Mixed semimartingales: Volatility estimation in the presence of rough noise

Carsten Chong*
Department of Statistics, Columbia University
and
Thomas Delerue
Department of Mathematics, Technical University of Munich
and
Guoying Li
Department of Statistics, Columbia University

Abstract

We consider the problem of estimating volatility based on high-frequency data when the observed price process is a continuous Itô semimartingale contaminated by microstructure noise. Assuming that the noise process is compatible across different sampling frequencies, we argue that it typically has a similar local behavior to fractional Brownian motion. For the resulting class of processes, which we call *mixed semimartingales*, we derive consistent estimators and asymptotic confidence intervals for the roughness parameter of the noise and the integrated price and noise volatilities, in all cases where these quantities are identifiable. Our model can explain key features of recent stock price data, most notably divergence rates in volatility signature plots that vary considerably over time and between assets.

Keywords: Central limit theorem, fractional noise, high-frequency data, Hurst parameter, market microstructure noise, mixed fractional Brownian

*We thank Jean Jacod and Serena Ng for valuable comments and suggestions which greatly improved the paper. The second author is partially supported by the Deutsche Forschungsgemeinschaft, project number KL 1041/7-2.

1 Introduction

One of the stylized features of high-frequency financial time series is the presence of market microstructure noise (Black 1986). In financial econometrics, the *observed (logarithmic) price process* Y of an asset is therefore often modeled as a sum

$$Y_t = X_t + Z_t, \tag{1.1}$$

where X, called the *fundamental* or *efficient price process*, reflects the value of the asset according to some economic theory and Z is a *microstructure noise process* that captures deviations of Y from X. Typical noise sources include bid–ask bounces, discreteness of prices, informational asymmetry or transaction costs. As both X and Z are of economic interest but not observable, a major challenge is to develop statistical procedures to disentangle the two based on observations of Y. For example, given a continuous Itô semimartingale

$$X_t = X_0 + \int_0^t a_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}B_s, \tag{1.2}$$

a key quantity of interest is the *integrated (price) volatility* $C_T = \int_0^T \sigma_s^2 \,\mathrm{d}s$ for some finite time horizon T. In the absence of noise, estimating C_T is a straightforward matter: given observations $\{X_{i\Delta_n} : i = 1, \dots, [T/\Delta_n]\}$, the *realized variance (RV)* defined by $\sum_{i=1}^{[T/\Delta_n]} (\Delta_i^n X)^2$, where $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$, is a consistent estimator of C_T as $\Delta_n \to 0$ (Andersen & Bollerslev 1998, Andersen et al. 2003, Barndorff-Nielsen & Shephard 2002).

However, in practice, RV typically explodes as the sampling frequency increases, indicating the presence of noise at high frequencies. This is well documented by, for example, the volatility signature plots of Andersen et al. (2000). In order to construct noise-robust estimators of C_T, a common approach in the literature is to model $(Z_t)_{t\geq 0}$ at the observation times $i\Delta_n$ as

$$Z_{i\Delta_n} = \varepsilon_i^n, \tag{1.3}$$

where for each n, $(\varepsilon_i^n)_{i=1}^{[T/\Delta_n]}$ is a discrete time series. Examples for ε_i^n include rounding noise (Jacod 1996, Li & Mykland 2007, Robert & Rosenbaum 2010, 2012, Rosenbaum 2009), white noise (Bandi & Russell 2006, Barndorff-Nielsen et al. 2008, Podolskij & Vetter 2009, Zhang et al. 2005), AR- or MA-type noise (Aït-Sahalia et al. 2011, Da & Xiu 2021, Hansen & Lunde 2006), and certain non-parametric extensions thereof (Jacod et al. 2009, 2017, Li et al. 2020, Li & Linton 2021).

Our contribution to this area of research starts from the following observation: in almost all microstructure noise models in the literature, the noise process is only specified at the observation times (as in (1.3)). This naturally raises the question of compatibility between different sampling frequencies: if $i\Delta_n = j\Delta_m$, do we have $\varepsilon_i^n = \varepsilon_j^m$, at least in distribution? If ε_i^n is a white noise, this certainly holds true. However, if ε_i^n is a serially dependent time series, ensuring compatibility is a non-trivial matter.

One may ask why we insist on compatibility when in practice, only one or a few selected frequencies are typically used. While this is certainly true, virtually all asymptotic results in high-frequency econometrics are based on taking the limit $\Delta_n \to 0$, which by default means considering higher and higher frequencies. Also, volatility signature plots, the most prominent visual device to detect microstructure noise, explicitly track the behavior of RV as the sampling frequency increases. Therefore, whenever infill asymptotics are employed, understanding the impact of noise compatibility conditions becomes important.

The main objective of the this work is two-fold: first, to examine how compatibility assumptions affect the structure of the noise process, and second, to develop statistical procedures that allow us to estimate volatility in this setting. We start by introducing our model in Section 1.1, compare it to existing microstructure noise models in Section 1.2 and describe our methods in Section 1.3.

1.1 Model

We will now formulate the precise compatibility assumptions on the noise and derive our final model from these abstract conditions. Both the noise and the efficient price process are defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ satisfying the usual conditions.

Assumption (C). *The noise process $(Z_t)_{t\geq 0}$ is a mean-zero stochastic process indexed by a continuous time parameter t.*

This is our main compatibility condition and merely asserts that we want to model Z as a continuous-time process rather than a discrete time series for each frequency as in (1.3). Assumption (C) clearly ensures that $Z_{i\Delta_n} = Z_{j\Delta_m}$ as soon as $i\Delta_n = j\Delta_m$. For simplicity, we temporarily extend $(Z_t)_{t\in\mathbb{R}}$ to the whole real line. Moreover, we will impose additional assumptions on the noise process that will allow us to deduce our final model.

Assumption (C1). *The noise process $(Z_t)_{t\in\mathbb{R}}$ is a mean-zero, L^2-continuous, second-order stationary, and purely non-deterministic (i.e., $\bigcap_{t\in\mathbb{R}} \overline{\operatorname{span}}\{Z_s : s \leq t\} = \{0\}$, where $\overline{\operatorname{span}}$ denotes the L^2-closure of the linear span) stochastic process.*

The (rather strong) stationarity assumption on the noise reduces technicalities in the subsequent exposition and will be relaxed in our final model, where a time-varying and possibly non-stationary stochastic noise volatility is permitted. By the Wold–Karhunen representation of second-order stationary processes (Doob 1953, Chapter XII, Theorem 5.3), Assumption (C1) implies that

$$Z_t = \int_{-\infty}^{t} g(t-s)\,\mathrm{d}M_s \tag{1.4}$$

for some kernel $g \in L^2((0,\infty))$ and some process $(M_t)_{t\in\mathbb{R}}$ with second-order stationary and orthogonal increments. Let us now consider the variance function

$$\gamma(t) = \mathbb{E}[(Z_{s+t} - Z_s)^2], \qquad t > 0, \tag{1.5}$$

which, by stationarity, does not depend on the value of s. Because Z is L^2-continuous by Assumption (C1), we necessarily have $\gamma(t) \to 0$ as $t \to 0$. Our next assumption quantifies the speed of convergence.

Assumption (C2). *As $t \to 0$, we have that $\gamma(t) \sim t^{2H} L(t)$ for some $H \in (0, \frac{1}{2})$ and slowly varying (at 0) function L that is continuous on $(0,\infty)$.*

The condition $H < \frac{1}{2}$ is not restrictive for the purpose of modeling microstructure noise: if $H = \frac{1}{2}$, then Z has the same smoothness as Brownian motion, so that, in general, there is no way to discern Z from the efficient price process X; if $H > \frac{1}{2}$, then Z is smoother than X and RV remains a consistent estimator of C_T.

By a simple covariance computation (Barndorff-Nielsen et al. 2011, Equation (4.14)), Assumptions (C1) and (C2) imply that the autocorrelation function (ACF) of noise increments satisfies

$$\Gamma_r^n = \mathrm{Corr}(\Delta_i^n Z, \Delta_{i+r}^n Z) \to \Gamma_r^H \tag{1.6}$$

for every $r \geq 0$, where

$$\Gamma_0^H = 1 \qquad \text{and} \qquad \Gamma_r^H = \frac{1}{2}\Big((r+1)^{2H} - 2r^{2H} + (r-1)^{2H}\Big), \quad r \geq 1. \tag{1.7}$$

The family of ACFs displayed in (1.7) can therefore be seen as *prototypical* for the increments of compatible noise processes. This observation motivates our final noise model.

Assumption (Z). *The process $(Z_t)_{t\geq 0}$ is given by*

$$Z_t = Z_0 + \int_0^t g(t-s)\rho_s \,\mathrm{d}W_s, \qquad t \geq 0, \tag{1.8}$$

where W is a d-dimensional standard $\mathbb{F}$-Brownian motion and $(\rho_t)_{t\geq 0}$ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d}$-valued process. The kernel $g\colon (0,\infty) \to \mathbb{R}$ is of the form

$$g(t) = K_H^{-1} t^{H-\frac{1}{2}} + g_0(t) \tag{1.9}$$

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

$$K_H = \frac{\sqrt{2H \sin(\pi H)\Gamma(2H)}}{\Gamma(H+\frac{1}{2})} \tag{1.10}$$

is a normalizing constant and $g_0\colon [0,\infty) \to \mathbb{R}$ is a smooth function with $g_0(0) = 0$.

In principle, the function γ in (1.5) might satisfy $\gamma(t) \sim t^{2H}L(t)$ with $H = 0$ and $L(t) \to 0$. In this case, $\Gamma_r^n \to \Gamma_r^0 = \mathbb{1}_{\{r=0\}} - \frac{1}{2}\mathbb{1}_{\{r=1\}}$, which is exactly the ACF of increments of white noise. Because the case $H = 0$ is special and, at least for white noise, has been extensively studied in the literature, we only consider $H > 0$ in the following.

As g_0 is smooth, the kernel g in (1.9) produces exactly the same limiting ACF as in (1.7). We dropped the slowly varying function L to simplify the subsequent analysis (and also because such an extension can hardly be distinguished statistically).

In the special case where $g_0 \equiv 0$ and $\rho_s \equiv \rho$ is a constant, Z is—up to a term of finite variation—simply a multiple of *fractional Brownian motion (fBM)*. If further $X_t = \sigma B_t$ with constant volatility σ, then the resulting observed process $Y_t = \sigma B_t + \rho Z_t$ is a *mixed fractional Brownian motion (mfBM)* as introduced by Cheridito (2001). Our model for the observed price process, as the sum of X in (1.2) and Z in (1.8), can be viewed as a non-parametric generalization of mfBM that allows for stochastic volatility in both its Brownian and its noise component. We do keep the parameter H, though, which we refer to as the *roughness parameter* of Z (or Y). In analogy with mfBM, we call

$$Y_t = X_t + Z_t = Y_0 + \int_0^t a_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}B_s + \int_0^t g(t-s)\rho_s \,\mathrm{d}W_s, \qquad t \geq 0, \tag{1.11}$$

the observed price process in our model, a *mixed semimartingale*.

Remark 1.1. It is important to note that fractional Brownian motion and other fractional models were also considered as asset price models in the literature, often in the context of long-range dependence; see Mandelbrot (1997), Bayraktar et al. (2004), Bender et al. (2011), Bianchi & Pianese (2018), for example. In those works, it is typically the behavior of the kernel g at $t = \infty$ that is of primary interest, as this determines whether the resulting process has short or long memory. Our concern, by contrast, is the behavior of this kernel around $t = 0$, which governs the local regularity, or *roughness*, of the noise process Z. In fact, on a finite time interval $[0, T]$, there is no way to distinguish between short- and long-range dependence (note that in our model, the behavior of g at $t = \infty$ is not specified by (1.9)). This is why in this work, we explicitly do *not* call H the Hurst parameter (as this is usually associated with long-range dependence) but rather call it the roughness parameter of Z. Of course, for fBM, both interpretations fall together, but for non-parametric generalizations as we consider them in (1.8), this distinction is crucial.

Remark 1.2. Mixed semimartingale models are also in line with no-arbitrage concepts in mathematical finance. Clearly, as non-semimartingales, they admit arbitrage in the FLVR sense; see Delbaen & Schachermayer (1994). However, as shown in Cherny (2008) and Guasoni et al. (2008) (see also Jarrow et al. (2009)), mfBM does not admit arbitrage in the presence of transaction costs, which are exactly one of the market inefficiencies that microstructure noise models are supposed to capture.

1.2 Relation to existing microstructure noise models

Comparing mixed semimartingales with other microstructure noise models in the literature, we first note that Z in (1.8) is a non-shrinking noise. At the same time, what sets Z apart from many non-shrinking noise models (cf. Aït-Sahalia et al. (2011), Da & Xiu (2021), Jacod et al. (2017), Li et al. (2020), Li & Linton (2021)) is the fact that it has shrinking noise *increments*, that is, $\sup_{i=1,\dots,[T/\Delta_n]} \mathrm{Var}(\Delta_i^n Z) \to 0$ as $n \to \infty$. In fact, for Y from (1.11), $\sup_{i=1,\dots,[T/\Delta_n]} \mathrm{Var}(\Delta_i^n Y) \leq C\Delta_n^{2H}$, so the *shrinkage rate* of price increments is $2H$, which varies in $(0, 1)$ if $H \in (0, \frac{1}{2})$. A related observation is that the RV of a mixed semimartingale explodes at a rate of $\Delta_n^{-(1-2H)}$ by Theorem 2.1 below. Thus, for $H \in (0, \frac{1}{2})$, the divergence rate $1 - 2H$ of volatility signature plots varies within $(-1, 0)$. As Figure 1 shows, this is well matched by recent quote data obtained from major US stocks.

In other non-shrinking noise models in the literature (such as modulated white noise or AR- or MA-models for (1.3)), the noise increments are not shrinking, so the shrinkage rate of noise increments is 0. For a similar reason, the divergence rate in volatility signature plots is -1 according to these models. While earlier work did find such exponents in empirical studies (see, for example, Aït-Sahalia et al. (2011)), in our 2019 sample, only a very small percentage of companies and days exhibits exponents around those values; see Figure 1. This is in line with Aït-Sahalia & Xiu (2019) who found that noise has decreased over time due to improvements in market efficiency. So far, the rich variety of scaling exponents as we see in Figure 1 could only be explained by shrinking noise models (e.g., Aït-Sahalia & Xiu (2019), Da & Xiu (2021), Kalnina & Linton (2008)). However, as Aït-Sahalia & Jacod (2014) point out in their Chapter 7, shrinking noise is bound to violate compatibility across frequencies.

Another empirical characteristic of our data sample (see Figure 2) that can be explained by the mixed semimartingale model is serial dependence of price increments. While colored

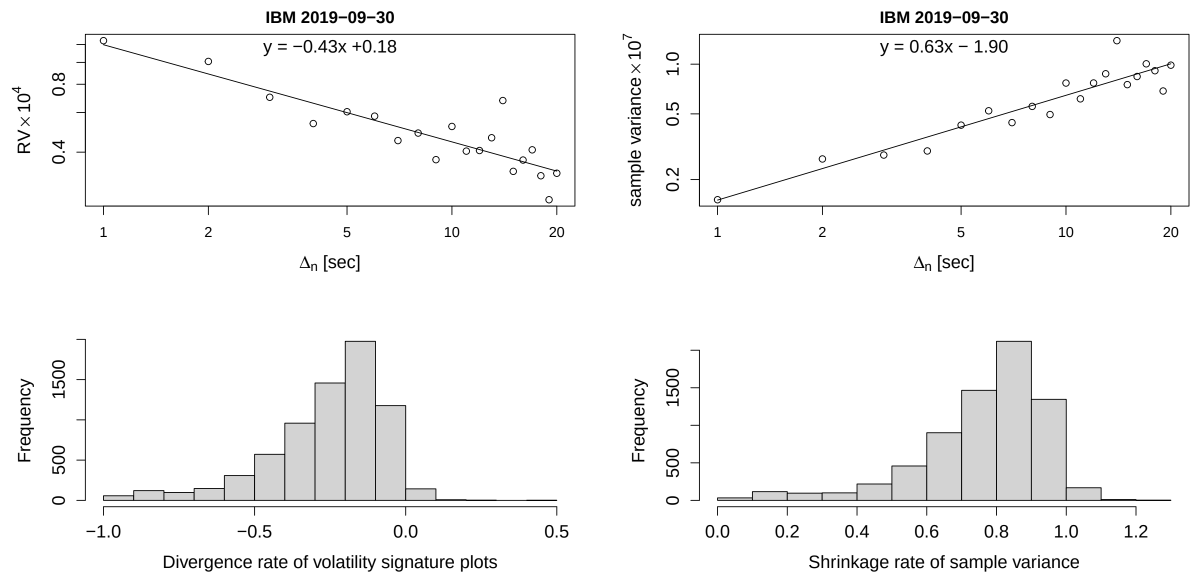


Figure 1: Top row: Volatility signature plot for IBM mid-quote data on September 30, 2019 (left), and sample variance of price increments as a function of Δ_n for the same asset and day (right). Both are shown on a log–log scale including least-square lines. Bottom row: Histogram of divergence rates in volatility signature plots (left) and histogram of shrinkage rates for sample variances of price increments (right). Each data point corresponds to one out of 29 DJIA companies and one trading day in 2019.

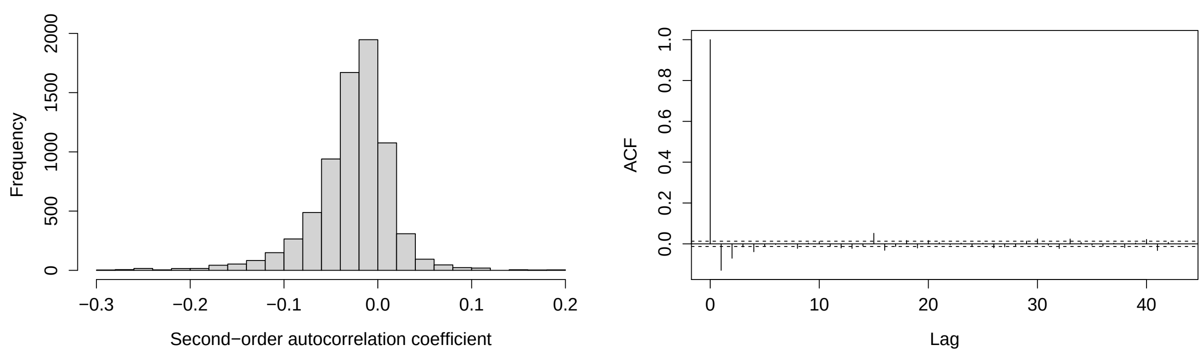


Figure 2: Left: Histogram of second-order autocorrelation coefficients of 1s increments for 29 DJIA stocks and all trading days in 2019. Right: ACF of 1s increments for IBM on September 30, 2019.

noise has been studied in many previous works (see the Introduction), because of (1.7), a specific property of mixed semimartingales are autocorrelations of increments that only decay polynomially at a rate of r^{-2H-2}, with an exponent strictly between -2 and -1 if $H \in (0, \frac{1}{2})$. Employing a non-shrinking noise model with non-shrinking noise increments, Jacod et al. (2017) found empirical support for polynomially decaying autocorrelations, with exponents between -0.5 and -1.7. In Da & Xiu (2021) and Li et al. (2020) and in Li & Linton (2021), the authors proved central limit theorems (CLTs) for noise models with decay exponents strictly smaller than -3 and -2, respectively.

Remark 1.3. In recent years, there has been growing interest in *rough volatility models*; see, Gatheral et al. (2018), El Euch & Rosenbaum (2019), for example. In these works, it is the volatility σ that is modeled by a rough stochastic process. In this paper, by contrast, we are concerned with roughness of observed prices, caused by market microstructure noise. It is important to note that roughness on the price level and roughness on the volatility level imply distinct features of asset returns and must therefore be modeled and analyzed separately. For instance, if the observed price is simply $\int_0^t \sigma_s \, \mathrm{d}B_s$, without noise but with a rough volatility σ, we will *not* see explosion of the RV measure in volatility signature plots as the sampling frequency increases. In fact, in the absence of microstructure noise, the asymptotic behavior of RV does *not* depend on the roughness of volatility (Jacod & Protter 2012, Theorem 5.4.2).

