^d \log^2 n\big)^{\frac{p}{2p+d}} \big(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\big) \xrightarrow{d} N\big(R(\boldsymbol{t}), \xi^2(\boldsymbol{t})\big),$$

where $\xi^2(\boldsymbol{t})$ *is as given in* (3.10).

4. Backfitting estimation when σ^2 is unknown. We have so far assumed that σ^2 in (1.1) is known. In this section, we will address the case where σ^2 is unknown and construct a slightly different estimator than $\widehat{H}(\boldsymbol{t})$.

The construction consists of three steps:

(i) First, consider a less efficient estimator, $\widehat{H}_1(\boldsymbol{t})$, that does not require the knowledge of σ^2 to crudely estimate $H(\boldsymbol{t})$.

(ii) Based on $\widehat{H}_1(\boldsymbol{t})$, estimate $\log \sigma^2$ by an estimator denoted by $\widehat{\log \sigma^2}$.

(iii) Finally, estimate $H(\boldsymbol{t})$ again by $\widehat{H}(\boldsymbol{t})$ by plugging in $\widehat{\log \sigma^2}$ for $\log \sigma^2$. The final estimator is denoted as $\widehat{H}_2(\boldsymbol{t})$.

The details are given below.

The estimator of $H(\boldsymbol{t})$ in step (i) is defined as

$$\widehat{H}_1(\boldsymbol{t}) = \frac{G(\widehat{H(\boldsymbol{t})}; n, 2\boldsymbol{h}) - G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})}{2 \log 2}, \tag{4.1}$$

where $G(\widehat{H(\boldsymbol{t})}; n, 2\boldsymbol{h})$ and $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ use the same bandwidth denoted as b_1. The proof of the following result is similar to that for Theorems 3.2 and 3.3.

THEOREM 4.1. *Assume that* [K], [A1]–[A3] *hold with b in* [A2] *replaced by b_1. Then, for any $\delta \in (0,1)$, we have uniformly for $\boldsymbol{t} \in \Omega_\delta$,*

$$\mathbb{E}\big(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\big) = O\big(b_1^p + \rho_n(\boldsymbol{t})\big)$$

and

$$\mathrm{Var}\big(\widehat{H}_1(\boldsymbol{t})\big) = O\big(T_2(n, b_1, \boldsymbol{t})\big),$$

where T_2 is defined in (3.3). *Additionally, if $2\psi(\boldsymbol{t}) \ge d$, $\widehat{H}_1(\boldsymbol{t}) - \mathbb{E}[H(\boldsymbol{t})]$ is asymptotically normal.*

REMARKS. Under the 5 scenarios listed in remark (i) following Theorem 3.4, Theorem 4.1 entails that $\widehat{H}_1(\boldsymbol{t})$ has asymptotic bias $O(b_1^p)$ and variance $O((nb_1)^{-d})$. Thus, taking $b_1 \sim n^{-\frac{d}{2p+d}}$ leads to the optimal rate $n^{-\frac{dp}{2p+d}}$. Note that this rate is slower than that of $\widehat{H}(\boldsymbol{t})$ for the case where σ^2 is known.

Next, we proceed to estimate σ^2. For some $m = 1, 2, \ldots$, obtain all of $\widehat{H}_1(\boldsymbol{t})$ for $\boldsymbol{t} \in \Omega_m = \{(i_1, i_2, \ldots, i_d)/m$, with $i_s = (j - 1/2)/m$ for $s = 1, \ldots, d$, $j = 1, \ldots, m\}$. Note that, in this step, we are still using the estimate $\widehat{H}_1$ obtained in the previous step based on data observed on Ω_n. Thus, m could be viewed as another tuning parameter. While it is possible to present our asymptotic results below for any choice of m, to streamline presentation we will fix $m \sim 1/b_1$ (which is the optimal choice in some sense) from this point on.

Define

$$\begin{aligned}\widetilde{G}(H; n, \boldsymbol{h}) &:= G(H; n, \boldsymbol{h}) - \log\sigma^2 \\ &= -2H\log n + \log g(H, \mathbf{0}, \boldsymbol{h}) + \mathbb{E}\log\chi_1^2.\end{aligned} \tag{4.2}$$

Intuitively, we could estimate $\log\sigma^2$ by $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) - \widetilde{G}(\widehat{H}_1(\boldsymbol{t}); n, \boldsymbol{h})$. However, while rare, numerically it is possible that $\widehat{H}_1(\boldsymbol{t}) \notin (0,1)$, in which case the computation of $\log g(\widehat{H}_1(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})$ would be problematic (e.g., $g(1, \mathbf{0}, \boldsymbol{h}) = 0$). By [A1], there exists a constant $\gamma > 0$ such that

$$\inf_{\boldsymbol{t}\in[0,1]^d} \big\{H(\boldsymbol{t}) \wedge \big(1 - H(\boldsymbol{t})\big)\big\} > \gamma. \tag{4.3}$$

Define the thresholded estimator

$$\widehat{H}_1^\gamma(\boldsymbol{t}) = \widehat{H}_1(\boldsymbol{t}) \cdot I\big(\widehat{H}_1(\boldsymbol{t}) \in [0, 1-\gamma/2]\big) + (1-\gamma/2)\cdot I\big(\widehat{H}_1(\boldsymbol{t}) > 1-\gamma/2\big),$$

and estimate $\log\sigma^2$ by

$$\begin{aligned}\widehat{\log\sigma^2} := \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta} &\big(G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) + 2\widehat{H}_1(\boldsymbol{t})\log n \\ &- \log g\big(\widehat{H}_1^\gamma(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}\big) - \mathbb{E}\log\chi_1^2\big).\end{aligned}$$

In the following results, let

$$T_1'(n, b_1) = \log n\big(b_1^p + (nb_1)^{-2\wedge p}\big) + n^{-1}\log^2 n + n^{-\bar{\psi}}\log n, \tag{4.4}$$

$$T_2'(n, b_1) = \begin{cases} (nb_1)^{-d} & \text{if } 2\bar{\psi} > d, \\ (nb_1)^{-d}\log(nb_1) & \text{if } 2\bar{\psi} = d, \\ (nb_1)^{-2\bar{\psi}} & \text{if } 2\bar{\psi} < d. \end{cases} \tag{4.5}$$

THEOREM 4.2. *Assume that the conditions* [K], [A1]–[A3] *hold with b in* [A2] *replaced by b_1. We have*

$$\mathbb{E}\big[\big(\widehat{\log\sigma^2} - \log\sigma^2\big)^2\big] = O\big(\big(T_1'(n,b_1)\big)^2 + T_2'(n,b_1)\big).$$

Furthermore, if $T_2'(n,b_1) = O(T_1'(n,b_1))$, then we have

$$\mathbb{E}\big(\widehat{\log\sigma^2} - \log\sigma^2\big) = O\big(T_1'(n,b_1)\big).$$

REMARKS. For the cases $2\bar{\psi} > d$, $d \le 2$ and $p \ge q \ge 2$ or $d = 3$ and $3 > p \ge q = 2$, it can be seen that the convergence rate of $\widehat{\log\sigma^2}$ based on Theorem 4.2 is $n^{-\frac{dp}{2p+d}}(\log n)^{\frac{d}{2p+d}}$ if $b_1 \sim (n^d \log^2 n)^{-\frac{1}{2p+d}}$.

Finally, we proceed to the final step and estimate $H(\boldsymbol{t})$ by

$$\widehat{H}_2(\boldsymbol{t}) = \widetilde{G}^{-1}\big(G\big(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\big) - \widehat{\log\sigma^2}; n, \boldsymbol{h}\big),$$

where the bandwidth used in $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ in this step is denoted as b_2.

THEOREM 4.3. *Assume that the conditions* [K], [A1]–[A3] *hold with b in* [A2] *replaced by b_1 and b_2. Then uniformly for $\boldsymbol{t} \in \Omega_\delta$, $\delta \in (0,1)$,*

$$\begin{aligned}
&\mathbb{E}\big(\widehat{H}_2(\boldsymbol{t}) - H(\boldsymbol{t})\big)^2 \\
&\quad = O\big(\big(\big(T_1'(n,b_1)\big)^2 + T_2'(n,b_1) + T_1^2(n,b_2,\boldsymbol{t}) + T_2(n,b_2,\boldsymbol{t})\big)/\log^2 n\big).
\end{aligned}$$

Furthermore, if b_1 and b_2 are picked so that $T_2'(n,b_1) = O(T_1'(n,b_1))$ and $T_2(n,b_2,\boldsymbol{t}) = O(T_1(n,b_2,\boldsymbol{t}))$, then we have

$$\mathbb{E}\big(\widehat{H}_2(\boldsymbol{t}) - H(\boldsymbol{t})\big) = O\big(\big(T_1'(n,b_1) + T_1(n,b_2,\boldsymbol{t})\big)/\log n\big).$$

REMARKS.

