# Rate-optimal estimation of mixed semimartingales

Chong, Carsten H.; Delerue, Thomas; Mies, Fabian

**DOI**
10.1214/24-AOS2461

**Publication date**
2025

**Document Version**
Final published version

**Published in**
Annals of Statistics

**Citation (APA)**
Chong, C. H., Delerue, T., & Mies, F. (2025). Rate-optimal estimation of mixed semimartingales. *Annals of Statistics*, *53*(1), 219-244. https://doi.org/10.1214/24-AOS2461

*The Annals of Statistics*
2025, Vol. 53, No. 1, 219–244
https://doi.org/10.1214/24-AOS2461


# RATE-OPTIMAL ESTIMATION OF MIXED SEMIMARTINGALES

BY CARSTEN H. CHONG[1,a], THOMAS DELERUE[2,b] AND FABIAN MIES[3,c]

[1]*Department of Information Systems, Business Statistics and Operations Management, The Hong Kong University of Science and Technology,* [a]*carstenchong@ust.hk*

[2]*Institute of Epidemiology, Helmholtz Munich,* [b]*thomas.delerue@gmail.com*

[3]*Delft Institute of Applied Mathematics, Delft University of Technology,* [c]*f.mies@tudelft.nl*

Consider the sum $Y = B + B(H)$ of a Brownian motion $B$ and an independent fractional Brownian motion $B(H)$ with Hurst parameter $H \in (0, 1)$. Even though $B(H)$ is not a semimartingale, it was shown by Cheridito (*Bernoulli* **7** (2001) 913–934) that $Y$ is a semimartingale if $H > 3/4$. Moreover, $Y$ is locally equivalent to $B$ in this case, so $H$ cannot be consistently estimated from local observations of $Y$. This paper pivots on another unexpected feature in this model: if $B$ and $B(H)$ become correlated, then $Y$ will never be a semimartingale, and $H$ can be identified, regardless of its value. This and other results will follow from a detailed statistical analysis of a more general class of processes called *mixed semimartingales*, which are semiparametric extensions of $Y$ with stochastic volatility in both the martingale and the fractional component. In particular, we derive consistent estimators and feasible central limit theorems for all parameters and processes that can be identified from high-frequency observations. We further show that our estimators achieve optimal rates in a minimax sense.

**1. Introduction.** Mixed fractional Brownian motions (mfBms) were introduced by [11] as the sum

$$Y_t = \sigma B_t + \rho B(H)_t, \quad t \geq 0, \tag{1.1}$$

of a standard Brownian motion $B$ and an independent fractional Brownian motion $B(H)$ with Hurst parameter $H \in (0, 1) \setminus \{\frac{1}{2}\}$. While this class of processes was originally introduced in mathematical finance to model long memory in asset prices, it poses nonstandard challenges from a statistical perspective: Even though the laws of $B$ and $B(H)$ on a finite time interval are mutually singular, the law of their superposition, $Y$, can be locally equivalent to that of $\sigma B$ (if $H > \frac{3}{4}$) or $\rho B(H)$ (if $H < \frac{1}{4}$); see [8, 11, 41]. In these cases, either $(\rho, H)$ or $\sigma$ cannot be consistently estimated on a finite time interval.

In this paper, we are interested in whether these results remain valid if $B$ and $B(H)$ are no longer assumed to be independent. The answer to this question is negative.

THEOREM 1.1. *Suppose that $Y$ is given by* (1.1) *with $\sigma, \rho > 0$ and*

$$B(H)_t = K_H^{-1} \int_{-\infty}^t \bigl((t-s)^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}}\bigr)\, \mathrm{d}\widetilde{B}_s, \quad t \geq 0, \tag{1.2}$$

*is the Mandelbrot–van Ness representation of standard fractional Brownian motion* (*see*, *e.g.*, [34], *Chapter* 1.3), *where $(B, \widetilde{B})$ is a two-dimensional Brownian motion with* $\mathrm{Var}(B_1) = \mathrm{Var}(\widetilde{B}_1) = 1$ *and* $\lambda = \mathrm{Cov}(B_1, \widetilde{B}_1) \in [0, 1]$ *and $K_H$ is the normalizing constant given in*

Received July 2022; revised September 2024.
*MSC2020 subject classifications.* Primary 60G22, 62G20, 62M09; secondary 60F05, 62P20.
*Key words and phrases.* Central limit theorem, high-frequency observations, Hurst parameter, KL divergence, minimax rate, mixed fractional Brownian motion, rough noise.

(2.3). *Then the process $Y$ has the same distribution as $\widetilde{Y}$ given by*

$$\widetilde{Y}_t = \sigma W_t + \sqrt{\frac{2\lambda\rho\sigma}{K_H(H+\frac{1}{2})}} W(\overline{H})_t + \rho W(H)_t, \tag{1.3}$$

*where $W$ is a standard Brownian motion, $W(H)$ and $W(\overline{H})$ are fractional Brownian motions with Hurst parameters $H$ and $\overline{H} = \frac{1}{2}(H + \frac{1}{2})$, respectively, and all three processes are independent. Moreover, if $\lambda > 0$, $Y$ is not a semimartingale and its distribution is locally singular to both $\sigma B$ and $\rho B(H)$ for all $H \in (0, 1) \setminus \{\frac{1}{2}\}$.*

The newly emerging fBm with Hurst parameter $\overline{H}$ significantly changes the properties of the model. In particular, for all values of $H$, we have $|\overline{H} - \frac{1}{2}| < \frac{1}{4}$, which has two consequences for the correlated case: first, the mixed process will never be locally equivalent to either of the two pure processes, and second, $\overline{H}$ (and therefore $H$) is identifiable from high-frequency observations in all cases. Our asymptotic results below show that this also upholds for negative correlation coefficients $\lambda$.

Fractional processes have a long history of applications in fields such as hydrology [23, 35, 37], telecommunications [28, 33], finance [11, 16, 21], turbulence [9, 17] among others. In these applications, superpositions of fractional processes arise naturally when multiple sources have a cumulative effect. For example, [42] describe a continuous GPS signal affected by both white noise and fractional noise. The same phenomenon is found by [43] in a range of astronomical data sets. Another example is found in hydrology, where fractional Brownian motion is commonly used to model river runoff, with varying Hurst parameters for different rivers [37]. In a system of multiple connected rivers, the runoffs add up downstream, which leads to a superposition of fBms with different Hurst parameters. Due to the spatial correlation of rainfall, these constituent fBms will be correlated, analogous to the mixed fBm model studied in this paper. The results of this paper highlight that correlation between the fBms is not negligible, as it alters the statistical properties significantly.

Our interest in mixed fractional Brownian motion of the specific form (1.1) is motivated by recent applications of mixed processes in financial econometrics. In [13], it is shown that for a large set of high-frequency stock return data, observed prices are contaminated by microstructure noise that locally resembles fractional Brownian motion with Hurst parameter $H \in (0, \frac{1}{2})$. Like in most microstructure noise models in the literature, the two innovation processes driving price and noise (which are $B$ and $B(H)$ in (1.1)) are assumed to be independent of each other. However, both economic theory [19] and empirical evidence [22] suggest that efficient price and microstructure noise should be contemporaneously cross-correlated.

Against this background, the main contribution of this paper is to develop an infill asymptotic theory for semiparametric extensions of the mfBm model (called *mixed semimartingales*) where $\sigma$ and $\rho$ can be stochastic and time-varying. The fundamental statistical question we address is the following: what parameters can be inferred from local observations of the sum of a martingale and a correlated fractional process, and what are the optimal rates of convergence as the sample size increases? To this end, we first derive in Section 2 a stable central limit theorem (CLT) for the empirical autocovariances of the increments of a mixed semimartingale process as the sampling frequency increases to infinity (Theorem 2.1). In line with Theorem 1.1, the population autocovariance consists of three terms with different scaling exponents. Most importantly, the leading order term only contains information about the parameter $\sigma$ (if $H > \frac{1}{2}$) or $(\rho, H)$ (if $H < \frac{1}{2}$). To optimally estimate all parameters, it is thus necessary to utilize the information in the asymptotically smaller contributions to the autocovariance. In particular, in Section 3, we combine the results of Theorem 2.1 with an optimal generalized methods of moments (GMM) procedure to construct consistent and asymptotically mixed normal estimators for all identifiable model parameters in Theorem 3.1. Interestingly, the rates of convergence are nonstandard and depend on the parameter value $H$. This phenomenon can be traced back to the utilization of autocovariance information of smaller asymptotic order than the leading term. To make the GMM method feasible, we exhibit in Corollary 3.3 consistent estimators of the asymptotic (co-)variances. In Section 4, we return to the parametric setting of mfBm and show in Theorem 4.1 that the rates of our estimators in Theorem 3.1 are optimal in the minimax sense. Section 5 presents Monte Carlo evidence for the estimators from Theorem 3.1. Sections 6 and 7 contain the proofs of our main results, except for Theorem 2.1, which is proved, in a more general setting, in Appendices A–D. Appendix E contains additional simulation results.

Since our principal aim is to analyze the impact of cross-correlation, we do not include jumps [1, 27], irregular observation times [6, 10, 24] or rounding errors [18, 30, 38] in our analysis. To simplify the exposition, we further refrain from studying mixed fractional models with more than two components. In what follows, $C$ denotes a constant in $(0,\infty)$, whose value may change from line to line. We also write $A \lesssim B$ if $A \leq CB$.

**2. Central limit theorem for sample autocovariances.** On a filtered space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions, consider a *mixed semimartingale*

$$Y_t = Y_0 + \int_0^t a_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}B_s + \int_0^t g(t-s)\rho_s \,\mathrm{d}B_s + \int_0^t g(t-s)\rho'_s \,\mathrm{d}B'_s, \tag{2.1}$$

where $B$ and $B'$ are two independent one-dimensional standard $\mathbb{F}$-Brownian motions and $a$, $\sigma$, $\rho$ and $\rho'$ are one-dimensional predictable processes. Moreover, we assume that $g\colon \mathbb{R} \to \mathbb{R}$ is a kernel of the form

$$g(t) = K_H^{-1} t^{H-\frac{1}{2}} \mathbb{1}_{\{t>0\}} + g_0(t), \quad t \in \mathbb{R}, \tag{2.2}$$

where $H \in (0,1) \setminus \{\frac{1}{2}\}$,

$$K_H = \sqrt{\frac{1}{2H} + \int_1^\infty \big(r^{H-\frac{1}{2}} - (r-1)^{H-\frac{1}{2}}\big)^2 \,\mathrm{d}r} = \frac{\Gamma(H+\frac{1}{2})}{\sqrt{\sin(\pi H)\Gamma(2H+1)}}, \tag{2.3}$$

and $g_0\colon \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function with $g_0(x) = 0$ for all $x \leq 0$. By (2.2), the kernel $g$ behaves as a power-law kernel around 0, but due to the addition of $g_0$, the behavior of $g$ outside of 0 is not further specified. In particular, because $g$ is not specified at infinity, the increments of $Y$ may have long or short memory, irrespective of the value of $H$.

Let $\Delta_n$ be a small time step such that $\Delta_n \to 0$ and define

$$\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n} \tag{2.4}$$

(and similarly for other processes). Our goal is to prove a CLT as $\Delta_n \to 0$ for the (normalized) autocovariances of the increments of $Y$, given by

$$V_{r,t}^n = \Delta_n^{1-2H} \sum_{i=1}^{[t/\Delta_n]-r} \Delta_i^n Y \Delta_{i+r}^n Y, \qquad \widehat{V}_{r,t}^n = \sum_{i=1}^{[t/\Delta_n]-r} \Delta_i^n Y \Delta_{i+r}^n Y$$

for $r \in \mathbb{N}_0$. This will then be used in Section 3 to derive feasible estimators of $H$ and

$$C_T = \int_0^T c_s \,\mathrm{d}s = \int_0^T \sigma_s^2 \,\mathrm{d}s, \qquad \Lambda_T = \int_0^T \lambda_s \,\mathrm{d}s = \int_0^T \sigma_s\rho_s \,\mathrm{d}s,$$

$$\Pi_T = \int_0^T (\rho_s^2 + \rho_s'^2) \,\mathrm{d}s,$$

whenever possible. We impose the following structural assumptions.

ASSUMPTION (CLT). Consider the process $Y$ from (2.1) and let

$$N(H) = \lfloor 1/|2H-1| \rfloor. \tag{2.5}$$

1. The kernel $g$ is of the form (2.2) where $H \in (0,1) \setminus \{\frac{1}{2}\}$ and $g_0 \in C^1(\mathbb{R})$ satisfies $g_0(x) = 0$ for all $x \leq 0$.
2. The drift process $a$ is locally bounded, $\mathbb{F}$-adapted and càdlàg. Moreover, $B$ is a standard $\mathbb{F}$-Brownian motion.
3. If $H > \frac{1}{2}$, the volatility process $\sigma$ takes the form

$$\sigma_t = \sigma_t^{(0)} + \int_0^t \widetilde{\sigma}_s \,\mathrm{d}B_s + \int_0^t \widetilde{\sigma}'_s \,\mathrm{d}B'_s + \int_0^t \widetilde{\sigma}''_s \,\mathrm{d}B''_s, \quad t \geq 0, \tag{2.6}$$

where

(a) $\sigma^{(0)}$ is an $\mathbb{F}$-adapted locally bounded process such that for all $T > 0$, there are $\gamma \in (\frac{1}{2}, 1]$ and $K_1 \in (0, \infty)$ with

$$\mathbb{E}\big[1 \wedge \big|\sigma_t^{(0)} - \sigma_s^{(0)}\big|\big] \leq K_1 |t-s|^\gamma, \quad s, t \in [0, T]; \tag{2.7}$$

(b) $\widetilde{\sigma}$, $\widetilde{\sigma}'$ and $\widetilde{\sigma}''$ are $\mathbb{F}$-adapted locally bounded processes such that for all $T > 0$, there are $\varepsilon \in (0,1)$ and $K_2 \in (0, \infty)$ with

$$\mathbb{E}\big[1 \wedge |\widetilde{\sigma}_t - \widetilde{\sigma}_s|\big] \leq K_2 |t-s|^\varepsilon, \quad s, t \in [0, T], \tag{2.8}$$

and an analogous bound for $\widetilde{\sigma}'$ and $\widetilde{\sigma}''$.

(c) $B''$ is a standard $\mathbb{F}$-Brownian motion that is independent of $(B, B')$.

If $H < \frac{1}{2}$, we have (2.6) but with $\sigma$, $\sigma^{(0)}$, $\widetilde{\sigma}$, $\widetilde{\sigma}'$ and $\widetilde{\sigma}''$ replaced by $\rho$ and some processes $\rho^{(0)}$, $\widetilde{\rho}$, $\widetilde{\rho}'$ and $\widetilde{\rho}''$ satisfying conditions analogous to (2.7) and (2.8).

4. If $H > \frac{1}{2}$, the process $\rho$ is $\mathbb{F}$-adapted and locally bounded. Moreover, for all $T > 0$, there is $K_3 \in (0, \infty)$ such that

$$\mathbb{E}\big[1 \wedge |\rho_t - \rho_s|\big] \leq K_3 |t-s|^{\frac{1}{2}}, \quad s, t \in [0, T]. \tag{2.9}$$

If $H < \frac{1}{2}$, we have the same condition but with $\rho$ replaced by $\sigma$.

These assumptions are fairly standard in high-frequency statistics (cf. [2, 25]). As usual, the most restrictive assumptions are (2.6) and (2.9), which essentially demand that the volatility processes $\sigma$ and $\rho$ be no rougher than a continuous Itô semimartingale. In particular, they do not cover the case of rough volatility (see [21]). However, we conjecture that the CLTs do remain valid even in the presence of rough volatility. This is due to the special structure of quadratic functions, which has been exploited for instance in [15] in a slightly different context.

For the following theorem, which is the main result of this section, we use the notation

$$\Gamma_0^H = 1 \quad \text{and} \quad \Gamma_r^H = \frac{1}{2}\big((r+1)^{2H} - 2r^{2H} + (r-1)^{2H}\big), \quad r \geq 1, \tag{2.10}$$

and

$$\Phi_0^H = \frac{2K_H^{-1}}{H + \frac{1}{2}}, \qquad \Phi_r^H = \frac{K_H^{-1}}{H + \frac{1}{2}}\big((r+1)^{H+\frac{1}{2}} - 2r^{H+\frac{1}{2}} + (r-1)^{H+\frac{1}{2}}\big), \quad r \geq 1. \tag{2.11}$$

Note that $(\Gamma_r^H)_{r \geq 0}$ is the autocovariance function of fractional Gaussian noise (i.e., of $(B(H)_{n+1} - B(H)_n)_{n \in \mathbb{N}}$) and $\Phi_r^H = \Phi_0^H \Gamma_r^{\overline{H}}$. We also use $\overset{\mathrm{st}}{\Longrightarrow}$ to denote functional stable convergence in law in the space of càdlàg functions equipped with the local uniform topology.