1.3 Methodology

On an abstract level, the statistical problem we are facing here is a deconvolution problem: given a semimartingale signal X and rough signal Z, how can we recover the two (or certain interesting components of the two, such as volatility) based on observing their sum $Y = X + Z$. The following result, due to Cheridito (2001) and van Zanten (2007), puts a constraint on the identifiability of the (smoother) semimartingale signal:

Proposition 1.4. *Assume that Y is an mfBM, that is, $Y = X + Z$ where $X = \sigma B$ and $Z = \rho B^H$ for some $\rho, \sigma \in (0, \infty)$, B is a Brownian motion and B^H is an independent fBM with Hurst parameter $H \in (0, \frac{1}{2})$. For any $T > 0$, the laws of $(Y_t)_{t\in[0,T]}$ and $(Z_t)_{t\in[0,T]}$ are mutually equivalent if $H \in (0, \frac{1}{4})$ and mutually singular if $H \in [\frac{1}{4}, \frac{1}{2})$.*

In other words, if $H \in (0, \frac{1}{4})$, due to the roughness of the noise, there is no way to consistently estimate σ on a finite time interval. This is conceptually similar to the fact that the finite-variation part of a semimartingale cannot be estimated consistently in finite time if there is a Brownian component. We will comment on possible pathways to estimate σ if $H < \frac{1}{4}$ in Section 6.

Remark 1.5. The case of white noise, which formally corresponds to $H = 0$ in terms of roughness, is special in this context: it is rougher than Z in (1.8), but $C_T = \int_0^T \sigma_s^2 \, \mathrm{d}s$ can still be recovered, for example, through subsampling (Zhang et al. 2005, Zhang 2006) or pre-averaging (Jacod et al. 2009, Podolskij & Vetter 2009, Hautsch & Podolskij 2013). In fact, if k_n is an increasing sequence and Z is a white noise, then $k_n^{-1} \sum_{j=0}^{k_n} Y_{(i+j)\Delta_n} \approx X_{i\Delta_n}$ by the law of large numbers. By contrast, if $H \in (0, \frac{1}{2})$, the process Z in (1.8) is *continuous* (and so is Y in (1.1)), which implies that $k_n^{-1} \sum_{j=0}^{k_n} Y_{(i+j)\Delta_n} \approx Y_{i\Delta_n}$, so pre-averaging does not remove the noise part at all! Even worse, if we average over increments of Y, then, by some variance computations (not shown here), this will actually remove the semimartingale part and not the noise component. Therefore, while classical noise-robust volatility estimators work well if Z is a modulated white noise ("$H = 0$"), they become inconsistent for C_T as soon as $H > 0$.

Against this background, we will first establish a CLT for variation functionals of mixed semimartingales in Section 2 and then use this CLT in Section 3 to derive consistent and asymptotically mixed normal estimators for H, $\int_0^T \sigma_s^2 \, \mathrm{d}s$ (if $H > \frac{1}{4}$) and $\int_0^T \rho_s^2 \, \mathrm{d}s$. A major challenge here is the subtle interplay between the semimartingale X and the noise process Z, leading to (a potentially large number of) intermediate limits between the law of large numbers (LLN) and the CLT if $H > \frac{1}{4}$. While the LLN and the CLT limits only depend on the noise, these intermediate limits depend on σ and H at the same time. On the one hand, this is desirable as it permits us to identify σ in the first place; on the other hand, this creates a complex dependence between the estimators of C_T and H, and we need to employ an iterative debiasing procedure to obtain rate-optimal estimators. Sections 4 and 5 contain a simulation and an empirical study, respectively. Section 6 concludes. The supplement contains the proof of the main results (Appendices A–E), the details of the iterative debiasing procedure (Appendix F) and our choice of tuning parameters for the simulation study (Appendix G).

2 Central limit theorem for variation functionals

As with most estimators in high-frequency statistics, ours are based on limit theorems for power variations and related functionals. More precisely, given $L, M \in \mathbb{N}$ and a test function $f\colon \mathbb{R}^{d\times L} \to \mathbb{R}^M$, our goal is to establish a CLT for *normalized variation functionals* of the form

$$V_f^n(Y,t) = \Delta_n \sum_{i=1}^{[t/\Delta_n]-L+1} f\Big(\frac{\underline{\Delta}_i^n Y}{\Delta_n^H}\Big),$$

where

$$\Delta_i^n Y = Y_{i\Delta_n} - Y_{(i-1)\Delta_n} \in \mathbb{R}^d, \quad \underline{\Delta}_i^n Y = (\Delta_i^n Y, \Delta_{i+1}^n Y, \dots, \Delta_{i+L-1}^n Y) \in \mathbb{R}^{d\times L}. \tag{2.1}$$

For semimartingales, this is a well studied topic; see Aït-Sahalia & Jacod (2014) and Jacod & Protter (2012) for in-depth treatments of this subject. For fractional Brownian motion or moving-average processes as in (1.8), the theory is similarly well understood; see Barndorff-Nielsen et al. (2011) and Brouste & Fukasawa (2018). Surprisingly, it turns out that the mixed case is more complicated than the "union" of the purely semimartingale and the purely fractional case. For instance, as we elaborate in Remark 2.4, already for power variations of even order, we may have a large number of higher-order bias terms.

2.1 The result

Our CLT will be proved under the following set of assumptions. In what follows, $\|\cdot\|$ denotes the Euclidean norm (in $\mathbb{R}^n$ if applied to vectors and in $\mathbb{R}^{nm}$ if applied to a matrix in $\mathbb{R}^{n\times m}$).

Assumption (CLT). *The observation process Y is given by the sum of X from* (1.2) *and Z from* (1.8) *with the following specifications:*

(i) The function $f\colon \mathbb{R}^{d\times L} \to \mathbb{R}^M$ is even and infinitely differentiable. Moreover, all its derivatives (including f itself) have at most polynomial growth.

(ii) The drift process a is d-dimensional, locally bounded and $\mathbb{F}$-adapted. The volatility process σ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d}$-valued process. Moreover, for every $T > 0$, there is $K_1 \in (0,\infty)$ such that for all $s,t \in [0,T]$,

$$\mathbb{E}\Big[1 \wedge \|\sigma_t - \sigma_s\|\Big] \le K_1 |t-s|^{\frac{1}{2}}. \tag{2.2}$$

(iii) Both B and W are independent d-dimensional standard $\mathbb{F}$-Brownian motions.

(iv) The noise volatility process ρ takes the form

$$\rho_t = \rho_t^{(0)} + \int_0^t \widetilde{b}_s \,\mathrm{d}s + \int_0^t \widetilde{\rho}_s \,\mathrm{d}\widetilde{W}_s, \qquad t \geq 0, \tag{2.3}$$

where

(a) $\rho^{(0)}$ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d}$-valued process such that for all $T > 0$,

$$\mathbb{E}\Big[1 \wedge \|\rho_t^{(0)} - \rho_s^{(0)}\|\Big] \leq K_2|t-s|^\gamma, \qquad s,t \in [0,T], \tag{2.4}$$

for some $\gamma \in (\frac{1}{2}, 1]$ and $K_2 \in (0,\infty)$;

(b) $\widetilde{b}$ is $d \times d$-dimensional, locally bounded and $\mathbb{F}$-adapted;

(c) $\widetilde{\rho}$ is an $\mathbb{F}$-adapted locally bounded $\mathbb{R}^{d\times d\times d}$-valued process (e.g., the (ij)th component of the stochastic integral in (2.3) equals $\sum_{k=1}^d \int_0^t \widetilde{\rho}_s^{ijk} \,\mathrm{d}\widetilde{W}_s^k$) such that for all $T > 0$, there exist $\varepsilon > 0$ and $K_3 \in (0,\infty)$ with

$$\mathbb{E}\Big[1 \wedge \|\widetilde{\rho}_t - \widetilde{\rho}_s\|\Big] \leq K_3|t-s|^\varepsilon, \qquad s,t \in [0,T]. \tag{2.5}$$

(d) $\widetilde{W}$ is a d-dimensional $\mathbb{F}$-Brownian motion that is jointly Gaussian with (B,W).

(v) The kernel g takes the form (1.9) with $H \in (0, \frac{1}{2})$ and some $g_0 \in C^\infty([0,\infty))$ with $g_0(0) = 0$.

To describe the CLT for $V_f^n(Y,t)$, we need some more notation. Define μ_f as the $\mathbb{R}^M$-valued function that maps $v = (v_{k\ell,k'\ell'}) \in (\mathbb{R}^{d\times L})^2$ to $\mathbb{E}[f(\mathcal{Z})]$ where $\mathcal{Z} \in \mathbb{R}^{d\times L}$ follows a multivariate normal distribution with mean 0 and $\mathrm{Cov}(\mathcal{Z}_{k\ell}, \mathcal{Z}_{k'\ell'}) = v_{k\ell,k'\ell'}$. Note that μ_f is infinitely differentiable because f is. Furthermore, if $\mathcal{Z}' \in \mathbb{R}^{d\times L}$ is such that $\mathcal{Z}$ and $\mathcal{Z}'$ are jointly Gaussian with mean 0, covariances $\mathrm{Cov}(\mathcal{Z}_{k\ell}, \mathcal{Z}_{k'\ell'}) = \mathrm{Cov}(\mathcal{Z}'_{k\ell}, \mathcal{Z}'_{k'\ell'}) = v_{k\ell,k'\ell'}$ and cross-covariances $\mathrm{Cov}(\mathcal{Z}_{k\ell}, \mathcal{Z}'_{k'\ell'}) = q_{k\ell,k'\ell'}$, we define

$$\gamma_{f_{m_1},f_{m_2}}(v,q) = \mathrm{Cov}\left(f_{m_1}(\mathcal{Z}), f_{m_2}(\mathcal{Z}')\right), \qquad m_1, m_2 = 1,\dots,M.$$

We further introduce a multi-index notation adapted to the definition of μ_f. For $\chi = (\chi_{k\ell,k'\ell'}) \in \mathbb{N}_0^{(d\times L)\times(d\times L)}$ and v as above, we let

$$|\chi| = \sum_{k,k'=1}^d \sum_{\ell,\ell'=1}^L \chi_{k\ell,k'\ell'}, \qquad \chi! = \prod_{k,k'=1}^d \prod_{\ell,\ell'=1}^L \chi_{k\ell,k'\ell'}!\,,$$

$$v^\chi = \prod_{k,k'=1}^d \prod_{\ell,\ell'=1}^L {v_{k\ell,k'\ell'}}^{\chi_{k\ell,k'\ell'}}, \qquad \partial^\chi \mu_f = \frac{\partial^{|\chi|}\mu_f}{\partial v_{11,11}^{\chi_{11,11}} \cdots \partial v_{dL,dL}^{\chi_{dL,dL}}}.$$

Finally, recalling (1.7), we define for all $k,k' \in \{1,\dots,d\}$, $\ell,\ell' \in \{1,\dots,L\}$ and $r \in \mathbb{N}_0$,

$$\pi_r(s)_{k\ell,k'\ell'} = (\rho_s\rho_s^T)_{kk'}\Gamma^H_{|\ell-\ell'+r|}, \quad c(s)_{k\ell,k'\ell'} = (\sigma_s\sigma_s^T)_{kk'}\mathbb{1}_{\{\ell=\ell'\}}, \quad \pi(s) = \pi_0(s). \tag{2.6}$$

The following CLT is our first main result. We use $\stackrel{\mathrm{st}}{\Longrightarrow}$ (resp., $\stackrel{L^1}{\Longrightarrow}$) to denote functional stable convergence in law (resp., convergence in L^1) in the space of càdlàg functions $[0,\infty) \to \mathbb{R}$ equipped with the local uniform topology. In the special case where Y follows the parametric model of an mfBM and the test function is $f(x) = x^2$, the CLT was obtained by Dozzi et al. (2015).

Theorem 2.1. *Grant Assumption (CLT) and let $N(H) = [1/(2-4H)]$. Then*

$$\Delta_n^{-\frac{1}{2}}\bigg\{V_f^n(Y,t) - \int_0^t \mu_f(\pi(s))\,\mathrm{d}s \\ - \sum_{j=1}^{N(H)} \Delta_n^{j(1-2H)} \sum_{|\chi|=j} \frac{1}{\chi!}\int_0^t \partial^\chi \mu_f(\pi(s))c(s)^\chi\,\mathrm{d}s\bigg\} \overset{\text{st}}{\Longrightarrow} \mathcal{Z}, \tag{2.7}$$

where $\mathcal{Z} = (\mathcal{Z}_t)_{t\geq 0}$ is an $\mathbb{R}^M$-valued continuous process defined on a very good filtered extension $(\overline{\Omega}, \overline{\mathcal{F}}, (\overline{\mathcal{F}}_t)_{t\geq 0}, \overline{\mathbb{P}})$ of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ which, conditionally on $\mathcal{F}$, is a centered Gaussian process with independent increments and such that the covariance function $\mathcal{C}_t^{m_1m_2} = \overline{\mathbb{E}}[\mathcal{Z}_t^{m_1}\mathcal{Z}_t^{m_2} \mid \mathcal{F}]$, for $m_1, m_2 = 1, \ldots, M$, is given by

$$\mathcal{C}_t^{m_1m_2} = \int_0^t \bigg\{\gamma_{f_{m_1},f_{m_2}}(\pi(s),\pi(s)) + \sum_{r=1}^\infty \Big(\gamma_{f_{m_1},f_{m_2}} + \gamma_{f_{m_2},f_{m_1}}\Big)(\pi(s),\pi_r(s))\bigg\}\,\mathrm{d}s. \tag{2.8}$$

Remark 2.2. In fact, it suffices to require f be $2(N(H)+1)$-times continuously differentiable with derivatives of at most polynomial growth. A decomposition as in (2.3) is standard for CLTs in high-frequency statistics. But here we need it for ρ (instead of σ) as the noise process dominates the efficient price process in the limit $\Delta_n \to 0$. Condition (2.2) on σ is satisfied if, for example, σ is itself a continuous Itô semimartingale.

Remark 2.3. Both the LLN limit

$$V_f(Y,t) = \int_0^t \mu_f(\pi(s))\,\mathrm{d}s \tag{2.9}$$

and the fluctuation process $\mathcal{Z}$ originate from the rough process Z. In other words, if $\sigma \equiv 0$ (i.e., in the pure fractional case), we would have (2.7) without the $\sum_{j=1}^{N(H)}$-expression; see Barndorff-Nielsen et al. (2011). Even if $\sigma \not\equiv 0$, in the case where $H < \frac{1}{4}$, no additional terms are present because $N(H) = 0$. This is in line with Proposition 1.4, which states that it is impossible to consistently estimate $C_t = \int_0^t \sigma_s^2\,\mathrm{d}s$ if $H < \frac{1}{4}$. If $H \in (\frac{1}{4}, \frac{1}{2})$, the "mixed" terms in the $\sum_{j=1}^{N(H)}$-expression will allow us to estimate C_t.

Remark 2.4. Let us consider the special case where $d = 1$ and $f(x) = x^{2p}$ for some $p \in \mathbb{N}$. Then (2.7) reads

$$\Delta_n^{-\frac{1}{2}}\bigg\{V_f^n(Y,t) - V_f(Y,t) - \sum_{j=1}^{N(H)} \Delta_n^{j(1-2H)}\mu_{2p}\binom{p}{j}\int_0^t \rho_s^{2p-2j}\sigma_s^{2j}\,\mathrm{d}s\bigg\} \overset{\text{st}}{\Longrightarrow} \mathcal{Z},$$

where μ_{2p} is the moment of order $2p$ of a standard normal variable. Typically, one is interested in estimating only one of the terms in the sum $\sum_{j=1}^{N(H)}$ at a time (e.g., $\int_0^t \sigma_s^{2p}\,\mathrm{d}s$ corresponding to $j = p$). All other terms (e.g., $j \neq p$) have to be considered as higher-order bias terms in this case. The appearance of (potentially many, if $N(H)$ is large) bias terms for test functions as simple as powers of even order neither happens in the pure semimartingale nor in the pure fractional setting.

Remark 2.5. The following values for H are special:

$$\mathcal{H} = \bigg\{\frac{1}{2} - \frac{1}{4n} : n \geq 1\bigg\} = \bigg\{\frac{1}{4}, \frac{3}{8}, \frac{5}{12}, \frac{7}{16}, \ldots\bigg\}. \tag{2.10}$$

Indeed, if $H \in \mathcal{H}$, then $N(H) = 1/(2-4H)$. In particular, the term in (2.7) that corresponds to $j = N(H)$ is exactly of order $\Delta_n^{1/2}$. So in this case, (2.7) can also be viewed as convergence to a non-central mixed normal distribution.

2.2 Overview of the proof of Theorem 2.1

In the following, we describe the main difficulties in the proof of Theorem 2.1 and defer the details to the supplementary material. In addition to the usual steps that are common to CLTs in high-frequency statistics, there are two new challenges in the present setting:

(i) The observation process Y is *not* a semimartingale (and not even close to one). This is because the rough component Z *dominates* the efficient price process X in the limit as $\Delta_n \to 0$ (which cannot be remedied by pre-averaging; see Remark 1.5). In particular, the increments of Y remain conditionally *dependent* as $\Delta_n \to 0$.

(ii) If H is close to (but smaller than) $\frac{1}{2}$, the semimartingale part is only marginally smoother than the noise part. So for the CLT, there will be an intricate interplay between the efficient price process and the noise process.

To overcome the first challenge, we employ a multiscale analysis: by suitably truncating the increments of Y, we can restore, to some degree (not on the finest scale Δ_n but on some intermediate scale $\theta_n\Delta_n$ where $\theta_n \to \infty$), asymptotic conditional independence between increments of Y (see Lemma C.1). This in turn gives $V^n_f(Y,t)$, as a process in t, a semimartingale-like structure on this intermediate scale, which is sufficient for deriving the CLT when we center by appropriate conditional expectations (see (C.6)). However, because increments are still correlated on the finest scale, the limiting process is not the usual one for semimartingales but the one for (modulated) fractional Brownian motion (see (2.8), in particular). Regarding the second challenge above, we find, to our surprise, that the semimartingale component *never* enters the CLT limit of $V^n_f(Y,t)$ when centered by conditional expectations (see Lemma C.2), no matter how close H is to $\frac{1}{2}$. By contrast, it does affect the limit behavior of these conditional expectations (Lemmas C.3–C.9), producing an H-dependent number of higher-order bias terms that neither appear in the pure semimartingale nor in the pure fractional setting.