(i) Under the 5 scenarios listed in the first remark after Theorem 3.4, if $2\bar{\psi} > d$ and

$$b_1 = b_2 \sim \big(n^d \log^2 n\big)^{-\frac{1}{2p+d}},$$

the convergence rate of $\widehat{H}_2(\boldsymbol{t})$ will be $(n^d \log^2 n)^{-\frac{p}{2p+d}}$, which is the same as the convergence rate of $\widehat{H}(\boldsymbol{t})$ when σ^2 is known.

(ii) To the best of our knowledge, up to this point, the most complete asymptotic theory in the existing literature for the estimation of H when σ^2 is unknown were established for the estimators QV and IR in [1] for $d = 1$. If $p = 2$, the rate obtained in their equation (4.14) is $O_p(n^{-2/5+\varepsilon})$ for any $\varepsilon > 0$. This is slower than the optimal rate described in (i) for $\widehat{H}_2$, which is $O_p(n^{-2/5}(\log n)^{-4/5})$. As explained before, [1] essentially does not take higher-order smoothness into account in defining the estimators.

5. Selection of q and b. In this section, we consider the selection of q and b in $\widehat{H}$.

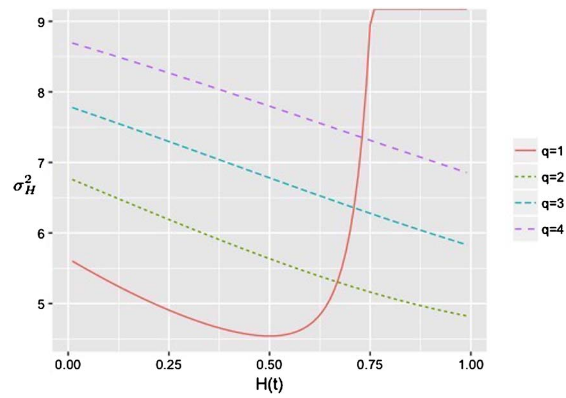

FIG. 1. *Plots of σ_H^2 for $q = 1, 2, 3, 4$.*

5.1. *Selection of q.* To consider the choice of q, we examine the asymptotic mean squared error of $\widehat{H}(\boldsymbol{t})$, or equivalently, the asymptotic bias and variance of $G(\widehat{H(\boldsymbol{t}); n, \boldsymbol{h}})$ given by Theorem 3.4 under the assumption that the bandwidth b is chosen optimally. We will focus on the 5 scenarios in the first remark after Theorem 3.4 and $p \in \mathbb{N}$.

Let $b \sim (n^d \log^2 n)^{-\frac{1}{2p+1}}$, which is the bandwidth that leads to the optimal rate for $\widehat{H}$. When $p \in \mathbb{N}$, it can be seen from (8.11) that the dominant term of the bias is $R(b, \boldsymbol{t})$ which does not depend on q. Thus, we only need to check the effect of q on variance In all of the 5 scenarios, the variance is proportional to $\sigma^2_{H(\boldsymbol{t})}$ by Theorem 3.3. As $\sigma^2_{H(\boldsymbol{t})}$ depends on both q and $H(\boldsymbol{t})$, the effect of q may be different for different values of $H(\boldsymbol{t})$.

In Figure 1, we present plots of σ_H^2 versus H for $d = 1$ and $q = 1, 2, 3, 4$, where the values of σ_H^2 are numerically computed using 80 terms in the expansion. One observes from this plot that the values of σ_H^2 for $q = 3, 4$ are uniformly larger than those for $q = 2$. For $q > 4$, σ_H^2 increases progressively but the plots for those are omitted for clarity of presentation. Consequently, there is no need to let $q > 2$. We also observe that for $H \leq 0.66$, σ_H^2 is the smallest when $q = 1$. However, for $H > 0.66$, σ_H^2 increases rapidly if $q = 1$. To conclude, $q = 2$ is generally a safe choice, but if $H(\boldsymbol{t}) < 0.66$ for most of $\boldsymbol{t}$, $q = 1$ may lead to better estimation results. The discussions for $d \geq 2$ are similar.

5.2. *Selection of b.* In this subsection, we aim to construct a bandwidth selection method for $\widehat{H}$ when $d = 1$ and $q = 2$. For a fixed $\delta > 0$, define

$$\mathrm{MISE}(b) = \mathbb{E}\left(\int_{\Omega_\delta} [\widehat{H}_{n,b}(t) - H(t)]^2 \, dt\right),$$

which is a common criterion for measuring goodness of the fit. Denote by b^* the optimal bandwidth based on MISE(b). By Theorem 3.4 and the remarks that follow, we conclude that $b^* := O(n \log^2 n)^{-\frac{1}{2p+1}}$. Instead of minimizing MISE(b) directly, minimizing a discretized version of MISE(b) is usually preferable for computations. Clearly, we can drop the term $H^2(t)$ which does not depend on b. These lead to the objective function

$$R(b) = \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{t \in \Omega_m \cap \Omega_\delta} \left(\widehat{H}^2_{n,b}(t) - 2\widehat{H}_{n,b}(t) H(t)\right)$$

for some large m where, as usual, $\Omega_m = \{(i - 1/2)/m, \, j = 1, \ldots, m\}$. Note that we highlight the dependence on data and bandwidth in the notion $\widehat{H}^2_{n,b}(t)$.

However, $R(b)$ is still not calculable as $H(t)$ is unknown. We follow an idea from [5] and replace $H(t)$ by an undersmoothed estimate. Specifically, let

$$\widehat{R}(b) = \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{t \in \Omega_m \cap \Omega_\delta} \big(\widehat{H}^2_{n,b}(t) - 2\widehat{H}_{n,b}(t)\widetilde{H}_{n,b^2}(t)\big),$$

where

$$\widetilde{H}_{n,b^2}(t) = \frac{1}{\#(B_{b^{2p}}(t) \cap \Omega_m \cap \Omega_\delta)} \sum_{t' \in B_{b^{2p}}(t) \cap \Omega_m \cap \Omega_\delta} \widehat{H}_{n,b^2}(t'),$$

and $B_{b^{2p}}(t)$ is the interval centered at t with radius b^{2p}. Now, we can select our bandwidth to be

$$\widehat{b}^* = \operatorname*{argmin}_{b \in E} \widehat{R}(b),$$

where E is the interval defined by

$$E = \big[\kappa_1 n^{-\frac{p+1}{4p+2}} (\log n)^{\frac{p}{2p+1}}, \kappa_2 \big(n \log^2 n\big)^{-\frac{1}{2p+1}}\big], \tag{5.1}$$

with κ_1, κ_2 being any two positive constants. We will refer to this as the LSCV approach, a terminology borrowed from the density estimation literature. We expect $\widehat{R}(\widehat{b}^*)$ to closely approximate $R(b^*)$ if m is chosen large enough. For $\widehat{H}_2$, with $b_1 = b_2$, we can adopt the same strategy to select the bandwidth.

6. A simulation study.

We primarily focus on the case $d = 1$ but will also briefly discuss the $d = 2$ case.

First, let $d = 1$ and assume that $H(t) = 0.5 + 0.4\sin(4\pi t)$, $t \in [0, 1]$, and $\sigma^2 = 1$. The estimators that we compare are $\widehat{H}$, $\widetilde{H}$, $\widehat{H}_1$, $\widehat{H}_2$, $\widetilde{H}_2$ plus the approaches QV and IR considered by [1, 5] and [3], where $\widetilde{H}$, $\widetilde{H}_2$ are $\widehat{H}$ and $\widehat{H}_2$ with bandwidth b selected by the LSCV approach in Section 5. Note that σ^2 is assumed known for $\widehat{H}$ and $\widetilde{H}$ but unknown for the other procedures. To implement our approaches, we set $p = 3$, $q = 2$ and use the Epanechnikov kernel. QV and IR do not produce estimation results in the boundary areas while all our approaches do. As such, all the ISEs are computed on the interval $[0.1, 0.9]$ to make the comparisons fair.

In the first set of simulations, we compare the four procedures that do not require the knowledge of σ^2, $\widehat{H}_1$, $\widehat{H}_2$ and QV and IR. The bandwidth parameter b for all estimators is taken from the set $0.3 \times 0.8^{0:19}$, and the results reported are based on the b's having the smallest MISEs. Obviously, this requires the knowledge of H and is not a data-driven bandwidth selector. The plots in Figure 2 are the empirical pointwise quartiles based on $n = 10{,}000$ and 1000 simulation runs; the true H is also plotted in red. Among the four approaches, $\widehat{H}_2$ is the clearly winner and actually substantially improves upon its preliminary procedure $\widehat{H}_1$. One could also see that the performance of IR significantly lags the other procedures for small values of $H(t)$.