THEOREM 2.1. *Suppose that Assumption (CLT) holds and let* $V_t^n = (V_{0,t}^n, \dots, V_{r-1,t}^n)^T$ *and* $\widehat{V}_t^n = (\widehat{V}_{0,t}^n, \dots, \widehat{V}_{r-1,t}^n)^T$ *for some fixed* $r \in \mathbb{N}$. *Further define* $\Gamma^H = (\Gamma_0^H, \dots, \Gamma_{r-1}^H)^T$, $\Phi^H = (\Phi_0^H, \dots, \Phi_{r-1}^H)^T$ *and* $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^r$.

1. *If* $H > \frac{1}{2}$, *then*

$$\Delta_n^{-\frac{1}{2}}\{\widehat{V}_t^n - e_1 C_t - \Delta_n^{H-\frac{1}{2}} \Phi^H \Lambda_t - \Delta_n^{2H-1} \Gamma^H \Pi_t\} \stackrel{\text{st}}{\Longrightarrow} \mathcal{Z}', \tag{2.12}$$

*where* $\mathcal{Z}' = (\mathcal{Z}'_t)_{t\geq 0}$ *is an* $\mathbb{R}^r$*-valued process, defined on a very good filtered extension* $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ *of the original probability space* $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ (*see* [25], *Chapter* 2.1.4), *that conditionally on* $\mathcal{F}$ *is a centered Gaussian process with independent increments and covariance process* $\mathcal{C}'_t = (\mathcal{C}_t'^{ij})_{i,j=0}^{r-1}$ *given by*

$$\mathcal{C}_t'^{ij} = \overline{\mathbb{E}}[\mathcal{Z}_t'^i \mathcal{Z}_t'^j \mid \mathcal{F}] = 2^{\mathbb{1}_{\{i=0\}}} \int_0^t \sigma_s^4 \, \mathrm{d}s \, \mathbb{1}_{\{i=j\}}. \tag{2.13}$$

2. *If* $H < \frac{1}{2}$, *then*

$$\begin{aligned} &\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t - \Delta_n^{\frac{1}{2}-H} \Phi^H \Lambda_t - e_1 \Delta_n^{1-2H} C_t\} \\ &\quad = \Delta_n^{\frac{1}{2}-2H}\{\widehat{V}_t^n - e_1 C_t - \Delta_n^{H-\frac{1}{2}} \Phi^H \Lambda_t - \Delta_n^{2H-1} \Gamma^H \Pi_t\} \stackrel{\text{st}}{\Longrightarrow} \mathcal{Z}, \end{aligned} \tag{2.14}$$

*where* $\mathcal{Z} = (\mathcal{Z}_t)_{t\geq 0}$ *is an* $\mathbb{R}^r$*-valued process defined on* $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ *that conditionally on* $\mathcal{F}$ *is a centered Gaussian process with independent increments and covariance process* $\mathcal{C}_t = (\mathcal{C}_t^{ij})_{i,j=0}^{r-1}$ *given by*

$$\mathcal{C}_t^{ij} = \overline{\mathbb{E}}[\mathcal{Z}_t'^i \mathcal{Z}_t'^j \mid \mathcal{F}] = \mathcal{C}^{ij} \int_0^t (\rho_s^2 + \rho_s'^2)^2 \, \mathrm{d}s, \qquad \mathcal{C}^{ij} = v_{ij}^{H,0} + \sum_{k=1}^{\infty} (v_{ij}^{H,k} + v_{ij}^{H,k}), \tag{2.15}$$

*with*

$$v_{ij}^{H,k} = \Gamma_k^H \Gamma_{|i-j+k|}^H + \Gamma_{|k-j|}^H \Gamma_{k+i}^H. \tag{2.16}$$

Theorem 2.1 is a special case of Theorems A.1 and A.2, which are stated and proved in the appendix. Note that if $H > \frac{3}{4}$, the term $\Delta_n^{2H-1}\Gamma^H \Pi_t$ in (2.12) is dominated by the Gaussian fluctuation process $\Delta_n^{1/2}\mathcal{Z}'$ and can thus be omitted in this case. The same comment applies to (2.14), where the term $e_1 C_t$ can be dropped if $H < \frac{1}{4}$.

**3. Semiparametric estimation of mixed semimartingales.** In order to construct rate-optimal estimators of $\Theta_T = (H, C_T, \Lambda_T, \Pi_T)$, we combine Theorem 2.1 with a GMM approach. The idea is to choose $r$ lags and, at stage $n$, a symmetric positive weight matrix $\mathcal{W}_n \in \mathbb{R}^{r\times r}$ (which can be a random statistic) and to obtain an estimator of $\Theta_T$ by solving the minimization problem

$$\underset{\theta=(H,C,\Lambda,\Pi)}{\operatorname{argmin}} \|\mathcal{W}_n^{1/2}(\widehat{V}_T^n - e_1 C - \Delta_n^{H-\frac{1}{2}} \Phi^H \Lambda - \Delta_n^{2H-1} \Gamma^H \Pi)\|_2^2, \tag{3.1}$$

where $\|\cdot\|_2$ is the Euclidean norm. More precisely, we construct an estimator $\widehat{\Theta}_T^n = (\widehat{H}^n, \widehat{C}_T^n, \widehat{\Lambda}_T^n, \widehat{\Pi}_T^n)$ of $\Theta_T$ by solving the *estimating equation*

$$F_n(\theta) = 0, \tag{3.2}$$

where $F_n(\theta) = \nabla_\theta \|\mathcal{W}_n^{1/2}(\widehat{V}_T^n - \mu_n(\theta))\|_2^2 = -2D_\theta \mu_n(\theta)^T \mathcal{W}_n(\widehat{V}_T^n - \mu_n(\theta)) \in \mathbb{R}^4$ and $\mu_n(\theta) = \mu_n(H, C, \Lambda, \Pi) = e_1 C + \Delta_n^{H-1/2}\Phi^H \Lambda + \Delta_n^{2H-1}\Gamma^H \Pi \in \mathbb{R}^r$. Using the theory of estimating equations (see [26, 32]), we can derive the asymptotic properties of this estimator.

THEOREM 3.1. *Suppose that conditions* 2–5 *in Assumption* (CLT') *hold for the processes specified in* (A.2). *Further assume that* $r \geq 5$ *and that* $\mathcal{W}$ *and* $\mathcal{W}_n$ *are* (*possibly random*) *symmetric positive definite matrices in* $\mathbb{R}^{r\times r}$ *such that* $\mathcal{W}_n \xrightarrow{\mathbb{P}} \mathcal{W}$. *If* $H < \frac{1}{2}$, *suppose that* $\Pi_T > 0$ *almost surely*; *if* $H > \frac{1}{2}$, *suppose that* $\Lambda_T \neq 0$ *almost surely.*

1. *If* $H \in (\frac{1}{4}, \frac{1}{2})$, *there exists a sequence* $\widehat{\Theta}^n_T = (\widehat{H}^n, \widehat{C}^n_T, \widehat{\Lambda}^n_T, \widehat{\Pi}^n_T)$ *of estimators of* $\Theta_T = (H, C_T, \Lambda_T, \Pi_T)$ *such that* $\mathbb{P}(F_n(\widehat{\Theta}^n_T) = 0) \xrightarrow{\mathbb{P}} 1$ *and*

$$D_n^{-1}\big(\widehat{\Theta}^n_T - \Theta_T\big) \xrightarrow{\text{st}} E^{-1}\big(\partial_H \Gamma^H \Pi_T, e_1, \Phi^H, \Gamma^H\big)^T \mathcal{W}\mathcal{Z}_T, \tag{3.3}$$

*where*

$$D_n = \begin{pmatrix} \Delta_n^{\frac{1}{2}} & 0 & 0 & 0 \\ 0 & \Delta_n^{2H-\frac{1}{2}} & 0 & 0 \\ 0 & 0 & \Delta_n^H & 0 \\ 2\Delta_n^{\frac{1}{2}} \log \Delta_n^{-1} \Pi_T & 0 & 0 & \Delta_n^{\frac{1}{2}} \end{pmatrix}, \tag{3.4}$$

$$E = \big(\partial_H \Gamma^H \Pi_T, e_1, \Phi^H, \Gamma^H\big)^T \mathcal{W}\big(\partial_H \Gamma^H \Pi_T, e_1, \Phi^H, \Gamma^H\big) \in \mathbb{R}^{4\times 4},$$

*and* $\mathcal{Z}$ *is the same process as in Theorem* 2.1. *The matrix* $E$ *in the last display is regular. Moreover, the sequence* $\widehat{\Theta}^n_T$ *is* locally unique *in the sense that any other sequence* $\widetilde{\Theta}^n_T$ *of estimators such that* $\mathbb{P}(F_n(\widetilde{\Theta}^n_T) = 0) \xrightarrow{\mathbb{P}} 1$ *and* $\mathbb{P}(\|\widetilde{\Theta}^n_T - \Theta_T\|_2 \leq 1/(\log \Delta_n^{-1})^2) \xrightarrow{\mathbb{P}} 1$ *satisfies* $\mathbb{P}(\widetilde{\Theta}^n_T = \widehat{\Theta}^n_T) \xrightarrow{\mathbb{P}} 1$.

2. *If* $H \in (0, \frac{1}{4})$, *define* $F_n^{(1)}(\theta^{(1)}) = \nabla_\theta \|\mathcal{W}_n^{1/2}(\widehat{V}^n_T - \mu_n^{(1)}(\theta^{(1)}))\|_2^2$, *where* $\mu_n^{(1)}(\theta^{(1)}) = \mu_n^{(1)}(H, \Lambda, \Pi) = \Delta_n^{H-1/2}\Phi^H \Lambda + \Delta_n^{2H-1}\Gamma^H \Pi$. *Then there is a locally unique sequence* $\widehat{\Theta}^{n,(1)}_T = (\widehat{H}^{n,(1)}, \widehat{\Lambda}^{n,(1)}_T, \widehat{\Pi}^{n,(1)}_T)$ *of estimators of* $\Theta^{(1)}_T = (H, \Lambda_T, \Pi_T)$ *with the property* $\mathbb{P}(F_n^{(1)}(\widehat{\Theta}^{n,(1)}_T) = 0) \xrightarrow{\mathbb{P}} 1$ *such that* (3.3) *remains valid for* $\widehat{\Theta}^{n,(1)}_T - \Theta^{(1)}_T$ *with the second row and column* (*out of four*) *deleted from all vectors and matrices appearing in* (3.3) *and* (3.4).

3. *If* $H \in (\frac{1}{2}, \frac{3}{4})$, *Part* 1 *of the theorem remains true if* (3.3) *and* (3.4) *are replaced by*

$$D_n^{-1}\big(\widehat{\Theta}^n_T - \Theta_T\big) \xrightarrow{\text{st}} E^{-1}\big(\partial_H \Phi^H \Lambda_T, e_1, \Phi^H, \Gamma^H\big)^T \mathcal{W}\mathcal{Z}'_T, \tag{3.5}$$

*and*

$$D_n = \begin{pmatrix} \Delta_n^{1-H} & 0 & 0 & 0 \\ 0 & \Delta_n^{\frac{1}{2}} & 0 & 0 \\ \Delta_n^{1-H} \log(\Delta_n^{-1})\Lambda_T & 0 & \Delta_n^{1-H} & 0 \\ 0 & 0 & 0 & \Delta_n^{\frac{3}{2}-2H} \end{pmatrix}, \tag{3.6}$$

$$E = \big(\partial_H \Phi^H \Lambda_T, e_1, \Phi^H, \Gamma^H\big)^T \mathcal{W}\big(\partial_H \Phi^H \Lambda_T, e_1, \Phi^H, \Gamma^H\big) \in \mathbb{R}^{4\times 4},$$

*respectively, and* $\mathcal{Z}'$ *is the same process as in Theorem* 2.1.

4. *If* $H \in (\frac{3}{4}, 1)$, *define* $F_n^{(2)}(\theta^{(2)}) = \nabla_\theta \|\mathcal{W}_n^{1/2}(\widehat{V}^n_T - \mu_n^{(2)}(\theta^{(2)}))\|_2^2$, *where* $\mu_n^{(2)}(\theta^{(2)}) = \mu_n^{(2)}(H, C, \Lambda) = e_1 C + \Delta_n^{H-1/2}\Phi^H \Lambda$. *Then there is a locally unique sequence* $\widehat{\Theta}^{n,(2)}_T = (\widehat{H}^{n,(2)}, \widehat{C}^{n,(2)}_T, \widehat{\Lambda}^{n,(2)}_T)$ *of estimators of* $\Theta^{(2)}_T = (H, C_T, \Lambda_T)$ *satisfying* $\mathbb{P}(F_n^{(2)}(\widehat{\Theta}^{n,(2)}_T) = 0) \xrightarrow{\mathbb{P}} 1$ *such that* (3.5) *remains valid for* $\widehat{\Theta}^{n,(2)}_T - \Theta^{(2)}_T$ *with the last row and column* (*out of four*) *deleted from all vectors and matrices appearing in* (3.5) *and* (3.6).

5. *In the setting of Part* 2 (*resp*., *Part* 4), *if a sequence of estimators* $\widehat{\Theta}^n_T = (\widehat{H}^n, \widehat{C}^n_T, \widehat{\Lambda}^n_T, \widehat{\Pi}^n_T)$ *satisfies* $\mathbb{P}(F_n(\widehat{\Theta}^n_T) = 0) \to 1$, *then the weak convergence statements in Part* 2 (*resp*., *Part* 4) *remain valid with* $\overline{\Theta}^n_T = (\widehat{H}^n, \widehat{\Lambda}^n_T, \widehat{\Pi}^n_T)$ *instead of* $\widehat{\Theta}^{n,(1)}_T$ (*resp*., $\overline{\Theta}^n_T = (\widehat{H}^n, \widehat{C}^n_T, \widehat{\Lambda}^n_T)$ *instead of* $\widehat{\Theta}^{n,(2)}_T$).

The proof will be given in Section 6. Note that a nondiagonal rate matrix also occurs in similar situations where a self-similarity parameter is estimated; see [7, 12, 32], for example.

In the case $H \in (\frac{1}{4}, \frac{3}{4})$, all parameters of the model are identifiable, and Parts 1 and 3 of Theorem 3.1 describe how the exact value of $H$ affects the asymptotic behavior of the estimators. In the case $H \notin (\frac{1}{4}, \frac{3}{4})$, the model is only partially identifiable: if $H < \frac{1}{4}$ (Part 2), we cannot consistently estimate $C_T$, while if $H > \frac{3}{4}$ (Part 4), we cannot consistently estimate $\Pi_T$. Parts 2 and 4 of Theorem 3.1 state that in these partially identifiable cases, one may obtain asymptotically normal estimators by reducing the GMM equations to only include identifiable parameters. However, these estimators are infeasible if the regime of $H$ is not known. Fortunately, by Part 5, the feasible GMM estimator (3.1) can still be employed in the regime $H \notin (\frac{1}{4}, \frac{3}{4})$ to derive asymptotically normal estimators for all identifiable parameters.

REMARK 3.2. The assumption $\Lambda_T \neq 0$ is important in Theorem 3.1 if $H > \frac{1}{2}$ to ensure identifiability of all parameters. For example, in the case of an mfBm as in (1.1), if $\lambda = 0$, then by [41] there is no way to consistently estimate $H$ and $\rho$ if $H > \frac{3}{4}$. Moreover, by Theorem 1.1, it will not be possible to asymptotically distinguish the model $(H, \sigma, \lambda, \rho) = (H_0, \sigma_0, \lambda_0, \rho_0) \in (\frac{3}{4}, 1) \times (0, \infty) \times (0, 1] \times (0, \infty)$ from the model $(H, \sigma, \lambda, \rho) = (H_1, \sigma_1, 0, \rho_1)$ if $H_1 = \frac{1}{2}(\frac{1}{2} + H_0)$, $\sigma_1 = \sigma_0$ and $\rho_1 = (2\lambda_0\rho_0\sigma_0/(K_{H_0}(H_0 + \frac{1}{2})))^{1/2}$. To see this, note that the process $Y$ in the model $(H_0, \sigma_0, \lambda_0, \rho_0)$ has the same law as $\sigma_0 W + \rho_1 W(H_1) + \rho_0 W(H_0)$ by Theorem 1.1. Moreover, by [41], the laws of $\sigma_0 W + \rho_0 W(H_0)$ and $\sigma_0 W$ are locally equivalent, so we deduce that the laws of $\sigma_0 W + \rho_0 W(H_0) + \rho_1 W(H_1)$ and $\sigma_0 W + \rho_1 W(H_1)$ are equivalent by noting that convolution of measures preserves equivalence. However, since $\sigma$ remains the same in both models, this local equivalence has no consequence for estimating $\sigma$, which often (e.g., in econometrics) is the main parameter of interest. We also note that if $\frac{1}{2} < H < \frac{3}{4}$ and $\Lambda_T = 0$, then the rate of convergence for estimating $H$ will be slower (equal to $\Delta_n^{3/2-2H}$, see [20]), as one can no longer rely on the fictitious fBm for inferring $H$.