3 Estimating the roughness parameter and integrated price and noise volatilities

In this section, we assume $d = 1$ for simplicity. We develop an estimation procedure for the roughness parameter of the noise and the integrated price (if $H > \frac{1}{4}$) and noise volatilities, that is, for H, $C_t = \int_0^t \sigma_s^2\,\mathrm{d}s$ and $\Pi_t = \int_0^t \rho_s^2\,\mathrm{d}s$. To avoid additional bias terms (cf. Remark 2.4), we use quadratic functionals only, that is, we consider

$$f_r(x) = x_1 x_{r+1}, \quad x = (x_1, \dots, x_{r+1}) \in \mathbb{R}^{r+1}, \quad r \in \mathbb{N}_0,$$

and the associated variation functionals

$$V^n_{r,t} = V^n_{f_r}(Y,t) = \Delta_n^{1-2H} \sum_{k=1}^{[t/\Delta_n]-r} \Delta^n_k Y \Delta^n_{k+r} Y.$$

Note that $V^n_{r,t}$ is not a statistic as it depends on the unknown parameter H. Therefore, we introduce $\widehat{V}^n_t = (\widehat{V}^n_{0,t}, \dots, \widehat{V}^n_{R,t})$, a non-normalized version of $V^n_{r,t}$ that is a statistic:

$$\widehat{V}^n_{r,t} = \widehat{V}^n_{f_r}(Y,t) = \sum_{k=1}^{[t/\Delta_n]-r} \Delta^n_k Y \Delta^n_{k+r} Y, \qquad r \in \mathbb{N}_0.$$

Clearly, $\Delta_n^{1-2H}\widehat{V}^n_{r,t} = V^n_{r,t}$, so our main CLT (Theorem 2.1) immediately yields:

Corollary 3.1. *Let $\widehat{V}^n_t = (\widehat{V}^n_{0,t}, \dots, \widehat{V}^n_{R,t})$ for a fixed but arbitrary $R \in \mathbb{N}_0$. For all $H \in (0, \frac{1}{2})$,*

$$\Delta_n^{-\frac{1}{2}}\left\{\Delta_n^{1-2H}\widehat{V}^n_t - \Gamma^H \int_0^t \rho_s^2\,\mathrm{d}s - e_1 \int_0^t \sigma_s^2\,\mathrm{d}s\,\Delta_n^{1-2H}\,\mathbb{1}_{[\frac{1}{4},\frac{1}{2})}(H)\right\} \overset{\mathrm{st}}{\Longrightarrow} \mathcal{Z}, \tag{3.1}$$

where $\Gamma^H = (\Gamma^H_0, \dots, \Gamma^H_R)$, $e_1 = (1, 0, \dots, 0) \in \mathbb{R}^{1+R}$ and $\mathcal{Z}$ is as in (2.7). The covariance process $\mathcal{C}^H(t) = (\mathcal{C}^H_{ij}(t))_{i,j=0,\dots,R}$ in (2.8) is given by

$$\mathcal{C}^H_{ij}(t) = \mathcal{C}^H_{ij} \int_0^t \rho_s^4\,\mathrm{d}s, \tag{3.2}$$

$$\mathcal{C}^H_{ij} = \Gamma^H_{|i-j|} + \Gamma^H_i\Gamma^H_j + \sum_{r=1}^{\infty}\Big(\Gamma^H_r\Gamma^H_{|i-j+r|} + \Gamma^H_{|r-j|}\Gamma^H_{i+r} + \Gamma^H_r\Gamma^H_{|j-i+r|} + \Gamma^H_{|r-i|}\Gamma^H_{j+r}\Big).$$

As we can see, if $H \in (\frac{1}{4}, \frac{1}{2})$, only RV ($r = 0$) contains information about $C_t = \int_0^t \sigma_s^2\,\mathrm{d}s$. But to first order, $V^n_{0,t} = \Delta_n^{1-2H}\widehat{V}^n_{0,t}$ estimates $\Pi_t = \int_0^t \rho_s^2\,\mathrm{d}s$, the integrated noise volatility. In order to obtain C_t, our strategy is to use $\widehat{V}^n_{r,t}$ for $r \geq 1$ to remove the first-order limit of $\widehat{V}^n_{0,t}$. But here is a caveat: both Δ_n^{1-2H} and Γ^H contain the unknown parameter H, so we need to estimate H first.

The most obvious estimator for H is obtained by calculating the rate of divergence in volatility signature plots, that is, by regressing $\log\Delta_n$ on $\log\widehat{V}^n_{0,t}$ (see also Rosenbaum (2011) for a more general but related concept). However, as noted by Dozzi et al. (2015) in their Remark 3.1, already in an mfBM model, this regression based estimator only has a logarithmic rate of convergence. Indeed, as our simulation study in Section 4 shows, this estimator systematically overestimates H unless H is very close to 0 or $\frac{1}{2}$. In the pure fractional case, rate-optimal estimators are given by so-called change-of-frequency or autocorrelation estimators (Barndorff-Nielsen et al. 2011, Corcuera et al. 2013). Both extract information about H by considering the ratio of (different combinations of) $\widehat{V}^n_{r,t}$ for different values of r. For example, the simplest autocorrelation estimator is

$$\widetilde{H}^n_{\mathrm{acf}} = \frac{1}{2}\left[1 + \log_2\left(\frac{\widehat{V}^n_{1,t}}{\widehat{V}^n_{0,t}} + 1\right)\right], \tag{3.3}$$

which is based on the fact that $\widehat{V}^n_{1,t}/\widehat{V}^n_{0,t} = V^n_{1,t}/V^n_{0,t} \xrightarrow{\mathbb{P}} \Gamma^H_1 = 2^{2H-1} - 1$. But due to the bias term that appears in (3.1) when $r = 0$, the convergence rate worsens and becomes suboptimal when (3.3) is applied to mixed semimartingales. The first rate-optimal estimator for H in the case of mfBM was constructed in Theorem 3.2 of Dozzi et al. (2015) by using a variant of (3.3) that cancels out the contribution from $\widehat{V}^n_{0,t}$. However, this estimator suffers from a large constant in the asymptotic variance (and another issue that we address in Section 3.2). In fact, in their Remark 3.2, Dozzi et al. (2015) do not recommend using it in practice even though it has a better convergence rate than the estimator based on volatility signature plots.

To do better, our strategy is to use linear combinations of $\widehat{V}^n_{r,t}$ for multiple values of r. To this end, we choose two weight vectors $a = a(R) = (a_0, \dots, a_R)$ and $b = b(R) = (b_0, \dots, b_R)$ in $\mathbb{R}^{1+R}$ and consider the statistic

$$\widetilde{H}^n = \varphi^{-1}\Big(\frac{\langle a, \widehat{V}^n_t\rangle}{\langle b, \widehat{V}^n_t\rangle}\Big) \qquad \text{with} \qquad \varphi(H) = \frac{\langle a, \Gamma^H\rangle}{\langle b, \Gamma^H\rangle}, \tag{3.4}$$

where $\langle\cdot,\cdot\rangle$ denotes the standard inner product on $\mathbb{R}^{1+R}$ and a and b are assumed to be such that φ is invertible. The further analysis is now dependent on whether $H \in (0, \frac{1}{4})$ or $H \in (\frac{1}{4}, \frac{1}{2})$ and, in the latter case, whether $a_0 = b_0 = 0$ or at least one of a_0 and b_0 is not zero.

3.1 Estimation without quadratic variation or if $H \in (0, \frac{1}{4})$

If $a_0 = b_0 = 0$, we exclude quadratic variation from our estimation procedure for H. This has the advantage that the term $e_1 \int_0^t \sigma_s^2 \,\mathrm{d}s\, \Delta_n^{1-2H}$ in (3.1), which is only non-zero for $r = 0$, disappears. The same holds true if $H < \frac{1}{4}$ (even if a_0 or b_0 is not zero): there is no asymptotic bias term in (3.1).

Theorem 3.2. *Assume that $H \in (0, \frac{1}{2})$ and choose $R \in \mathbb{N}$ and $a, b \in \mathbb{R}^{1+R}$ such that φ from (3.4) is invertible. If $H \in (\frac{1}{4}, \frac{1}{2})$, further assume that $a_0 = b_0 = 0$.*

(i) The estimator $\widetilde{H}^n$ introduced in (3.4) satisfies

$$\Delta_n^{-\frac{1}{2}}(\widetilde{H}^n - H) \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{H,0} \frac{\int_0^t \rho_s^4 \,\mathrm{d}s}{(\int_0^t \rho_s^2 \,\mathrm{d}s)^2}\Big), \tag{3.5}$$

where $\mathcal{Z}$ is the same as in (3.1) and

$$\begin{aligned}\mathrm{Var}_{H,0} &= \mathrm{Var}_{H,0}(R, a, b, H) \\ &= \Big(\frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle}\Big)^2 \{a^T - \varphi(H)b^T\}\mathcal{C}^H\{a - \varphi(H)b\}.\end{aligned} \tag{3.6}$$

(ii) If $H \in (\frac{1}{4}, \frac{1}{2})$, choose $c \in \mathbb{R}^{1+R}$ and define

$$\widehat{C}_t^n = \Big\{\widehat{V}_{0,t}^n - \frac{\langle c, \widehat{V}_t^n\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle}\Big\}\Big(1 - \frac{c_0}{\langle c, \Gamma^{\widetilde{H}^n}\rangle}\Big)^{-1}. \tag{3.7}$$

Then

$$\Delta_n^{\frac{1}{2}-2H}\{\widehat{C}_t^n - C_t\} \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_C \int_0^t \rho_s^4 \,\mathrm{d}s\Big), \tag{3.8}$$

where

$$\mathrm{Var}_C = \mathrm{Var}_C(R, a, b, c, H) = u^T \mathcal{C}^H u, \tag{3.9}$$

$$u = \Big(e_1 - \frac{c}{\langle c, \Gamma^H\rangle} + \frac{\langle c, \partial_H \Gamma^H\rangle}{\langle c, \Gamma^H\rangle} \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle}(a - \varphi(H)b)\Big)\Big(1 - \frac{c_0}{\langle c, \Gamma^H\rangle}\Big)^{-1},$$

and $\partial_H \Gamma^H = (\partial_H \Gamma_0^H, \ldots, \partial_H \Gamma_R^H)$ with $\partial_H \Gamma_0^H = 0$ and

$$\partial_H \Gamma_r^H = \log(r+1)(r+1)^{2H} - 2\log(r) r^{2H} + \log(r-1)(r-1)^{2H}, \quad r \geq 1. \tag{3.10}$$

(iii) The estimator

$$\widehat{\Pi}_t^n = \Delta_n^{1-2\widetilde{H}^n} \frac{\langle a, \widehat{V}_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \tag{3.11}$$

satisfies

$$\frac{\Delta_n^{-\frac{1}{2}}}{|\log \Delta_n|}(\widehat{\Pi}_t^n - \Pi_t) \xrightarrow{\text{st}} \mathcal{N}\Big(0, 4\,\mathrm{Var}_{H,0} \int_0^t \rho_s^4 \,\mathrm{d}s\Big). \tag{3.12}$$

Remark 3.3. To construct the estimator $\widehat{C}_t^n$, we allow for the possibility of choosing a new weight vector c. Therefore, a and b should be thought of as weights that one can choose to, for example, minimize $\mathrm{Var}_{H,0}(R,a,b,H)$, while c can then be chosen to minimize $\mathrm{Var}_C(R,a,b,c,H)$. Alternatively, one may decide to choose a, b and c to minimize $\mathrm{Var}_C(R,a,b,c,H)$ directly (if $H > \frac{1}{4}$).

Remark 3.4. According to work in progress by F. Mies (private communication), the rates of $\widetilde{H}^n$, $\widehat{C}_t^n$ and $\widehat{\Pi}_t^n$, as estimators of H, C_t and Π_t, respectively, are optimal in the parametric setting of an mfBM.

In order to obtain feasible CLTs, we replace the unknown quantities in $\mathrm{Var}_{H,0}$ and Var_C by consistent estimators thereof. To this end, consider $f(x) = x^4$ and

$$Q_t^n = V_f^n(Y,t) = \Delta_n \sum_{i=1}^{[t/\Delta_n]} \left| \frac{\Delta_i^n Y}{\Delta_n^H} \right|^4, \qquad \widehat{Q}_t^n = \sum_{i=1}^{[t/\Delta_n]} (\Delta_i^n Y)^4. \tag{3.13}$$

By Theorem 2.1, we have the LLN

$$Q_t^n \stackrel{L^1}{\Longrightarrow} 3 \int_0^t \rho_s^4 \, \mathrm{d}s. \tag{3.14}$$

Therefore, the following theorem is a direct consequence of Theorem 3.2 and well-known properties of stable convergence in law (Jacod & Protter 2012, Equation (2.2.5)).

Theorem 3.5. *Grant the assumptions of Theorem 3.2. For (3.16) below, further assume that $H \in (\frac{1}{4}, \frac{1}{2})$. Then*

$$\Delta_n^{-\frac{1}{2}} (\widetilde{H}^n - H) \sqrt{\frac{3\Delta_n (\widehat{V}_{0,t}^n)^2}{\mathrm{Var}_{H,0}(R,a,b,\widetilde{H}^n)\widehat{Q}_t^n}} \stackrel{\text{st}}{\longrightarrow} \mathcal{N}(0,1), \tag{3.15}$$

$$\Delta_n^{\frac{1}{2}-2\widetilde{H}^n} (\widehat{C}_t^n - C_t) \sqrt{\frac{3\Delta_n^{4\widetilde{H}^n-1}}{\mathrm{Var}_C(R,a,b,c,\widetilde{H}^n)\widehat{Q}_t^n}} \stackrel{\text{st}}{\longrightarrow} \mathcal{N}(0,1), \tag{3.16}$$

$$\frac{\Delta_n^{-\frac{1}{2}}}{|\log \Delta_n|} (\widehat{\Pi}_t^n - \Pi_t) \sqrt{\frac{3\Delta_n^{4\widetilde{H}^n-1}}{4\,\mathrm{Var}_{H,0}(R,a,b,\widetilde{H}^n)\widehat{Q}_t^n}} \stackrel{\text{st}}{\longrightarrow} \mathcal{N}(0,1). \tag{3.17}$$

3.2 Estimation with quadratic variation if $H \in (\frac{1}{4}, \frac{1}{2})$

The estimators based on weight vectors a and b with $a_0 = b_0 = 0$ were easy to construct but suffer from a serious shortcoming: If the observed price process is simply given by $Y = \sigma B$ for some constant $\sigma > 0$ (i.e., there is no noise), then, by standard CLTs for Brownian motion, the ratio $\langle a, \widehat{V}_t^n \rangle / \langle b, \widehat{V}_t^n \rangle$ converges stably in law to the ratio Z_1/Z_2 of two centered (possibly correlated) normals that are independent of B. In particular, because Z_1/Z_2 has a density supported on $\mathbb{R}$, the asymptotic probability that $\widetilde{H}^n$ from (3.4) falls into any non-empty open subinterval of $(0,1)$ is non-zero. So based on $\widetilde{H}^n$ only, it is impossible to tell whether there is evidence for rough noise or whether an estimate produced by $\widetilde{H}^n$ is simply the result of chance! This shortcoming is shared by the estimator proposed by Dozzi et al. (2015).

To solve this problem, we have to include lag 0 in our estimation of H. If $H \in (\frac{1}{4}, \frac{1}{2})$, this significantly complicates the estimation procedure: By the discussion at the beginning of Section 3, in order to estimate C_t, we need to estimate H first. At the same time, as Corollary 3.1 shows, using $\widehat{V}^n_{0,t}$ to estimate H induces an asymptotic bias term coming from the $\int_0^t \sigma_s^2\,\mathrm{d}s$ term, which can only be corrected with an estimator of C_t. Resolving this circular dependence necessitates a complex iterated estimation procedure for H and C_t that we describe in Appendix F. In particular, as $H \uparrow \frac{1}{2}$, we obtain an increasing number of higher-order bias terms as a result of the interdependence between the H- and the C_t-estimators. The final result we obtain after the debiasing procedure described in Appendix F is as follows (for the proof, combine (3.14), Proposition F.5 and Theorem F.6):

Theorem 3.6. *Assume that $H \in (\frac{1}{4}, \frac{1}{2})$. Choose $R \geq 1$, $m \geq 2$ and $a, b, c \in \mathbb{R}^{1+R}$ such that $b_0 = 0$ and φ from (3.4) is invertible (now, a_0 need not be 0 anymore). Further choose $a^0, b^0 \in \mathbb{R}^{1+R}$ such that $a^0_0 = b^0_0 = 0$. The estimators $\widehat{H}^n$, $\widehat{C}^n_t$ and $\widehat{\Pi}^n_t$, defined in (F.20), (F.22) and (F.23), respectively, satisfy*

$$\Delta_n^{-\frac{1}{2}}(\widehat{H}^n - H)\sqrt{\frac{3\Delta_n(\widehat{V}^n_{0,t})^2}{\mathrm{Var}_H(R, a, b, \widehat{H}^n)\widehat{Q}^n_t}} \xrightarrow{\text{st}} \mathcal{N}(0,1), \tag{3.18}$$

$$\Delta_n^{\frac{1}{2}-2\widehat{H}^n}(\widehat{C}^n_t - C_t)\sqrt{\frac{3\Delta_n^{4\widehat{H}^n-1}}{\mathrm{Var}_C(R, a, b, c, \widehat{H}^n)\widehat{Q}^n_t}} \xrightarrow{\text{st}} \mathcal{N}(0,1), \tag{3.19}$$

$$\frac{\Delta_n^{-\frac{1}{2}}}{|\log \Delta_n|}(\widehat{\Pi}^n_t - \Pi_t)\sqrt{\frac{3\Delta_n^{4\widehat{H}^n-1}}{4\,\mathrm{Var}_H(R, a, b, \widehat{H}^n)\widehat{Q}^n_t}} \xrightarrow{\text{st}} \mathcal{N}(0,1), \tag{3.20}$$

where $\widehat{Q}^n_t$ is defined in (3.13) and the functions Var_H and Var_C are defined in (F.21) and (F.26), respectively.

4 Simulation study

All results reported in this section are based on 5,000 simulations from the mfBM

$$Y_t = X_t + Z_t = \sigma B_t + \rho B^H_t, \qquad t \in [0, T],$$

where $\sigma = 0.01$, $\rho = 0.001$, B and B^H are independent and $T = 1$ or $T = 20$ trading days, each consisting of 6.5 hours or $n = 23{,}400$ seconds. Accordingly, we choose $\Delta_n = 1/n = 1/23{,}400$. The values of H will be taken from the set

$$H \in \{0.05, 0.1, 0.15, 0.2, 0.25, 0.275, 0.3, 0.325, 0.35, 0.4, 0.45\}. \tag{4.1}$$

We additionally consider the cases "$H = 0.5$" (i.e., $\rho = 0$) and "$H = 0$" (i.e., $(B^0_t)_{t\in[0,T]}$ is a collection of independent standard normal noise variables). The choice of the tuning parameters is described in Section G in the supplement.

4.1 Performance of estimators

For H, we first compare our estimator $\widetilde{H}^{n,0} = \widetilde{H}^n$ from (3.4), constructed with a^0 and b^0 from (G.2), with four variants of $\widehat{H}^n$ from (F.20), denoted by $\widehat{H}^{n,i}$ for $i = 0, 1, 2, 3$. For each i, $\widehat{H}^{n,i}$ is defined in the same way as $\widehat{H}^n$ in (F.20) except that $N(\widetilde{H}^n)$ in (F.14) and

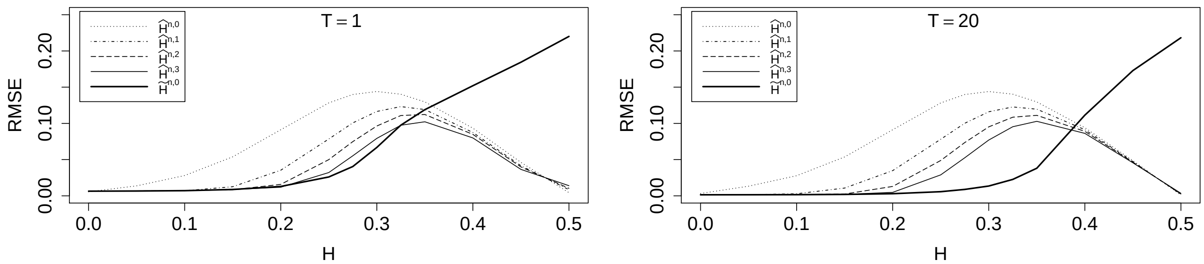


Figure 3: RMSE of $\widehat{H}^{n,i}$, $i = 0, 1, 2, 3$ and $\widetilde{H}^{n,0}$ for $T = 1$ (top) and $T = 20$ (bottom).

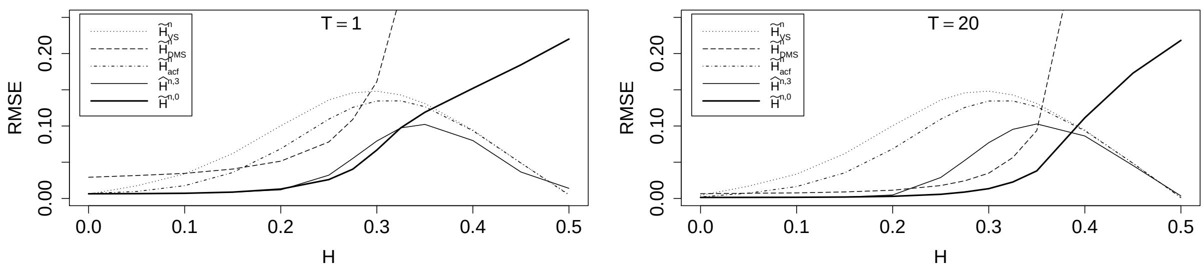


Figure 4: RMSE of $\widehat{H}^{n,3}$, $\widetilde{H}^{n,0}$, $\widetilde{H}^{n}_{\mathrm{VS}}$, $\widetilde{H}^{n}_{\mathrm{DMS}}$ and $\widetilde{H}^{n}_{\mathrm{acf}}$ for $T = 1$ (top) and $T = 20$ (bottom).