To visualize the convergence rates of all eight estimators, we let the number of observations vary from 1000 to 10,000 and compute the $\sqrt{\text{MISE}}$s based on 1000 simulation runs. The results are displayed in Table 1 and Figure 3, where Figure 3 contains the plots for the log MISEs. The bandwidths for $\widehat{H}$, $\widehat{H}_1$, $\widehat{H}_2$, QV and IR are selected optimally as described in the previous paragraph while $\widetilde{H}$ and $\widetilde{H}_2$ use data-driven bandwidth determined by LCSV. By conducting linear regression on the $\log\sqrt{\text{MISE}}$s, we arrive at very crude estimates of the convergence rates, which are $n^{-0.54}$, $n^{-0.52}$, $n^{-0.44}$, $n^{-0.49}$, $n^{-0.47}$, $n^{-0.39}$ and $n^{-0.37}$ for $\widehat{H}$, $\widetilde{H}$, $\widehat{H}_1$, $\widehat{H}_2$, $\widetilde{H}_2$, QV and IR, respectively. It is important to note that although $\widehat{H}$ and $\widetilde{H}$

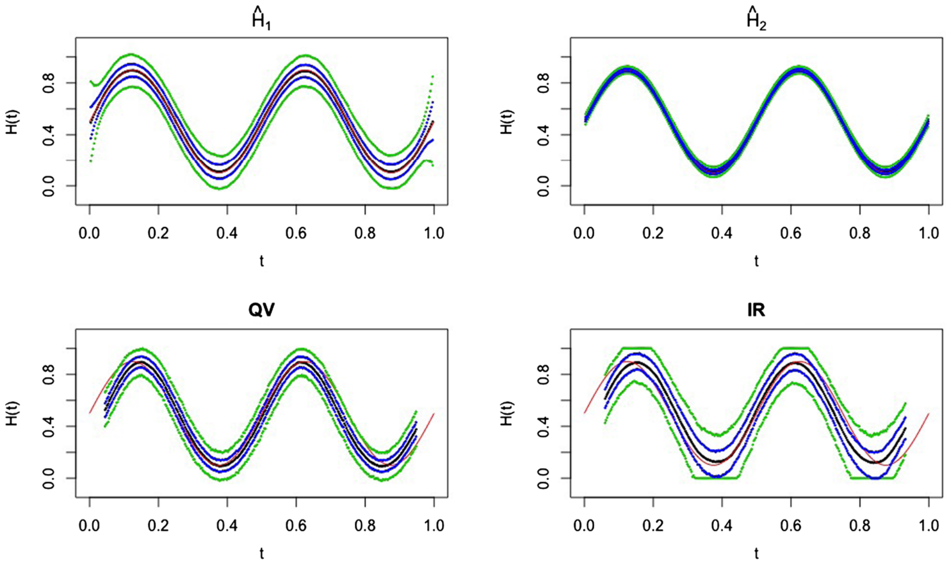

FIG. 2. *Empirical quantiles of* $\widehat{H}_1(t)$, $\widehat{H}_2(t)$, QV(t), IR(t) *with* $n = 10{,}000$ *based on* 1000 *simulation runs*; *green curves are* 5% *and* 95% *pointwise empirical quantiles*, *respectively*; *blue curves are* 25% *and* 75% *empirical quantiles*, *respectively*; *black curves are empirical means*.

performs the best in this experiment assuming the true σ^2, they will have a large bias when σ^2 is misspecified. The performance of all other methods are not affected by the change of σ^2.

To evaluate the bandwidth selector LCSV in detail, we focus on $\widehat{H}$ with $n = 1000$ and select b from the set $0.3 \times 0.8^{0:8}$. This is done 1000 times. We compare our selection approach with the "oracle" bandwidth, which is obtained by assuming we know the true $H(t)$ and select the bandwidth with smallest MISE for each run. Table 2 contains the number of times the individual bandwidths in $0.3 \times 0.8^{0:8}$ are selected by oracle and LSCV. Figure 4 contains the empirical histograms of the integrated squared errors of $\widehat{H}$ and $\widetilde{H}$. It can be seen that the bandwidths selected by two approaches are fairly close. The result for $\widetilde{H}_2$ are very similar, and we omit them here. However, the bandwidth selection method does not work well for $\widetilde{H}_1$ and tends to select the smallest bandwidth in the range of candidate bandwidths.

Finally, we present some simulation results for the case $d = 2$ to illustrate the performance of our three estimators $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$. The model that we consider is the mBm on $[0, 1] \times [0, 1]$ with Hurst function $H(x, y) = 0.5 + 0.4 \sin(4\pi x) \sin(2\pi y)$, where we assume that the

TABLE 1
$\sqrt{\text{MISE}}$ *for all estimators with a range of sample sizes*

n	QV	IR	$\widehat{H}_1$	$\widehat{H}_2$	$\widehat{H}$	$\widetilde{H}_2$	$\widetilde{H}$
1000	0.1849	0.1792	0.1602	0.0576	0.01574	0.0604	0.0186
2000	0.1451	0.1388	0.1131	0.0408	0.01060	0.0431	0.0131
3000	0.1214	0.1201	0.0951	0.0328	0.00845	0.0365	0.0106
4000	0.1085	0.1071	0.0845	0.0290	0.00736	0.0308	0.0091
5000	0.0993	0.0987	0.0774	0.0263	0.00656	0.0285	0.0082
6000	0.0926	0.0932	0.0720	0.0243	0.00595	0.0264	0.0074
7000	0.0869	0.0866	0.0671	0.0221	0.00545	0.0241	0.0068
8000	0.0821	0.0834	0.0628	0.0212	0.00504	0.0236	0.0064
9000	0.0790	0.0800	0.0595	0.0193	0.00473	0.0216	0.0059
10,000	0.0751	0.0758	0.0558	0.0183	0.00449	0.0206	0.0057

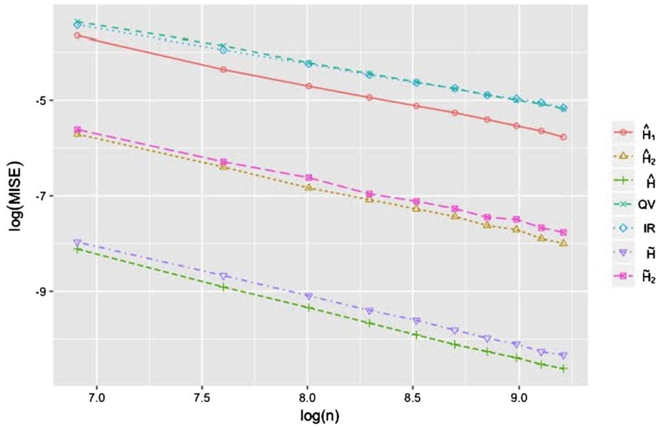

FIG. 3. *Plot of* log(MISE) *versus* $\log n$.

process is observed on a regular grid of size 166×166. For ease of presentation, we only show the estimation results on the circle $(x-0.5)^2+(y-0.5)^2=0.25^2$. As before, we let $p=3$ and $q=2$ where the direction of differencing is taken as $\boldsymbol{h}=(0,1)$. With bandwidths equal to $0.15 \times 0.8^{0:5}$, 300 runs of simulations were carried out. Figure 5 presents the the empirical quantiles of the estimation results for this simulation experiment using the optimal bandwidths based on MISE. One can see that $\widehat{H}$ continues to perform quite well assuming that σ^2 is known, whereas for $\widehat{H}_1$, $\widehat{H}_2$ that do not require a priori knowledge of σ^2, $\widehat{H}_2$ outperforms $\widehat{H}_1$ by a wide margin.

7. Discussions and extensions. In this section, we briefly discuss the problems of relaxing the assumptions of gridded data and constant variance.

7.1. *Nongridded data.* So far, we have focused on the estimation of the Hurst function of the mBm based on gridded data. In some specialized non-gridded settings, our approach can be readily modified to produce estimators that have essentially the same rates of convergence as for the gridded case. To demonstrate, we focus on $d=1$ and follow the approach of [12]. Assume that we have observations $X(t_i)$, $1 \le i \le n$, with $t_i = \varphi((i-1/2)/n)$, where φ satisfies the following condition:

[B] $\varphi \in \mathcal{H}_{p+1}([0,1])$, $p \ge 1$, is a strictly monotone and surjective mapping from $[0,1]$ to $[0,1]$, with its first order derivative $\varphi^{(1)}$ bounded away from 0.