In order to make the CLTs of Theorem 3.1 feasible, we adapt the results of [29] to construct consistent estimators of the involved asymptotic covariance matrices. This further allows us to choose an optimal weight matrix $\mathcal{W}_n$. We choose two integer sequences $k_n$ and $\ell_n$ and define

(3.7)

$$\widehat{\Sigma}_n = \widehat{\Sigma}^{(0)}_n + \sum_{\ell=1}^{\ell_n} w(\ell, \ell_n)\big(\widehat{\Sigma}^{(\ell)}_n + (\widehat{\Sigma}^{(\ell)}_n)^T\big),$$

$$\widehat{\Sigma}^{(\ell)}_n = \Delta_n \sum_{i=\ell+1}^{[T/\Delta_n]-r+1} \psi^{(i)}\big(\psi^{(i-\ell)}\big)^T \in \mathbb{R}^{r\times r}, \qquad \psi^{(i)} = \big(\psi^{(i)}_0, \dots, \psi^{(i)}_{r-1}\big)^T,$$

$$\psi^{(i)}_j = \Delta^n_i Y \Delta^n_{i+j} Y - \widehat{m}^{n,j}_i, \qquad \widehat{m}^{n,j}_i = \frac{1}{k_n}\sum_{k=0}^{k_n-1} \Delta^n_{i+k} Y \Delta^n_{i+k+j} Y,$$

$$\widehat{\eta}_n = \big(\Delta_n^{2\widehat{H}^n} \widehat{\Pi}^n_T\big(\partial_H \Gamma^{\widehat{H}^n} - 2\big(\log \Delta_n^{-1}\big)\Gamma^{\widehat{H}^n}\big) + \Delta_n^{\widehat{H}^n+1/2} \widehat{\Lambda}^n_T\big(\partial_H \Phi^{\widehat{H}^n} - \big(\log \Delta_n^{-1}\big)\Phi^{\widehat{H}^n}\big),$$
$$\Delta_n e_1, \Delta_n^{\widehat{H}^n+1/2}\Phi^{\widehat{H}^n}, \Delta_n^{2\widehat{H}^n}\Gamma^{\widehat{H}^n}\big) \in \mathbb{R}^{r\times 4}$$

for some deterministic weight function $w$.

COROLLARY 3.3. *Assume the conditions of (any part of) Theorem 3.1 and suppose that* $k_n, \ell_n \to \infty$ *with* $\ell_n/\sqrt{k_n} \to 0$ *and* $\ell_n\sqrt{k_n\Delta_n} \to 0$. *Further assume that* $w$ *is uniformly bounded and satisfies* $w(\ell, \ell_n) \to 1$ *for every* $\ell \geq 1$. *If we denote the diagonal entries of*

$$\mathcal{V}_n = \Delta_n(\widehat{\eta}_n^T \mathcal{W}_n \widehat{\eta}_n)^{-1} \widehat{\eta}_n^T \mathcal{W}_n \widehat{\Sigma}_n \mathcal{W}_n \widehat{\eta}_n (\widehat{\eta}_n^T \mathcal{W}_n \widehat{\eta}_n)^{-1} \in \mathbb{R}^{4\times 4} \tag{3.8}$$

*by* $\mathcal{V}_n^H$, $\mathcal{V}_n^C$, $\mathcal{V}_n^\Lambda$ *and* $\mathcal{V}_n^\Pi$, *then asymptotic* $\gamma$*-confidence intervals for* $H$, $C_T$ *(if* $H > \frac{1}{4}$*),* $\Lambda_T$ *and* $\Pi_T$ *(if* $H < \frac{3}{4}$*) for* $\gamma \in (0, 1)$ *are given by*

$$\big[\widehat{H}^n \pm \Phi^{-1}((1-\gamma)/2)\sqrt{\mathbb{V}_n^H}\big], \qquad \big[\widehat{C}_T^n \pm \Phi^{-1}((1-\gamma)/2)\sqrt{\mathbb{V}_n^C}\big],$$

$$\big[\widehat{\Lambda}_T^n \pm \Phi^{-1}((1-\gamma)/2)\sqrt{\mathbb{V}_n^\Lambda}\big], \qquad \big[\widehat{\Pi}_T^n \pm \Phi^{-1}((1-\gamma)/2)\sqrt{\mathbb{V}_n^\Pi}\big],$$

*respectively, where* $\Phi$ *is the cumulative distribution function of the standard normal law and* $\widehat{\Theta}_T^n = (\widehat{H}^n, \widehat{C}_T^n, \widehat{\Lambda}_T^n, \widehat{\Pi}_T^n)$ *is the solution to* (3.2).

REMARK 3.4 (Optimal GMM). If we choose $\mathcal{W}_n = \widehat{\Sigma}_n^{-1}$ and we have $H \in (\frac{1}{4}, \frac{1}{2})$ (resp., $H \in (\frac{1}{2}, \frac{3}{4})$), then the right-hand side of (3.3) (resp., (3.5)) has a centered Gaussian distribution with mean 0 and covariance matrix $E^{-1}$ with $\mathcal{W} = \mathcal{C}_T^{-1}$ (resp., $\mathcal{W} = (\mathcal{C}_T')^{-1}$). Analogous statements hold if $H \in (0, \frac{1}{4})$ and $H \in (\frac{3}{4}, 1)$.

## 4. Statistical lower bounds.

To derive a statistical lower bound, we consider the parametric setup of an mfBm

$$Y_t = \int_0^t \sigma \,\mathrm{d}B_s + \rho \int_{-\infty}^t h_H(t,s)\,\mathrm{d}B_s + \rho' \int_{-\infty}^t h_H(t,s)\,\mathrm{d}B_s', \tag{4.1}$$

where $h_H(t,s) = K_H^{-1}[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}]$, $\sigma > 0$, $\rho, \rho' \in \mathbb{R}$, and $B$ and $B'$ are two independent standard Brownian motions. Note that this model is a special case of (A.1) but with $a_s = (H - \frac{1}{2})K_H^{-1}\int_{-\infty}^0 (s-r)^{H-3/2}(\rho\,\mathrm{d}B_r + \rho'\,\mathrm{d}B_r')$, which is unbounded as $s \downarrow 0$ if $H < \frac{1}{2}$. Nevertheless, one can show that any small neighborhood around 0 only has a negligible impact on the asymptotics in $V_t^n$, so Theorem 2.1 remains valid.

The methods presented in Section 3 therefore yield estimators of the parameters $H$, $\sigma^2$, $\Pi = \rho^2 + \rho'^2$ and $\Lambda = \rho\sigma$, with rates given in Table 1. We show that these rates are optimal, by establishing matching minimax lower bounds. To this end, consider model (4.1) with parameter vector $\theta = (H, \sigma^2, \Lambda, \Pi) \subset \Theta$ for the open parameter set

$$\Theta = \left\{(H, \sigma^2, \Lambda, \Pi) \in \mathbb{R}^4 : H \in (0,1) \setminus \left\{\frac{1}{2}\right\}, \sigma^2 > 0,\ \Lambda \neq 0,\ \Pi > 0,\ \Lambda^2 < \sigma^2\Pi\right\}.$$

We use the notation $H(\theta) = \theta_1$, $\sigma^2(\theta) = \theta_2$ etc. For $\theta_0 \in \Theta$, define the local parameter set

$$\mathcal{D}_n(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq 1/|\log \Delta_n|\}$$

TABLE 1
*Rates of convergence of the estimators presented in Section 3*

| Parameter | $H \in (0, \frac{1}{2})$ | $H \in (\frac{1}{2}, 1)$ |
|---|---|---|
| $H$ | $\Delta_n^{\frac{1}{2}}$ | $\Delta_n^{1-H}$ |
| $\sigma^2$ | $\Delta_n^{2H-\frac{1}{2}}$ (if $H > \frac{1}{4}$) | $\Delta_n^{\frac{1}{2}}$ |
| $\Lambda$ | $\Delta_n^H$ | $\Delta_n^{1-H}\lvert\log\Delta_n\rvert$ |
| $\Pi$ | $\Delta_n^{\frac{1}{2}}\lvert\log\Delta_n\rvert$ | $\Delta_n^{\frac{3}{2}-2H}$ (if $H < \frac{3}{4}$) |

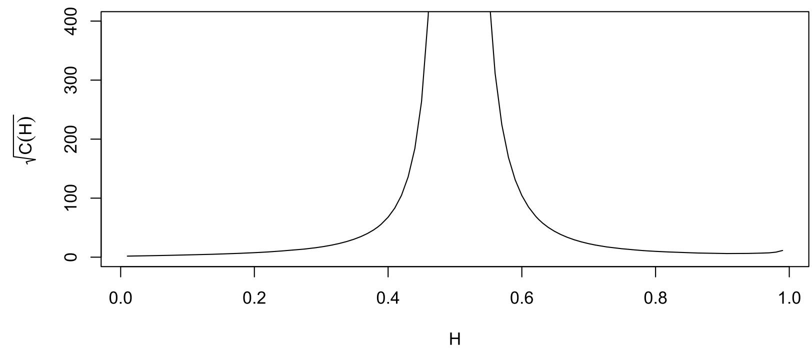


FIG. 1. *Constant in the asymptotic variance of the optimal GMM estimator of $H$ in an mfBm model.*

and the rate vector

$$R_n(\theta_0) = \begin{cases} \big(\Delta_n^{\frac{1}{2}}, \Delta_n^{2H-\frac{1}{2}}, \Delta_n^H, \Delta_n^{\frac{1}{2}}/|\log \Delta_n|\big) & \text{if } H(\theta_0) < \dfrac{1}{2}, \\ \big(\Delta_n^{1-H}, \Delta_n^{\frac{1}{2}}, \Delta_n^{1-H}/|\log \Delta_n|, \Delta_n^{\frac{3}{2}-2H}\big) & \text{if } H(\theta_0) > \dfrac{1}{2}. \end{cases}$$

For the regime $H \in (0, \frac{1}{4})$, the parameter $\sigma^2$ is not identifiable, as evidenced by the deteriorating rates of convergence. The same holds for $\Pi$ in the regime $H \in (\frac{3}{4}, 1)$.

THEOREM 4.1. *Let $\theta_0 \in \Theta$ be such that $\Lambda(\theta_0) \neq 0$. If $H(\theta_0) \in (\frac{1}{4}, \frac{3}{4}) \setminus \{\frac{1}{2}\}$, there exists some $c > 0$ such that*

$$\limsup_{n\to\infty} \inf_{\widehat{\theta}_n} \sup_{\theta\in\mathcal{D}_n(\theta_0)} \mathbb{P}_\theta\big(\big|(\widehat{\theta}_n - \theta)_k\big| \geq c\, R_n(\theta_0)_k\big) > 0, \quad k = 1, 2, 3, 4, \tag{4.2}$$

*where the infimum is taken among all measurable functions $\widehat{\theta}_n$ of $\{Y_{i\Delta_n} : i = 1, \dots, [1/\Delta_n]\}$. If $H \in (0, \frac{1}{4}]$ (resp., $H \in [\frac{3}{4}, 1)$), then (4.2) remains true for $k = 1, 3, 4$ (resp., $k = 1, 2, 3$).*

While our current methods do not allow us to determine the sharp value of the constant $c$ in (4.2), we conjecture that the GMM estimators from the previous section (if the weight matrix is chosen as in Remark 3.4) are close to being asymptotically efficient. Figure 1 plots the value of the constant $\sqrt{C(H)}$ as a function of $H$, where $C(H)$ is the asymptotic variance of the optimal GMM estimator $\widehat{H}^n$ from Theorem 3.1 in the mfBm model (1.1), that is, $C(H)$ is the constant that satisfies $\Delta_n^{-1/2}(\widehat{H}^n - H) \xrightarrow{d} N(0, C(H)/T)$ if $H < \frac{1}{2}$ and $\Delta_n^{H-1}(\widehat{H}^n - H) \xrightarrow{d} N(0, C(H)(\frac{\sigma}{\lambda\rho})^2/T))$ if $H > \frac{1}{2}$.

As we can see, while the rate of convergence of $\widehat{H}^n$ for $H > \frac{1}{2}$ improves as $H \downarrow \frac{1}{2}$, the associated constant deteriorates. This happens because in the limit as $H \downarrow \frac{1}{2}$, mfBm converges in distribution to Brownian motion, with the consequence that the fBm and the BM parts can no longer be separately identified. The same argument clearly applies when $H \uparrow \frac{1}{2}$, except that here the convergence rate itself does not change with $H$. The divergence of $C(H)$ as $H \to \frac{1}{2}$ will have an impact on the finite-sample performance of our estimators, as we shall see momentarily in a Monte Carlo simulation. Besides the singular behavior at $H \approx \frac{1}{2}$ we also find the asymptotic variance to be very large in absolute terms. This demonstrates the intrinsic difficulty of the statistical problem.

**5. Simulation study.** In order to evaluate the performance of the estimators from Theorem 3.1, we simulate $\{Y_{i\Delta_n} : i = 1, \dots, [T/\Delta_n]\}$ from the mfBm model (1.1) with $\Delta_n = 1/23{,}400$ and $T = 20$, which in a typical financial context corresponds to sampling every second and aggregating one month of data. We consider $H \in \{0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8, 0.9\}$ and $\lambda \in \{-0.9, -0.5, 0, 0.5, 0.9\}$ and further take $\sigma = 0.02$. In order to simulate from a setting that is representative of the magnitude of noise in high-frequency financial data (see, e.g., [3]), we fix the signal-to-noise ratio to $2:1$, that is, we assume that the increments of Brownian motion are responsible for $2/3$ of the variance of $\Delta_i^n Y$ and compute $\rho$ accordingly. We also include $H = 0.5$, in which case we let $\rho = \lambda = 0$. In Appendix E of the Supplementary Material [14], we present additional simulation results where we fix $\sigma$ and $\rho$ instead of the signal-to-noise ratio.

Regarding the tuning parameters, we choose $r = 31$ and $k_n = 300 \approx 2\Delta_n^{-1/2}$ in (3.7), which corresponds to considering autocorrelations up to half a minute and computing the local autocovariance estimates $\widehat{m}_i^{n,j}$ over 5-minute blocks (if $\Delta_n$ has the meaning of one second). We also experimented with $r = 16$ and obtained similar results. In (3.7), we further choose the Parzen kernel $w(\ell, \ell_n) = w(\ell/(\ell_n + 1))$, where $w(x) = (1 - 6x^2 + 6x^3)\mathbb{1}_{\{x \le 1/2\}} + 2(1-x)^3\mathbb{1}_{\{x>1/2\}}$ and the sequence $\ell_n$ is chosen according to the optimal procedure by [36] (with the details given in Table I B and Table II C of the reference). With this choice, $\widehat{\Sigma}_n$ is positive semidefinite in finite samples and $\ell_n$ is of order $\Delta_n^{-1/5}$ and hence satisfies $\ell_n/\sqrt{k_n} \to 0$ and $\ell_n\sqrt{k_n\Delta_n} \to 0$ if $k_n$ is order $\Delta_n^{-1/2}$.

For every simulated path, we first use a classical Ljung–Box test statistic $Q_T^n = \lfloor T/\Delta_n \rfloor(\lfloor T/\Delta_n \rfloor + 2)\sum_{\ell=1}^{r-1}(R_\ell^n)^2/(\lfloor T/\Delta_n \rfloor - \ell)$, where $R_\ell^n$ is the sample autocorrelation coefficient of $\{\Delta_i^n Y : i = 1, \dots, \lfloor T/\Delta_n \rfloor\}$ at lag $\ell$, to discriminate whether $H = \frac{1}{2}$ or not. Indeed, if $H = \frac{1}{2}$, we have $S_T^n \xrightarrow{d} \chi_{r-1}^2$, while $S_T^n \xrightarrow{\mathbb{P}} \infty$ in the case $H \neq \frac{1}{2}$ because $R_\ell^n \xrightarrow{\mathbb{P}} \Gamma_\ell^H \neq 0$. If the Brownian case $H = \frac{1}{2}$ is not rejected, we let $(H_n, C_n, \Lambda_n, \Pi_n) = (\frac{1}{2}, \widehat{V}_{0,T}^n, 0, 0)$ be the estimated parameter vector. This initial test is necessary because as $H \to \frac{1}{2}$, the mfBm model collapses to Brownian motion. Thus, if $H = \frac{1}{2}$ (or close to $\frac{1}{2}$), simply minimizing (3.1) typically produces a value of $H$ which is very close to $\frac{1}{2}$ and arbitrary splits the total variance among $C$, $\Pi$ and $\Lambda$. This is in line with the behavior of the asymptotic variance as $H \to \frac{1}{2}$ depicted in Figure 1. Therefore, the initial test we perform can be seen as shrinking our estimators towards the Brownian model to circumvent the weakly identified regime.