(F.15) and $N(\widehat{H}^{n}_{k-1})$ in (F.17) are replaced by the fixed number i. In particular, if n is large, then with high probability,

$$\widehat{H}^n = \begin{cases} \widehat{H}^{n,0} & \text{if } H \in (0, 0.25), \\ \widehat{H}^{n,1} & \text{if } H \in (0.25, 0.375), \end{cases} \qquad \widehat{H}^n = \begin{cases} \widehat{H}^{n,2} & \text{if } H \in (0.375, 0.417), \\ \widehat{H}^{n,3} & \text{if } H \in (0.417, 0.4375). \end{cases} \tag{4.2}$$

We do not include four or more correction terms as it becomes increasingly intractable to compute higher-order derivatives of composite functions like φ^{-1} and ψ in (F.3) or (F.10).

As Figure 3 shows, $\widehat{H}^{n,3}$ has a lower root-mean-square error (RMSE) than $\widehat{H}^{n,i}$ for any $i = 0, 1, 2$, although, according to the theory in Section 3.2, it suffices for asymptotic normality to only include i corrections and consider $\widehat{H}^{n,i}$ in the ranges of H specified in (4.2). Moreover, if $T = 1$ (resp., $T = 20$), $\widetilde{H}^{n,0}$ is superior to $\widehat{H}^{n,3}$ in terms of RMSE if $H \leq 0.3$ (resp., $H \leq 0.35$) and inferior to $\widehat{H}^{n,3}$ if $H \geq 0.325$ (resp., $H \geq 0.4$). This is in line with our previous observation that $\widetilde{H}^{n,0}$ fails to estimate H if $H = \frac{1}{2}$. Also, taking $T = 20$ instead of $T = 1$ significantly reduces the RMSE of $\widetilde{H}^{n,0}$ for $H \leq 0.35$, but the RMSE of $\widehat{H}^{n,i}$ is largely unaffected.

In Figure 4, we further compare $\widehat{H}^{n,3}$ and $\widetilde{H}^{n,0}$ with

- the estimator $\widetilde{H}^{n}_{\mathrm{VS}} = \frac{1}{2}(\widetilde{\beta}^{n}_{\mathrm{VS}} + 1)$ based on volatility signature plots, where $\widetilde{\beta}^{n}_{\mathrm{VS}}$ is the slope estimate in a linear regression of $\log \widehat{V}^{n/i}_{0,t}$ on $\log i$ for $i = 1, \dots, 10$;
- the estimator $\widetilde{H}^{n}_{\mathrm{DMS}} = \frac{1}{2}(1 + \log_{2+}[(\widehat{V}^{n/4}_{0,t} - \widehat{V}^{n/2}_{0,t})/(\widehat{V}^{n/2}_{0,t} - \widehat{V}^{n}_{0,t})])$ from Dozzi et al. (2015), where $\log_{2+} x = \log_2 x$ if $x > 0$ and $\log_{2+} x = 0$ otherwise;
- the autocorrelation estimator $\widetilde{H}^{n}_{\mathrm{acf}}$ from (3.3).

For small values of H (i.e., $H \leq 0.3$ if $T = 1$ and $H \leq 0.35$ if $T = 20$), the best estimator is $\widetilde{H}^{n,0}$. For large values of H (i.e., $H \geq 0.325$ if $T = 1$ and $H \geq 0.4$ if $T = 20$), the best

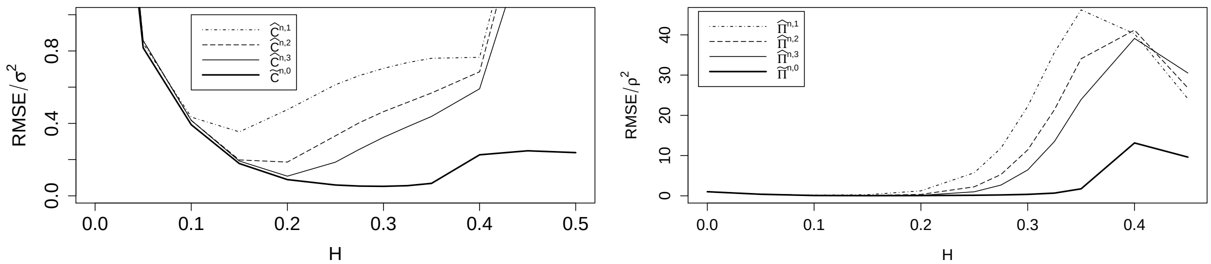


Figure 5: Top: RMSE of $\widehat{C}^{n,1}$, $\widehat{C}^{n,2}$, $\widehat{C}^{n,3}$ and $\widetilde{C}^{n,0}$. Bottom: RMSE of $\widehat{\Pi}^{n,1}$, $\widehat{\Pi}^{n,2}$, $\widehat{\Pi}^{n,3}$ and $\widetilde{\Pi}^{n,0}$. Negative volatility estimates were replaced by 0 in the evaluation.

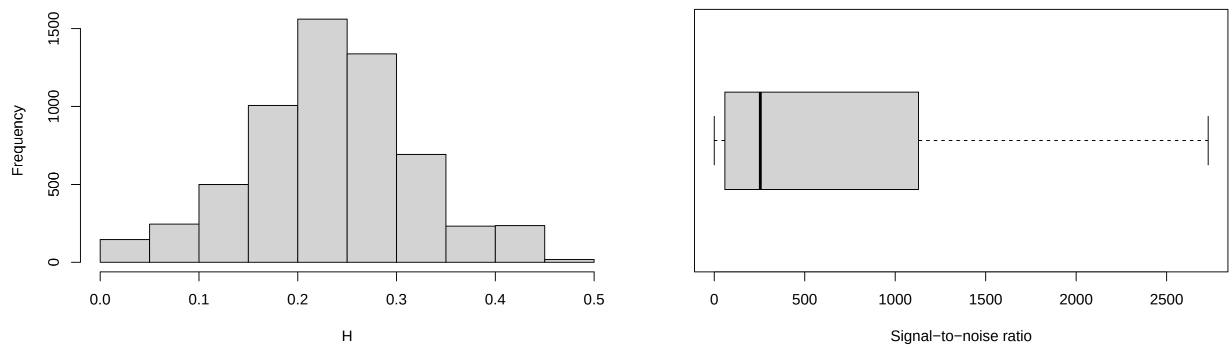


Figure 6: Histogram of estimates for H and boxplot of signal–to–noise ratios. Each data point corresponds to one company and day. Signal–to–noise ratios where the volatility or the noise volatility estimate is negative are omitted. Outliers in the boxplot are not shown.

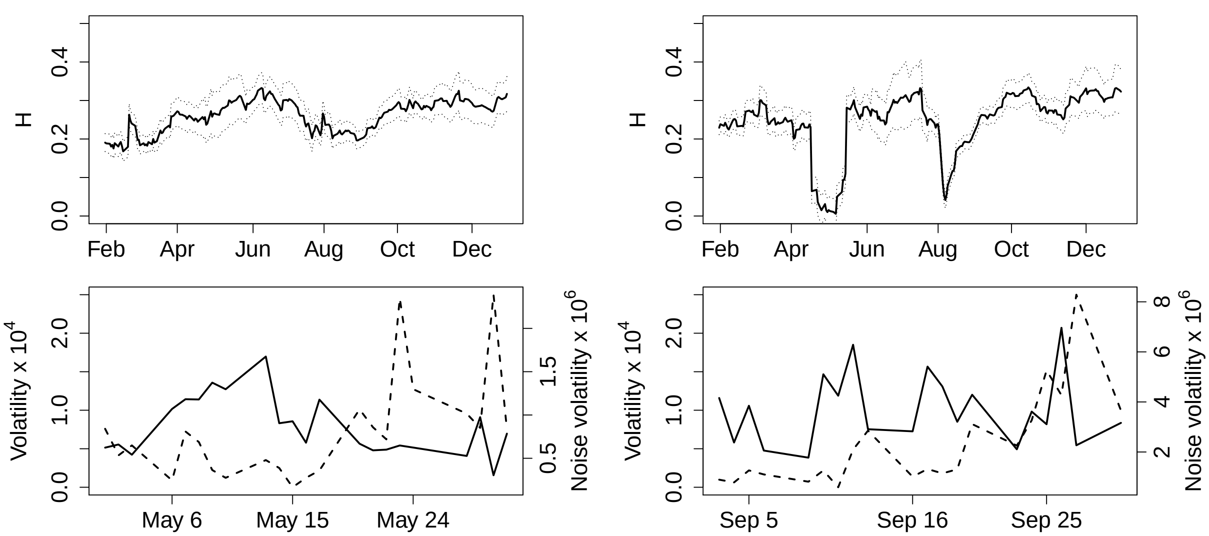
estimator is $\widehat{H}^{n,3}$ (except at $H = 0.5$, where the performance of $\widehat{H}^{n,3}$ is marginally worse than some of the other estimators). Therefore, the best strategy is to combine $\widetilde{H}^{n,0}$ and $\widehat{H}^{n,3}$ by using the former if H is small and the latter if H is large. We refer to Section 5 for one way of implementing this strategy.

Finally, we study the performance of our volatility estimators. To this end, we implement $\widetilde{C}^{n,0} = \widehat{C}^n_{20} - \widehat{C}^n_{19}$ and $\widetilde{\Pi}^{n,0} = \widehat{\Pi}^n_{20} - \widehat{\Pi}^n_{19}$ from (3.7) and (3.11) on the last out of 20 simulated trading days, using the estimator $\widetilde{H}^{n,0} = \widetilde{H}^n$ from (3.4) that is based on the whole simulated period. Similarly, for $i = 1, 2, 3$, we consider $\widehat{C}^{n,i} = \widehat{C}^n_{20} - \widehat{C}^n_{19}$ and $\widehat{\Pi}^{n,i} = \widehat{\Pi}^n_{20} - \widehat{\Pi}^n_{19}$ from (F.22) and (F.23) using, instead of $\widehat{H}^n$, the estimator $\widehat{H}^{n,i}$ from above (computed again based on the whole period of 20 simulated days). From Figure 5, we find that $\widetilde{C}^{n,0}$ shows a good performance for all $H \geq 0.15$, which covers the whole interval on which H is identifiable according to Proposition 1.4. Our best estimator for the integrated noise volatility is $\widetilde{\Pi}^{n,0}$, which works well if $H \leq 0.3$ but exhibits a large RMSE as H gets closer to $\frac{1}{2}$. In Section 6, we will comment on possible ways of improving this estimator for large H.

5 Empirical analysis

We apply the estimators from Theorems 3.2 and 3.6 to (logarithmic) mid-quote data for each of the 29 stocks that were constituents of the DJIA index for the whole year of 2019. The data source is the TAQ database. For each trading day in 2019, we collect all quotes on the NYSE and NASDAQ from 9:00 am until 4:00 pm Eastern Time and preprocess them using the `quotesCleanup()` function from the R package `highfrequency`. We sample in calendar time every second.

To reduce the variability of the resulting estimates, we calculate, for each trading day from January 31 to December 31, the estimators $\widehat{H}^{n,3}$ and $\widetilde{H}^{n,0}$ based on the previous 20 trading days. Afterwards, based on the insights from the simulation study, we calculate an estimate of H using $\widehat{H}^{n,3}$ if its asymptotic 95%-confidence interval contains 0.5 or is a subset of $(0.4, 0.5)$; otherwise, we report the estimate produced by $\widetilde{H}^{n,0}$. Correspondingly, we either take $\widehat{C}^{n,3}$ or $\widetilde{C}^{n,0}$ (resp., $\widehat{\Pi}^{n,3}$ or $\widetilde{\Pi}^{n,0}$) to estimate the daily integrated volatility (resp., noise volatility). Figure 6 shows the empirical distribution of the daily estimators of H and a boxplot of the daily signal–to–noise ratios (i.e., of $\widehat{C}^{n,3}/\widehat{\Pi}^{n,3}$ or $\widetilde{C}^{n,0}/\widetilde{\Pi}^{n,0}$).



(a) AXP 2019 (b) IBM 2019

Figure 7: Top row: Estimates of H with asymptotic 95%-confidence intervals. Bottom row: Volatility (solid line) and noise volatility (dashed line) estimates for May 2019 (AXP) and September 2019 (IBM), respectively.

The top row of Figure 7 shows the daily H-estimates for two individual stocks, American Express (AXP) and IBM, including 95%-confidence intervals. In the bottom row, we show the daily volatility and noise volatility estimates for the month of May (AXP) and September (IBM), respectively (where the H-estimates are all above 0.25).

6 Conclusion and future directions

Volatility estimation based on high-frequency return observations is often impeded by the presence of market microstructure noise. In this paper, we show that under natural compatibility conditions on the noise, the observed price process typically follows what we call a mixed semimartingale, that is, the sum of a continuous Itô semimartingale and a rougher component that locally resembles fractional Brownian motion. Given that traditional noise-robust volatility estimators are no longer consistent in the mixed semimartingale setting (see Remark 1.5), we combine central limit theorems for variation functionals and an iterative debiasing procedure to construct consistent and asymptotically normal estimators for the roughness parameter H of the noise and the integrated price and noise volatilities,whenever these quantities are identifiable.

In a simulation study, we find that our estimators of H outperform existing ones in the literature for all values of H. We further identify an estimator for the integrated price volatility C_T that shows good performance throughout the region of H in which C_T is identifiable. In light of Remark 1.5, an interesting open problem is to investigate whether subsampling or pre-averaging techniques can improve our estimators for the noise volatility, which currently work well only when H is not close to $\frac{1}{2}$. Because C_T is not identifiable when $H < \frac{1}{4}$, another promising direction is to analyze whether taking a simultaneous small noise limit helps identify the price volatility in such cases.

Applying our estimators to 2019 quote data of DJIA stocks, we find strong empirical evidence for asset- and time-dependent values of H. Besides desirable properties such as serial dependence of increments and stochastic volatility for both price and noise, mixed semimartingales constitute non-shrinking noise models that can explain the rich variety of the scaling exponents found in volatility signature plots (see Figure 1).

In this first paper, we do not examine the effect of jumps (Aït-Sahalia & Jacod 2009*a*,*b*, Jacod & Todorov 2014, 2018) or irregular observation times (Barndorff-Nielsen & Shephard 2005, Chen et al. 2020, Hayashi & Yoshida 2008, Jacod et al. 2017) on our estimators. While the estimators of H and noise volatility from Theorem 3.5 might not be affected by jumps too much, as they do not use quadratic variation, certainly the volatility estimators and all estimators from Theorem 3.6 are. We leave it to future research to develop estimators that are fully robust to jumps and asynchronous sampling. Similarly, the current mixed semimartingale model does not capture rounding effects in observed prices (Aït-Sahalia & Jacod 2014, Jacod 1996, Robert & Rosenbaum 2010, 2012). It remains open how to incorporate rounding in such a way that the scaling exponents exhibited in Figure 1 are preserved. Pure and mixed rounding, in the way they are currently considered in the literature, do not give rise to such a rich distribution of exponents.

References

Aït-Sahalia, Y. & Jacod, J. (2009*a*), 'Estimating the degree of activity of jumps in high frequency data', *The Annals of Statistics*, **37**, 2202–2244.

Aït-Sahalia, Y. & Jacod, J. (2009*b*), 'Testing for jumps in a discretely observed process', *The Annals of Statistics*, **37**, 184–222.

Aït-Sahalia, Y. & Jacod, J. (2014), *High-Frequency Financial Econometrics*, Princeton University Press, Princeton.

Aït-Sahalia, Y., Mykland, P. A. & Zhang, L. (2011), 'Ultra high frequency volatility estimation with dependent microstructure noise', *Journal of Econometrics*, **160**, 160–175.

Aït-Sahalia, Y. & Xiu, D. (2019), 'A Hausman test for the presence of market microstructure noise in high frequency data', *Journal of Econometrics*, **211**, 176–205.

Andersen, T., Bollerslev, T., Diebold, F. & Labys, P. (2000), 'Great realizations', *Risk*, **13**, 105–108.

Andersen, T. G. & Bollerslev, T. (1998), 'Answering the skeptics: Yes, standard volatility models do provide accurate forecasts', *International Economic Review*, **39**, 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Labys, P. (2003), 'Modeling and forecasting realized volatility', *Econometrica*, **71**, 579–625.

Bandi, F. M. & Russell, J. R. (2006), 'Separating microstructure noise from volatility', *Journal of Financial Economics*, **79**, 655–692.

Barndorff-Nielsen, O., Corcuera, J. & Podolskij, M. (2011), 'Multipower variation for Brownian semistationary processes', *Bernoulli*, **17**, 1159–1194.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. & Shephard, N. (2008), 'Designing realized kernels to measure the ex post variation of equity prices in the presence of noise', *Econometrica*, **76**, 1481–1536.

Barndorff-Nielsen, O. E. & Shephard, N. (2002), 'Econometric analysis of realized volatility and its use in estimating stochastic volatility models', *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, **64**, 253–280.

Barndorff-Nielsen, O. E. & Shephard, N. (2005), 'Power variation and time change', *Teoriya Veroyatnosteĭ i ee Primeneniya*, **50**, 115–130.

Bayraktar, E., Poor, H. V. & Sircar, K. R. (2004), 'Estimating the fractal dimension of the S&P 500 index using wavelet analysis', *International Journal of Theoretical and Applied Finance*, **7**, 615–643.

Bender, C., Sottinen, T. & Valkeila, E. (2011), Fractional processes as models in stochastic finance, *in* 'Advanced mathematical methods for finance', Springer, Heidelberg, pp. 75–103.

Bianchi, S. & Pianese, A. (2018), 'Time-varying Hurst-Hölder exponents and the dynamics of (in)efficiency in stock markets', *Chaos, Solitons & Fractals*, **109**, 64–75.

Black, F. (1986), 'Noise', *The Journal of Finance*, **41**, 528–543.

Brouste, A. & Fukasawa, M. (2018), 'Local asymptotic normality property for fractional Gaussian noise under high-frequency observations', *The Annals of Statistics*, **46**, 2045–2061.

Chen, D., Mykland, P. A. & Zhang, L. (2020), 'The five trolls under the bridge: principal component analysis with asynchronous and noisy high frequency data', *Journal of the American Statistical Association*, **115**, 1960–1977.

Cheridito, P. (2001), 'Mixed fractional Brownian motion', *Bernoulli*, **7**, 913–934.

Cherny, A. (2008), 'Brownian moving averages have conditional full support', *The Annals of Applied Probability*, **18**, 1825–1830.

Corcuera, J. M., Hedevang, E., Pakkanen, M. S. & Podolskij, M. (2013), 'Asymptotic theory for Brownian semi-stationary processes with application to turbulence', *Stochastic Processes and their Applications*, **123**, 2552–2574.

Da, R. & Xiu, D. (2021), 'When moving-average models meet high-frequency data: Uniform inference on volatility', *Econometrica*. forthcoming.

Delbaen, F. & Schachermayer, W. (1994), 'A general version of the fundamental theorem of asset pricing', *Mathematische Annalen*, **300**, 463–520.

Doob, J. L. (1953), *Stochastic Processes*, John Wiley & Sons, Inc., New York; Chapman & Hall, Limited, London.

Dozzi, M., Mishura, Y. & Shevchenko, G. (2015), 'Asymptotic behavior of mixed power variations and statistical estimation in mixed models', *Statistical Inference for Stochastic Processes*, **18**, 151–175.

El Euch, O. & Rosenbaum, M. (2019), 'The characteristic function of rough Heston models', *Mathematical Finance*, **29**, 3–38.

Gatheral, J., Jaisson, T. & Rosenbaum, M. (2018), 'Volatility is rough', *Quantitative Finance*, **18**, 933–949.

Guasoni, P., Rásonyi, M. & Schachermayer, W. (2008), 'Consistent price systems and face-lifting pricing under transaction costs', *The Annals of Applied Probability*, **18**, 491–520.

Hansen, P. R. & Lunde, A. (2006), 'Realized variance and market microstructure noise', *Journal of Business & Economic Statistics*, **24**, 127–218.

Hautsch, N. & Podolskij, M. (2013), 'Preaveraging-based estimation of quadratic variation in the presence of noise and jumps: theory, implementation, and empirical evidence', *Journal of Business & Economic Statistics*, **31**, 165–183.