Observe that [B] guarantees that the gaps between neighboring t_i's are of the order n^{-1}. Denote the modified estimators of $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$ for this setting as $\widehat{H}'$, $\widehat{H}_1'$ and $\widehat{H}_2'$, respectively, whose definitions involve some new notation to be introduced below. For any given $t \in (0,1)$ and $q=1,2,\dots,$ define $x_i(t)=\varphi(\varphi^{-1}(t)+i/n)$, $i=0,\dots,q$, and $a_i=(-1)^i\binom{q}{i}$. Also, let the symbol $\Delta^q_{1/n}\mathfrak{X}(t)$ stand for the quantity $\sum_{i=0}^q a_i X(x_i(t))$. Observe that $x_i(t_j)=t_{i+j}$,

TABLE 2
Frequencies with which various bandwidths are picked by LSCV and oracle for $\widehat{H}$ *based on* 1000 *simulation runs*

	0.3×0.8^3	0.3×0.8^4	0.3×0.8^5	0.3×0.8^6	0.3×0.8^7	0.3×0.8^8
LSCV	434	413	115	29	4	5
Oracle	60	584	328	27	1	0

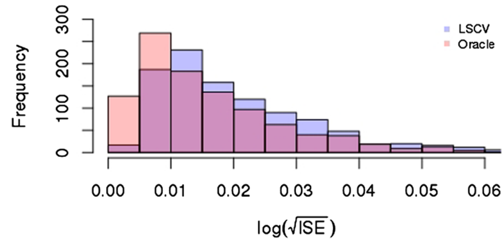

FIG. 4. *Empirical histograms of* $\log(\sqrt{\mathrm{ISE}})$ *based on* $\widehat{H}$ *with the oracle bandwidth and that selected by LSCV.*

which shows that $\Delta^q_{1/n}\mathfrak{X}(t_j)$ can be computed from data. It can be shown (cf. (2.5)) that

$$2\mathbb{E}\log\bigl|\Delta^q_{1/n}\mathfrak{X}(t)\bigr| \approx G\bigl(H(t); n, \varpi(t)\bigr),$$

where $\varpi(t) := \varphi^{(1)}(\varphi^{-1}(t))$. Thus, $G(H(t); n, \varpi(t))$ can be estimated nonparametrically by the local polynomial estimator based on the log-transformed differenced data $2\log|\Delta^q_{1/n}\mathfrak{X}(t_j)|$, as was done for the gridded case. This motivates the new estimators $\widehat{H}'$, $\widehat{H}'_1$ and $\widehat{H}'_2$.

Specifically, the definition of $\widehat{H}'_1(t)$ is unchanged from $\widehat{H}_1(t)$ since the definition does not involve inversion of $G(H; n, h)$. For $\widehat{H}'$ and $\widehat{H}'_2$, however, since we need to solve $G(H; \widehat{n, \varpi}(t))$ for H, the function $\varpi(t)$ must be estimated. The condition [B] implies that $\varpi(t)$ is differentiable, from which it is easily concluded that $\varpi(t)$ can be estimated with precision $O(1/n)$ using the pairs $(i/n, t_i)$; for instance, a smoothed version of the naive estimator $\sum_i n(t_{i+1} - t_i)I_{(t_i,t_{i+1}]}(t)$ will suffice. Thus, to define $\widehat{H}'$ and $\widehat{H}'_2$, we first estimate $\varpi(t)$ by some $\widehat{\varpi}(t)$ and then proceed to estimate $H(t)$ in much the same way as in $\widehat{H}$ and $\widehat{H}_2$ using $\widehat{\varpi}(t)$ and the $\Delta^q_{1/n}\mathfrak{X}(t_j)$. Section S.6 in the Supplementary Material [15] shows that the rates of convergence of $\widehat{H}'$, $\widehat{H}'_1$ and $\widehat{H}'_2$ are largely unchanged from their gridded-data

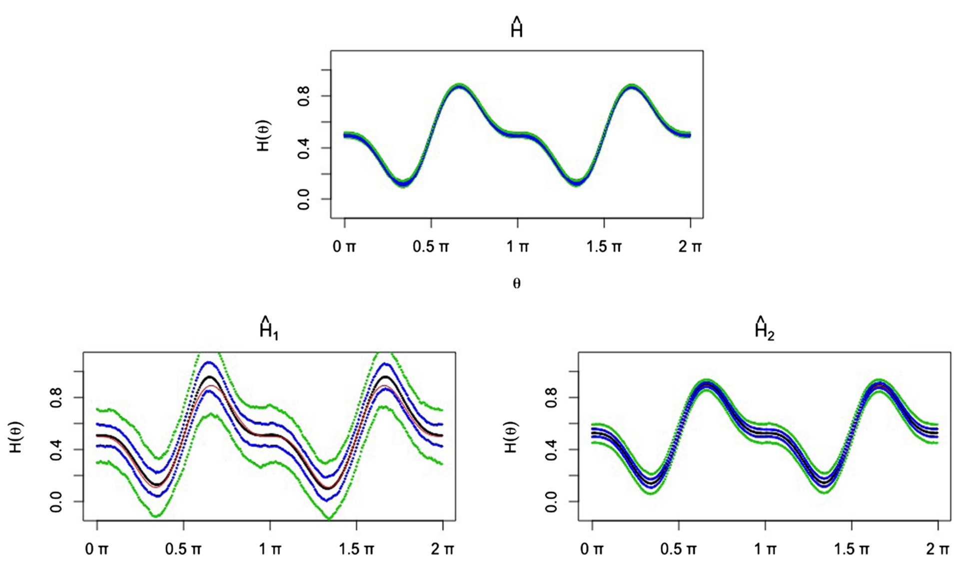

FIG. 5. *Empirical quantiles of* $\widehat{H}$, $\widehat{H}_1$, $\widehat{H}_2$ *on the circle* $(x-0.5)^2+(y-0.5)^2 = 0.25^2$ *based on* 300 *simulation runs*; $\theta = \arctan((y-0.5)/(x-0.5))$; *red curves are the true function*; *green curves are* 5% *and* 95% *pointwise empirical quantiles*, *respectively*; *blue curves are* 25% *and* 75% *empirical quantiles*, *respectively*; *black curves are empirical means*.

counterparts $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$. A simulation study is also included there to demonstrate this numerically.

We could go beyond the setting of [B] and consider data locations that are less regular; for instance, the t_i's are distributed as i.i.d. uniform [0,1] or belong to a Poisson point process. In that case, one might be able to exclude those data for which the gaps are not of the order n^{-1} and also improvise in some ways in $\widehat{H}'$, $\widehat{H}_1'$ and $\widehat{H}_2'$ to still obtain meaningful estimators. However, much of what we know so far in that regard is still preliminary and will require more investigations.

7.2. *Nonconstant variance.* Consider the model

$$Y(\boldsymbol{t}) = \sigma(\boldsymbol{t})X(\boldsymbol{t}),$$

where X is mBm as defined in Definition 1.1 with $\sigma = 1$, and $\sigma(\boldsymbol{t})$ a function that determines the variance of $Y(\boldsymbol{t})$. Allowing $\sigma(\boldsymbol{t})$ to vary with $\boldsymbol{t}$ broadens the mBm process considerably. As before, assume that Y is observed on a grid.

First, consider the case that $\sigma(\boldsymbol{t})$ is known. Define

$$G(H; n, \boldsymbol{h}, \sigma) := -2H \log n + \log \sigma^2 + \log g(H, \mathbf{0}, \boldsymbol{h}) + \mathbb{E} \log \chi_1^2.$$

As the constant variance case, one can first obtain $G(\widehat{H(\boldsymbol{t}); n, \boldsymbol{h}}, \sigma(\boldsymbol{t}))$ by local polynomial regression based on data $\Delta^q_{\boldsymbol{h}/n} Y(\boldsymbol{t}_i)$ and then estimate $H(\boldsymbol{t})$ by

$$\widehat{H}'(\boldsymbol{t}) := G^{-1}\big(G\big(\widehat{H(\boldsymbol{t}); n, \boldsymbol{h}}, \sigma(\boldsymbol{t})\big); n, \boldsymbol{h}, \sigma(\boldsymbol{t})\big). \tag{7.1}$$

Intuitively, the properties of $\widehat{H}'(\boldsymbol{t})$ should be quite similar to those of $\widehat{H}(\boldsymbol{t})$ in the constant variance case so long as $\sigma(\boldsymbol{t})$ is sufficiently smooth. To that end, we define

[S] $\sigma(\cdot) \in \mathcal{H}_p((0,1)^d)$ and $\inf_{\boldsymbol{t}} \sigma(\boldsymbol{t}) > \delta$ for some $\delta > 0$.

Propositions S.7.1 and S.7.2 of the Supplementary Material [15] show that the essence of Theorem 8.2 and Theorem 3.2 established for the constant variance assumption continues to hold for the nonconstant variance case under [S]. In particular, it can be seen that $\widehat{H}'(\boldsymbol{t})$ continues to achieve the minimax convergence rate of $(n^d \log^2 n)^{-\frac{p}{2p+d}}$ under the two scenarios:

- $d \le 2$, $p \ge q \ge 2$ and any H,
- $d = 3$, $3 > p \ge q = 2$ and any H.