If $H = \frac{1}{2}$ is rejected, we let $(H_n, C_n, \Lambda_n, \Pi_n)$ be a numerical solution to (3.1) found in the following way: For each candidate $H$, we first we run a weighted linear regression of $\mathcal{W}_n^{1/2}\widehat{V}_T^n$ on $\mathcal{W}_n^{1/2}e_1$, $\Delta_n^{H-1/2}\mathcal{W}_n^{1/2}\Phi^H$ and $\Delta_n^{2H-1}\mathcal{W}_n^{1/2}\Gamma^H$ (with intercept forced to be 0 and the optimal weight matrix $\mathcal{W}_n = \widehat{\Sigma}_n^{-1}$) with the constraint that $C$, $\Lambda$ and $\Pi$ must satisfy $\Pi \ge 0$, $C \ge 0$ and $\Lambda^2 \le \Pi C$. This is to reflect the fact that $\rho^2 \ge 0$, $\sigma^2 \ge 0$ and $\Lambda_T^2 \le \Pi_T C_T$. Denoting the resulting coefficients by $C_n(H)$, $\Lambda_n(H)$ and $\Pi_n(H)$, we construct $H_n$ by minimizing the objective function $H \mapsto \text{score}(H)$ on the interval $(0, 1)$, where $\text{score}(H)$ is the sum of squared residuals in the regression analysis associated with $H$ but set to $\text{score}(H) = \infty$ if $C_n(H)/(\Pi_n(H)\Delta_n^{2H-1} + \Phi_0^H\Lambda_n(H)\Delta_n^{H-1/2} + C_n(H)) > 0.99$. In the latter case, we consider the fractional part as practically absent if the Brownian motion part accounts for more than 99% of the variance of increments. Indeed, if $H$ is close to or equal to $\frac{1}{2}$, it can happen that adding a tiny fractional component with $H$ very close to 0 can achieve a higher score value, which is undesirable in practice.

Finally, we define the remaining estimators by $C_n = C_n(H_n)$, $\Lambda_n = \Lambda_n(H_n)$ and $\Pi_n = \Pi_n(H_n)$. If $\text{score}(H) = \infty$ for all $H$, we let $H_n = 0.5$ and $(H_n, C_n, \Lambda_n, \Pi_n) = (\frac{1}{2}, \widehat{V}_{0,T}^n, 0, 0)$. Comparing with Theorem 3.1, we note that $(H_n, C_n, \Lambda_n, \Pi_n)$ is equal to $(\widehat{H}^n, \widehat{C}_T^n, \widehat{\Lambda}_T^n, \widehat{\Pi}_T^n)$ unless one of the constraints or exceptions above occur (which happen with asymptotically vanishing probability). Having obtained these estimators using $T = 20$ days of data, we further estimate integrated volatility on the last day, that is, $C_{20} - C_{19} = \sigma^2$, by rerunning the

TABLE 2
*Median and interquartile range of $H_n$ based on* 1000 *simulated paths*

| | $\lambda$ | | | | |
|---|---|---|---|---|---|
| $H$ | −0.9 | −0.5 | 0 | 0.5 | 0.9 |
| 0.1 | 0.0995 | 0.0989 | 0.0960 | 0.0976 | 0.1492 |
| | [0.0939, 0.1058] | [0.0860, 0.1133] | [0.0467, 0.1364] | [0.0582, 0.1951] | [0.1123, 0.2499] |
| 0.2 | 0.2000 | 0.1996 | 0.1975 | 0.2104 | 0.2590 |
| | [0.1959, 0.2038] | [0.1868, 0.2121] | [0.1139, 0.2443] | [0.1769, 0.3203] | [0.2237, 0.3646] |
| 0.3 | 0.2999 | 0.2992 | 0.2902 | 0.3292 | 0.3603 |
| | [0.2953, 0.3044] | [0.2806, 0.3166] | [0.2071, 0.3615] | [0.2887, 0.4081] | [0.3184, 0.4198] |
| 0.4 | 0.3994 | 0.3981 | 0.3789 | 0.4019 | 0.4148 |
| | [0.3818, 0.4084] | [0.3192, 0.4157] | [0.3493, 0.4287] | [0.3699, 0.4389] | [0.3775, 0.4679] |
| 0.5 | | | 0.5000 | | |
| | | | [0.5000,0.5000] | | |
| 0.6 | 0.5999 | 0.5986 | 0.5954 | 0.5977 | 0.5818 |
| | [0.5942, 0.6133] | [0.5888, 0.6441] | [0.5726, 0.6463] | [0.5663, 0.6255] | [0.5428, 0.6151] |
| 0.7 | 0.6994 | 0.6991 | 0.6983 | 0.7044 | 0.6629 |
| | [0.6922, 0.7071] | [0.6888, 0.7107] | [0.6739, 0.7336] | [0.6298, 0.7332] | [0.6012, 0.6962] |
| 0.8 | 0.7978 | 0.7975 | 0.7981 | 0.7975 | 0.7879 |
| | [0.7916, 0.8055] | [0.7897, 0.8072] | [0.7856, 0.8126] | [0.7688, 0.8364] | [0.7354, 0.8081] |
| 0.9 | 0.8918 | 0.8913 | 0.8905 | 0.8895 | 0.8885 |
| | [0.8819, 0.9058] | [0.8801, 0.9062] | [0.8778, 0.9074] | [0.8727, 0.9118] | [0.8635, 0.9067] |

weighted and constrained linear regression mentioned above, but using only data of the last day and with $H$ fixed at $H_n$ (which was previously obtained using 20 days of data). We denote this estimator by $C_n^{[19,20]}$. In a financial context, this mimics daily estimation of integrated volatility based on an estimate of $H$ obtained from a moving window of one month.

Table 2 summarizes the results for $H_n$. If $\lambda = -0.9$ or $\lambda = -0.5$, the estimator $H_n$ is centered around the true value of $H$ with only low to moderate dispersion uniformly for all considered values of $H$. At $\lambda = 0$ or $\lambda = 0.5$, $H_n$ is still relatively centered around its true value but there is a noticeable increase in the variability of the estimates. At $\lambda = 0.9$, the estimator $H_n$ exhibits a clear upward (resp., downward) bias for $H < 0.5$ (resp., $H > 0.5$). It is interesting that this bias together with an increase in the spread only appears at the positive end but not at the negative end of $\lambda$. In fact, additional plots (not shown here to save space) reveal that for negative values of $\lambda$, the empirical distribution of $H_n$ has a symmetrical bell shape around the true value of $H$, while for positive values of $\lambda$, the empirical distribution of $H_n$ becomes bimodal, with one local maximum around $H$ and another one at some value not far from $\frac{1}{2}(\frac{1}{2} + H)$. We believe that this is due to the fact that the original fBm with Hurst index $H$ and the fictitious one with Hurst parameter $\frac{1}{2}(H + \frac{1}{2})$ in the case of positive (resp., negative) $\lambda$ introduce return autocorrelations with the same (resp., opposite) sign, making it harder (resp., easier) to separate them.

The situation is different for the volatility estimator $C_n^{[19,20]}$, as Table 3 reveals. Here the results vary much less with $\lambda$ but mainly with the value of $H$ itself. Both the median bias and the interquartile range tend to increase as $H$ gets closer to but does not reach $\frac{1}{2}$. This is because the fBm, the fictitious fBm and the Brownian motion collapse to one in the limit as $H \to \frac{1}{2}$, making it harder to distinguish the three as $H$ approaches $\frac{1}{2}$ (see the discussion at the end of Section 4). It is interesting to note that the volatility estimator performs quite well even if $H < \frac{1}{4}$, even though in theory volatility is not identifiable in this case. This, of course, is due to our simulation setup, in which we fix the signal-to-noise ratio rather than $\sigma$ and $\rho$. Indeed, if we fix $\sigma$ and $\rho$ instead, then the signal-to-noise ratio decreases sharply with $H$ (e.g.,

TABLE 3
*Median and interquartile range of* $C_n^{[19,20]}/\sigma^2$ *based on* 1000 *simulated paths*

| $H$ | $\lambda$: −0.9 | −0.5 | 0 | 0.5 | 0.9 |
|---|---|---|---|---|---|
| 0.1 | 1.0005 | 1.0009 | 1.0064 | 1.0199 | 1.0229 |
| | [0.9614, 1.0390] | [0.9539, 1.0461] | [0.9585, 1.0553] | [0.9728, 1.0738] | [0.9698, 1.0940] |
| 0.2 | 0.9994 | 0.9984 | 1.0103 | 1.0442 | 1.0465 |
| | [0.9532, 1.0491] | [0.9325, 1.0753] | [0.9506, 1.1020] | [0.9774, 1.1547] | [0.9604, 1.1942] |
| 0.3 | 1.0004 | 1.0029 | 1.0302 | 1.0759 | 1.0722 |
| | [0.9232, 1.0882] | [0.8592, 1.1809] | [0.9414, 1.2249] | [0.9576, 1.3396] | [0.9246, 1.4233] |
| 0.4 | 0.9092 | 0.8077 | 1.0487 | 1.1634 | 1.2228 |
| | [0.4252, 1.3912] | [0.5862, 1.4440] | [0.8745, 1.4470] | [0.9696, 1.6896] | [0.9840, 2.0461] |
| 0.5 | | | 1.0007 | | |
| | | | [0.9942,1.1068] | | |
| 0.6 | 0.8722 | 0.8084 | 1.0727 | 1.1545 | 1.1850 |
| | [0.3602, 1.7096] | [0.4232, 1.5999] | [0.8752, 1.6210] | [0.9570, 1.5893] | [0.9590, 1.7388] |
| 0.7 | 1.0109 | 1.0158 | 1.0128 | 1.0727 | 1.0874 |
| | [0.7115, 1.3561] | [0.7783, 1.2695] | [0.9008, 1.2201] | [0.9859, 1.2302] | [0.9627, 1.3477] |
| 0.8 | 1.0240 | 1.0058 | 0.9995 | 1.0215 | 1.0297 |
| | [0.8769, 1.1752] | [0.8721, 1.1419] | [0.9220, 1.0882] | [0.9783, 1.0730] | [0.9862, 1.0846] |
| 0.9 | 1.0305 | 1.0005 | 0.9879 | 1.0035 | 1.0208 |
| | [0.9144, 1.1296] | [0.8994, 1.1160] | [0.9251, 1.0768] | [0.9647, 1.0469] | [0.9875, 1.0553] |

if $\sigma = \rho$ and $\Delta_n = 23{,}400$, the signal-to-noise ratio will be in the range 1:3000 to 1:3200, depending on $\lambda$, at $H = 0.1$). In additional simulations in Appendix E of the Supplementary Material [14], where we fix $\sigma$ and $\rho$, we do see that the variance of $C_n^{[19,20]}$ increases as $H$ becomes smaller.

Next, we report simulation results for $\Lambda_n/\sqrt{C_n \Pi_n}$ as an estimator of $\lambda$ in Table 4. Here the picture is closer to what we observed for $H_n$. For nonpositive values of $\lambda$, the estimator

TABLE 4
*Median and interquartile range of* $\Lambda_n/\sqrt{C_n \Pi_n}$ *(defined as* 0 *if* $\Pi_n = 0$*) based on* 1000 *simulated paths*

| $H$ | $\lambda$: −0.9 | −0.5 | 0 | 0.5 | 0.9 |
|---|---|---|---|---|---|
| 0.1 | −0.9005 | −0.5007 | 0.0104 | 0.5690 | 0.5619 |
| | [−0.9069, −0.8949] | [−0.5248, −0.4749] | [−0.1297, 0.2866] | [0.0162, 0.9707] | [−0.0461, 0.7417] |
| 0.2 | −0.9001 | −0.4999 | 0.0113 | 0.5377 | 0.4909 |
| | [−0.9050, −0.8945] | [−0.5382, −0.4569] | [−0.1839, 0.4411] | [−0.1937, 0.6355] | [−0.3396, 0.6186] |
| 0.3 | −0.8996 | −0.4970 | 0.0162 | 0.4539 | 0.4700 |
| | [−0.9089, −0.8903] | [−0.5864, −0.3905] | [−0.4076, 0.6350] | [−0.4873, 0.5873] | [−0.4875, 0.6017] |
| 0.4 | −0.9040 | −0.4844 | 0.5423 | 0.5490 | 0.5130 |
| | [−0.9265, −0.8240] | [−0.6839, 0.8631] | [−0.4863, 0.6871] | [−0.3783, 0.6134] | [−0.5045, 0.9777] |
| 0.6 | −0.9209 | −0.5482 | −0.1145 | 0.6914 | 0.6545 |
| | [−0.9997, −0.9035] | [−0.6673, 0.9978] | [−0.4754, 0.8584] | [−0.3639, 0.7694] | [−0.3771, 0.9811] |
| 0.7 | −0.9020 | −0.5082 | −0.0196 | 0.6289 | 0.4570 |
| | [−0.9171, −0.8824] | [−0.5955, −0.3894] | [−0.2687, 0.3519] | [−0.2415, 0.8973] | [−0.4257, 0.8507] |
| 0.8 | −0.9046 | −0.5117 | −0.0089 | 0.4797 | 0.8148 |
| | [−0.9166, −0.8866] | [−0.5536, −0.4564] | [−0.1019, 0.1021] | [0.2148, 1.0000] | [0.2354, 0.9998] |
| 0.9 | −0.9071 | −0.5226 | −0.0317 | 0.4544 | 0.8219 |
| | [−0.9174, −0.8929] | [−0.5552, −0.4771] | [−0.0952, 0.0486] | [0.3233, 0.6273] | [0.5653, 1.0000] |

TABLE 5
*Median and interquartile range of $\Pi_n/(20\rho^2)$ based on 1000 simulated paths*

| | $\lambda$ | | | | |
|---|---|---|---|---|---|
| $H$ | $-0.9$ | $-0.5$ | 0 | 0.5 | 0.9 |
| 0.1 | 0.9868 | 0.9726 | 0.9040 | 0.8409 | 3.8838 |
| | [0.8626, 1.1509] | [0.7019, 1.3861] | [0.2543, 2.5844] | [0.2994, 14.5667] | [1.5175, 84.0271] |
| 0.2 | 0.9998 | 0.9858 | 0.9354 | 1.1647 | 6.7238 |
| | [0.9008, 1.1021] | [0.7025, 1.3560] | [0.0912, 3.5754] | [0.4954, 44.1307] | [2.5029, 323.7322] |
| 0.3 | 0.9954 | 0.9829 | 0.8002 | 2.0526 | 8.1592 |
| | [0.8723, 1.1385] | [0.5547, 1.6563] | [0.0457, 8.3227] | [0.6617, 81.7545] | [2.5609, 228.6105] |
| 0.4 | 0.9831 | 0.9140 | 0.1906 | 0.8187 | 2.8029 |
| | [0.4771, 1.4473] | [0.0331, 2.0469] | [0.0851, 4.9195] | [0.2751, 13.0497] | [0.6495, 97.6562] |
| 0.6 | 1.0076 | 1.0299 | 1.1053 | 0.8441 | 1.5601 |
| | [0.8997, 1.0551] | [0.7188, 1.1265] | [0.5266, 1.5510] | [0.6166, 3.2883] | [1.1024, 6.5762] |
| 0.7 | 0.9935 | 0.9992 | 1.0033 | 1.0671 | 1.3798 |
| | [0.9484, 1.0526] | [0.9406, 1.0637] | [0.9400, 1.0914] | [0.8951, 1.2127] | [1.0304, 1.9491] |
| 0.8 | 0.9669 | 0.9662 | 0.9681 | 0.9780 | 0.9236 |
| | [0.8898, 1.0709] | [0.8805, 1.0889] | [0.8674, 1.1212] | [0.8290, 1.1729] | [0.8048, 1.0568] |
| 0.9 | 0.8405 | 0.8333 | 0.8348 | 0.8388 | 0.8157 |
| | [0.6926, 1.1298] | [0.6825, 1.1369] | [0.6746, 1.1419] | [0.6578, 1.1673] | [0.6453, 1.1054] |

performs relatively well (with exceptions), while biases start to show up as $\lambda$ moves into the positive range. In fact, comparing the results between $\lambda = 0.5$ and $\lambda = 0.9$, it seems that the estimator $\Lambda_n/\sqrt{C_n\Pi_n}$ has a hard time distinguishing between these two cases. We do not have a plausible explanation for this behavior.

Finally, we consider the simulation results for $\Pi_n$, which can be found in Table 5. Here the impact of $\lambda$ is particularly striking: the estimator shows a relatively good performance for negative values of $\lambda$ and is practically useless for positive values of $\lambda$. We conjecture that the cause is the same as before: with positive values of $\lambda$, both the original and the fictitious fBm components lead to autocorrelations of the same sign, making it difficult to separate them.

In summary, we draw the following conclusions from the simulation study for a scenario where the signal-to-noise ratio is fixed:

- The statistical estimation of mixed semimartingale models is a hard task in general. Even in a parametric mfBm model, the behavior of the estimators $H_n$, $\Pi_n$, $C_n^{[19,20]}$ and $\Lambda_n/\sqrt{\Pi_n C_n}$ is quite different for different true parameter values.
- If $\lambda$ is negative and $H$ is bounded away from $\frac{1}{2}$, all estimators $H_n$, $\Pi_n$, $C_n^{[19,20]}$ and $\Lambda_n/\sqrt{\Pi_n C_n}$ perform relatively well. This is exactly the case where the statistical properties of the fractional component, the fictitious fractional component and the Brownian motion component are sufficiently distinct to separate them from each other.
- In all other cases, at least two of three are statistically similar, making it intrinsically difficult to disentangle them.