Hayashi, T. & Yoshida, N. (2008), 'Asymptotic normality of a covariance estimator for non-synchronously observed diffusion processes', *Annals of the Institute of Statistical Mathematics*, **60**, 367–406.

Jacod, J. (1996), 'La variation quadratique du brownien en présence d'erreurs d'arrondi', *Astérisque*, **236**, 155–161.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M. & Vetter, M. (2009), 'Microstructure noise in the continuous case: the pre-averaging approach', *Stochastic Processes and their Applications*, **119**, 2249–2276.

Jacod, J., Li, Y. & Zheng, X. (2017), 'Statistical properties of microstructure noise', *Econometrica*, **85**, 1133–1174.

Jacod, J. & Protter, P. (2012), *Discretization of Processes*, Vol. 67 of *Stochastic Modelling and Applied Probability*, Springer, Heidelberg.

Jacod, J. & Todorov, V. (2014), ‘Efficient estimation of integrated volatility in presence of infinite variation jumps’, *The Annals of Statistics*, **42**, 1029–1069.

Jacod, J. & Todorov, V. (2018), ‘Limit theorems for integrated local empirical characteristic exponents from noisy high-frequency data with application to volatility and jump activity estimation’, *The Annals of Applied Probability*, **28**, 511–576.

Jarrow, R. A., Protter, P. & Sayit, H. (2009), ‘No arbitrage without semimartingales’, *The Annals of Applied Probability*, **19**, 596–616.

Kalnina, I. & Linton, O. (2008), ‘Estimating quadratic variation consistently in the presence of endogenous and diurnal measurement error’, *Journal of Econometrics*, **147**, 47–59.

Li, Y. & Mykland, P. A. (2007), ‘Are volatility estimators robust with respect to modeling assumptions?’, *Bernoulli*, **13**, 601–622.

Li, Z. M., Laeven, R. J. A. & Vellekoop, M. H. (2020), ‘Dependent microstructure noise and integrated volatility estimation from high-frequency data’, *Journal of Econometrics*, **215**, 536–558.

Li, Z. M. & Linton, O. (2021), ‘A ReMeDI for microstructure noise’, *Econometrica*. forthcoming.

Mandelbrot, B. B. (1997), *Fractals and Scaling in Finance*, Selected Works of Benoit B. Mandelbrot, Springer-Verlag, New York.

Podolskij, M. & Vetter, M. (2009), ‘Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps’, *Bernoulli*, **15**, 634–658.

Robert, C. Y. & Rosenbaum, M. (2010), ‘A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones’, *Journal of Financial Econometrics*, **9**, 344–366.

Robert, C. Y. & Rosenbaum, M. (2012), ‘Volatility and covariation estimation when microstructure noise and trading times are endogenous’, *Mathematical Finance*, **22**, 133–164.

Rosenbaum, M. (2009), ‘Integrated volatility and round-off error’, *Bernoulli*, **15**, 687–720.

Rosenbaum, M. (2011), ‘A new microstructure noise index’, *Quantitative Finance*, **11**, 883–899.

van Zanten, H. (2007), ‘When is a linear combination of independent fBm’s equivalent to a single fBm?’, *Stochastic Processes and their Applications*, **117**, 57–70.

Zhang, L. (2006), ‘Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach’, *Bernoulli*, **12**, 1019–1043.

Zhang, L., Mykland, P. A. & Aït-Sahalia, Y. (2005), ‘A tale of two time scales: determining integrated volatility with noisy high-frequency data’, *Journal of the American Statistical Association*, **100**, 1394–1411.

Supplement to "Mixed semimartingales: Volatility estimation in the presence of rough noise"

Carsten Chong
Department of Statistics, Columbia University
and
Thomas Delerue
Department of Mathematics, Technical University of Munich
and
Guoying Li
Department of Statistics, Columbia University

Abstract

This supplement contains the proof of Theorem 2.1 (Appendices A–D), the proof of Theorem 3.2 (Appendix E), the details of the iterative debiasing procedure of Section 3.2 (Appendix F) and our choice of tuning parameters for the simulation study (Appendix G).

A Size estimates

We use the notation from the main paper. In addition, we write $A \lesssim B$ if there is a constant C that is independent of any quantity of interest such that $A \leq CB$. In the following, we repeatedly make use of so-called *standard size estimates* (cf. Chong (2020*c*), Appendix D). Under the strengthened hypotheses of Assumption (CLT'), consider for fixed $j, k \in \{1, \dots, d\}$ and $\ell \in \{1, \dots, L\}$ an expression like

$$\begin{aligned} S_n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]} h(\zeta_i^n) \Bigg(\frac{\Delta_{i+\ell-1}^n A^k}{\Delta_n^H} + \frac{1}{\Delta_n^H} \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \left(\sigma_s^{kj} - \sigma_{(i-\theta_n'')\Delta_n}^{kj} \right) \mathrm{d}B_s^j \\ + \int_0^\infty \frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H} \left(\rho_s^{kj} - \rho_{(i-\theta_n)\Delta_n}^{kj} \right) \mathbb{1}_{((i-\theta_n)\Delta_n, (i-\theta_n')\Delta_n)}(s) \, \mathrm{d}W_s^j \Bigg), \end{aligned} \tag{A.1}$$

where $\theta_n = [\Delta_n^{-\theta}]$, $\theta_n' = [\Delta_n^{-\theta'}]$, $\theta_n'' = [\Delta_n^{-\theta''}]$ and $-\infty \leq \theta', \theta'' < \theta \leq \infty$. In addition, h is a function such that $|h(x)| \lesssim 1 + \|x\|^p$ for some $p > 1$, and ζ_i^n are random variables with

$$\sup_{n \in \mathbb{N}} \sup_{i=1,\dots,[T/\Delta_n]} \mathbb{E}[\|\zeta_i^n\|^p] < \infty.$$

For any $q \geq 1$, because a is uniformly bounded by Assumption (CLT'), Minkowski's integral inequality yields

$$\mathbb{E}\left[\left\| \frac{\Delta_{i+\ell-1}^n A}{\Delta_n^H} \right\|^q \right]^{\frac{1}{q}} \leq \frac{1}{\Delta_n^H} \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \mathbb{E}[\|a_s\|^q]^{\frac{1}{q}} \, \mathrm{d}s \lesssim \Delta_n^{1-H}. \tag{A.2}$$

Similarly, by the Burkholder–Davis–Gundy (BDG) inequality and Assumption (CLT'),

$$\mathbb{E}\left[\left| \frac{1}{\Delta_n^H} \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \left(\sigma_s^{kj} - \sigma_{(i-\theta_n'')\Delta_n}^{kj} \right) \mathrm{d}B_s^j \right|^q \right]^{\frac{1}{q}} \lesssim (\theta_n'' \Delta_n)^{\frac{1}{2}} \Delta_n^{\frac{1}{2}-H}. \tag{A.3}$$

Combining Assumption (CLT') with Lemma B.1, we deduce that

$$\begin{aligned} &\mathbb{E}\left[\left| \int_0^\infty \frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H} \left(\rho_s^{kj} - \rho_{(i-\theta_n)\Delta_n}^{kj} \right) \mathbb{1}_{((i-\theta_n)\Delta_n, (i-\theta_n')\Delta_n)}(s) \, \mathrm{d}W_s^j \right|^q \right]^{\frac{1}{q}} \\ &\quad \lesssim (\theta_n \Delta_n)^{\frac{1}{2}} \left(\frac{1}{\Delta_n^{2H}} \int_0^{(i-\theta_n')\Delta_n} \Delta_{i+\ell-1}^n g(s)^2 \, \mathrm{d}s \right)^{\frac{1}{2}} \lesssim (\theta_n \Delta_n)^{\frac{1}{2}} \Delta_n^{\theta'(1-H)}. \end{aligned} \tag{A.4}$$

Finally, using Hölder's inequality to separate $h(\zeta_i^n)$ from the subsequent expression in (A.1), we have shown that

$$\begin{aligned} \mathbb{E}\left[\sup_{t \leq T} |S_n(t)| \right] &\lesssim \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[T/\Delta_n]} \left\{ \Delta_n^{1-H} + \Delta_n^{1-H} (\theta_n'')^{\frac{1}{2}} + (\theta_n \Delta_n)^{\frac{1}{2}} \Delta_n^{\theta'(1-H)} \right\} \\ &\lesssim \Delta_n^{\frac{1}{2}-H} + \Delta_n^{\frac{1}{2}-H-\frac{\theta''}{2}} + \Delta_n^{\theta'(1-H)-\theta}. \end{aligned} \tag{A.5}$$

The upshot of this example is that the absolute moments of sums and products of more or less complicated expressions can always be bounded term by term: for example, in (A.1), the terms

$$\sum_{i=\theta_n+1}^{[t/\Delta_n]}, \qquad h(\zeta_i^n), \qquad \Delta_{i+\ell-1}^n A^k, \qquad \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} (\cdots)\,\mathrm{d}B_s^j, \qquad \sigma_s^{kj} - \sigma_{(i-\theta''_n)\Delta_n}^{kj},$$
$$\int_0^{(i-\theta'_n)\Delta_n} \frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H}(\cdots)\,\mathrm{d}W_s^j, \qquad \rho_s^{kj} - \rho_{(i-\theta_n)\Delta_n}^{kj}$$

have *sizes* (i.e., the L^q-moments, for any q, are uniformly bounded by a constant times)

$$\Delta_n^{-1}, \qquad 1, \qquad \Delta_n, \qquad \sqrt{\Delta_n}, \qquad (\theta''_n\Delta_n)^{\frac{1}{2}}, \qquad \Delta_n^{\theta'(1-H)}, \qquad (\theta_n\Delta_n)^{\frac{1}{2}},$$

respectively. The final estimate (A.5) is then obtained by combining these bounds. Clearly, size estimates can be applied to variants of (A.1), too, for example, when the stochastic integral in (A.1) is squared, when we have products of integrals, when $S_n(t)$ is matrix-valued, etc.

Even though size estimates are optimal in general, better estimates may be available in specific cases. One such case occurs when sums have a martingale structure. To illustrate this, let $\mathcal{F}_i^n = \mathcal{F}_{i\Delta_n}$ and consider

$$S'_n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=1}^{[t/\Delta_n]-L+1} \varpi_i^n$$

with random variables ϖ_i^n that are $\mathcal{F}_i^n$-measurable and satisfy $\mathbb{E}[\varpi_i^n \mid \mathcal{F}_{i-\theta'''_n}^n] = 0$, where $\theta'''_n = [\Delta_n^{-\theta'''}]$ for some $0 < \theta''' < 1$. Suppose that $\mathbb{E}[|\varpi_i^n|^2]^{1/2} \lesssim \Delta_n^{\varpi}$ uniformly in i and n for some $\varpi > 0$. Writing

$$S'_n(t) = \sum_{j=1}^{\theta'''_n} S'_{n,j}(t), \qquad S'_{n,j}(t) = \Delta_n^{\frac{1}{2}} \sum_{k=1}^{[([t/\Delta_n]-L+1)/\theta'''_n]} \varpi_{j+(k-1)\theta'''_n}^n,$$

we observe that each $S'_{n,j}$ is a martingale in t (albeit relative to different filtrations), so the BDG inequality and the triangle inequality yield

$$\mathbb{E}\left[\sup_{t\leq T}|S'_n(t)|\right] \lesssim (\theta'''_n)^{\frac{1}{2}}\Delta_n^{\varpi}. \tag{A.6}$$

Very often, ϖ_i^n will actually only be $\mathcal{F}_{i+L-1}^n$-measurable. However, a shift by L increments will not change the value of the above estimate. Following Chong (2020*b*), Section 4, we refer to (A.6) as a *martingale size estimate*.

B Estimates for fractional kernels

Here we gather some useful results about the kernel $g(t) = K_H^{-1}t^{H-1/2}$ introduced in (1.9) (we consider the case $g_0 \equiv 0$ here).

Lemma B.1. *Recall the notations introduced in* (1.10)*,* (1.7) *and* (C.2)*.*

(i) For any $k, n \in \mathbb{N}$,

$$\int_0^\infty \Delta_k^n g(t)^2\,\mathrm{d}t = K_H^{-2}\left\{\frac{1}{2H} + \int_1^k \left(r^{H-\frac{1}{2}} - (r-1)^{H-\frac{1}{2}}\right)^2 \mathrm{d}r\right\}\Delta_n^{2H} \leq \Delta_n^{2H}. \tag{B.1}$$

(ii) *For any* $k, \ell, n \in \mathbb{N}$ *with* $k < \ell$,

$$\int_{-\infty}^{\infty} \Delta_k^n g(t) \Delta_\ell^n g(t) \, \mathrm{d}t = \Delta_n^{2H} \Gamma_{\ell-k}^H \lesssim \Delta_n^{2H} \overline{\Gamma}_{\ell-k}^H, \tag{B.2}$$

where $\overline{\Gamma}_1^H = \Gamma_1^H$ *and* $\overline{\Gamma}_r^H = (r-1)^{-2(1-H)}$ *for* $r \geq 2$.

(iii) *For any* $\theta \in (0,1)$, *setting* $\theta_n = [\Delta_n^{-\theta}]$, *we have for any* $i > \theta_n$ *and* $r \in \mathbb{N}$,

$$\int_{-\infty}^{(i-\theta_n)\Delta_n} \Delta_i^n g(s) \Delta_{i+r}^n g(s) \, \mathrm{d}s \lesssim \Delta_n^{2H} \Delta_n^{2\theta(1-H)}. \tag{B.3}$$

Proof. Let $k \leq \ell$. By direct calculation,

$$\begin{aligned}
&\int_0^\infty \Delta_k^n g(t) \Delta_\ell^n g(t) \, \mathrm{d}t \\
&\quad = \Delta_n^{2H} K_H^{-2} \int_0^k \left(r^{H-\frac{1}{2}} - (r-1)_+^{H-\frac{1}{2}} \right) \left((r + (\ell-k))^{H-\frac{1}{2}} - (r + (\ell-k) - 1)_+^{H-\frac{1}{2}} \right) \mathrm{d}r,
\end{aligned}$$

which shows (B.1) by setting $k = \ell$. Next, let $(B^H)_{t \geq 0}$ be a fractional Brownian motion with Hurst index H. Then B^H has the Mandelbrot–van Ness representation

$$B_t^H = K_H^{-1} \int_{\mathbb{R}} \left((t-s)_+^{H-\frac{1}{2}} - (-s)_+^{H-\frac{1}{2}} \right) \mathrm{d}\overline{B}_s, \quad t \geq 0,$$

where $\overline{B}$ is a two-sided standard Brownian motion. Moreover, $\Delta_i^n B^H = \int_{\mathbb{R}} \Delta_i^n g(s) \, \mathrm{d}\overline{B}_s$ for any i. Therefore, by well-known properties of fractional Brownian motion,

$$\begin{aligned}
\int_{-\infty}^{\infty} \Delta_k^n g(s) \Delta_\ell^n g(s) \, \mathrm{d}s &= \mathbb{E}[\Delta_k^n B^H \Delta_\ell^n B^H] = \mathbb{E}[B_{\Delta_n}^H B_{(\ell-k+1)\Delta_n}^H] - \mathbb{E}[B_{\Delta_n}^H B_{(\ell-k)\Delta_n}^H] \\
&= \frac{1}{2} \Big\{ \Delta_n^{2H} + ((\ell-k+1)\Delta_n)^{2H} - ((\ell-k)\Delta_n)^{2H} \\
&\qquad\qquad - \Delta_n^{2H} - ((\ell-k)\Delta_n)^{2H} + ((\ell-k-1)\Delta_n)^{2H} \Big\} \\
&= \Delta_n^{2H} \Gamma_{\ell-k}^H,
\end{aligned}$$

which is the equality in (B.2). Next, use the mean-value theorem twice on Γ_r^H in order to obtain for all $r \geq 2$,

$$\begin{aligned}
\Gamma_r^H &= \frac{1}{2} \Big(\{ (r+1)^{2H} - r^{2H} \} - \{ r^{2H} - (r-1)^{2H} \} \Big) \leq \frac{1}{2} (2H) \Big((r+1)^{2H-1} - (r-1)^{2H-1} \Big) \\
&\leq H(2H-1)(r-1)^{2H-2},
\end{aligned}$$

which shows the inequality in (B.2). Finally,

$$\begin{aligned}
&\int_{-\infty}^{(i-\theta_n)\Delta_n} \Delta_i^n g(s) \Delta_{i+r}^n g(s) \, \mathrm{d}s \\
&\quad = \Delta_n^{2H} K_H^{-2} \int_{\theta_n}^\infty \left(t^{H-\frac{1}{2}} - (t-1)^{H-\frac{1}{2}} \right) \left((t+r)^{H-\frac{1}{2}} - (t+r-1)^{H-\frac{1}{2}} \right) \mathrm{d}t \\
&\quad \lesssim \Delta_n^{2H} \int_{\theta_n}^\infty \left(t^{H-\frac{1}{2}} - (t-1)^{H-\frac{1}{2}} \right)^2 \mathrm{d}t \lesssim \Delta_n^{2H} \int_{\theta_n}^\infty (t-1)^{2H-3} \, \mathrm{d}t \lesssim \Delta_n^{2H} \Delta_n^{\theta(2-2H)},
\end{aligned}$$

which yields (B.3). ◻

C Proof of Theorem 2.1

Throughout the proof, by a standard localization argument (cf. Lemma 4.4.9 in Jacod & Protter (2012)), we may and will assume a strengthened version of Assumption (CLT):

Assumption (CLT'). *In addition to Assumption (CLT), there is $C > 0$ such that*

$$\sup_{(\omega,t)\in\Omega\times[0,\infty)} \Big\{\|a_t(\omega)\| + \|\sigma_t(\omega)\| + \|\rho_t(\omega)\| + \|\rho_t^{(0)}(\omega)\| + \|\widetilde{b}_t(\omega)\| + \|\widetilde{\rho}_t(\omega)\|\Big\} < C.$$

Moreover, for every $p > 0$, there is $C_p > 0$ such that for all $s,t > 0$,

$$\begin{aligned} &\mathbb{E}[\|\sigma_t - \sigma_s\|^p]^{\frac{1}{p}} \leq C_p|t-s|^{\frac{1}{2}}, \qquad \mathbb{E}[\|\rho_t^{(0)} - \rho_s^{(0)}\|^p]^{\frac{1}{p}} \leq C_p|t-s|^{\gamma},\\ &\mathbb{E}[\|\widetilde{\rho}_t - \widetilde{\rho}_s\|^p]^{\frac{1}{p}} \leq C_p|t-s|^{\varepsilon}. \end{aligned} \tag{C.1}$$

Proof of Theorem 2.1. Except for (C.6) below, we may and will assume that $M = 1$. Recalling the decomposition (1.9), since g_0 is smooth with $g_0(0) = 0$, we can use the stochastic Fubini theorem (see Protter (2005), Chapter IV, Theorem 65) to write

$$\int_0^t g_0(t-r)\rho_r \,\mathrm{d}W_r = \int_0^t \Big(\int_r^t g_0'(s-r)\,\mathrm{d}s\Big)\rho_r\,\mathrm{d}W_r = \int_0^t \Big(\int_0^s g_0'(s-r)\rho_r\,\mathrm{d}W_r\Big)\mathrm{d}s.$$

This is a finite variation process and can be incorporated in the drift process in (1.11). So without loss of generality, we may assume $g_0 \equiv 0$ and $g(t) = K_H^{-1}t^{H-1/2}$ in the following. Then $Y_t = A_t + M_t + Z_t$, where $A_t = \int_0^t a_s\,\mathrm{d}s$ and $M_t = \int_0^t \sigma_s\,\mathrm{d}B_s$, and we have $\underline{\Delta}_i^n Y = \underline{\Delta}_i^n A + \underline{\Delta}_i^n M + \underline{\Delta}_i^n Z$ in the notation of (2.1). Writing $g(t) = 0$ for $t \leq 0$, we also define for all $s,t \geq 0$ and $i,n \in \mathbb{N}$,

$$\begin{aligned} \Delta_i^n g(s) &= g(i\Delta_n - s) - g((i-1)\Delta_n - s),\\ \underline{\Delta}_i^n g(s) &= (\Delta_i^n g(s),\ldots,\Delta_{i+L-1}^n g(s)), \end{aligned} \tag{C.2}$$

such that, in matrix notation,

$$\underline{\Delta}_i^n Z = \Big(\int_0^\infty \Delta_i^n g(s)\rho_s\,\mathrm{d}W_s,,\ldots,\int_0^\infty \Delta_{i+L-1}^n g(s)\rho_s\,\mathrm{d}W_s\Big) = \int_0^\infty \rho_s\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s).$$

The first step in our proof is to shrink the domain of integration for each $\underline{\Delta}_i^n Z$. Let

$$\theta \in (\tfrac{1}{4(1-H)},\tfrac{1}{2}), \tag{C.3}$$

which is always possible for $H \in (0,\frac{1}{2})$, and set $\theta_n = [\Delta_n^{-\theta}]$. Further define

$$\underline{\Delta}_i^n Y^{\mathrm{tr}} = \underline{\Delta}_i^n A + \underline{\Delta}_i^n M + \xi_i^n, \qquad \xi_i^n = \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \rho_s\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s). \tag{C.4}$$

Lemma C.1. *If θ is chosen according to (C.3), then*

$$\Delta_n^{-\frac{1}{2}}\Big\{V_f^n(Y,t) - \Delta_n \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} f\Big(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\Big)\Big\} \overset{L^1}{\Longrightarrow} 0.$$

The last sum can be further decomposed into three parts:

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) = V^n(t) + U^n(t) + \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n\right], \quad \text{(C.5)}$$

where

$$V^n(t) = \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \Xi_i^n, \qquad \Xi_i^n = \Delta_n^{\frac{1}{2}}\left(f\left(\frac{\xi_i^n}{\Delta_n^H}\right) - \mathbb{E}\left[f\left(\frac{\xi_i^n}{\Delta_n^H}\right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n\right]\right),$$

$$U^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \left\{f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) - f\left(\frac{\xi_i^n}{\Delta_n^H}\right) - \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) - f\left(\frac{\xi_i^n}{\Delta_n^H}\right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n\right]\right\}.$$

Lemma C.2. *For all $H < \frac{1}{2}$, we have that $U^n \stackrel{L^1}{\Longrightarrow} 0$.*

In other words, in the limit $\Delta_n \to 0$, the impact of the semimartingale component is negligible, except for its contributions to the conditional expectations in (C.5). As we mentioned above, this is somewhat surprising: It is true that the L^2-norm of the semimartingale increment $\underline{\Delta}_i^n A + \underline{\Delta}_i^n M$, divided by Δ_n^H, converges to 0. But the rate $\Delta_n^{1/2-H}$ at which this takes place can be arbitrarily slow if H is close to $\frac{1}{2}$. So Lemma C.2 implies that there is a big gain in convergence rate if one considers the sum of the centered differences $f(\underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H) - f(\xi_i^n/\Delta_n^H)$. In the proof, we will need for the first time that f has at least $2(N(H)+1)$ continuous derivatives.