Next, we discuss the case where $\sigma(\boldsymbol{t})$ is unknown. First, we consider $\widehat{H}_1$ which does not require knowing $\sigma(\boldsymbol{t})$. Following the same steps as in the proof of Theorem 4.1 and Proposition S.7.2, it is straightforward to conclude that the optimal convergence rate of $\widehat{H}_1$ based on Y is the same as that based on X when $d \le 3$ and $q \ge 2$, which is $n^{-\frac{dp}{2p+d}}$. This is somewhat worse than the minimax rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$. One possible way to improve upon $\widehat{H}_1$ is to consider a modified version of $\widehat{H}_2$ obtained as follows. Step 1: estimate $H(\boldsymbol{t})$ by $\widehat{H}_1(\boldsymbol{t})$, step 2: based on $\widehat{H}_1$, estimate $\sigma(\boldsymbol{t})$ nonparametrically by some $\widehat{\sigma}(\boldsymbol{t})$, and step 3: based on $\widehat{\sigma}(\boldsymbol{t})$, estimate $H(\boldsymbol{t})$ by $\widehat{H}(\boldsymbol{t})$. There are a few challenges in making this approach work. First, it involves the choices of three smoothing parameters, one for each of $\widehat{H}_1$, $\widehat{\sigma}$ and $\widehat{H}$. This makes the implementation difficult. Also, in order for $\widehat{H}_2(\boldsymbol{t})$ to achieve the rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$, it is necessary that $\widehat{\sigma}(\boldsymbol{t})$ first achieves the rate $n^{-\frac{dp}{2p+d}}(\log n)^{-\frac{d}{2p+d}}$. We conjecture that this is possible only if $H(\boldsymbol{t}) \le 0.5$; see additional discussions and simulation results in Section S.7 of the Supplementary Material [15]. A complete solution of this problem is beyond the scope of this paper.

8. Proofs. Due to space limitation, we only include proofs of Theorem 3.1, Theorem 3.2 and Theorem 3.3 in this section. Proofs for other results can be found in the Supplementary Material [15].

8.1. *Proof of Theorem* 3.1. The proof is based on the general approach introduced in Theorem 2.5 of [19], which, in our context, can be described as follows. For candidate Hurst functions H_1 and H_2, let $K(H_1\|H_2)$ be the Kullback–Leibler divergence between mBms with Hurst functions H_1 and H_2 observed on $\Omega_n = \{(j-1/2)/n,\ j=1,\dots,n\}$. Suppose that $H_{i\delta}$, $i=0,1,\dots,N_\delta$ are $N_\delta+1$ candidate Hurst functions in $\mathcal{H}_p((0,1),M)$ satisfying

$$\|H_{j\delta}-H_{k\delta}\|_q \ge 2\delta > 0, \quad 0\le j,k\le N_\delta, \tag{8.1}$$

and

$$\frac{1}{N_\delta}\sum_{i=1}^{N_\delta} K(H_{i\delta}\|H_{0\delta}) \le \alpha \log(N_\delta), \tag{8.2}$$

where $\alpha\in(0,1/8)$. Then

$$\inf_{\widetilde{H}_n}\sup_{H\in\mathcal{H}_p((0,1),M)} \mathbb{P}_H\big(\|\widetilde{H}_n-H\|_q\ge\delta\big) \ge \frac{\sqrt{N_\delta}}{1+\sqrt{N_\delta}}\left(1-2\alpha-\sqrt{\frac{2\alpha}{\log N_\delta}}\right) > 0.$$

We proceed to construct a set of functions that satisfy the conditions (8.1) and (8.2) above where δ and N_δ depend on n, and δ corresponds to the rate in the theorem.

First, let $H_{0,\delta}\equiv 1/2$, which is the Hurst function of the standard Brownian motion. Define

$$\kappa(x) = \begin{cases} \exp\left\{-\dfrac{1}{x+\frac{1}{2}}-\dfrac{1}{x-\frac{1}{2}}\right\} & -\dfrac{1}{2}<x<\dfrac{1}{2}, \\ 0 & \text{otherwise.}\end{cases}$$

It is clear that κ is a nonnegative function with support $(-1/2,1/2)$ and has derivatives of all orders. Let

$$M_1 = \sup_x \big|\kappa^{(\lfloor p\rfloor+1)}(x)\big|/M \quad \text{and} \quad \phi(x) = \kappa(x)/M_1.$$

Also, for some sequence $m=m_n$, set

$$\phi_{mj}(x) = m^{-p}\phi(mx-j), \quad j=1,\dots,m-1.$$

We will specify m later and define δ and N_δ in terms of m.

It is clear that the support of ϕ_{mj} is $((j-1/2)/m,(j+1/2)/m)$, and therefore the ϕ_{mj}'s have nonoverlapping supports. For any vector $\boldsymbol{a}=(a_1,\dots,a_{m-1})$ with each $a_j=0$ or 1, define

$$H_{\boldsymbol{a}}(t) = \frac{1}{2}+\sum_{j=1}^{m-1} a_j\phi_{mj}(x). \tag{8.3}$$

Clearly, for sufficiently large m and all $\boldsymbol{a}$,

$$\begin{aligned}
&\sup_{x,y\in(0,1)} \frac{|H_{\boldsymbol{a}}^{(\lfloor p\rfloor)}(x)-H_{\boldsymbol{a}}^{(\lfloor p\rfloor)}(y)|}{|x-y|^{p-\lfloor p\rfloor}} \\
&\quad\le \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|\phi^{(\lfloor p\rfloor)}(mx-1)-\phi^{(\lfloor p\rfloor)}(my-1)|}{|x-y|^{p-\lfloor p\rfloor}}
\end{aligned}$$

$$\leq M \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|mx-my|}{|x-y|^{p-\lfloor p\rfloor}}$$

$$\leq M \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|mx-my|^{p-\lfloor p\rfloor}}{|x-y|^{p-\lfloor p\rfloor}} = M.$$

For different $\boldsymbol{a}$ and $\boldsymbol{a}'$, we have

$$\|H_{\boldsymbol{a}} - H_{\boldsymbol{a}'}\|_q = \left\|\sum_j (a_j - a_j')\phi_{mj}\right\|_q = m^{-p-1/q}\left(\sum_j |a_j - a_j'|\right)^{1/q} \|\phi\|_q.$$

Denote by D_1 the set of all vectors $\boldsymbol{a}$ such that if $\boldsymbol{a}, \boldsymbol{a}' \in D_1$,

$$\sum_j |a_j - a_j'| > \frac{m-1}{4}.$$

By the Varshamov–Gilbert bound, we have

$$\#(D_1) \geq \exp\{(m-1)/8\}.$$

Now take the $H_{i\delta}$'s to be the $H_{\boldsymbol{a}}, a \in D_1$, and $N_\delta = \#(D_1)$. It follows that

$$\|H_{i\delta} - H_{j\delta}\|_q \geq m^{-p}\left(\frac{m-1}{4m}\right)^{1/q} \|\phi\|_q.$$

Then let

$$\delta = m^{-p}\left(\frac{m-1}{4m}\right)^{1/q} \|\phi\|_q = O(m^{-p}).$$

It follows from Proposition 8.1 below with $a_n = O(m^{-p})$ and $b_n = O(m^{-(p-1)})$ that there exists some $C \in (0, \infty)$ such that, uniformly in i,

$$K(H_{i\delta}\|H_{0\delta}) \leq C(m^{-2(p-1)} + nm^{-2p})\log^2 n$$

for large n. Fix any $\alpha \in (0, 1/8)$. To ensure that (8.2) holds, it suffices to find m such that

$$\alpha \log(N_\delta) \geq \alpha \frac{m-1}{8} \geq C(m^{-2(p-1)} + nm^{-2p})\log^2 n.$$

To achieve the slowest rate δ, we pick the smallest m for each n such that this inequality holds, in which case

$$m \sim (C/\alpha)(n \log^2 n)^{\frac{1}{2p+1}}.$$

The proof is complete. $\square$

The proof of Theorem 3.1 now hinges on the computation of $K(H_{i\delta}\|H_{0\delta})$, which is accomplished in the proposition below.

PROPOSITION 8.1. *Let $H_0(t) \equiv 1/2$ and $H(t) = 1/2 + \phi(t)$, where $\sup_t |\phi(t)| \leq a_n$ and $\sup_t |\phi'(t)| \leq b_n$ with $(a_n + b_n)\log n \to 0$. Then there is a finite (universal) constant C such that*

$$\limsup_{n\to\infty} \{(b_n^2 + na_n^2)\log^2 n\}^{-1} K(H\|H_0) \leq C. \tag{8.4}$$

A detailed proof for this proposition can be found in the Supplementary Material [15]. The general idea is the following.