## 6. Proof of Theorem 1.1, Theorem 3.1, Corollary 3.3 and Theorem 4.1.

PROOF OF THEOREM 1.1. One could derive the first statement of the theorem from general results about multivariate fractional Brownian motion [4]. For the sake of completeness, we give a short direct proof here. Since both $Y$ and $\widetilde{Y}$ are centered Gaussian processes with stationary increments, it suffices to study the variance of increments. On the one hand, $\mathbb{E}[(\widetilde{Y}_{t+h} - \widetilde{Y}_t)^2] = \sigma^2 h + 2\lambda\rho\sigma(K_H(H+\frac{1}{2}))^{-1}h^{2\overline{H}} + \rho^2 h^{2H}$; on the other hand, we can compute $\mathbb{E}[(Y_{t+h} - Y_t)^2]$ via Itô's isometry. Writing $h_H(t,s) = K_H^{-1}[(t-s)_+^{H-1/2} - (-s)_+^{H-1/2}]$, we have that $B(H)_t = \int_{-\infty}^t h_H(t,s)\,\mathrm{d}\widetilde{B}_s$ and

$$\mathbb{E}[(Y_{t+h}-Y_t)^2] = \sigma^2 h + \rho^2 h^{2H} + 2\lambda\rho\sigma \int_t^{t+h} [h_H(t+h,s) - h_H(t,s)]\,\mathrm{d}s, \tag{6.1}$$

where the last integral equals $\int_0^h [h_H(h,s) - h_H(0,s)]\,\mathrm{d}s = K_H^{-1} \int_0^h (h-s)^{H-\frac{1}{2}}\,\mathrm{d}s = h^{2\overline{H}}/(K_H(H+\frac{1}{2}))$. Thus, $\mathbb{E}[(Y_{t+h}-Y_t)^2] = \mathbb{E}[(\widetilde{Y}_{t+h}-\widetilde{Y}_t)^2]$ and $(Y_t)_{t\geq 0} \overset{d}{=} (\widetilde{Y}_t)_{t\geq 0}$.

It remains to show that $Y$ is not a semimartingale for $\lambda > 0$. To this end, we may employ the same arguments as [11]: If $H < \frac{1}{2}$, the process has an infinite quadratic variation, which contradicts the semimartingale property. If $H \in (\frac{3}{4}, 1)$, then the law of $\sigma W + \rho W(H)$ is locally equivalent to that of $\sigma W$. Thus, $Y$ is locally equivalent to $\sigma W + (2\lambda\rho\sigma/(K_H(H+\frac{1}{2})))^{1/2} W(\overline{H})$. Since $\overline{H} \in (\frac{1}{2}, \frac{3}{4})$, [11], Theorem 1.7, shows that $Y$ is not a semimartingale. If $H \in (\frac{1}{2}, \frac{3}{4}]$, then $Y$ is the sum of three independent (fractional) Brownian motions. The proof given in [11], Section 4, which is presented for two components, straightforwardly generalizes to the case of three fractional Brownian motions, showing that $Y$ is not a semimartingale. □

PROOF OF THEOREM 3.1. We want to apply [32], Theorem A.2, so we verify the associated conditions called (E.1) and (E.2)'. Note that the latter theory is formulated for classical weak convergence but readily extends to stable convergence. We only consider the case $H \in (\frac{1}{4}, \frac{1}{2}) \cup (\frac{1}{2}, \frac{3}{4})$ as the other two cases $H \in (0, \frac{1}{4})$ and $H \in (\frac{3}{4}, 1)$ are analogous. For (E.1), define the matrices $A_n = A_n(\theta)$, $B_n = B_n(\theta) \in \mathbb{R}^{4\times 4}$ as

$$\begin{cases} A_n = \Delta_n^{\frac{1}{2}-2H} B_n, \qquad B_n = \begin{pmatrix} \Delta_n^{1-2H} & 0 & 0 & 2\Delta_n^{1-2H}\log(\Delta_n^{-1})\Pi \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \Delta_n^{\frac{1}{2}-H} & 0 \\ 0 & 0 & 0 & \Delta_n^{1-2H} \end{pmatrix} & \text{for } H < \dfrac{1}{2}, \\ A_n = \Delta_n^{-\frac{1}{2}} B_n, \qquad B_n = \begin{pmatrix} \Delta_n^{\frac{1}{2}-H} & 0 & \Delta_n^{\frac{1}{2}-H}\log(\Delta_n^{-1})\Lambda & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \Delta_n^{\frac{1}{2}-H} & 0 \\ 0 & 0 & 0 & \Delta_n^{1-2H} \end{pmatrix} & \text{for } H > \dfrac{1}{2}. \end{cases}$$

Since $D_\theta \mu_n(\theta) = (\Delta_n^{H-\frac{1}{2}}(\partial_H \Phi^H - \Phi^H \log \Delta_n^{-1})\Lambda + \Delta_n^{2H-1}(\partial_H \Gamma^H - 2\Gamma^H \log \Delta_n^{-1})\Pi, e_1, \Delta_n^{H-\frac{1}{2}} \Phi^H, \Delta_n^{2H-1}\Gamma^H) \in \mathbb{R}^{r\times 4}$, we have by Theorem 2.1 that

$$\begin{cases} A_n(\Theta_T) F_n(\Theta_T) \xrightarrow{\text{st}} -2(\partial_H \Gamma^H \Pi, e_1, \Phi^H, \Gamma^H)^T \mathcal{W} \mathcal{Z}_T & \text{for } H \in \left(\dfrac{1}{4}, \dfrac{1}{2}\right), \\ A_n(\Theta_T) F_n(\Theta_T) \xrightarrow{\text{st}} -2(\partial_H \Phi^H \Lambda, e_1, \Phi^H, \Gamma^H)^T \mathcal{W} \mathcal{Z}'_T & \text{for } H \in \left(\dfrac{1}{2}, \dfrac{3}{4}\right). \end{cases}$$

This proves (E.1) in [32].

For (E.2)', note that continuous differentiability of $F_n(\theta)$ around $\Theta_T$ is clear. Next, we observe that

$$D_\theta F_n(\theta) = -2D_\theta^2 \mu_n(\theta)^T \mathcal{W}_n(\widehat{V}_T^n - \mu_n(\theta)) + 2D_\theta \mu_n(\theta)^T \mathcal{W}_n D_\theta \mu_n(\theta).$$

A straightforward computation shows that $2B_n(\Theta_T)[D_\theta \mu_n(\theta)^T \mathcal{W}_n D_\theta \mu_n(\theta)] B_n(\Theta_T)^T \xrightarrow{\mathbb{P}} 2E$, locally uniformly in a shrinking neighborhood of size $r_n = 1/(\log \Delta_n^{-1})^2$ around $\Theta_T$.

Applying Theorem 2.1, we further have $B_n(\Theta_T)D_\theta^2\mu_n(\theta)^T\mathcal{W}_n(\widehat{V}_T^n-\mu_n(\theta))B_n(\Theta_T)^T\xrightarrow{\mathbb{P}}0$ locally uniformly. Hence,

$$\sup_{\theta:|\theta-\Theta_T|\le 1/(\log\Delta_n^{-1})^2}\|B_n(\Theta_T)D_\theta F_n(\theta)B_n(\Theta_T)^T-2E\|\xrightarrow{\mathbb{P}}0. \tag{6.2}$$

Finally, by (3.4) and (3.6), we can check that $\|B_n(\Theta_T)^T\|\|B_n(\Theta_T)A_n(\Theta_T)^{-1}\|/r_n\xrightarrow{\mathbb{P}}0$ in the range of $H$ we consider, which proves (E.2)' in [32], with $C_n=B_n^T$ and $W=2E$.

The matrix $E$ is regular because the vectors $e_1,\partial_H\Gamma^H,\Phi^H,\Gamma^H\in\mathbb{R}^r$ are linearly independent. For $r\to\infty$, this is evident as all four vectors have different decay rates. For $r\ge 5$ fixed, we can check that the $3\times 3$ submatrix consisting of the entries two, three and five (i.e., lags 1, 2, 4) of $\partial_H\Gamma^H,\Phi^H,\Gamma^H$ has a nonzero determinant. We have verified that this is the case for $H\neq\frac{1}{2}$, using a computer algebra system. Analogously, we have verified the regularity of the same matrix based on $\partial_H\Phi^H,\Phi^H,\Gamma^H$. Thus, [32], Theorem A.2, yields, for $H\in(\frac{1}{4},\frac{1}{2})$,

$$A_n(\Theta_T)B_n(\Theta_T)^{-1}E\big(B_n(\Theta_T)^{-1}\big)^T\big(\widehat{\Theta}_T^n-\Theta_T\big)\xrightarrow{\mathrm{st}}\big(\partial_H\Gamma^H\Pi_T,e_1,\Phi^H,\Gamma^H\big)^T\mathcal{W}\mathcal{Z}_T,$$

which is equivalent to (3.3). For $H\in(\frac{1}{2},\frac{3}{4})$, we obtain (3.5).

Part 2 (resp., Part 4) of the theorem can be derived along the same lines, but with the second (resp., fourth) row and column (out of four) deleted from all vectors and matrices. For Part 5, we restrict ourselves to the setting of Part 3, where $H\in(\frac{3}{4},1)$ (the proof in the setting of Part 2 is similar). On the event $F_n(\widehat{\Theta}_T^n)=0$, which happens with probability converging to 1, we have

$$0=F_n(\widehat{\Theta}_T^n)=-2D_\theta\mu_n(\widehat{\Theta}_T^n)^T\mathcal{W}_n(\widehat{V}_T^n-\mu_n(\widehat{\Theta}_T^n)).$$

Next, we introduce the matrix $G_n(\theta)=((\partial_H\Gamma^H-2\Gamma^H\log\Delta_n^{-1})\Delta_n^{2H-1}\Pi,0,0)\in\mathbb{R}^{r\times 3}$ and denote the restriction of a matrix $M\in\mathbb{R}^{4\times 4}$ to the upper left $3\times 3$-corner by $M^{(2)}$ and the restriction of a vector $v\in\mathbb{R}^4$ to the first three entries by $v^{(2)}$. Now, applying $(\cdot)^{(2)}$ to both sides of the previous display and multiplying the result by $A_n^{(2)}(\Theta_T^{(2)})$ (recall the definition of $A_n(\theta)$ and $B_n(\theta)$ from the beginning of this proof), we obtain

$$\begin{aligned}0&=-2A_n^{(2)}(\Theta_T^{(2)})\big(D_{\theta^{(2)}}\mu_n^{(2)}(\overline{\Theta}_T^n)+G_n(\widehat{\Theta}_T^n)\big)^T\mathcal{W}_n\big(\widehat{V}_T^n-\mu_n^{(2)}(\overline{\Theta}_T^n)-\Delta_n^{2\widehat{H}^n-1}\Gamma^{\widehat{H}^n}\widehat{\Pi}_T^n\big)\\&=-2A_n^{(2)}(\Theta_T^{(2)})D_{\theta^{(2)}}\mu_n^{(2)}(\overline{\Theta}_T^n)^T\mathcal{W}_n\big(\widehat{V}_T^n-\mu_n^{(2)}(\overline{\Theta}_T^n)\big)+o_{\mathbb{P}}(1)\\&=A_n^{(2)}(\Theta_T^{(2)})F_n^{(2)}(\overline{\Theta}_T^n)+o_{\mathbb{P}}(1).\end{aligned}$$

For the second equality, note that $H\in(\frac{3}{4},1)$ and that $A_n^{(2)}(\Theta_T^{(2)})$, $A_n^{(2)}(\Theta_T^{(2)})G_n(\widehat{\Theta}_T^n)^T$ and $D_{\theta^{(2)}}\mu_n^{(2)}(\overline{\Theta}_T^n)$ have matrix norms of order $\Delta_n^{-1/2}$, $\Delta_n^{H-1}\log\Delta_n^{-1}$ and 1, respectively, while $\widehat{V}_T^n-\mu_n^{(2)}(\overline{\Theta}_T^n)=O_{\mathbb{P}}(\Delta_n^{1/2})$ by Theorem 2.1. With high probability, $F_n^{(2)}(\widehat{\Theta}_T^{n,(2)})=0$, so in this case, Taylor's theorem gives us some $\widetilde{\Theta}_T^n$ between $\overline{\Theta}_T^n$ and $\widehat{\Theta}_T^{n,(2)}$ such that

$$\begin{aligned}0&=A_n^{(2)}(\Theta_T^{(2)})D_{\theta^{(2)}}F_n^{(2)}(\widetilde{\Theta}_T^n)\big(\overline{\Theta}_T^n-\widehat{\Theta}_T^{n,(2)}\big)+o_{\mathbb{P}}(1)\\&=A_n^{(2)}(\Theta_T^{(2)})B_n^{(2)}(\Theta_T^{(2)})^{-1}\big[B_n^{(2)}(\Theta_T^{(2)})D_{\theta^{(2)}}F_n^{(2)}(\widetilde{\Theta}_T^n)B_n^{(2)}(\Theta_T^{(2)})^T\big]\\&\quad\times\big(B_n^{(2)}(\Theta_T^{(2)})^T\big)^{-1}\big(\overline{\Theta}_T^n-\widehat{\Theta}_T^{n,(2)}\big)+o_{\mathbb{P}}(1).\end{aligned}$$

As in (6.2), one can show that $B_n^{(2)}(\Theta_T^{(2)})D_{\theta^{(2)}}F_n^{(2)}(\widetilde{\Theta}_T^n)B_n^{(2)}(\Theta_T^{(2)})^T\xrightarrow{\mathbb{P}}2E^{(2)}$. Since $A_n^{(2)}(\Theta_T^{(2)})B_n^{(2)}(\Theta_T^{(2)})^{-1}=\Delta_n^{-1/2}\mathrm{Id}_3$, $\Delta_n^{1/2}B_n^{(2)}(\Theta_T^{(2)})=D_n^{(2)}$ and $E^{(2)}$ is regular, we conclude that $(D_n^{(2)})^{-1}(\overline{\Theta}_T^n-\widehat{\Theta}_T^{n,(2)})\xrightarrow{\mathbb{P}}0$. $\square$

PROOF OF COROLLARY 3.3. As before, we only consider one case, namely when $H \in (\frac{1}{2}, \frac{3}{4})$. The other cases are similar. Let

$$\overline{m}_i^{n,j} = \sigma^2_{(i-1)\Delta_n} \mathbb{1}_{\{j=0\}} + \sigma_{(i-1)\Delta_n}\rho_{(i-1)\Delta_n}\Phi^H_j \Delta_n^{H-1/2} + \big(\rho^2_{(i-1)\Delta_n} + \rho'^2_{(i-1)\Delta_n}\big)\Gamma^H_j \Delta_n^{2H-1}$$

and $\overline{\psi}_j^{(i)}$ (resp., $\overline{\Sigma}_n^{(\ell)}$, $\overline{\Sigma}_n$) be defined in the same way as $\psi_j^{(i)}$ (resp., $\Sigma_n^{(\ell)}$, $\Sigma_n$) but with $\overline{m}_i^{n,j}$ (resp., $\overline{\psi}^{(i)}$, $\overline{\Sigma}_n^{(\ell)}$) substituted for $\widehat{m}_i^{n,j}$ (resp., $\psi^{(i)}$, $\widehat{\Sigma}_n^{(\ell)}$). Then, because $\sigma$, $\rho$ and $\rho'$ are at least $\frac{1}{2}$-Hölder continuous in $L^2$, one can borrow from classical results concerning spot volatility estimation (e.g., [25], Chapter 13.3) to show that $\Delta_n^{-1}(m_i^{n,j} - \overline{m}_i^{n,j}) = O_{\mathbb{P}}(k_n^{-1/2} \vee \sqrt{k_n\Delta_n})$ uniformly in $j$, which implies $\Delta_n^{-2}(\widehat{\Sigma}_n - \overline{\Sigma}_n) = O_{\mathbb{P}}(\ell_n(k_n^{-1/2} \vee \sqrt{k_n\Delta_n})) \xrightarrow{\mathbb{P}} 0$ by our assumptions on $\ell_n$ and $k_n$. Next, again because $\sigma$, $\rho$ and $\rho'$ are $\frac{1}{2}$-Hölder continuous in $L^2$, we have $\Delta_n^{-2}\overline{\Sigma}_n = \int_0^T [c_n^{(0)}(s) + \sum_{\ell=1}^{\ell_n} w(\ell,\ell_n)(c_n^{(\ell)}(s) + c_n^{(\ell)}(s)^T)]\,\mathrm{d}s + O_{\mathbb{P}}(\ell_n\sqrt{\Delta_n})$, where

$$\begin{aligned}
c_n^{(\ell)}(s)_{jj'} &= \sigma_s^4(2\mathbb{1}_{\{\ell=j=j'=0\}} + \mathbb{1}_{\{\ell=0,j=j'>0\}}) \\
&\quad + \sigma^2\rho^2\big(\Phi^H_\ell\Phi^H_{|j'-j-\ell|} + \Phi^H_{|j'-\ell|}\Phi^H_{j+\ell}\big)\Delta_n^{2H-1} \\
&\quad + \big(\rho_s^2 + \rho_s'^2\big)^2\big(\Gamma^H_\ell\Gamma^H_{|j'-j-\ell|} + \Gamma^H_{|j'-\ell|}\Gamma^H_{j+\ell}\big)\Delta_n^{4H-2} \\
&\quad + \sigma_s^3\rho_s\big(\Phi^H_{|j'-j|}(\mathbb{1}_{\{\ell=0\}} + \mathbb{1}_{\{\ell=j'-j\}}) + \Phi^H_{j+j'}(\mathbb{1}_{\{\ell=j'\}} + \mathbb{1}_{\{\ell=j=0\}})\big)\Delta_n^{H-1/2} \\
&\quad + \sigma_s^2\big(\rho_s^2 + \rho_s'^2\big) \\
&\quad \times \big(\Gamma^H_{|j'-j|}(\mathbb{1}_{\{\ell=0\}} + \mathbb{1}_{\{\ell=j'-j\}}) + \Gamma^H_{j+j'}(\mathbb{1}_{\{\ell=j'\}} + \mathbb{1}_{\{\ell=j=0\}})\big)\Delta_n^{2H-1} \\
&\quad + \sigma_s\rho_s\big(\rho_s^2 + \rho_s'^2\big) \\
&\quad \times \big(\Phi^H_\ell\Gamma^H_{|j'-j-\ell|} + \Gamma^H_\ell\Phi^H_{|j'-j-\ell|} + \Phi^H_{|j'-\ell|}\Gamma^H_{j+\ell} + \Gamma^H_{|j'-\ell|}\Phi^H_{j+\ell}\big)\Delta_n^{3H-1/2}.
\end{aligned}$$