The process V^n only contains the fractional part and is responsible for the limit $\mathcal{Z}$ in (2.7). For the sake of brevity, we borrow a result from Chong (2020*a*): For each $m \in \mathbb{N}$, consider the sums

$$V^{n,m,1}(t) = \sum_{j=1}^{J^{n,m}(t)} V_j^{n,m}, \qquad V_j^{n,m} = \sum_{k=1}^{m\theta_n} \Xi^n_{(j-1)((m+1)\theta_n+L-1)+k},$$

$$V^{n,m,2}(t) = \sum_{j=1}^{J^{n,m}(t)} \sum_{k=1}^{\theta_n+L-1} \Xi^n_{(j-1)((m+1)\theta_n+L-1)+m\theta_n+k},$$

$$V^{n,m,3}(t) = \sum_{j=((m+1)\theta_n+L-1)J^{n,m}(t)+1}^{[t/\Delta_n]-L+1} \Xi_j^n,$$

where $J^{n,m}(t) = [([t/\Delta_n]-L+1)/((m+1)\theta_n+L-1)]$. We then have $V^n(t) = \sum_{i=1}^3 V^{n,m,i}(t)$. This is very similar to the decomposition on p. 1161 in Chong (2020*a*). With essentially the same proof, we infer that $V^n(t) \stackrel{\mathrm{st}}{\Longrightarrow} \mathcal{Z}$ and, hence,

$$\Delta_n^{\frac{1}{2}}\left\{\sum_{i=1}^{[t/\Delta_n]-L+1} f\left(\frac{\underline{\Delta}_i^n Y}{\Delta_n^H}\right) - \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n\right]\right\} \stackrel{\mathrm{st}}{\Longrightarrow} \mathcal{Z}, \quad \text{(C.6)}$$

where $\mathcal{Z}$ is exactly as in (2.7). Therefore, in order to complete the proof of Theorem 2.1, it remains to show that (recall $N(H) = [1/(2-4H)]$)

$$\Delta_n^{-\frac{1}{2}}\left\{\Delta_n \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n\right] - \int_0^t \mu_f(\pi(s))\,\mathrm{d}s \right.$$
$$\left. - \sum_{j=1}^{N(H)} \Delta_n^{j(1-2H)} \sum_{|\chi|=j} \frac{1}{\chi!}\int_0^t \partial^\chi \mu_f(\pi(s)) c(s)^\chi\,\mathrm{d}s\right\} \stackrel{L^1}{\Longrightarrow} 0.$$

To this end, we will discretize the volatility processes σ and ρ in $\underline{\Delta}_i^n Y^{\mathrm{tr}}$. The proof is technical (as it involves another multiscale analysis) and will be divided into further smaller steps in Appendix D.

Lemma C.3. *Assuming* (C.3)*, we have that*

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \left\{ \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H} \right) \,\Big|\, \mathcal{F}_{i-\theta_n}^n \right] - \mu_f(\Upsilon^{n,i}) \right\} \stackrel{L^1}{\Longrightarrow} 0,$$

where $\Upsilon^{n,i} \in (\mathbb{R}^{d\times L})^2$ *is defined by*

$$\begin{aligned}(\Upsilon^{n,i})_{k\ell,k'\ell'} &= c((i-1)\Delta_n)_{k\ell,k'\ell'}\,\Delta_n^{1-2H} \\ &\quad + (\rho_{(i-1)\Delta_n}\rho_{(i-1)\Delta_n}^T)_{kk'} \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \frac{\Delta_{i+\ell-1}^n g(s)\Delta_{i+\ell'-1}^n g(s)}{\Delta_n^{2H}}\,\mathrm{d}s. \end{aligned} \tag{C.7}$$

The last part of the proof consists of evaluating

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mu_f(\Upsilon^{n,i}).$$

This is the place where the asymptotic bias terms arise and which is different from the pure (semimartingale or fractional) cases. Roughly speaking, the additional terms are due to the fact that in the LLN limit (2.9), there is a contribution of magnitude $\Delta_n^{1-2H}c(s)$ coming from the semimartingale part that is negligible on first order but not at a rate of $\sqrt{\Delta_n}$. Expanding $\mu_f(\Upsilon^{n,i})$ in a Taylor sum up to order $N(H)$, we obtain

$$\begin{aligned}\mu_f(\Upsilon^{n,i}) &= \mu_f(\pi((i-1)\Delta_n)) + \sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{1}{\chi!}\partial^\chi \mu_f(\pi((i-1)\Delta_n))(\Upsilon^{n,i} - \pi((i-1)\Delta_n))^\chi \\ &\quad + \sum_{|\chi|=N(H)+1} \frac{1}{\chi!}\partial^\chi \mu_f(\upsilon_i^n)(\Upsilon^{n,i} - \pi((i-1)\Delta_n))^\chi,\end{aligned}$$

where υ_i^n is a point between $\Upsilon^{n,i}$ and $\pi((i-1)\Delta_n)$. The next lemma shows two things: first, the term of order $N(H)+1$ is negligible, and second, for $j=1,\ldots,N(H)$, we may replace $\Upsilon^{n,i} - \pi((i-1)\Delta_n)$ by $\Delta_n^{1-2H}c((i-1)\Delta_n)$.

Lemma C.4. *We have that* $\mathbb{X}_1^n \stackrel{L^1}{\Longrightarrow} 0$ *and* $\mathbb{X}_2^n \stackrel{L^1}{\Longrightarrow} 0$*, where*

$$\begin{aligned}\mathbb{X}_1^n(t) &= \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n-L+1} \sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{1}{\chi!}\partial^\chi \mu_f(\pi((i-1)\Delta_n)) \\ &\qquad\qquad \times \left\{ (\Upsilon^{n,i} - \pi((i-1)\Delta_n))^\chi - \Delta_n^{j(1-2H)} c((i-1)\Delta_n)^\chi \right\}, \\ \mathbb{X}_2^n(t) &= \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=N(H)+1} \frac{1}{\chi!}\partial^\chi \mu_f(\upsilon_i^n)(\Upsilon^{n,i} - \pi((i-1)\Delta_n))^\chi.\end{aligned} \tag{C.8}$$

In a final step, we remove the discretization of σ and ρ.

Lemma C.5. *If θ is chosen according to (C.3), then*

$$\Delta_n^{-\frac{1}{2}}\left\{\Delta_n \sum_{i=\lambda_n+1}^{[t/\Delta_n]-L+1} \mu_f(\pi((i-1)\Delta_n)) - \int_0^t \mu_f(\pi(s))\,\mathrm{d}s\right\} \stackrel{L^1}{\Longrightarrow} 0 \tag{C.9}$$

and

$$\begin{aligned}\Delta_n^{-\frac{1}{2}}\Bigg\{\Delta_n &\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{j=1}^{N(H)} \Delta_n^{j(1-2H)} \sum_{|\chi|=j} \frac{1}{\chi!}\partial^\chi \mu_f(\pi((i-1)\Delta_n))c((i-1)\Delta_n)^\chi \\ &- \int_0^t \sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{1}{\chi!}\partial^\chi \mu_f(\pi(s))\Delta_n^{j(1-2H)} c(s)^\chi\,\mathrm{d}s\Bigg\} \stackrel{L^1}{\Longrightarrow} 0.\end{aligned} \tag{C.10}$$

By the properties of stable convergence in law (see Equation (2.2.5) in Jacod & Protter (2012)), the CLT in (2.7) follows by combining Lemmas C.1–C.5. □

D Details for the proof of Theorem 2.1

Assumption (CLT') is in force throughout this section.

Proof of Lemma C.1. By the calculations in (A.2)–(A.4), we have $\mathbb{E}[\|\underline{\Delta}_i^n Y/\Delta_n^H\|^p]^{1/p} \lesssim 1$ for all $p \geq 1$. As f grows at most polynomially, we see that $\mathbb{E}[|f(\underline{\Delta}_i^n Y/\Delta_n^H)|]$ is of size 1. Hence, $\mathbb{E}[|\Delta_n^{1/2}\sum_{i=1}^{\theta_n} f(\underline{\Delta}_i^n Y/\Delta_n^H)|] \lesssim \Delta_n^{1/2-\theta}$, which implies $\Delta_n^{1/2}\sum_{i=1}^{\theta_n} f(\underline{\Delta}_i^n Y/\Delta_n^H) \to 0$ in L^1 since $\theta < \frac{1}{2}$ by (C.3). As a result, omitting the first θ_n terms in the definition of $V_f^n(Y,t)$ does no harm asymptotically. Next, we define

$$\Lambda_i^n = f\left(\frac{\underline{\Delta}_i^n Y}{\Delta_n^H}\right) - f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\right), \qquad \overline{\Lambda}_i^n = \Lambda_i^n - \mathbb{E}[\Lambda_i^n \mid \mathcal{F}_{i-\theta_n}^n]. \tag{D.1}$$

By our choice (C.3) of θ and since $H < \frac{1}{2}$, the lemma is proved once

$$\mathbb{E}\left[\sup_{t\leq T}\left|\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \overline{\Lambda}_i^n\right|\right] \lesssim \Delta_n^{\theta(\frac{1}{2}-H)}, \tag{D.2}$$

$$\mathbb{E}\left[\sup_{t\leq T}\left|\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}[\Lambda_i^n|\mathcal{F}_{i-\theta_n}^n]\right|\right] \lesssim \Delta_n^{\theta(\frac{1}{2}-H)} + \Delta_n^{2\theta(1-H)-\frac{1}{2}} \tag{D.3}$$

are established. To this end, let $\lambda_i^n = \underline{\Delta}_i^n Y - \underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H = \int_0^{(i-\theta_n)\Delta_n} \rho_s\,\mathrm{d}W_s\,\frac{\underline{\Delta}_i^n g(s)}{\Delta_n^H}$. By Assumption (CLT), we have $|f(z)-f(z')| \lesssim (1+\|z\|^{p-1}+\|z'\|^{p-1})\|z-z'\|$. In addition, $\mathbb{E}[|f(\underline{\Delta}_i^n Y/\Delta_n^H)|]$ is of size 1, so $\mathbb{E}[(\overline{\Lambda}_i^n)^2] \lesssim \mathbb{E}[(\Lambda_i^n)^2] \lesssim \mathbb{E}[\|\lambda_i^n\|^2] \lesssim \Delta_n^{2\theta(1-H)}$, where we used (A.4) for the last estimation. By construction, $\overline{\Lambda}_i^n$ is $\mathcal{F}_{i+L-1}^n$-measurable and has conditional expectation 0 given $\mathcal{F}_{i-\theta_n}^n$. Therefore, we can further use an estimate of the kind (A.6) to show that the left-hand side of (D.2) is bounded, up to constant, by $\sqrt{\theta_n}\Delta_n^{\theta(1-H)} \lesssim \Delta_n^{\theta(1-H)-\theta/2} = \Delta_n^{\theta(1/2-H)}$.

Next, let $\psi_i^n = \sigma_{(i-\theta_n)\Delta_n}\underline{\Delta}_i^n B + \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \rho_{(i-\theta_n)\Delta_n}\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s)$. Since f is smooth, applying Taylor's theorem twice yields $\Lambda_i^n = \Lambda_i^{n,1} + \Lambda_i^{n,2} + \Lambda_i^{n,3}$, where

$$\Lambda_i^{n,1} = \sum_{|\chi|=1} \partial^\chi f\Big(\frac{\psi_i^n}{\Delta_n^H}\Big)(\lambda_i^n)^\chi, \qquad \Lambda_i^{n,2} = \sum_{|\chi|,|\chi'|=1} \partial^{\chi+\chi'} f(\widetilde{\eta}_i^n)\Big(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}} - \psi_i^n}{\Delta_n^H}\Big)^{\chi'}(\lambda_i^n)^\chi,$$
$$\Lambda_i^{n,3} = \sum_{|\chi|=2} \frac{\partial^\chi(\eta_i^n)}{\chi!}(\lambda_i^n)^\chi$$

and $\chi, \chi' \in \mathbb{N}_0^{d\times L}$ are multi-indices and η_i^n (resp., $\widetilde{\eta}_i^n$) is a point on the line between $\underline{\Delta}_i^n Y/\Delta_n^H$ and $\underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H$ (resp., $\underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H$ and ψ_i^n/Δ_n^H). Accordingly, we split

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}[\Lambda_i^n \mid \mathcal{F}_{i-\theta_n}^n] = \sum_{j=1}^3 \mathbb{L}_j^n(t), \qquad \mathbb{L}_j^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \mathbb{E}[\Lambda_i^{n,j} \mid \mathcal{F}_{i-\theta_n}^n].$$

Note that $\mathbb{E}[\partial^\chi f(\frac{\psi_i^n}{\Delta_n^H})(\lambda_i^n)^\chi \mid \mathcal{F}_{i-\theta_n}^n] = (\lambda_i^n)^\chi\,\mathbb{E}[\partial^\chi f(\frac{\psi_i^n}{\Delta_n^H}) \mid \mathcal{F}_{i-\theta_n}^n] = 0$ because λ_i^n is $\mathcal{F}_{i-\theta_n}^n$-measurable, ψ_i^n is centered normal given $\mathcal{F}_{i-\theta_n}^n$ and f has odd partial derivatives of first orders (since f is even). It follows that $\mathbb{L}_1^n(t) = 0$ identically. Writing

$$\mathbb{1}_i^n(s) = (\mathbb{1}_{((i-1)\Delta_n, i\Delta_n)}(s), \ldots, \mathbb{1}_{((i+L-2)\Delta_n,(i+L-1)\Delta_n)}(s)),$$

we can decompose $\underline{\Delta}_i^n Y^{\mathrm{tr}} - \psi_i^n$ as

$$\underline{\Delta}_i^n A + \int_0^t (\sigma_s - \sigma_{(i-\theta_n)\Delta_n})\,\mathrm{d}B_s\,\mathbb{1}_i^n(s) + \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} (\rho_s - \rho_{(i-\theta_n)\Delta_n})\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s).$$

By a standard size estimate, it follows that

$$\begin{aligned}
\mathbb{E}\Big[\sup_{t\le T}|\mathbb{L}_2^n(t)|\Big] &\lesssim (\Delta_n^{\frac{1}{2}}\Delta_n^{-1})(\Delta_n^{1-H} + \theta_n^{\frac{1}{2}}\Delta_n^{1-H} + (\theta_n\Delta_n)^{\frac{1}{2}})\Delta_n^{\theta(1-H)} \\
&\lesssim \Delta_n^{-\frac{1}{2}}\Delta_n^{\theta(1-H)}(\theta_n\Delta_n)^{\frac{1}{2}} = \Delta_n^{\theta(\frac{1}{2}-H)}, \\
\mathbb{E}\Big[\sup_{t\le T}|\mathbb{L}_3^n(t)|\Big] &\lesssim \Delta_n^{-\frac{1}{2}}(\Delta_n^{\theta(1-H)})^2 = \Delta_n^{2\theta(1-H)-\frac{1}{2}},
\end{aligned}$$

proving (D.3) and thus the lemma. □

Proof of Lemma C.2. Let $\xi_i^{n,\mathrm{dis}} = \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \rho_{(i-\theta_n)\Delta_n}\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s)$ and recall the definition of ξ_i^n from (C.4). In a first step, we show that U^n can be approximated by

$$\begin{aligned}
\overline{U}^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \bigg\{ & f\Big(\frac{\sigma_{(i-1)\Delta_n}\underline{\Delta}_i^n B + \xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big) - f\Big(\frac{\xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big) \\
& - \mathbb{E}\Big[f\Big(\frac{\sigma_{(i-1)\Delta_n}\underline{\Delta}_i^n B + \xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big) - f\Big(\frac{\xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big) \;\Big|\; \mathcal{F}_{i-\theta_n}^n\Big]\bigg\}.
\end{aligned}$$

By (C.1) and a size estimate as in (A.4), the difference $\xi_i^n - \xi_i^{n,\mathrm{dis}}$ is of size $(\theta_n\Delta_n)^{1/2}$. Together with (A.2) and (A.3), we further have that $\underline{\Delta}_i^n Y^{\mathrm{tr}} - \sigma_{(i-1)\Delta_n}\underline{\Delta}_i^n B - \xi_i^{n,\mathrm{dis}}$ is of size $\Delta_n + \sqrt{\Delta_n} + (\theta_n\Delta_n)^{1/2}$. By the mean-value theorem, these size bounds imply that

$$\mathbb{E}\bigg[\Big|f\Big(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H}\Big) - f\Big(\frac{\sigma_{(i-1)\Delta_n}\underline{\Delta}_i^n B + \xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big)\Big|^p + \Big|f\Big(\frac{\xi_i^n}{\Delta_n^H}\Big) - f\Big(\frac{\xi_i^{n,\mathrm{dis}}}{\Delta_n^H}\Big)\Big|^p\bigg]^{\frac{1}{p}} \lesssim (\theta_n\Delta_n)^{\frac{1}{2}}$$

for any $p > 0$. Moreover, the ith term in the definition of $\overline{U}^n(t)$ is $\mathcal{F}^n_{i+L-1}$-measurable with zero mean conditionally on $\mathcal{F}^n_{i-\theta_n}$. Therefore, employing a martingale size estimate as in (A.6), we obtain $\mathbb{E}[\sup_{t\leq T}|U^n(t) - \overline{U}^n(t)|] \lesssim \sqrt{\theta_n}(\theta_n\Delta_n)^{1/2} \leq \Delta_n^{1/2-\theta}$, which converges to 0 by (C.3).