Denote by X the mBm with Hurst function H. To simplify notation, we assume that the data is observed on $\{i/n, i=1,\dots,n\}$ instead of on Ω_n. This does not change the conclusion in the result. Define

$$Y_i = \sqrt{n}\left(X\left(\frac{i}{n}\right) - X\left(\frac{i-1}{n}\right)\right), \quad i=1,\dots,n.$$

Let Σ be covariance matrix of $(Y_1,\dots,Y_n)$. It is well known that

$$K(H\|H_0) = \frac{1}{2}\big[-\log|\Sigma| - n + \operatorname{tr}(\Sigma)\big]. \tag{8.5}$$

Let $D=(d_{ij})_{n\times n} = \Sigma - I$ and λ_i the eigenvalues of D. Intuitively, D will be a matrix shrinking to 0-matrix with n increasing. Under some conditions for the shrinking speed, we have

$$-\log|\Sigma| = -\sum_i \log(1+\lambda_i) = -\operatorname{tr}(D) + O\big(\operatorname{tr}(D^2)\big). \tag{8.6}$$

Now by (8.5), we can obtain

$$K(H\|H_0) = O\big(\operatorname{tr}(D^2)\big) = O\big(\|D\|_F^2\big).$$

Therefore, to prove this proposition, we need to carefully study how fast D converges to 0, for which a through proof is given in the Supplementary Material [15].

8.2. *Proofs of Theorem* 3.2 *and Theorem* 3.3. We begin by developing a few technical results needed for the proof. For convenience, let

$$W_n(\boldsymbol{t},\boldsymbol{h}) := \sigma^{-1} n^{H(\boldsymbol{t})} \Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}),$$

$$C_n(\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) := \operatorname{Cov}\big(W_n(\boldsymbol{t},\boldsymbol{h}), W_n(\boldsymbol{s},\boldsymbol{h})\big).$$

Clearly,

$$C_n(\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \sigma^{-2} n^{H(\boldsymbol{t})+H(\boldsymbol{s})} \Delta^q_{\boldsymbol{h}/n,\boldsymbol{t}} \Delta^q_{\boldsymbol{h}/n,\boldsymbol{s}} C(\boldsymbol{t},\boldsymbol{s}),$$

where $\Delta^q_{\boldsymbol{h}/n,\boldsymbol{t}}$ and $\Delta^q_{\boldsymbol{h}/n,\boldsymbol{s}}$ denote differencing with respect to $\boldsymbol{t}$ and $\boldsymbol{s}$, respectively.

The following theorem addresses the behavior of the covariance of $W_n(\boldsymbol{t},\boldsymbol{h})$ for different ranges of the gap. There are some similarities between this theorem and Theorem 1 in [10] or Propositions 1 and 2 in [1]. However, the results here focus more on the mBm and provides the level of precision needed for our results. As noted before, the differencing direction $\boldsymbol{h}$ in this paper is generally assumed to be a unit vector that parallels an axis. The following theorem is one instance where this is not assumed. The proof is included in the Supplementary Material [15].

THEOREM 8.2. *In the following, let* $\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n \in \Omega_\delta$. *Under assumptions* [A1]–[A3], *for any* $\delta\in(0,1)$ *and* $M>0$, *there exist finite constants* C_δ, $C_{\delta,M}$ *such that uniformly in* $n,\boldsymbol{t}$:

(i) $|C_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}/n,\boldsymbol{h}) - g(H(\boldsymbol{t}),\boldsymbol{u},\boldsymbol{h})| \le C_{\delta,M}\rho_n(\boldsymbol{t})$ *for* $|\boldsymbol{u}|\le M$;
(ii) $|C_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}/n,\boldsymbol{h})| \le C_\delta |\boldsymbol{u}|^{-\psi(\boldsymbol{t})}$ *for* $\boldsymbol{u}$ *satisfying* $q|\boldsymbol{h}|+1<|\boldsymbol{u}|<2bn$;
(iii) $|C_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}/n,\boldsymbol{h})| \le C_\delta |\boldsymbol{u}|^{-\bar{\psi}}$ *for* $\boldsymbol{u}$ *satisfying* $|\boldsymbol{u}|>q|\boldsymbol{h}|+1$;
(iv) *for* $|\boldsymbol{u}|\to\infty$ *and* $|\boldsymbol{u}|<2bn$,

$$C_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}/n,\boldsymbol{h}) = \big(c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{u}) + O(b\log b)\big)|\boldsymbol{u}|^{-\psi(\boldsymbol{t})},$$

where $c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{u})$ *is defined in Theorem* 3.3.

Our approach for proving Theorems 3.2 and 3.3 starts from the following decomposition:

$$\begin{aligned} G\big(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}\big) &= \sum_i 2s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log\big|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i)\big| \\ &= I_1(\boldsymbol{t}, n, b) + \widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b) + \log \sigma^2, \end{aligned} \tag{8.7}$$

where

$$I_1(\boldsymbol{t}, n, b) = -2 \log n \sum_i s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) H(\boldsymbol{t}_i)$$

and

$$\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b) = 2 \sum_i s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log\big|W_n(\boldsymbol{t}_i, \boldsymbol{h})\big|.$$

Denote

$$\widetilde{g}(H, \boldsymbol{h}) = \log\big(g(H, \mathbf{0}, \boldsymbol{h})\big) + \mathbb{E} \log \chi_1^2. \tag{8.8}$$

It is easy to see if $H(\boldsymbol{t})$ bounded away from 1 and $H(\boldsymbol{t}) \in \mathcal{H}_p((0,1)^d)$, then $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h}) \in \mathcal{H}_p((0,1)^d)$. Clearly, $I_1(\boldsymbol{t}, n, b)$ approximates $-2(\log n)H(\boldsymbol{t})$. Similarly, the following lemma suggests that $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ could be a reasonable estimator of $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$.

LEMMA 8.3. *Under assumption* [A3], *we have*

$$\mathbb{E}\big[2 \log\big|W_n(\boldsymbol{t}, \boldsymbol{h})\big|\big] = \widetilde{g}\big(H(\boldsymbol{t}), \boldsymbol{h}\big) + O\big(\rho_n(\boldsymbol{t})\big).$$

Thus, the asymptotic properties of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ can be understood by conducting a careful analysis of $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ and $I_1(\boldsymbol{t}, n, b)$, where both $I_1(\boldsymbol{t}, n, b)$ and $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ contribute to the asymptotic bias whereas $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ determines the asymptotic variance and asymptotic distribution. This will be done in the following two lemmas whose proofs along with the proof of Lemma 8.3 are given in the Supplementary Material [15].

LEMMA 8.4. *Assume that* [K], [A1], [A2] *and* [A3] *hold. We have*, *uniformly for all* $\boldsymbol{t} \in [0,1]^d$,

$$I_1(\boldsymbol{t}, n, b) = -2 \log(n)\big(H(\boldsymbol{t}) + R(\boldsymbol{t}, b)\big) + O\left(\frac{\log(n)}{(nb)^{2\wedge p}}\right)$$

and

$$R(\boldsymbol{t}, b) = O\big(b^p\big).$$

Furthermore, *if* $p \in \mathbb{N}$, *we have*

$$\begin{aligned} R(\boldsymbol{t}, b) = \mathbf{e}_1^T &\left(\int_{D_{\boldsymbol{t},b}} K(\mathbf{z}) \mathbf{A}(\mathbf{z}) \mathbf{A}^T(\mathbf{z})\, d\mathbf{z}\right)^{-1} \\ &\times \int_{D_{\boldsymbol{t},b}} K(\mathbf{z}) \mathbf{A}(\mathbf{z}) \left(\sum_{|\boldsymbol{\alpha}|=p} R_{\boldsymbol{\alpha}}(\boldsymbol{t}) (b\mathbf{z})^{\boldsymbol{\alpha}}\right) d\mathbf{z}, \end{aligned}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)^T$, $\mathbf{z}^{\boldsymbol{\alpha}}$ *is as defined in* (2.6),

$$D_{\boldsymbol{t},b} = \{\mathbf{z} : \boldsymbol{t} + b\mathbf{z} \in \Omega_\delta\} \cap [0,1]^d,$$

and

$$R_{\boldsymbol{\alpha}}(\boldsymbol{t}) = \frac{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}!} \int_0^1 (1-s)^{|\boldsymbol{\alpha}|-1} D^{\boldsymbol{\alpha}} H(\boldsymbol{t} + sb\mathbf{z})\, ds.$$

LEMMA 8.5. *For* $\boldsymbol{t} \in (0,1)^d$, *define*

$$\mathbf{Z}(\boldsymbol{t}, n) := (nb)^{-\frac{d}{2}} \left\{ \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left(2\log\left|W_n(\boldsymbol{t}_i, \boldsymbol{h})\right| - \widetilde{g}(H(\boldsymbol{t}_i), \boldsymbol{h})\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\right\}.$$