Note that all terms defining $c_n^{(\ell)}(s)_{jj'}$ are summable in $\ell$ because $H < \frac{3}{4}$ and $\Gamma_r^H = O(r^{2H-2})$ and $\Phi_r^H = O(r^{H-3/2})$ as $r \to \infty$. Therefore, we have $\Delta_n^{-2}\overline{\Sigma}_n = \operatorname{diag}(2,1,\dots,1)\int_0^T \sigma_s^4\,\mathrm{d}s + O_{\mathbb{P}}(\ell_n\sqrt{\Delta_n} \vee \Delta_n^{H-1/2}) \xrightarrow{\mathbb{P}} \mathcal{C}'_T$. Since $\Delta_n^{-3/2}\widehat{\eta}_n D_n \xrightarrow{\mathbb{P}} (\partial_H\Phi^H\Lambda_T, e_1, \Phi^H, \Gamma^H)$, we obtain

$$\begin{aligned}
D_n^{-1}\mathcal{V}_n(D_n^{-1})^T &= \Delta_n\big(D_n^T\widehat{\eta}_n^T\mathcal{W}_n\widehat{\eta}_nD_n\big)^{-1}D_n^T\widehat{\eta}_n^T\mathcal{W}_n\widehat{\Sigma}_n\mathcal{W}_n\widehat{\eta}_nD_n\big(D_n^T\widehat{\eta}_n^T\mathcal{W}_n\widehat{\eta}_nD_n\big)^{-1} \\
&\xrightarrow{\mathbb{P}} E^{-1}\big(\partial_H\Phi^H\Lambda_T, e_1, \Phi^H, \Gamma^H\big)^T\mathcal{W}\mathcal{C}'_T\mathcal{W}\big(\partial_H\Phi^H\Lambda_T, e_1, \Phi^H, \Gamma^H\big)E^{-1}.
\end{aligned}$$

Recalling (3.5), we have shown that $\mathcal{V}_n$ consistently estimates the asymptotic covariance matrix of $\widehat{\Theta}_T^n$, which is the claim of the corollary. □

For the proof of Theorem 4.1, we assume $\Delta_n = \frac{1}{n}$ to simplify notation. Since $Y_0 = \widetilde{Y}_0 = 0$, observing $\{Y_{i/n} : i = 1,\dots,n\}$ is equivalent to observing the increments $\{\Delta_i^n Y : i = 1,\dots,n\}$. These increments constitute a stationary centered Gaussian time series with some covariance matrix $\widetilde{\Sigma}_n(\theta) \in \mathbb{R}^{n\times n}$. Noting that (6.1) is also valid for $\lambda < 0$, we find that

$$\widetilde{\Sigma}_n(\theta) = \sigma^2 n^{-1} I_n + \Pi n^{-2H}\Sigma_n(H) + \Lambda b(H) n^{-2\overline{H}}\Sigma_n(\overline{H}), \tag{6.3}$$

where $\overline{H} = \frac{1}{2}(H + \frac{1}{2})$, $b(H) = 2/\Gamma(H + \frac{3}{2})$ and $\Sigma_n(H) = (\Gamma^H_{|i-j|})_{j,k=1}^n$ is the covariance matrix of $n$ consecutive normalized increments of a fractional Brownian motion with Hurst parameter $H$. Given $\theta_0 \in \Theta$ and four nonnegative sequences $r_{1,n}, r_{2,n}, r_{3,n}, r_{4,n} \to 0$, we define $\theta_n \in \Theta$ by

$$H(\theta_n) = H + r_{1,n}, \qquad \sigma^2(\theta_n) = \sigma^2 + r_{2,n},$$
$$\Lambda(\theta_n) = \Lambda \frac{b(H)}{b(H + r_{1,n})} n^{r_{1,n}}(1 + r_{3,n}), \qquad \Pi(\theta_n) = \Pi n^{2r_{1,n}}(1 + r_{4,n}), \tag{6.4}$$

abbreviating $(H, \sigma, \Lambda, \Pi) = (H, \sigma, \Lambda, \Pi)(\theta_0)$. The parameter $\theta_n$ is chosen carefully such that

$$\begin{aligned}\widetilde{\Sigma}_n(\theta_n) &= (\sigma^2 + r_{2,n})n^{-1} I_n + (1 + r_{4,n})\Pi n^{-2H} \Sigma_n(H + r_{1,n}) \\ &\quad + (1 + r_{3,n})\Lambda b(H) n^{-2\overline{H}} \Sigma_n\left(\overline{H} + \frac{r_{1,n}}{2}\right).\end{aligned} \tag{6.5}$$

For now, we only assume that $r_{i,n} \to 0$ as $n \to \infty$ for $i = 1, 2, 3, 4$.

Following the general approach outlined in [40], Chapter 2, we shall prove Theorem 4.1 by deriving sharp KL divergence estimates. Recall that for two covariance matrices $\Sigma_1, \Sigma_2 \in \mathbb{R}^{n\times n}$, the KL divergence of the corresponding centered Gaussian distributions is given by

$$\mathrm{KL}(\Sigma_1 \| \Sigma_2) = \mathrm{KL}(\mathcal{N}(0, \Sigma_1) \| \mathcal{N}(0, \Sigma_0)) = \frac{1}{2}\left\{\mathrm{tr}(\Sigma_0^{-1}\Sigma_1) - n + \log \frac{\det \Sigma_0}{\det \Sigma_1}\right\}.$$

In the next proposition, which is the main technical estimate in the proof of Theorem 4.1, we establish an upper bound on the KL divergence

$$\mathrm{KL}(\theta_n \| \theta_0) = \mathrm{KL}(\widetilde{\Sigma}_n(\theta_n) \| \widetilde{\Sigma}_n(\theta_0)).$$

We give the proof in Section 7.

PROPOSITION 6.1. *Suppose that $H = H(\theta_0) \in (0, \frac{1}{2})$. For any $\delta > 0$ sufficiently small, there exists a $C = C(\theta_0, \delta)$ such that*

$$\mathrm{KL}(\theta_n \| \theta_0) \le C \begin{cases} r_{2,n}^2 n + (r_{1,n}^2 + r_{3,n}^2) n^{3-4\overline{H}} + r_{4,n}^2 n^{\delta} & \text{if } H \in \left[\frac{3}{4}, 1\right), \\ r_{2,n}^2 n + (r_{1,n}^2 + r_{3,n}^2) n^{3-4\overline{H}} + r_{4,n}^2 n^{3-4H} & \text{if } H \in \left(\frac{1}{2}, \frac{3}{4}\right), \\ r_{2,n}^2 n^{4H-1} + (r_{1,n}^2 + r_{4,n}^2) n + r_{3,n}^2 n^{1-4(\overline{H}-H)} & \text{if } H \in \left(\frac{1}{4}, \frac{1}{2}\right), \\ r_{2,n}^2 n^{\delta} + (r_{1,n}^2 + r_{4,n}^2) n + r_{3,n}^2 n^{1-4(\overline{H}-H)} & \text{if } H \in \left(0, \frac{1}{4}\right]. \end{cases}$$

PROOF OF THEOREM 4.1. For any $\theta_0 = (H, \sigma^2, \Lambda, \Pi) \in \Theta$, we define $\theta_n$ as in (6.4) and $r_{i,n} = r_0 n^{\alpha_i(H)}$ as shown in Table 6. In view of Proposition 6.1, these rates are chosen such

TABLE 6
*Choice of $r_{1,n}, \dots, r_{4,n}$ in the proof of Theorem 4.1*

| | $H \in (0, \frac{1}{4})$ | $H \in (\frac{1}{4}, \frac{1}{2})$ | $H \in (\frac{1}{2}, \frac{3}{4})$ | $H \in (\frac{3}{4}, 1)$ |
|---|---|---|---|---|
| $r_{1,n}$ | $n^{-\frac{1}{2}}$ | $n^{-\frac{1}{2}}$ | $n^{2\overline{H}-\frac{3}{2}}$ | $n^{2\overline{H}-\frac{3}{2}}$ |
| $r_{2,n}$ | $0$ | $n^{\frac{1}{2}-2H}$ | $n^{-\frac{1}{2}}$ | $n^{-\frac{1}{2}}$ |
| $r_{3,n}$ | $n^{2(\overline{H}-H)-\frac{1}{2}}$ | $n^{2(\overline{H}-H)-\frac{1}{2}}$ | $n^{2\overline{H}-\frac{3}{2}}$ | $n^{2\overline{H}-\frac{3}{2}}$ |
| $r_{4,n}$ | $n^{-\frac{1}{2}}$ | $n^{-\frac{1}{2}}$ | $n^{2H-\frac{3}{2}}$ | $0$ |

that $\mathrm{KL}(\theta_n\|\theta_0) \le r_0^2 C(\theta_0)$. Upon setting $r_0$ small enough, we find that

$$\mathrm{KL}(\theta_n\|\theta_0) \le \frac{1}{9}. \tag{6.6}$$

Moreover, from (6.4), it is simple to derive a lower bound on the errors $\theta_n - \theta_0$, component by component. In particular, since $r_{1,n} = o(1/\log n)$, we have

$$\big|\Lambda(\theta_n) - \Lambda(\theta_0)\big| = \Omega(r_{1,n}\log n + r_{3,n}), \qquad \big|\Pi(\theta_n) - \Pi(\theta_0)\big| = \Omega(r_{1,n}\log n + r_{4,n}),$$

where $\Omega$ denotes an asymptotic lower bound, that is, $a_n = \Omega(b_n)$ if $b_n = O(a_n)$. The resulting bounds on $\theta_n - \theta_0$ are exactly the rates $R_n(\theta)$ of Theorem 4.1 (listed in Table 1).

In order to translate these KL estimates into statistical lower bounds, we follow [40], Chapter 2. Intuitively speaking, if $\theta_n$ satisfies (6.6), we cannot consistently decide whether $\theta_n$ or $\theta_0$ is the true parameter. Hence, no estimator can converge towards $\theta_0$ faster than $\theta_n$. To make this mathematically precise, let $\widehat{\theta}_n$ be any measurable function of $\{Y_{i/n} : i = 1, \dots, n\}$. Note that $\theta_n \in \mathcal{D}_n(\theta_0)$ for $n$ large enough. Then, for any $c > 0$,

$$\begin{aligned}
&\sup_{\theta\in\mathcal{D}_n(\theta_0)} \mathbb{P}_\theta\big(\big|(\widehat{\theta}_n - \theta)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) \\
&\quad \ge \frac{1}{2}\mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_0)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) + \frac{1}{2}\mathbb{P}_{\theta_n}\big(\big|(\widehat{\theta}_n - \theta_n)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) \\
&\quad \ge \frac{1}{2}\mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_0)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) + \frac{1}{2}\mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_n)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) - \frac{1}{3}.
\end{aligned}$$

In the last step, we used that the fact that the total variation distance between $\mathbb{P}_{\theta_n}$ and $\mathbb{P}_{\theta_0}$ is upper bounded by $\sqrt{\mathrm{KL}(\theta_n\|\theta_0)/2} \le \sqrt{\mathrm{KL}(\theta_n\|\theta_0)} \le \frac{1}{3}$. Now use the union bound to obtain

$$\begin{aligned}
&\mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_0)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) + \mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_n)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) \\
&\quad \ge \mathbb{P}_{\theta_0}\big(\big|(\widehat{\theta}_n - \theta_0)_k\big| \vee \big|(\widehat{\theta}_n - \theta_n)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) \\
&\quad \ge \mathbb{P}_{\theta_0}\big(\big|(\theta_n - \theta_0)_k\big| \ge 2c\big|(\theta_n - \theta_0)_k\big|\big) = 1,
\end{aligned}$$

where the last equality holds for $c \le \frac{1}{2}$. Hence,

$$\sup_{\theta\in\mathcal{D}_n(\theta_0)} \mathbb{P}_\theta\big(\big|(\widehat{\theta}_n - \theta)_k\big| \ge c\big|(\theta_n - \theta_0)_k\big|\big) \ge \frac{1}{6}.$$

Because $|(\theta_n - \theta_0)_k| = \Omega(R_n(\theta_0)_k)$, this establishes the claimed minimax rates. □

**7. Proof of Proposition 6.1.** Let $A(h) = h\widetilde{\Sigma}_n(\theta_n) + (1-h)\widetilde{\Sigma}_n(\theta_0)$ and $\delta_n = \widetilde{\Sigma}_n(\theta_n) - \widetilde{\Sigma}_n(\theta_0)$. By Taylor expansion,

$$\mathrm{KL}\big(A(1)\|A(0)\big) = \frac{d}{dh}\mathrm{KL}\big(A(h)\|A(0)\big)|_{h=0} + \frac{1}{2}\int_0^1\int_0^s \frac{d^2}{dh^2}\mathrm{KL}\big(A(h)\|A(0)\big)|_{h=v}\,\mathrm{d}v\,\mathrm{d}s,$$

where

$$\begin{aligned}
&\frac{d}{dh}\mathrm{KL}\big(A(h)\|A(0)\big) = \frac{1}{2}\operatorname{tr}\big(A(0)^{-1}\delta_n\big) - \frac{1}{2}\operatorname{tr}\big(A(h)^{-1}\delta_n\big) = \frac{1}{2}\operatorname{tr}\big(\big[A(0)^{-1} - A(h)^{-1}\big]\delta_n\big), \\
&\frac{d^2}{dh^2}\mathrm{KL}\big(A(h)\|A(0)\big) = \frac{1}{2}\operatorname{tr}\big(A(h)^{-1}\delta_n A(h)^{-1}\delta_n\big).
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{KL}\big(A(1)\|A(0)\big) &= \frac{1}{4}\int_0^1\int_0^s \operatorname{tr}\big(A(h)^{-1}\delta_n A(h)^{-1}\delta_n\big)\,\mathrm{d}h\,\mathrm{d}s \\
&\le \frac{1}{4}\sup_{h\in[0,1]} \operatorname{tr}\big(A(h)^{-1}\delta_n A(h)^{-1}\delta_n\big).
\end{aligned} \tag{7.1}$$

In order to find an upper bound for (7.1), we will use the following technical lemma.

LEMMA 7.1. *Let $B$ be a symmetric matrix and $A$ and $A_0$ be symmetric positive semidefinite matrices such that $A - A_0$ is positive semidefinite, too. Then*

$$\operatorname{tr}(A_0BA_0B) \le \operatorname{tr}(ABA_0B) = \operatorname{tr}(A_0BAB) \le \operatorname{tr}(ABAB).$$

PROOF. Denote $C = BA_0B$, which is symmetric positive semidefinite. Von Neumann's trace inequality yields $\operatorname{tr}((A - A_0)C) \ge \sum_{i=1}^n a_i b_{n-i+1}$, where $a_i$ and $b_i$ are the descending eigenvalues of $A - A_0$ and $C$, respectively. Since $a_i, b_i \ge 0$, we conclude that $\operatorname{tr}(A_0C) \le \operatorname{tr}(AC)$, which proves the first inequality and, consequently, the second. □

To bound (7.1), we therefore need a lower bound on $A(h)$ and an upper bound on $\delta_n$.