Next, because B and W are independent, we can apply Itô's formula with $\xi^{n,\mathrm{dis}}_i$ as starting point and write

$$
\begin{aligned}
&f\Big(\frac{\sigma_{(i-1)\Delta_n}\underline{\Delta}^n_i B + \xi^{n,\mathrm{dis}}_i}{\Delta_n^H}\Big) - f\Big(\frac{\xi^{n,\mathrm{dis}}_i}{\Delta_n^H}\Big)\\
&\quad= \Delta_n^{-H}\sum_{j,k=1}^d\sum_{\ell=1}^L \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \frac{\partial}{\partial z_{k\ell}} f\Big(\frac{\underline{\Delta}Y^{n,\mathrm{dis}}_i(s)}{\Delta_n^H}\Big)\sigma^{kj}_{(i-1)\Delta_n}\,\mathrm{d}B^j_s \\
&\qquad + \frac{1}{2}\Delta_n^{-2H}\sum_{k,k'=1}^d\sum_{\ell=1}^L \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \frac{\partial^2}{\partial z_{k\ell}\partial z_{k'\ell}} f\Big(\frac{\underline{\Delta}Y^{n,\mathrm{dis}}_i(s)}{\Delta_n^H}\Big)(\sigma\sigma^T)^{kk'}_{(i-1)\Delta_n}\,\mathrm{d}s,
\end{aligned}
\tag{D.4}
$$

where $\underline{\Delta}Y^{n,\mathrm{dis}}_i(s) = \int_{(i-1)\Delta_n}^s \sigma_{(i-1)\Delta_n}\,\mathrm{d}B_r\,\mathbb{1}^n_i(r) + \xi^{n,\mathrm{dis}}_i$. Clearly, the stochastic integral is $\mathcal{F}^n_{i+L-1}$-measurable and conditionally centered given $\mathcal{F}^n_{i-1}$. Therefore, by a martingale size estimate, its contribution to $\overline{U}^n(t)$ is of magnitude $\Delta_n^{1/2-H}$, which is negligible because $H < \frac{1}{2}$. For the Lebesgue integral, we apply Itô's formula again and write

$$
\begin{aligned}
&\frac{\partial^2}{\partial z_{k\ell}\partial z_{k'\ell}} f\Big(\frac{\underline{\Delta}Y^{n,\mathrm{dis}}_i(s)}{\Delta_n^H}\Big) = \frac{\partial^2}{\partial z_{k\ell}\partial z_{k'\ell}} f\Big(\frac{\xi^{n,\mathrm{dis}}_i}{\Delta_n^H}\Big)\\
&\quad + \Delta_n^{-H}\sum_{j_2,k_2=1}^d\sum_{\ell_2=1}^L \int_{(i+\ell_2-2)\Delta_n}^{s\wedge(i+\ell_2-1)\Delta_n} \frac{\partial^3}{\partial z_{k\ell}\partial z_{k'\ell}\partial z_{k_2\ell_2}} f\Big(\frac{\underline{\Delta}Y^{n,\mathrm{dis}}_i(r)}{\Delta_n^H}\Big)\sigma^{k_2j_2}_{(i-1)\Delta_n}\,\mathrm{d}B^{j_2}_r\\
&\quad + \frac{\Delta_n^{-2H}}{2}\sum_{k_2,k_2'=1}^d\sum_{\ell_2=1}^L \int_{(i+\ell_2-2)\Delta_n}^{s\wedge(i+\ell_2-1)\Delta_n} \frac{\partial^4}{\partial z_{k\ell}\partial z_{k\ell}\partial z_{k_2\ell_2}\partial z_{k_2'\ell_2}} f\Big(\frac{\underline{\Delta}Y^{n,\mathrm{dis}}_i(r)}{\Delta_n^H}\Big)(\sigma\sigma^T)^{k_2k_2'}_{(i-1)\Delta_n}\,\mathrm{d}r.
\end{aligned}
$$

By the same reason as before, the stochastic integral (even after we plug it into the drift in (D.4)) is $\mathcal{F}^n_{i+L-1}$-measurable with zero $\mathcal{F}^n_{i-1}$-conditional mean and therefore negligible. The Lebesgue integral is essentially of the same form as the one in (D.4). Because f is smooth, we can repeat this procedure as often as we want. What is important, is that we gain a net factor of Δ_n^{1-2H} in each step (we have Δ_n^{-2H} times a Lebesgue integral over an interval of length at most Δ_n). After N applications of Itô's formula, the final drift term yields a contribution of size $\sqrt{\theta_n}\Delta_n^{N(1-2H)}$ to $\overline{U}^n(t)$. As $\theta < \frac{1}{2}$, it suffices to take $N = N(H) + 1$ to make this convergent to 0. ☐

Proof of Lemma C.3. We begin by discretizing ρ on a finer scale and let

$$
\Theta^n_i = \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \sum_{k=1}^Q \rho_{(i-\theta_n^{(q-1)})\Delta_n}\mathbb{1}_{((i-\theta_n^{(q-1)})\Delta_n,(i-\theta_n^{(q)})\Delta_n)}(s)\,\mathrm{d}W_s\,\underline{\Delta}^n_i g(s),
\tag{D.5}
$$

where $\theta_n^{(q)} = [\Delta_n^{-\theta^{(q)}}]$ for $q = 0,\dots,Q-1$, $\theta_n^{(Q)} = -(L-1)$ and the numbers $\theta^{(q)}$, $q = 0,\dots,Q-1$ for some $Q\in\mathbb{N}$, are chosen such that $\theta = \theta^{(0)} > \cdots > \theta^{(Q-1)} > \theta^{(Q)} = 0$ and

$$
\theta^{(q)} > \frac{\gamma}{1-H}\theta^{(q-1)} - \frac{\gamma - \frac{1}{2}}{1-H}, \qquad q = 1,\dots,Q,
\tag{D.6}
$$

where γ describes the regularity of the volatility process $\rho^{(0)}$ in (2.3). Because $H < \frac{1}{2}$ and we can make γ arbitrarily close to $\frac{1}{2}$ if we want, there is no loss of generality to assume that $\gamma/(1-H) < 1$. In this case, the fact that a choice as in (D.6) is possible can be verified by solving the associated linear recurrence equation. Defining $\underline{\Delta}_i^n Y^{\mathrm{dis}} = \sigma_{(i-1)\Delta_n} \underline{\Delta}_i^n B + \Theta_i^n$, we will show in Lemma D.1 below that

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \left\{ \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{tr}}}{\Delta_n^H} \right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n \right] - \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{dis}}}{\Delta_n^H} \right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n \right] \right\} \stackrel{L^1}{\Longrightarrow} 0. \tag{D.7}$$

Next, we define another matrix $\Upsilon_i^{n,0} \in (\mathbb{R}^{d\times L})^2$ by

$$\begin{aligned}(\Upsilon_i^{n,0})_{k\ell,k'\ell'} = c((i-1)\Delta_n)\Delta_n^{1-2H} + \sum_{q=1}^{Q} \Big(\rho_{(i-\theta_n^{(q-1)})\Delta_n} \rho^T_{(i-\theta_n^{(q-1)})\Delta_n} \Big)_{kk'} \\ \times \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \frac{\Delta_{i+\ell-1}^n g(s) \Delta_{i+\ell'-1}^n g(s)}{\Delta_n^{2H}} \mathbb{1}_{((i-\theta_n^{(q-1)})\Delta_n, (i-\theta_n^{(q)})\Delta_n)}(s)\, \mathrm{d}s.\end{aligned} \tag{D.8}$$

If c and ρ are deterministic, this is the covariance matrix of $\underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H$. Also notice that the only difference to Υ_i^n are the discretization points of ρ. Next, we show that

$$\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \left\{ \mathbb{E}\left[f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{dis}}}{\Delta_n^H} \right) \;\middle|\; \mathcal{F}_{i-\theta_n}^n \right] - \mu_f(\mathbb{E}[\Upsilon_i^{n,0} \mid \mathcal{F}_{i-\theta_n}^n]) \right\} \stackrel{L^1}{\Longrightarrow} 0, \tag{D.9}$$

where μ_f is the mapping defined after Assumption (CLT). This will be achieved through successive conditioning in Lemma D.2. Finally, as we show in Lemma D.3, we have

$$\mathbb{E}\left[\sup_{t\le T} \left| \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \left\{ \mu_f\Big(\mathbb{E}[\Upsilon_i^{n,0} \mid \mathcal{F}_{i-\theta_n}^n] \Big) - \mu_f(\Upsilon_i^{n,0}) \right\} \right| \right] \to 0, \tag{D.10}$$

$$\mathbb{E}\left[\sup_{t\le T} \left| \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \{ \mu_f(\Upsilon_i^{n,0}) - \mu_f(\Upsilon^{n,i}) \} \right| \right] \to 0, \tag{D.11}$$

which completes the proof of the current lemma. □

Lemma D.1. *The convergence* (D.7) *holds true.*

Proof. By Taylor's theorem, the left-hand side of (D.7) is $\mathbb{Q}_1^n(t) + \mathbb{Q}_2^n(t)$ with

$$\mathbb{Q}_1^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=1} \mathbb{E}\left[\partial^\chi f\left(\frac{\underline{\Delta}_i^n Y^{\mathrm{dis}}}{\Delta_n^H} \right) (\kappa_i^n)^\chi \;\middle|\; \mathcal{F}_{i-\theta_n}^n \right],$$

$$\mathbb{Q}_2^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=2} \frac{1}{\chi!} \mathbb{E}[\partial^\chi f(\overline{\kappa}_i^n)(\kappa_i^n)^\chi \mid \mathcal{F}_{i-\theta_n}^n],$$

where $\kappa_i^n = (\underline{\Delta}_i^n Y^{\mathrm{tr}} - \underline{\Delta}_i^n Y^{\mathrm{dis}})/\Delta_n^H$ and $\overline{\kappa}_i^n$ is some point on the line between $\underline{\Delta}_i^n Y^{\mathrm{tr}}/\Delta_n^H$ and $\underline{\Delta}_i^n Y^{\mathrm{dis}}/\Delta_n^H$. By definition,

$$\begin{aligned}(\kappa_i^n)_{k\ell} = \frac{\Delta_{i+\ell-1}^n A^k}{\Delta_n^H} + \frac{1}{\Delta_n^H} \int_{(i+\ell-2)\Delta_n}^{(i+\ell-1)\Delta_n} \sum_{\ell'=1}^{d'} \Big(\sigma_s^{k\ell'} - \sigma_{(i-1)\Delta_n}^{k\ell'} \Big) \mathrm{d}B_s^{\ell'} \\ + \sum_{q=1}^{Q} \int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n} \frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H} \sum_{\ell'=1}^{d'} \Big(\rho_s^{k\ell'} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k\ell'} \Big) \mathrm{d}W_s^{\ell'}.\end{aligned} \tag{D.12}$$

Using Hölder's inequality, the estimates (A.2), (A.3) and (A.4) and the polynomial growth assumption on $\partial^\chi f$, we see that $\underline{\Delta}_i^n Y^{\mathrm{dis}}/\Delta_n^H$ is of size one and, since $0<\theta^{(q)}<\frac{1}{2}$,

$$\mathbb{E}\Big[\sup_{t\leq T}|\mathbb{Q}_2^n(t)|\Big]\lesssim \Delta_n^{-\frac{1}{2}}\Big(\Delta_n^{2(1-H)}+\Delta_n^{2(1-H)}+\sum_{q=1}^Q\Delta_n^{(1-\theta^{(q-1)})+2\theta^{(q)}(1-H)}\Big)\to 0. \tag{D.13}$$

Next, we further split $\mathbb{Q}_1^n(t)=\mathbb{Q}_{11}^n(t)+\mathbb{Q}_{12}^n(t)+\mathbb{Q}_{13}^n(t)$ into three terms according to the decomposition (D.12). Using again (A.2) and (A.3), we see that both $\mathbb{Q}_{11}^n(t)$ and $\mathbb{Q}_{12}^n(t)$ are of size $\Delta_n^{-1/2+(1-H)}=\Delta_n^{1/2-H}$. We first tackle the term $\mathbb{Q}_{13}^n(t)$, which requires a more careful analysis. Here we need assumption (2.3) on the noise volatility ρ. Since $t\mapsto\int_0^t\widetilde{b}_s\,\mathrm{d}s$ satisfies a better regularity condition than (C.1), we may incorporate the drift term in $\rho^{(0)}$ for the remainder of the proof. Then we further write $\mathbb{Q}_{13}^n(t)=\mathbb{R}_1^n(t)+\mathbb{R}_2^n(t)$ where $\mathbb{R}_1^n(t)$ and $\mathbb{R}_2^n(t)$ correspond to taking only $\rho^{(0)}$ and $\int_0^t\widetilde{\rho}_s\,\mathrm{d}\widetilde{W}_s$ instead of ρ, respectively. By (2.4), (A.4) and (D.6), $\mathbb{R}_1^n(t)$ is of size

$$\sum_{q=1}^Q\Delta_n^{-\frac{1}{2}+\gamma(1-\theta^{(q-1)})+\theta^{(q)}(1-H)}\to 0. \tag{D.14}$$

For $\mathbb{R}_2^n(t)$, we write $\mathbb{R}_2^n(t)=\sum_{|\chi|=1}(\mathbb{R}_{21}^{n,\chi}(t)+\mathbb{R}_{22}^{n,\chi}(t)+\mathbb{R}_{23}^{n,\chi}(t))$, where, if $\chi_{k\ell}=1$,

$$\begin{aligned}
\mathbb{R}_{21}^{n,\chi}(t)&=\Delta_n^{\frac{1}{2}}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1}\sum_{\ell',\ell''=1}^d\mathbb{E}\Big[\partial^\chi f\Big(\frac{\underline{\Delta}_i^nY^{\mathrm{dis}}}{\Delta_n^H}\Big)\sum_{q=1}^Q\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H}\\
&\qquad\times\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\Big(\widetilde{\rho}_r^{k,\ell',\ell''}-\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\Big)\,\mathrm{d}\widetilde{W}_r^{\ell''}\,\mathrm{d}W_s^{\ell'}\;\Big|\;\mathcal{F}_{i-\theta_n}^n\Big],\\
\mathbb{R}_{22}^{n,\chi}(t)&=\Delta_n^{\frac{1}{2}}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1}\sum_{q=1}^Q\sum_{\ell',\ell''=1}^d\mathbb{E}\Big[\Big\{\partial^\chi f\Big(\frac{\underline{\Delta}_i^nY^{\mathrm{dis}}}{\Delta_n^H}\Big)-\partial^\chi f\Big(\frac{\underline{\Delta}_i^nY^{\mathrm{dis},q}}{\Delta_n^H}\Big)\Big\}\\
&\qquad\times\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H}\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\,\mathrm{d}\widetilde{W}_r^{\ell''}\,\mathrm{d}W_s^{\ell'}\;\Big|\;\mathcal{F}_{i-\theta_n}^n\Big],\\
\mathbb{R}_{23}^{n,\chi}(t)&=\Delta_n^{\frac{1}{2}}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1}\sum_{q=1}^Q\sum_{\ell',\ell''=1}^d\mathbb{E}\Big[\partial^\chi f\Big(\frac{\underline{\Delta}_i^nY^{\mathrm{dis},q}}{\Delta_n^H}\Big)\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H}\\
&\qquad\times\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\,\mathrm{d}\widetilde{W}_r^{\ell''}\,\mathrm{d}W_s^{\ell'}\;\Big|\;\mathcal{F}_{i-\theta_n}^n\Big]
\end{aligned}$$

and $\underline{\Delta}_i^nY^{\mathrm{dis},q}=\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i+L-1)\Delta_n}\rho_{(i-\theta_n^{(q-1)})\Delta_n}\,\mathrm{d}W_s\,\underline{\Delta}_i^n g(s)$. Using the BDG and Minkowski integral inequality alternatingly, we obtain, for any $p\geq 2$,

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)}{\Delta_n^H}\Big(\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\Big(\widetilde{\rho}_r^{k,\ell',\ell''}-\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\Big)\,\mathrm{d}\widetilde{W}_r^{\ell''}\Big)\mathrm{d}W_s^{\ell'}\Big|^p\Big]^{\frac{1}{p}}\\
&\quad\lesssim\Big(\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)^2}{\Delta_n^{2H}}\mathbb{E}\Big[\Big|\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\Big(\widetilde{\rho}_r^{k,\ell',\ell''}-\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\Big)\,\mathrm{d}\widetilde{W}_r^{\ell''}\Big|^{2p}\Big]^{\frac{1}{p}}\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad\lesssim\Big(\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)^2}{\Delta_n^{2H}}\int_{(i-\theta_n^{(q-1)})\Delta_n}^s\mathbb{E}\Big[\Big|\widetilde{\rho}_r^{k,\ell',\ell''}-\widetilde{\rho}_{(i-\theta_n^{(q-1)})\Delta_n}^{k,\ell',\ell''}\Big|^{2p}\Big]^{\frac{1}{p}}\,\mathrm{d}r\,\mathrm{d}s\Big)^{\frac{1}{2}}\\
&\quad\lesssim(\theta_n^{(q-1)}\Delta_n)^{\frac{1}{2}(1+2\varepsilon')}\Big(\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\Delta_{i+\ell-1}^n g(s)^2}{\Delta_n^{2H}}\,\mathrm{d}s\Big)^{\frac{1}{2}}\lesssim\Delta_n^{(\frac{1}{2}+\varepsilon')(1-\theta^{(q-1)})+\theta^{(q)}(1-H)},
\end{aligned}$$

where ε' is as in (2.5). Thus, $\mathbb{R}_{21}^{n,\chi}(t)$ is of size $\sum_{q=1}^{Q}\Delta_n^{-\frac{1}{2}+(\frac{1}{2}+\varepsilon')(1-\theta^{(q-1)})+\theta^{(q)}(1-H)}$, which is almost the same as (D.14); the only difference is that γ is replaced by $\frac{1}{2}+\varepsilon'$. Since we can assume without loss of generality that $\frac{1}{2}+\varepsilon'<\gamma$, the formula (D.6) implies that we have $-\frac{1}{2}+(\frac{1}{2}+\varepsilon')(1-\theta^{(q-1)})+\theta^{(q)}(1-H)>0$ for all $q=1,\dots,Q$, which means that $\mathbb{R}_{21}^{n,\chi}(t)$ is asymptotically negligible.

Next, using Lemma B.1 (iii) and a similar estimate to the previous display, we see that $(\Theta_i^n-\underline{\Delta}_i^n Y^{\mathrm{dis},q})/\Delta_n^H$ is of size $\Delta_n^{\theta^{(q-1)}(1-H)}+\Delta_n^{(1-\theta^{(q-1)})/2}$. Hence, with the two estimates (A.2) and (A.3) at hand, we deduce that $\mathbb{R}_{22}^{n,\chi}(t)$ is of size

$$\sum_{q=1}^{Q}\Delta_n^{-\frac{1}{2}}(\Delta_n^{\frac{1}{2}-H}+\Delta_n^{\theta^{(q-1)}(1-H)}+\Delta_n^{\frac{1}{2}(1-\theta^{(q-1)})})\Delta_n^{\theta^{(q)}(1-H)+\frac{1}{2}(1-\theta^{(q-1)})}$$
$$\leq\sum_{q=1}^{Q}\Big(\Delta_n^{\frac{1}{2}-H-(\gamma-\frac{1}{2})(1-\theta^{(q-1)})}+\Delta_n^{(\gamma+\frac{1}{2}-H)\theta^{(q-1)}-(\gamma-\frac{1}{2})}+\Delta_n^{\theta^{(q)}(1-H)+(\frac{1}{2}-\theta^{(q-1)})}\Big).$$

The last term clearly goes to 0 because $\theta^{(q-1)}\leq\theta<\frac{1}{2}$ by (C.3). Without loss of generality, we can assume that $\gamma>\frac{1}{2}$ is sufficiently close to $\frac{1}{2}$ such that the first term is negligible as well. With this particular value, we then make sure that

$$\frac{\gamma-\frac{1}{2}}{\gamma+\frac{1}{2}-H}<\theta^{(Q-1)}<\frac{\gamma-\frac{1}{2}}{\gamma},$$

which, on the one hand, is in line with (D.6) and, on the other hand, guarantees that the second term in the preceding display tends to 0 for all $q=1,\dots,Q$.