Assume that [A1], [A2] *and* [A3] *hold. If* $2\psi(\boldsymbol{t}) > d$, *then*

$$\mathbf{Z}(\boldsymbol{t}, n) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2_{H(\boldsymbol{t})} \int_{\mathbb{R}^d} \mathbf{f}(\mathbf{z})\mathbf{f}^T(\mathbf{z})\, d\mathbf{z}\right),$$

where $\mathbf{f}(\mathbf{z}) = K(\mathbf{z})\mathbf{A}(\mathbf{z})$ *and* $\sigma^2_{H(\boldsymbol{t})}$ *is defined by* (3.9); *if* $2\psi(\boldsymbol{t}) = d$, *then*

$$\log^{-\frac{1}{2}}(nb)\mathbf{Z}(\boldsymbol{t}, n) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \frac{V_d \mathbf{f}(\mathbf{0})\mathbf{f}^T(\mathbf{0})}{g^2(H(\boldsymbol{t}, \mathbf{0}, \boldsymbol{h}))} \int_{B_1(\mathbf{0})} \frac{c^2_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{x})}{|\boldsymbol{x}|^{d-1}}\, d\boldsymbol{x};$$

if $2\psi(\boldsymbol{t}) < d$, *then the characteristic function for* $(nb)^{-\frac{d}{2}+\psi(\boldsymbol{t})}\mathbf{a}^T\mathbf{Z}(\boldsymbol{t}, n)$ *converges to*

$$\phi(u) = \exp\left\{\frac{1}{2}\sum_{k=2}^{\infty} \frac{(2iu)^k}{k} S_k\right\},$$

where

$$S_k = \int_{(B_1(\mathbf{0}))^{\otimes k}} \frac{\prod_{i=1}^k c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i - \mathbf{z}_{i+1})|\mathbf{z}_i - \mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})} f_{\boldsymbol{a}}(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})}\, d\mathbf{z}_1 \cdots d\mathbf{z}_k,$$

in which $\mathbf{z}_{k+1} := \mathbf{z}_1$ *and* $f_{\boldsymbol{a}} = \boldsymbol{a}^T\mathbf{f}$. *In addition, for each* $\delta \in (0,1)$ *and* $\mathbf{a} \in \mathbb{R}^d$, *there exist constants* C_δ *and* $C_{\delta,\boldsymbol{a}}$ *such that all* $\boldsymbol{t} \in \Omega_\delta$,

$$\left|\mathbb{E}\left[\mathbf{Z}(\boldsymbol{t}, n)\right]\right| \le C_\delta (nb)^{\frac{d}{2}} \rho_n(\boldsymbol{t}),$$

and

$$\operatorname{Var}\left(\mathbf{a}^T\mathbf{Z}(\boldsymbol{t}, n)\right) \le \begin{cases} C_{\delta,\mathbf{a}} & \text{if } 2\psi(\boldsymbol{t}) > d, \\ C_{\delta,\mathbf{a}} \log(nb) & \text{if } 2\psi(\boldsymbol{t}) = d, \\ C_{\delta,\mathbf{a}} (nb)^{d-2\psi(\boldsymbol{t})} & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases}$$

PROOFS FOR THEOREMS 3.2 AND 3.3. To simplify notation, write $\widetilde{g}(\boldsymbol{t}) := \widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$, where $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$ is defined by (8.8), and

$$\widetilde{\boldsymbol{\beta}} := \operatorname*{argmin}_{\beta_{j_1}, \dots, \beta_{j_S}} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left\{2\log\left|W_n(\boldsymbol{t}_i, \boldsymbol{h})\right| - \sum_{m=1}^{S} \beta_{j_m}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^{j_m}\right\}^2.$$

Thus we have

$$\sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left\{2\log\left|W_n(\boldsymbol{t}_i, \boldsymbol{h})\right| - \widetilde{\boldsymbol{\beta}}^T\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\right\}\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) = 0.$$

Plug this equation into the definition of $\mathbf{Z}(\boldsymbol{t}, n)$, basically replacing $2\log|W_n(\boldsymbol{t}_i, \boldsymbol{h})|$ by $\widetilde{\boldsymbol{\beta}}^T\mathbf{A}$ there, and we obtain

$$\mathbf{Z}(\boldsymbol{t}, n) = (nb)^{-\frac{d}{2}}\left\{\sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left(\widetilde{\boldsymbol{\beta}}^T\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) - \widetilde{g}(\boldsymbol{t}_i)\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\right\}.$$

If $p \in \mathbb{N}$, denote the Taylor expansion of $\widetilde{g}(\boldsymbol{t}+b\boldsymbol{v})$ at $\boldsymbol{t}$ up to degree $p-1$ by $\breve{g}(\boldsymbol{t}, b\boldsymbol{v})$. Then we have

$$\breve{g}(\boldsymbol{t}, b\boldsymbol{v}) - \widetilde{g}(\boldsymbol{t}+b\boldsymbol{v}) = -\frac{b^p}{p!}\sum_{|\boldsymbol{i}|=p}\big[\boldsymbol{v}^{\boldsymbol{i}} D^{\boldsymbol{i}}\widetilde{g}(\boldsymbol{t})\big] + o\big((b\boldsymbol{v})^p\big).$$

Define

$$\mathbf{M}_n(b, \boldsymbol{t}) = (nb)^{-d}\sum_{\boldsymbol{v}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_{\boldsymbol{t}}} \mathbf{A}(\boldsymbol{v})\mathbf{A}(\boldsymbol{v})^T K(\boldsymbol{v}),$$

where $\Omega_{\boldsymbol{t}} = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^d, \boldsymbol{t}+\boldsymbol{i}/n \in (0,1)^d\}$. On one hand,

$$\begin{aligned}
&(nb)^{-d}\sum_{\boldsymbol{v}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_t} \mathbf{A}(\boldsymbol{v})K(\boldsymbol{v})\big(\widetilde{\boldsymbol{\beta}}^T\mathbf{A}(\boldsymbol{v}) - \breve{g}(\boldsymbol{t}, b\boldsymbol{v})\big)\\
&\quad = (nb)^{-d}\sum_{\boldsymbol{v}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_t} \mathbf{A}(\boldsymbol{v})\mathbf{A}(\boldsymbol{v})^T K(\boldsymbol{v})\begin{pmatrix}\widetilde{\beta}_1 - \dfrac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!}\dfrac{\partial^{\boldsymbol{j}_1}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_1}}\\ \vdots \\ \widetilde{\beta}_S - \dfrac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!}\dfrac{\partial^{\boldsymbol{j}_S}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_S}}\end{pmatrix}\\
&\quad = \mathbf{M}_n(b, \boldsymbol{t})\begin{pmatrix}\widetilde{\beta}_1 - \dfrac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!}\dfrac{\partial^{\boldsymbol{j}_1}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_1}}\\ \vdots \\ \widetilde{\beta}_S - \dfrac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!}\dfrac{\partial^{\boldsymbol{j}_S}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_S}}\end{pmatrix}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&(nb)^{-d}\sum_{\boldsymbol{v}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_t} \mathbf{A}(\boldsymbol{v})K(\boldsymbol{v})\big(\widetilde{\boldsymbol{\beta}}^T\mathbf{A}(\boldsymbol{v}) - \breve{g}(\boldsymbol{t}, b\boldsymbol{v})\big)\\
&\quad = (nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t}, n) + B(\boldsymbol{t}, b) + o\big(b^p\big),
\end{aligned}$$

where

$$B(\boldsymbol{t}, b) = -\frac{b^p}{p!}\sum_{|\boldsymbol{i}|=p}\Bigg[(nb)^{-d}\sum_{\boldsymbol{v}\in\{\frac{\boldsymbol{j}}{nb}, \boldsymbol{j}\in\Omega_t\}} \mathbf{A}(\boldsymbol{v})K(\boldsymbol{v})\boldsymbol{v}^{\boldsymbol{i}} D^{\boldsymbol{i}}\widetilde{g}(\boldsymbol{t})\Bigg] = O\big(b^p\big).$$

If $p \notin \mathbb{N}$, denote the Taylor expansion of $\widetilde{g}(\boldsymbol{t}+b\boldsymbol{v})$ at $\boldsymbol{t}$ up to degree $\lfloor p \rfloor$ by $\breve{g}(\boldsymbol{t}, b\boldsymbol{v})$. Then we have

$$\breve{g}(\boldsymbol{t}, b\boldsymbol{v}) - \widetilde{g}(\boldsymbol{t}+b\boldsymbol{v}) = O\big((b\boldsymbol{v})^p\big).$$

From this, we can still get $B(\boldsymbol{t}, b) = O(b^p)$, although an explicit form is not available. Thus

$$\begin{pmatrix}\widetilde{\beta}_1 - \dfrac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!}\dfrac{\partial^{\boldsymbol{j}_1}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_1}}\\ \vdots \\ \widetilde{\beta}_S - \dfrac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!}\dfrac{\partial^{\boldsymbol{j}_S}\widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_S}}\end{pmatrix} = \mathbf{M}_n(b, \boldsymbol{t})^{-1}\big((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t}, n) + B(\boldsymbol{t}, b) + o\big(b^p\big)\big).$$