LEMMA 7.2. *For any $\theta_0 \in \Theta$ and $\delta > 0$, there exists $c = c(\theta_0, \delta) > 0$ such that*

$$A(h) \ge c\big[n^{-1}I_n + n^{-2H}\Sigma_n(H - \delta)\big], \quad h \in [0, 1]. \tag{7.2}$$

*Here, $\ge$ denotes the Loewner partial order on the cone of positive semidefinite matrices.*

PROOF OF LEMMA 7.2. Since $A(h)$ is a convex combination, it suffices to establish the lower bounds for $A(1)$ and $A(0)$. Note that by definition

$$\begin{aligned}&\sqrt{\Pi\sigma^2}b(H)n^{-2\overline{H}}\Sigma_n(\overline{H})_{jk}\\&\quad = \operatorname{Cov}\bigg(|\sigma|\frac{\Delta_j^n B}{\Delta_n^{1/2}}, \sqrt{\Pi}\frac{\Delta_k^n B(H)}{\Delta_n^H}\bigg) + \operatorname{Cov}\bigg(|\sigma|\frac{\Delta_k^n B}{\Delta_n^{1/2}}, \sqrt{\Pi}\frac{\Delta_j^n B(H)}{\Delta_n^H}\bigg),\end{aligned}$$

where $B(H)_t = \int_{-\infty}^t h_H(t, s)\,\mathrm{d}B_s$. Since $XY^T + YX^T \le XX^T + YY^T$ for $X, Y \in \mathbb{R}^{n\times n}$, it follows that $\sqrt{\Pi\sigma^2}b(H)n^{-2\overline{H}}\Sigma_n(\overline{H}) \le \sigma^2 n^{-1}I_n + \Pi n^{-2H}\Sigma_n(H)$ and therefore, by (6.3),

$$\widetilde{\Sigma}_n(\theta) \ge \bigg(1 - \frac{|\Lambda|}{\sqrt{\Pi\sigma^2}}\bigg)\big(\sigma^2 n^{-1}I_n + \Pi n^{-2H}\Sigma_n(H)\big).$$

If we apply this to $\theta = \theta_0$, we immediately obtain (7.2) for $A(0) = \widetilde{\Sigma}_n(\theta_0)$. To derive the estimate for $A(1) = \widetilde{\Sigma}_n(\theta_n)$, we apply the above to $\theta = \theta_n$, which yields

$$A(1) \ge \frac{1}{2}\bigg(1 - \frac{|\Lambda|}{\sqrt{\Pi\sigma^2}}\bigg)\big(\sigma^2 n^{-1}I_n + \Pi n^{-2H}\Sigma_n(H + r_{1,n})\big)$$

for all sufficiently large $n$. Since $\Lambda^2 < \Pi\sigma^2$, this implies (7.2) for $h = 1$. □

We proceed to finding an upper bound on $\delta_n$.

LEMMA 7.3. *For any $\theta_0 \in \Theta$ and $\delta > 0$, there exists $C = C(\theta_0, \delta) > 0$ such that $\delta_n \le CB_n(\theta_0, \delta)$, where*

$$B_n(\theta_0, \delta) = r_{2,n}n^{-1}I_n + (r_{1,n} + r_{4,n})n^{-2H}\Sigma_n(H + \delta) + (r_{1,n} + r_{3,n})n^{-2\overline{H}}\Sigma_n(\overline{H} + \delta).$$

PROOF OF LEMMA 7.3. By [39], Proposition 7.2.9, we have that

$$\Sigma_n(H) = T_n(f_H),$$

where

$$T_n(f)_{j,k} = \int_{-\pi}^{\pi} f(\lambda)e^{-i\lambda|j-k|}\,\mathrm{d}\lambda \tag{7.3}$$

and the spectral density $f_H$ is given by

$$f_H(\lambda) = \frac{\Gamma(2H+1)\sin(\pi H)}{\pi}(1-\cos\lambda)\sum_{k\in\mathbb{Z}}|\lambda+2k\pi|^{-2H-1}.$$

In particular, by (6.5),

$$\begin{aligned}\delta_n &= \widetilde{\Sigma}_n(\theta_n) - \widetilde{\Sigma}_n(\theta) \\ &= r_{2,n}n^{-1}I_n + \Pi n^{-2H}\big[(1+r_{4,n})\Sigma_n(H+r_{1,n}) - \Sigma_n(H)\big] \\ &\quad + \Lambda b(H)n^{-2\overline{H}}\bigg[(1+r_{3,n})\Sigma_n\bigg(\overline{H}+\frac{r_{1,n}}{2}\bigg) - \Sigma_n(\overline{H})\bigg] = T_n(\widetilde{g}_n),\end{aligned}$$

where

$$\begin{aligned}\widetilde{g}_n &= r_{2,n}\frac{n^{-1}}{2\pi} + \Pi n^{-2H}\big[(1+r_{4,n})f_{H+r_{1,n}} - f_H\big] + \Lambda b(H)n^{-2\overline{H}}\big[(1+r_{3,n})f_{\overline{H}+\frac{r_{1,n}}{2}} - f_{\overline{H}}\big] \\ &\le C(\theta_0)\big[r_{2,n}n^{-1} + n^{-2H}r_{4,n}|f_{H+r_{1,n}}| + n^{-2H}|f_{H+r_{1,n}} - f_H| \\ &\quad + n^{-2\overline{H}}r_{3,n}|f_{\overline{H}+\frac{r_{1,n}}{2}}| + n^{-2\overline{H}}|f_{\overline{H}+\frac{r_{1,n}}{2}} - f_{\overline{H}}|\big]\end{aligned}$$

and $C(\theta_0)$ may change its value from line to line. For large $n$, we have that

$$\begin{aligned}\big|f_{H+r_{1,n}}(\lambda)\big| &\le C(\theta_0)|\lambda|^{1-2H-2r_{1,n}} \le C(\theta_0)|\lambda|^{1-2H-2\delta}, \\ \big|f_{\overline{H}+\frac{r_{1,n}}{2}}(\lambda)\big| &\le C(\theta_0)|\lambda|^{1-2\overline{H}-r_{1,n}} \le C(\theta_0)|\lambda|^{1-2\overline{H}-2\delta}, \\ \big|f_{H+r_{1,n}}(\lambda) - f_H(\lambda)\big| &\le C(\theta_0)r_{1,n}|\lambda|^{1-2H-2r_{1,n}}\big|\log|\lambda|\big| \le C(\theta_0)r_{1,n}|\lambda|^{1-2H-2\delta}, \\ \big|f_{\overline{H}+\frac{r_{1,n}}{2}}(\lambda) - f_{\overline{H}}(\lambda)\big| &\le C(\theta_0)r_{1,n}|\lambda|^{1-2\overline{H}-r_{1,n}}\big|\log|\lambda|\big| \le C(\theta_0)r_{1,n}|\lambda|^{1-2\overline{H}-2\delta}\end{aligned}$$

for all $\lambda \in [-\pi,\pi]$. Hence, $\widetilde{g}_n(\lambda) \le C(\theta_0)\widehat{g}_n(\lambda)$ for all $\lambda \in [-\pi,\pi]$, where

$$\widetilde{g}_n \le C(\theta_0)\bigg[r_{2,n}\frac{n^{-1}}{2\pi} + (r_{1,n}+r_{4,n})n^{-2H}f_{H+\delta} + (r_{1,n}+r_{3,n})n^{-2\overline{H}}f_{\overline{H}+\delta}\bigg],$$

which yields the claim. $\square$

LEMMA 7.4. *Let $g\colon [-\pi,\pi] \to \mathbb{R}$ be a symmetric function such that $g(\lambda) = O(|\lambda|^{-\beta})$ as $\lambda \to 0$ for some $\beta \in [0,1)$. Then, for all $\delta > 0$ and $H \in (0,1)$,*

$$\begin{aligned}\operatorname{tr}\big(T_n(g)T_n(g)\big) &= O\big(n \vee n^{2\beta+\delta}\big), \\ \operatorname{tr}\big(\Sigma_n(H)^{-1}T_n(g)\Sigma_n(H)^{-1}T_n(g)\big) &= O\big(n \vee n^{2(\beta-2H+1)+\delta}\big).\end{aligned}$$

PROOF OF LEMMA 7.4. We apply [31], Theorem 5 in the full version, to the spectral densities $f = f_H$ and $g$ with $\alpha = \alpha(\theta) = 2H-1$ and $\beta$ as above. $\square$

PROOF OF PROPOSITION 6.1. By (7.1) and Lemmas 7.2 and 7.3, we obtain for $\delta > 0$,

$$\mathrm{KL}(\theta_n\|\theta_0) \le \begin{cases} C(\theta_0)n^2\operatorname{tr}(B_nB_n) & \text{if } H > \frac{1}{2}, \\ C(\theta_0)n^{4H}\operatorname{tr}\big(\Sigma_n(H-\delta)^{-1}B_n\Sigma_n(H-\delta)^{-1}B_n\big) & \text{if } H < \frac{1}{2},\end{cases}$$

where $B_n = B_n(\theta_0,\delta)$ as in Lemma 7.3.

If $H > \frac{1}{2}$, the Cauchy–Schwarz inequality yields

$$\operatorname{tr}(B_n B_n) \leq 3r_{2,n}^2 n^{-2} \operatorname{tr}(I_n^2) + 3(r_{1,n} + r_{4,n})^2 n^{-4H} \operatorname{tr}(\Sigma_n(H+\delta)^2) + 3(r_{1,n} + r_{3,n})^2 n^{-4\overline{H}} \operatorname{tr}(\Sigma_n(\overline{H}+\delta)^2).$$

Clearly, $\operatorname{tr}(I_n^2) = n$. Moreover, since $\Sigma_n(\overline{H}+\delta)$ and $\Sigma_n(H+\delta)$ satisfy the conditions of Lemma 7.4 with $\beta = 2\overline{H} + 2\delta - 1$ and $\beta = 2H + 2\delta - 1$, respectively, we have that

$$\operatorname{tr}(\Sigma_n(\overline{H}+\delta)^2) = O(n \vee n^{4\overline{H}+5\delta-2}), \quad \operatorname{tr}(\Sigma_n(H+\delta)^2) = O(n \vee n^{4H+5\delta-2}).$$

Since $\delta > 0$ was arbitrary, we find that $\operatorname{tr}(B_n B_n) = O(z_n)$, where

$$z_n = r_{2,n}^2 n^{-1} + (r_{1,n} + r_{4,n})^2 (n^{1-4H} \vee n^{\delta-2}) + (r_{1,n} + r_{3,n})^2 (n^{1-4\overline{H}} \vee n^{\delta-2}).$$

This establishes the KL upper bound for the cases $H \in (\frac{1}{2}, \frac{3}{4})$ and $H \in [\frac{3}{4}, 1)$.

If $H < \frac{1}{2}$, we may again apply Lemma 7.4 to find that

$$\begin{aligned}
\operatorname{tr}[(\Sigma_n(H-\delta)^{-1} I_n)^2] &= O(n \vee n^{2-4H+5\delta}), \\
\operatorname{tr}[(\Sigma_n(H-\delta)^{-1} \Sigma_n(\overline{H}+\delta))^2] &= O(n \vee n^{4(\overline{H}-H)+5\delta}), \\
\operatorname{tr}[(\Sigma_n(H-\delta)^{-1} \Sigma_n(H+\delta))^2] &= O(n \vee n^{5\delta}).
\end{aligned}$$

Since $\delta > 0$ was arbitrary, we find that $\operatorname{tr}[\Sigma_n(H-\delta)^{-1} B_n \Sigma_n(H-\delta)^{-1} B_n] = O(w_n)$, where

$$w_n = r_{2,n}^2 (n^{-1} \vee n^{-4H+\delta}) + (r_{1,n} + r_{3,n})^2 (n^{1-4\overline{H}} \vee n^{-4H+\delta}) + (r_{1,n} + r_{4,n})^2 (n^{1-4H} \vee n^{-4H+\delta}).$$

This yields the KL upper bound for the remaining cases $H \in (\frac{1}{4}, \frac{1}{2})$ and $H \in (0, \frac{1}{4}]$. $\square$

## APPENDIX A: CENTRAL LIMIT THEOREM FOR GENERAL VARIATION FUNCTIONALS

In this section, we state and prove a CLT for general variation functionals of multivariate mixed semimartingale processes. To this end, we consider a $d$-dimensional mixed semimartingale of the form

$$Y_t = Y_0 + \int_0^t a_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}B_s + \int_0^t g(t-s)\rho_s \, \mathrm{d}B_s, \tag{A.1}$$

where now $B$ is a $d'$-dimensional Brownian motion, $a$ (resp., $\sigma$ and $\rho$) is $\mathbb{R}^d$-valued (resp., $\mathbb{R}^{d \times d'}$-valued) and predictable and $g: \mathbb{R} \to \mathbb{R}$ is given by (2.2). Since $d'$ may be larger than $d$, (A.1) includes the case where the martingale and the fractional part of $Y$ are driven by correlated (and not necessarily identical) Brownian motions. In fact, the situation considered in Section 2 can be embedded into the current setting by defining $d' = 2$ and

$$B^{(A.1)} = (B, B'), \qquad \sigma^{(A.1)} = (\sigma, 0), \qquad \rho^{(A.1)} = (\rho, \rho') \tag{A.2}$$

(the superscript stands for "from equation (A.1)"). In particular, Theorem 2.1 then becomes a special case of Theorems A.1 and A.2 below.

Let $f: \mathbb{R}^{d \times L} \to \mathbb{R}^M$ for some $L, M \in \mathbb{N}$. For $Y$ and similarly for other $d$-dimensional processes, we define

$$\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n} \in \mathbb{R}^d, \qquad \underline{\Delta}_i^n Y = (\Delta_i^n Y, \Delta_{i+1}^n Y, \ldots, \Delta_{i+L-1}^n Y) \in \mathbb{R}^{d \times L}. \tag{A.3}$$

Our goal is to formulate and prove a CLT for normalized variation functionals of the form

$$V_f^n(Y,t)=\Delta_n \sum_{i=1}^{[t/\Delta_n]-L+1} f\bigg(\frac{\underline{\Delta}_i^n Y}{\Delta_n^{H\wedge(1/2)}}\bigg),\quad t\geq 0, \tag{A.4}$$

where $[\cdot]$ denotes the floor function. In what follows, $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^n$ for any $n\in\mathbb{N}$. Also, if $z$ is some matrix in $\mathbb{R}^{n\times m}$ for any $n,m\in\mathbb{N}$, then $\|z\|$ is defined by viewing $z$ as a vector in $\mathbb{R}^{nm}$. We introduce the following assumptions.

ASSUMPTION (CLT'). Consider the process $Y$ from (A.1) and recall $N(H)$ from (2.5). We make the following assumptions:

1. The function $f\colon \mathbb{R}^{d\times L}\to\mathbb{R}^M$ is even (i.e., satisfies $f(-x)=f(x)$ for all $x$) and belongs to $C^{2N(H)+1}(\mathbb{R}^{d\times L},\mathbb{R}^M)$, with all partial derivatives up to order $2N(H)+1$ (including $f$ itself) being of polynomial growth.
2. The kernel $g$ is of the form (2.2) where $H\in(0,1)\setminus\{\frac{1}{2}\}$ and $g_0\in C^1(\mathbb{R})$ satisfies $g_0(x)=0$ for all $x\leq 0$.
3. The drift process $a$ is $d$-dimensional, locally bounded, $\mathbb{F}$-adapted and càdlàg. Moreover, $B$ is a $d'$-dimensional standard $\mathbb{F}$-Brownian motion.
4. If $H>\frac{1}{2}$, the volatility process $\sigma$ takes the form

$$\sigma_t=\sigma_t^{(0)}+\int_0^t \widetilde{\sigma}_s\,\mathrm{d}B_s,\quad t\geq 0, \tag{A.5}$$

where

(a) $\sigma^{(0)}$ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d'}$-valued process such that for all $T>0$, there are $\gamma\in(\frac{1}{2},1]$ and $K_1\in(0,\infty)$ with

$$\mathbb{E}\big[1\wedge\|\sigma_t^{(0)}-\sigma_s^{(0)}\|\big]\leq K_1|t-s|^{\gamma},\quad s,t\in[0,T]; \tag{A.6}$$

(b) $\widetilde{\sigma}$ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d'\times d'}$-valued process such that for all $T>0$, there are $\varepsilon\in(0,1)$ and $K_2\in(0,\infty)$ with

$$\mathbb{E}\big[1\wedge\|\widetilde{\sigma}_t-\widetilde{\sigma}_s\|\big]\leq K_2|t-s|^{\varepsilon},\quad s,t\in[0,T]. \tag{A.7}$$

If $H<\frac{1}{2}$, we have (A.5) but with $\sigma$, $\sigma^{(0)}$ and $\widetilde{\sigma}$ replaced by $\rho$ and some processes $\rho^{(0)}$ and $\widetilde{\rho}$ satisfying conditions analogous to (A.6) and (A.7).

5. If $H>\frac{1}{2}$, the process $\rho$ is $\mathbb{F}$-adapted, locally bounded and $\mathbb{R}^{d\times d'}$-valued. Moreover, for all $T>0$, there is $K_3\in(0,\infty)$ such that

$$\mathbb{E}\big[1\wedge\|\rho_t-\rho_s\|\big]\leq K_3|t-s|^{\frac{1}{2}},\quad s,t\in[0,T]. \tag{A.8}$$

If $H<\frac{1}{2}$, we have the same condition but with $\rho$ replaced by $\sigma$.