It is easily shown that

$$\mathbf{M}_n(b,\boldsymbol{t}) \to \mathbf{M} := \int_{B_1(\mathbf{0})} \mathbf{A}(\boldsymbol{v})\mathbf{A}(\boldsymbol{v})^T K(\boldsymbol{v})\, d\boldsymbol{v}$$

uniformly. Therefore,

$$\widetilde{\beta}_1 = \widetilde{g}(\boldsymbol{t}) + \bigl(1+o(1)\bigr)\mathbf{e}_1^T \mathbf{M}^{-1}\bigl((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n) + B(\boldsymbol{t},b)\bigr). \tag{8.9}$$

Combining this with equation (8.7) and Lemma 8.4, we obtain

$$\begin{aligned}
&G\bigl(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\bigr)\\
&\quad= \widetilde{g}\bigl(H(\boldsymbol{t}),\boldsymbol{h}\bigr) - 2\log(n)H(\boldsymbol{t}) + \log\sigma^2 - 2\log(n)R(\boldsymbol{t},b) + O\biggl(\frac{\log(n)}{(nb)^{2\wedge p}}\biggr)\\
&\qquad + \mathbf{e}_1^T\mathbf{M}^{-1}\bigl((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n) + B(\boldsymbol{t},b)\bigr)\bigl(1+o(1)\bigr)\\
&\quad= G\bigl(H(\boldsymbol{t}); n, \boldsymbol{h}\bigr) - 2\log(n)R(\boldsymbol{t},b) + O\biggl(\frac{\log(n)}{(nb)^{2\wedge p}}\biggr)\\
&\qquad + \mathbf{e}_1^T\mathbf{M}^{-1}\bigl((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n) + B(\boldsymbol{t},b)\bigr)\bigl(1+o(1)\bigr),
\end{aligned} \tag{8.10}$$

where the last line uses the definitions of G and $\widetilde{g}$ in equations (2.4) and (8.8). As such, the bias of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ is

$$\begin{aligned}
&\operatorname{Bias}\bigl(G\bigl(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\bigr)\bigr)\\
&\quad= O\biggl(\log(n)R(\boldsymbol{t},b) + \frac{\log(n)}{(nb)^{2\wedge p}} + (nb)^{-\frac{d}{2}}\bigl|\mathbb{E}\mathbf{Z}(\boldsymbol{t},n)\bigr| + \mathbf{e}_1^T\mathbf{M}^{-1}B(\boldsymbol{t},b)\biggr)\\
&\quad= O\bigl(\log(n)\bigl(b^p + (nb)^{-2\wedge p}\bigr) + \rho_n(\boldsymbol{t})\bigr) = O\bigl(T_1(n,b,\boldsymbol{t})\bigr),
\end{aligned} \tag{8.11}$$

where the last line use facts that $R(\boldsymbol{t},b), B(\boldsymbol{t},b) = O(b^p)$ and $|\mathbb{E}\mathbf{Z}(\boldsymbol{t},n)| = O((nb)^{\frac{d}{2}}\rho_n(\boldsymbol{t}))$ in Lemma 8.5. Similarly,

$$\begin{aligned}
\operatorname{Var}\bigl(G\bigl(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}\bigr)\bigr) &= O\bigl((nb)^{-d}\operatorname{Var}\bigl(\mathbf{e}_1^T\mathbf{M}^{-1}\mathbf{Z}(\boldsymbol{t},n)\bigr)\bigr)\\
&= O\bigl(T_2(n,b,\boldsymbol{t})\bigr),
\end{aligned} \tag{8.12}$$

where the last step uses the result on the variance of $\mathbf{Z}$ in Lemma 8.5. The two bounds, (8.11) and (8.12), complete the proof for Theorem 3.2.

Finally, in view of (8.10), $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h}) - \mathbb{E}G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ has the same asymptotic distribution as that of $\mathbf{e}_1^T\mathbf{M}^{-1}(\mathbf{Z}(\boldsymbol{t},n) - \mathbb{E}\mathbf{Z}(\boldsymbol{t},n))$. Applying Lemma 8.5, it is straightforward to obtain those asymptotic distributions listed in Theorem 3.3. This completes the proofs of Theorem 3.2 and Theorem 3.3. $\square$

Due to space limitation, the proofs of Theorem 3.4 and Corollary 3.5 are deferred to the Supplementary Material [15]. However, the general idea behind those is the simple fact $\partial G(H; n, \boldsymbol{h})/\partial H = -2\log n + O(1)$. Intuitively, $G(H; n, \boldsymbol{h})$ behaves increasingly like a linear function with slope $-2\log n$ as n increases. Therefore, solving the inversion problem in obtaining $\widehat{H}$ will have the effect of shrinking the estimation error of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$ by the rate $\log n$, and the asymptotic linearity of $G(H; n, \boldsymbol{h})$ also allows $\widehat{H}$ to preserve the asymptotic distribution of $G(\widehat{H(\boldsymbol{t}); n}, \boldsymbol{h})$. Detailed justifications of these can be found in the Supplementary Material [15].

Acknowledgments. The authors wish to sincerely thank the referees and the Associate Editor for their very helpful comments that led to a considerable refinement and generalization of the results.

SUPPLEMENTARY MATERIAL

Supplement to "Hurst function estimation" (DOI: 10.1214/19-AOS1825SUPP; .pdf). This supplemental material includes all the proofs not included in the paper.

REFERENCES

[1] BARDET, J.-M. and SURGAILIS, D. (2013). Nonparametric estimation of the local Hurst function of multifractional Gaussian processes. *Stochastic Process. Appl.* **123** 1004–1045. MR3005013 https://doi.org/10.1016/j.spa.2012.11.009

[2] BENASSI, A., JAFFARD, S. and ROUX, D. (1997). Elliptic Gaussian random processes. *Rev. Mat. Iberoam.* **13** 19–90. MR1462329 https://doi.org/10.4171/RMI/217

[3] BERTRAND, P. R., FHIMA, M. and GUILLIN, A. (2013). Local estimation of the Hurst index of multifractional Brownian motion by increment ratio statistic method. *ESAIM Probab. Stat.* **17** 307–327. MR3066382 https://doi.org/10.1051/ps/2011154

[4] BREUER, P. and MAJOR, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13** 425–441. MR0716933 https://doi.org/10.1016/0047-259X(83)90019-2

[5] COEURJOLLY, J.-F. (2005). Identification of multifractional Brownian motion. *Bernoulli* **11** 987–1008. MR2188838 https://doi.org/10.3150/bj/1137421637

[6] COHEN, S. (1999). From self-similarity to local self-similarity: The estimation problem. In *Fractals: Theory and Applications in Engineering* 3–16. Springer, London. MR1726364

[7] FALCONER, K. J. (2002). Tangent fields and the local structure of random fields. *J. Theoret. Probab.* **15** 731–750. MR1922445 https://doi.org/10.1023/A:1016276016983

[8] FAN, J. and GIJBELS, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. CRC Press, London. MR1383587

[9] HERBIN, E. (2006). From *N* parameter fractional Brownian motions to *N* parameter multifractional Brownian motions. *Rocky Mountain J. Math.* **36** 1249–1284. MR2274895 https://doi.org/10.1216/rmjm/1181069415

[10] HSING, T., BROWN, T. and THELEN, B. (2016). Local intrinsic stationarity and its inference. *Ann. Statist.* **44** 2058–2088. MR3546443 https://doi.org/10.1214/15-AOS1402

[11] LÉVY-VÉHEL, J. and PELTIER, R. F. (1995). Multifractional Brownian motion: Definition and preliminary results. Rapport de Recherche de L'INRIA N2645.

[12] LOH, W.-L. (2015). Estimating the smoothness of a Gaussian random field from irregularly spaced data via higher-order quadratic variations. *Ann. Statist.* **43** 2766–2794. MR3405611 https://doi.org/10.1214/15-AOS1365

[13] MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437. MR0242239 https://doi.org/10.1137/1010093

[14] NOURDIN, I. (2012). *Selected Aspects of Fractional Brownian Motion. Bocconi & Springer Series* **4**. Springer, Milan. MR3076266 https://doi.org/10.1007/978-88-470-2823-4

[15] SHEN, J. and HSING, T. (2020). Supplement to "Hurst function estimation." https://doi.org/10.1214/19-AOS1825SUPP.

[16] STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging. Springer Series in Statistics*. Springer, New York. MR1697409 https://doi.org/10.1007/978-1-4612-1494-6

[17] STOEV, S. A. and TAQQU, M. S. (2006). How rich is the class of multifractional Brownian motions? *Stochastic Process. Appl.* **116** 200–221. MR2197974 https://doi.org/10.1016/j.spa.2005.09.007

[18] TAQQU, M. S. (1974/75). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrsch. Verw. Gebiete* **31** 287–302. MR0400329 https://doi.org/10.1007/BF00532868

[19] TSYBAKOV, A. B. (2009). *Introduction to Nonparametric Estimation. Springer Series in Statistics*. Springer, New York. Revised and extended from the 2004 French original. Translated by Vladimir Zaiats. MR2724359 https://doi.org/10.1007/b13794