To state the CLT, we need more additional notation. For suitable $v=(v_{k\ell,k'\ell'})_{k,k',\ell,\ell'=1}^{d,d,L,L}$ and $q=(q_{k\ell,k'\ell'})_{k,k',\ell,\ell'=1}^{d,d,L,L}\in\mathbb{R}^{(d\times L)\times(d\times L)}$, define

$$\begin{aligned}&\mu_f(v)=\big(\mathbb{E}\big[f_m(Z)\big]\big)_{m=1}^M\in\mathbb{R}^M,\\ &\gamma_f(v,q)=\big(\mathrm{Cov}\big(f_m(Z),f_{m'}(Z')\big)\big)_{m,m'=1}^M\in\mathbb{R}^{M\times M},\quad \gamma_f(v)=\gamma_f(v,v),\end{aligned} \tag{A.9}$$

where $(Z,Z')\in(\mathbb{R}^{d\times L})^2$ is multivariate normal with mean 0 and $\mathrm{Cov}(Z_{k\ell},Z_{k'\ell'})=\mathrm{Cov}(Z'_{k\ell},Z'_{k'\ell'})=v_{k\ell,k'\ell'}$ and $\mathrm{Cov}(Z_{k\ell},Z'_{k'\ell'})=q_{k\ell,k'\ell'}$. Furthermore, given a multi-index

$\chi = (\chi_{k\ell,k'\ell'})_{k,k',\ell,\ell'=1}^{d,d,L,L} \in \mathbb{N}_0^{(d\times L)\times(d\times L)}$, we define

$$
\begin{aligned}
|\chi| &= \sum_{k,k'=1}^{d}\sum_{\ell,\ell'=1}^{L} \chi_{k\ell,k\ell'}, \qquad \chi! = \prod_{k,k'=1}^{d}\prod_{\ell,\ell'=1}^{L} \chi_{k\ell,k'\ell'}!,\\
v^{\chi} &= \prod_{k,k'=1}^{d}\prod_{\ell,\ell'=1}^{L} v_{k\ell,k'\ell'}^{\chi_{k\ell,k'\ell'}}
\end{aligned}
\tag{A.10}
$$

and the partial derivatives

$$
\partial^{\chi}\mu_f(v) = \frac{\partial^{|\chi|}\mu_f}{\partial v_{11,11}^{\chi_{11,11}}\cdots\partial v_{dL,dL}^{\chi_{dL,dL}}}(v) \in \mathbb{R}^M. \tag{A.11}
$$

For $s \geq 0$, we also define

$$
\begin{aligned}
c(s)_{k\ell,k'\ell'} &= \big(\sigma_s\sigma_s^T\big)_{kk'}\mathbf{1}_{\{\ell=\ell'\}},\\
\pi_r(s)_{k\ell,k'\ell'} &= \big(\rho_s\rho_s^T\big)_{kk'}\Gamma^H_{|\ell-\ell'+r|}, \qquad \pi(s)_{k\ell,k'\ell'} = \pi_0(s)_{k\ell,k'\ell'},\\
\lambda(s)_{k\ell,k'\ell'} &= 2^{-\mathbb{1}_{\{\ell=\ell'\}}}\big(\sigma_s\rho_s^T\mathbb{1}_{\{\ell\leq\ell'\}} + \rho_s\sigma_s^T\mathbb{1}_{\{\ell\geq\ell'\}}\big)_{kk'}\Phi^H_{|\ell-\ell'|}
\end{aligned}
\tag{A.12}
$$

for all $k, k' \in \{1,\dots,d\}$ and $\ell,\ell' \in \{1,\dots,L\}$, where $\Gamma^H_r$ and $\Lambda^H_r$ are defined in (2.10) and (2.11), respectively.

THEOREM A.1. *If Assumption (CLT') holds with $H > \frac{1}{2}$, then*

$$
\Delta_n^{-\frac{1}{2}}\big\{V_f^n(Y,t) - V_f(Y,t) - \mathcal{A}_t'^n\big\} \stackrel{\text{st}}{\Longrightarrow} \mathcal{Z}', \tag{A.13}
$$

*where $V_f(Y,t) = \int_0^t \mu_f(c(s))\,\mathrm{d}s$ and*

$$
\mathcal{A}_t'^n = \sum_{j=1}^{N(H)}\sum_{|\chi|=j}\frac{1}{\chi!}\int_0^t \partial^{\chi}\mu_f\big(c(s)\big)\big(\Delta_n^{H-\frac{1}{2}}\lambda(s) + \Delta_n^{2H-1}\pi(s)\big)^{\chi}\,\mathrm{d}s \tag{A.14}
$$

*and $\mathcal{Z}' = (\mathcal{Z}_t')_{t\geq 0}$ is an $\mathbb{R}^M$-valued process, defined on a very good filtered extension $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ of the original probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ that conditionally on $\mathcal{F}$ is a centered Gaussian process with independent increments and covariance function*

$$
\mathcal{C}_t' = \big(\overline{\mathbb{E}}[\mathcal{Z}_t'^m\mathcal{Z}_t'^{m'} \mid \mathcal{F}]\big)_{m,m'=1}^M = \int_0^t \gamma_f\big(c(s)\big)\,\mathrm{d}s. \tag{A.15}
$$

THEOREM A.2. *If Assumption (CLT') holds with $H < \frac{1}{2}$, then*

$$
\Delta_n^{-\frac{1}{2}}\left\{V_f^n(Y,t) - \int_0^t \mu_f\big(\pi(s)\big)\,\mathrm{d}s - \mathcal{A}_t^n\right\} \stackrel{\text{st}}{\Longrightarrow} \mathcal{Z}, \tag{A.16}
$$

*where*

$$
\mathcal{A}_t^n = \sum_{j=1}^{N(H)}\sum_{|\chi|=j}\frac{1}{\chi!}\int_0^t \partial^{\chi}\mu_f\big(\pi(s)\big)\big(\Delta_n^{\frac{1}{2}-H}\lambda(s) + \Delta_n^{1-2H}c(s)\big)^{\chi}\,\mathrm{d}s \tag{A.17}
$$

*and $\mathcal{Z} = (\mathcal{Z}_t)_{t\geq 0}$ is an $\mathbb{R}^M$-valued process defined on $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ that conditionally on $\mathcal{F}$ is a centered Gaussian process with independent increments and covariance function*

$$
\begin{aligned}
\mathcal{C}_t &= \big(\overline{\mathbb{E}}[\mathcal{Z}_t^m\mathcal{Z}_t^{m'} \mid \mathcal{F}]\big)_{m,m'=1}^M\\
&= \int_0^t \left\{\gamma_f\big(c(s)\big) + \sum_{r=1}^{\infty}\big(\gamma_f\big(\pi(s),\pi_r(s)\big) + \gamma_f\big(\pi(s),\pi_r(s)\big)^T\big)\right\}\mathrm{d}s.
\end{aligned}
\tag{A.18}
$$

The proof of the two results is given in Appendices B–D of the Supplementary Material [14]. Apart from the term $\mathcal{A}_t'^n$, Theorem A.1 is exactly the CLT of semimartingale variation functionals (see [25]). Similarly, if we ignore $\mathcal{A}_t^n$, Theorem A.2 is the CLT for variation functionals of a fractional process with roughness parameter $H$ (cf. [5, 17]). The two processes $\mathcal{A}_t'^n$ and $\mathcal{A}_t^n$ respectively play the role of higher-order bias terms (for the estimation of the integrated volatility or noise volatility functional). However, these processes are also key to identifying all other quantities of interest in $Y$. If $\lambda \equiv 0$ and $H < \frac{1}{2}$, Theorem A.2 reduces to [13], Theorem 3.1. If $\lambda \not\equiv 0$, additional terms involving $\lambda$ appear in both $\mathcal{A}_t'^n$ and $\mathcal{A}_t^n$.

**Acknowledgments.** We would like to thank the Associate Editor and four referees for their careful reading of the paper. Their constructive comments have led to significant improvements of the paper.

**Funding.** The first author is supported by ECS project 26301724.
The second author is partially supported by the DFG, project number KL 1041/7-2.

## SUPPLEMENTARY MATERIAL

**Supplement to "Rate-optimal estimation of mixed semimartingales"** (DOI: 10.1214/24-AOS2461SUPP; .pdf). The supplement [14] contains Appendices B–E.

## REFERENCES

[1] AÏT-SAHALIA, Y. and JACOD, J. (2009). Testing for jumps in a discretely observed process. *Ann. Statist.* **37** 184–222. MR2488349 https://doi.org/10.1214/07-AOS568

[2] AÏT-SAHALIA, Y. and JACOD, J. (2014). *High-Frequency Financial Econometrics*. Princeton Univ. Press, Princeton, NJ.

[3] AÏT-SAHALIA, Y. and YU, J. (2009). High frequency market microstructure noise estimates and liquidity measures. *Ann. Appl. Stat.* **3** 422–457. MR2668714 https://doi.org/10.1214/08-AOAS200

[4] AMBLARD, P.-O., COEURJOLLY, J.-F., LAVANCIER, F. and PHILIPPE, A. (2013). Basic properties of the multivariate fractional Brownian motion. In *Self-Similar Processes and Their Applications*. *Sémin. Congr.* **28** 63–84. Soc. Math. France, Paris. MR3203519

[5] BARNDORFF-NIELSEN, O. E., CORCUERA, J. M. and PODOLSKIJ, M. (2011). Multipower variation for Brownian semistationary processes. *Bernoulli* **17** 1159–1194. MR2854768 https://doi.org/10.3150/10-BEJ316

[6] BARNDORFF-NIELSEN, O. E. and SHEPHARD, N. (2005). Power variation and time change. *Teor. Veroyatn. Primen.* **50** 115–130. MR2222740 https://doi.org/10.1137/S0040585X97981482

[7] BROUSTE, A. and FUKASAWA, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high-frequency observations. *Ann. Statist.* **46** 2045–2061. MR3845010 https://doi.org/10.1214/17-AOS1611

[8] CAI, C., CHIGANSKY, P. and KLEPTSYNA, M. (2016). Mixed Gaussian processes: A filtering approach. *Ann. Probab.* **44** 3032–3075. MR3531685 https://doi.org/10.1214/15-AOP1041

[9] CHEN, D., CHENG, Y., CHONG, C. H., GENTINE, P., JIA, W., MONIER, B. and SHEN, S. (2022). Pre-averaging fractional processes contaminated by noise, with an application to turbulence. Available at arXiv:2212.00867.

[10] CHEN, D., MYKLAND, P. A. and ZHANG, L. (2020). The five trolls under the bridge: Principal component analysis with asynchronous and noisy high frequency data. *J. Amer. Statist. Assoc.* **115** 1960–1977. MR4189770 https://doi.org/10.1080/01621459.2019.1672555

[11] CHERIDITO, P. (2001). Mixed fractional Brownian motion. *Bernoulli* **7** 913–934. MR1873835 https://doi.org/10.2307/3318626

[12] CHIGANSKY, P. and KLEPTSYNA, M. (2023). Estimation of the Hurst parameter from continuous noisy data. *Electron. J. Stat.* **17** 2343–2385. MR4649984 https://doi.org/10.1214/23-ejs2156

[13] CHONG, C. H., DELERUE, T. and LI, G. (2024). When frictions are fractional: Rough noise in high-frequency data. *J. Amer. Statist. Assoc.* Forthcoming. https://doi.org/10.1080/01621459.2024.2428466

[14] CHONG, C. H., DELERUE, T. and MIES, F. (2025). Supplement to "Rate-optimal estimation of mixed semimartingales." https://doi.org/10.1214/24-AOS2461SUPP

[15] CHONG, C. H., HOFFMANN, M., LIU, Y., ROSENBAUM, M. and SZYMANSKI, G. (2024). Statistical inference for rough volatility: Central limit theorems. *Ann*. *Appl*. *Probab*. **34** 2600–2649. MR4757488 https://doi.org/10.1214/23-aap2002

[16] COMTE, F. and RENAULT, E. (1998). Long memory in continuous-time stochastic volatility models. *Math*. *Finance* **8** 291–323. MR1645101 https://doi.org/10.1111/1467-9965.00057

[17] CORCUERA, J. M., HEDEVANG, E., PAKKANEN, M. S. and PODOLSKIJ, M. (2013). Asymptotic theory for Brownian semi-stationary processes with application to turbulence. *Stochastic Process*. *Appl*. **123** 2552–2574. MR3054536 https://doi.org/10.1016/j.spa.2013.03.011

[18] DELATTRE, S. and JACOD, J. (1997). A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli* **3** 1–28. MR1466543 https://doi.org/10.2307/3318650

[19] DIEBOLD, F. X. and STRASSER, G. (2013). On the correlation structure of microstructure noise: A financial economic approach. *Rev*. *Econ*. *Stud*. **80** 1304–1337. MR3129949 https://doi.org/10.1093/restud/rdt008

[20] DOZZI, M., MISHURA, Y. and SHEVCHENKO, G. (2015). Asymptotic behavior of mixed power variations and statistical estimation in mixed models. *Stat*. *Inference Stoch*. *Process*. **18** 151–175. MR3348583 https://doi.org/10.1007/s11203-014-9106-5

[21] GATHERAL, J., JAISSON, T. and ROSENBAUM, M. (2018). Volatility is rough. *Quant*. *Finance* **18** 933–949. MR3805308 https://doi.org/10.1080/14697688.2017.1393551

[22] HANSEN, P. R. and LUNDE, A. (2006). Realized variance and market microstructure noise. *J*. *Bus*. *Econom*. *Statist*. **24** 127–218. MR2234447 https://doi.org/10.1198/073500106000000071

[23] HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Trans*. *Amer*. *Soc*. *Civ*. *Eng*. **116** 770–799.

[24] JACOD, J., LI, Y. and ZHENG, X. (2017). Statistical properties of microstructure noise. *Econometrica* **85** 1133–1174. MR3681766 https://doi.org/10.3982/ECTA13085

[25] JACOD, J. and PROTTER, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Springer, Heidelberg. MR2859096 https://doi.org/10.1007/978-3-642-24127-7

[26] JACOD, J. and SØRENSEN, M. (2018). A review of asymptotic theory of estimating functions. *Stat*. *Inference Stoch*. *Process*. **21** 415–434. MR3824976 https://doi.org/10.1007/s11203-018-9178-8

[27] JACOD, J. and TODOROV, V. (2014). Efficient estimation of integrated volatility in presence of infinite variation jumps. *Ann*. *Statist*. **42** 1029–1069. MR3210995 https://doi.org/10.1214/14-AOS1213

[28] LELAND, W., TAQQU, M., WILLINGER, W. and WILSON, D. (1994). On the self-similar nature of Ethernet traffic (extended version). *IEEE/ACM Trans*. *Netw*. **2** 1–15.

[29] LI, J. and XIU, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica* **84** 1613–1633. MR3537165 https://doi.org/10.3982/ECTA12306

[30] LI, Y. and MYKLAND, P. A. (2015). Rounding errors and volatility estimation. *J*. *Financ*. *Econom*. **13** 478–504.

[31] LIEBERMAN, O., ROSEMARIN, R. and ROUSSEAU, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory* **28** 457–470. MR2913638 https://doi.org/10.1017/S0266466611000399

[32] MIES, F. and PODOLSKIJ, M. (2023). Estimation of mixed fractional stable processes using high-frequency data. *Ann*. *Statist*. **51** 1946–1964. MR4678791 https://doi.org/10.1214/23-aos2312

[33] MIKOSCH, T., RESNICK, S., ROOTZÉN, H. and STEGEMAN, A. (2002). Is network traffic approximated by stable Lévy motion or fractional Brownian motion? *Ann*. *Appl*. *Probab*. **12** 23–68. MR1890056 https://doi.org/10.1214/aoap/1015961155

[34] MISHURA, Y. S. (2008). *Stochastic Calculus for Fractional Brownian Motion and Related Processes*. *Lecture Notes in Math*. **1929**. Springer, Berlin. MR2378138 https://doi.org/10.1007/978-3-540-75873-0

[35] MOLZ, F. J., LIU, H. H. and SZULGA, J. (1997). Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions. *Water Resour*. *Res*. **33** 2273–2286.

[36] NEWEY, W. K. and WEST, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Rev*. *Econ*. *Stud*. **61** 631–653. MR1299308 https://doi.org/10.2307/2297912

[37] PANDEY, G., LOVEJOY, S. and SCHERTZER, D. (1998). Multifractal analysis of daily river flows including extremes for basins of five to two million square kilometres, one day to 75 years. *J*. *Hydrol*. **208** 62–81.

[38] ROSENBAUM, M. (2009). Integrated volatility and round-off error. *Bernoulli* **15** 687–720. MR2555195 https://doi.org/10.3150/08-BEJ170

[39] SAMORODNITSKY, G. and TAQQU, M. S. (1994). *Stable Non-Gaussian Random Processes*: *Stochastic Models with Infinite Variance*. *Stochastic Modeling*. CRC Press, New York. MR1280932

[40] 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

[41] VAN ZANTEN, H. (2007). When is a linear combination of independent fBm's equivalent to a single fBm? *Stochastic Process. Appl.* **117** 57–70. MR2287103 https://doi.org/10.1016/j.spa.2006.05.013

[42] WILLIAMS, S. D., BOCK, Y., FANG, P., JAMASON, P., NIKOLAIDIS, R. M., PRAWIRODIRDJO, L., MILLER, M. and JOHNSON, D. J. (2004). Error analysis of continuous GPS position time series. *J. Geophys. Res.* **109** B03412.

[43] XU, C. (2017). Detecting periodic oscillations in astronomy data: Revisiting wavelet analysis with coloured and white noise. *Mon. Not. R. Astron. Soc.* **466** 3827–3833.