Finally, to compute $\mathbb{R}_{23}^{n,\chi}(t)$, we first condition on $\mathcal{F}^n_{i-\theta_n^{(q-1)}}$. Because f is even and $\underline{\Delta}_i^n Y^{\mathrm{dis},q}/\Delta_n^H$ has a centered normal distribution given $\mathcal{F}^n_{i-\theta_n^{(q-1)}}$, if follows that $\partial^\chi f(\Theta_i^{n,q}/\Delta_n^H)$ is an element of the direct sum of all odd-order Wiener chaoses. At the same time, the double stochastic integrals in $\mathbb{R}_{23}^{n,\chi}(t)$ belongs to the second Wiener chaos; see Proposition 1.1.4 in Nualart (2006). Since Wiener chaoses are mutually orthogonal, we obtain $\mathbb{R}_{23}^{n,\chi}(t)=0$. Because this reasoning is valid for all multi-indices with $|\chi|=1$, we have shown that $\mathbb{R}_2^n(t)$ is asymptotically negligible. □

Lemma D.2. *The convergence* (D.9) *holds true.*

Proof. For $r=0,\dots,Q$ (where Q is as in Lemma D.1), define

$$\mathbb{Y}_i^{n,r}=\int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n}\Big(\sum_{q=1}^{r}\rho_{(i-\theta_n^{(q-1)})\Delta_n}\mathbb{1}_{((i-\theta_n^{(q-1)})\Delta_n,(i-\theta_n^{(q)})\Delta_n)}(s)\Big)\,\mathrm{d}W_s\,\frac{\underline{\Delta}_i^n g(s)}{\Delta_n^H},$$
$$\Upsilon_i^{n,r}=c((i-1)\Delta_n)\Delta_n^{1-2H}+\sum_{q=r+1}^{Q}(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n}\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n}\frac{\underline{\Delta}_i^n g(s)^T\underline{\Delta}_i^n g(s)}{\Delta_n^{2H}}\,\mathrm{d}s.$$

Note that $\mathbb{Y}_i^{n,r}\in\mathbb{R}^{d\times L}$, $\Upsilon_i^{n,r}\in\mathbb{R}^{(d\times L)\times(d\times L)}$ and that $\mathbb{Y}_i^{n,Q}=\Theta_i^n/\Delta_n^H$ by (D.5). In order to show (D.9), we need the following approximation result for each $r=1,\dots,Q-1$:

$$\Delta_n^{\frac{1}{2}}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1}\mathbb{E}\Big[\mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\Upsilon_i^{n,r,r})-\mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\Upsilon_i^{n,r,r-1})\mid\mathcal{F}^n_{i-\theta_n}\Big]\overset{L^1}{\Longrightarrow}0,\qquad\text{(D.15)}$$

where $\Upsilon_i^{n,r,q} = \mathbb{E}[\Upsilon_i^{n,r} \mid \mathcal{F}_{i-\theta_n^{(q)}}]$. Let us proceed with the proof of (D.9), taking the previous statement for granted. Defining

$$\overline{\mathbb{Y}}_i^n = \int_{(i-\theta_n)\Delta_n}^{(i-1)\Delta_n} \sum_{q=1}^{Q} \rho_{(i-\theta_n^{(q-1)})\Delta_n} \mathbb{1}_{((i-\theta_n^{(q-1)})\Delta_n,(i-\theta_n^{(q)})\Delta_n)}(s)\, \mathrm{d}W_s \, \frac{\underline{\Delta}_i^n g(s)}{\Delta_n^H},$$

we can use the tower property of conditional expectation to derive

$$\begin{aligned}
&\mathbb{E}\Big[f\Big(\frac{\underline{\Delta}_i^n Y^{\mathrm{dis}}}{\Delta_n^H}\Big) \;\Big|\; \mathcal{F}_{i-\theta_n}\Big] = \mathbb{E}\Big[\mathbb{E}\Big[f\Big(\frac{\underline{\Delta}_i^n Y^{\mathrm{dis}}}{\Delta_n^H}\Big) \;\Big|\; \mathcal{F}_{i-1}\Big] \;\Big|\; \mathcal{F}_{i-\theta_n}\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\mu_{f(\overline{\mathbb{Y}}_i^n+\cdot)}\Big(c((i-1)\Delta_n)\Delta_n^{1-2H} \\
&\qquad + (\rho\rho^T)_{(i-\theta_n^{(Q-1)})\Delta_n} \int_{(i-1)\Delta_n}^{(i+L-1)\Delta_n} \frac{\underline{\Delta}_i^n g(s)^T \underline{\Delta}_i^n g(s)}{\Delta_n^{2H}}\,\mathrm{d}s\Big) \;\Big|\; \mathcal{F}_{i-\theta_n^{(Q-1)}}\Big] \;\Big|\; \mathcal{F}_{i-\theta_n}\Big] \\
&\quad = \mathbb{E}\Big[\mu_{f(\mathbb{Y}_i^{n,Q-1}+\cdot)}(\Upsilon_i^{n,Q-1}) \mid \mathcal{F}_{i-\theta_n}\Big].
\end{aligned}$$

Thanks to (D.15), we can replace $\Upsilon_i^{n,Q-1} = \Upsilon_i^{n,Q-1,Q-1}$ in the last line by $\Upsilon_i^{n,Q-1,Q-2}$. We can then further compute

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\Big[\mu_{f(\mathbb{Y}_i^{n,Q-1}+\cdot)}\Big(\Upsilon_i^{n,Q-1,Q-2}\Big) \mid \mathcal{F}_{i-\theta_n^{(Q-2)}}\Big] \mid \mathcal{F}_{i-\theta_n}\Big] \\
&\quad = \mathbb{E}\Big[\mu_{f(\mathbb{Y}_i^{n,Q-2}+\cdot)}\Big(\Upsilon_i^{n,Q-2,Q-2}\Big) \mid \mathcal{F}_{i-\theta_n}\Big] \\
&\quad = \mathbb{E}\Big[\mathbb{E}\Big[\mu_{f(\mathbb{Y}_i^{n,Q-2}+\cdot)}\Big(\Upsilon_i^{n,Q-2,Q-2}\Big) \mid \mathcal{F}_{i-\theta_n^{(Q-3)}}\Big] \mid \mathcal{F}_{i-\theta_n}\Big].
\end{aligned} \tag{D.16}$$

Again by (D.15), we may replace $\Upsilon_i^{n,Q-2,Q-2}$ by $\Upsilon_i^{n,Q-2,Q-3}$ in (D.16). Repeating this procedure Q times, we obtain $\mu_{f(\mathbb{Y}_i^{n,0}+\cdot)}(\mathbb{E}[\Upsilon_i^{n,0} \mid \mathcal{F}_{i-\theta_n^{(0)}}]) = \mu_f(\mathbb{E}[\Upsilon_i^{n,0} \mid \mathcal{F}_{i-\theta_n}])$ in the end, which shows (D.9).

It remains to prove (D.15). For $(u,v) \mapsto \mu_{f(u+\cdot)}(v)$, we use $\partial^{\chi'}$ to denote differentiation with respect to u (where $\chi' \in \mathbb{N}_0^{d\times L}$) and $\partial^{\chi''}$ to denote differentiation with respect to v (where $\chi'' \in \mathbb{N}_0^{(d\times L)\times(d\times L)}$). By a Taylor expansion of $\mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\cdot)$ around the point $(\mathbb{Y}_i^{n,r}, \Upsilon_i^{n,r,r-1})$, the difference inside $\mathbb{E}[\cdot \mid \mathcal{F}_{i-\theta_n}^n]$ in (D.15) equals

$$\begin{aligned}
&\Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi''|=1} \mathbb{E}\Big[\partial^{\chi''}\mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\Upsilon_i^{n,r,r-1})(\Upsilon_i^{n,r,r} - \Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n\Big] \\
&\quad + \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi''|=2} \frac{1}{\chi''!}\mathbb{E}\Big[\partial^{\chi''}\mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\overline{\upsilon}_i^n)(\Upsilon_i^{n,r,r} - \Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n\Big]
\end{aligned} \tag{D.17}$$

for some $\overline{\upsilon}_i^n$ between $\Upsilon_i^{n,r,r}$ and $\Upsilon_i^{n,r,r-1}$. Write

$$\begin{aligned}
&\mathbb{E}\Big[(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} \mid \mathcal{F}_{i-\theta_n^{(r)}}^n\Big] - \mathbb{E}\Big[(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} \mid \mathcal{F}_{i-\theta_n^{(r-1)}}^n\Big] \\
&\quad = \mathbb{E}\Big[(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} - (\rho\rho^T)_{(i-\theta_n^{(r-1)})\Delta_n} \mid \mathcal{F}_{i-\theta_n^{(r)}}^n\Big] \\
&\qquad - \mathbb{E}\Big[(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} - (\rho\rho^T)_{(i-\theta_n^{(r-1)})\Delta_n} \mid \mathcal{F}_{i-\theta_n^{(r-1)}}^n\Big],
\end{aligned} \tag{D.18}$$

and note that, because of Assumption (CLT') and the identity

$$xy - x_0y_0 = y_0(x-x_0) + x_0(y-y_0) + (x-x_0)(y-y_0), \tag{D.19}$$

the two conditional expectations on the right-hand side of (D.18) are both of size $(\theta_n^{(r-1)}\Delta_n)^{1/2}$. The same holds true if we replace $\rho_{(i-\theta_n^{(q-1)})\Delta_n}$ by $\sigma_{(i-1)\Delta_n}$. Therefore,

$$\mathbb{E}\Big[\|\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1}\|^p\Big]^{\frac{1}{p}} \lesssim (\theta_n^{(r-1)}\Delta_n)^{\frac{1}{2}}. \tag{D.20}$$

Thus, the second expression in (D.17) is of size $\Delta_n^{-1/2}((\theta_n^{(r-1)}\Delta_n)^{1/2})^2 = \Delta_n^{1/2-\theta^{(r-1)}}$ which goes to 0 as $n\to\infty$ since all numbers $\theta^{(r)}$ are chosen to be smaller than $\frac{1}{2}$; see (D.6).

Next, we expand $\partial^\chi \mu_{f(\mathbb{Y}_i^{n,r}+\cdot)}(\cdot)$ around $(0,\Upsilon_i^{n,r,r-1})$ and write the first expression in (D.17) as $\mathbb{S}_1^n(t)+\mathbb{S}_2^n(t)+\mathbb{S}_3^n(t)$, where

$$\mathbb{S}_1^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi''|=1} \mathbb{E}\Big[\partial^{\chi''}\mu_f(\Upsilon_i^{n,r,r-1})(\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n\Big],$$

$$\mathbb{S}_2^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi'|=|\chi''|=1} \mathbb{E}\Big[\partial^{\chi'}\partial^{\chi''}\mu_f(\Upsilon_i^{n,r,r-1})(\mathbb{Y}_i^{n,r})^{\chi'}(\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n\Big],$$

$$\begin{aligned}\mathbb{S}_3^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi'|=2,\,|\chi''|=1} \frac{1}{\chi'!}\mathbb{E}\Big[&\partial^{\chi'}\partial^{\chi''}\mu_{f(\varsigma_i^n+\cdot)}(\Upsilon_i^{n,r,r-1}) \\ &\times (\mathbb{Y}_i^{n,r})^{\chi'}(\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n\Big],\end{aligned}$$

and ς_i^n is a point between 0 and $\mathbb{Y}_i^{n,r}$. Observe that $\partial^{\chi''}\mu_f(\Upsilon_i^{n,r,r-1})$ is $\mathcal{F}_{i-\theta_n^{(r-1)}}$-measurable and that the $\mathcal{F}^n_{i-\theta_n^{(r-1)}}$-conditional expectation of $\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1}$ is 0. Hence,

$$\mathbb{E}[\partial^{\chi''}\mu_f(\Upsilon_i^{n,r,r-1})(\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1})^{\chi''} \mid \mathcal{F}_{i-\theta_n}^n] = 0$$

and it follows that $\mathbb{S}_1^n(t)$ vanishes. Next, by Chong (2020*c*), Equation (D.46), given $|\chi'| = |\chi''| = 1$, there are $\alpha,\beta,\gamma \in \{1,\ldots,d\}\times\{1,\ldots,L\}$ such that

$$\partial^{\chi'}\partial^{\chi''}\mu_{f(u+\cdot)}(v) = \frac{\partial\mu_{f(u+\cdot)}}{\partial u_\gamma \partial v_{\alpha,\beta}}(v) = \frac{1}{2^{\mathbb{1}_{\{\alpha=\beta\}}}}\mu_{\partial_{\alpha\beta\gamma}f(u+\cdot)}(v).$$

If $u=0$, since f has odd third derivatives, we have that $\mu_{\partial_{\alpha\beta\gamma}f}(v)=0$. Therefore, the $\partial^{\chi'}\partial^{\chi''}\mu_f$-expression in $\mathbb{S}_2^n(t)$ is equal to 0, so $\mathbb{S}_2^n(t)$ vanishes as well. Finally, we use the generalized Hölder inequality and the estimates (D.20) and (A.4) to see that

$$\begin{aligned}\mathbb{E}\Big[\sup_{t\le T}|\mathbb{S}_3^n(t)|\Big] &\lesssim \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[T/\Delta_n]-L+1} \mathbb{E}[\|\mathbb{Y}_i^{n,r}\|^4]^{\frac{1}{2}}\mathbb{E}\Big[\|\Upsilon_i^{n,r,r}-\Upsilon_i^{n,r,r-1}\|^4\Big]^{\frac{1}{4}} \\ &\lesssim \Delta_n^{-\frac{1}{2}}\Delta_n^{2\theta^{(r)}(1-H)}(\theta_n^{(r-1)}\Delta_n)^{\frac{1}{2}}.\end{aligned}$$

This converges to 0 as $n\to\infty$ if $2\theta^{(r)}(1-H)-\frac{1}{2}\theta^{(r-1)}>0$ for all $r=1,\ldots,Q-1$, which is equivalent to $\theta^{(r)} > \frac{1}{4(1-H)}\theta^{(r-1)}$. Because $\frac{1}{4(1-H)}<1$, this condition means that $\theta^{(r)}$ must not decrease to 0 too fast. By adding more intermediate θ's between $\theta^{(0)}$ and $\theta^{(Q-1)}$ if necessary, which does no harm to (D.6), we can make sure this is satisfied. □

Lemma D.3. *The convergences* (D.10) *and* (D.11) *hold true.*

Proof. By Taylor's theorem, $\mu_f(\Upsilon_i^{n,0}) - \mu_f(\Upsilon_i^{n,0,0})$ is equal to

$$\sum_{|\chi|=1} \partial^\chi \mu_f(\Upsilon_i^{n,0,0})(\Upsilon_i^{n,0} - \Upsilon_i^{n,0,0})^\chi + \sum_{|\chi|=2} \frac{1}{\chi!} \partial^\chi \mu_f(\widetilde{\upsilon}_i^n)(\Upsilon_i^{n,0} - \Upsilon_i^{n,0,0})^\chi \tag{D.21}$$

for some $\widetilde{\upsilon}_i^n$ on the line between $\Upsilon_i^{n,0}$ and $\Upsilon_i^{n,0,0}$. The expression $\Upsilon_i^{n,0} - \Upsilon_i^{n,0,0}$ contains the difference $(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} - \mathbb{E}[(\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} \mid \mathcal{F}_{i-\theta_n}^n]$ and a similar one with $\rho_{(i-\theta_n^{(q-1)})\Delta_n}$ replaced by $\sigma_{(i-1)\Delta_n}$. Inserting $\rho\rho^T$ or $\sigma\sigma^T$ at $(i-\theta_n)\Delta_n$ artificially (cf. (D.18)), we can use (D.19) and Assumption (CLT') to find that the said difference is of size at most $(\theta_n\Delta_n)^{1/2}$. This immediately leads to the bound $\mathbb{E}[\|\Upsilon_i^{n,0} - \Upsilon_i^{n,0,0}\|^2]^{1/2} \lesssim (\theta_n\Delta_n)^{1/2}$, which in turn shows that the second-order term in (D.21) is $o_{\mathbb{P}}(\sqrt{\Delta_n})$ by (C.3). Therefore, in (D.10), it remains to consider $\sqrt{\Delta_n}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=1} \partial^\chi \mu_f(\Upsilon_i^{n,0,0})(\Upsilon_i^{n,0} - \Upsilon_i^{n,0,0})^\chi$. For each i, the $\sum_{|\chi|=1}$-expression is $\mathcal{F}_i^n$-measurable and has a vanishing conditional expectation given $\mathcal{F}_{i-\theta_n}^n$. Thus, by a martingale size estimate of the type (A.6), the whole term is of size $\sqrt{\theta_n}(\theta_n\Delta_n)^{1/2}$ at most, which tends to 0 by (C.3). This proves (D.10).

For (D.11), recall $\Upsilon^{n,i}$ from (C.7) and note that the difference $(\Upsilon^{n,i} - \Upsilon_i^{n,0})_{k\ell,k'\ell'}$ equals

$$\sum_{q=1}^{Q} \Big((\rho\rho^T)_{(i-1)\Delta_n} - (\rho\rho^T)_{(i-\theta_n^{(q-1)})\Delta_n} \Big)_{kk'} \int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n} \frac{\Delta_{i+\ell-1}^n g(s) \Delta_{i+\ell'-1}^n g(s)}{\Delta_n^{2H}} \,\mathrm{d}s$$

for all $k, k' = 1, \dots, d$ and $\ell, \ell' = 1, \dots, L$. Thus, if we expand

$$\begin{aligned} \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \{\mu_f(\Upsilon^{n,i}) - \mu_f(\Upsilon_i^{n,0})\} &= \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=1} \partial^\chi \mu_f(\Upsilon^{n,i})(\Upsilon^{n,i} - \Upsilon^{n,0})^\chi \\ &\quad + \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=2} \frac{1}{\chi!} \partial^\chi \mu_f(\widehat{\upsilon}_i^n)(\Upsilon^{n,i} - \Upsilon^{n,0})^\chi, \end{aligned} \tag{D.22}$$

where $\widehat{\upsilon}_i^n$ is some point between $\Upsilon^{n,i}$ and $\Upsilon_i^{n,0}$, Hölder's inequality together with the identity (D.19) as well as the moment and regularity assumptions on ρ shows that the last sum in the above display is of size $\Delta_n^{-1/2}\sum_{q=1}^{Q}(\theta_n^{(q-1)}\Delta_n)\Delta_n^{4\theta^{(q)}(1-H)}$, which goes to 0 as $n \to \infty$; cf. (D.13). Next, recall the decomposition (2.3). As before, we incorporate the drift $t \mapsto \int_0^t \widetilde{b}_s \,\mathrm{d}s$ into $\rho^{(0)}$ so that $\rho = \rho^{(0)} + \rho^{(1)}$ with $\rho_t^{(1)} = \int_0^t \widetilde{\rho}_s \,\mathrm{d}\widetilde{W}_s$. By (D.19),

$$\begin{aligned} &\rho_{(i-1)\Delta_n}^{k\ell}\rho_{(i-1)\Delta_n}^{k'\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k\ell}\rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k'\ell} \\ &\quad = \Big(\rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k\ell} \Big\{ \rho_{(i-1)\Delta_n}^{(0),k'\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{(0),k'\ell} \Big\} + \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k'\ell} \Big\{ \rho_{(i-1)\Delta_n}^{(0),k\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{(0),k\ell} \Big\} \Big) \\ &\qquad + \Big(\rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k\ell} \Big\{ \rho_{(i-1)\Delta_n}^{(1),k'\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{(1),k'\ell} \Big\} + \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k'\ell} \Big\{ \rho_{(i-1)\Delta_n}^{(1),k\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{(1),k\ell} \Big\} \Big) \\ &\qquad + \Big(\rho_{(i-1)\Delta_n}^{k\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k\ell} \Big) \Big(\rho_{(i-1)\Delta_n}^{k'\ell} - \rho_{(i-\theta_n^{(q-1)})\Delta_n}^{k'\ell} \Big). \end{aligned}$$

The remaining term $\Delta_n^{1/2}\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=1} \partial^\chi \mu_f(\Upsilon^{n,i})(\Upsilon^{n,i} - \Upsilon^{n,0})^\chi$ in (D.22) can thus be written as $\mathbb{T}_1^n(t) + \mathbb{T}_2^n(t) + \mathbb{T}_3^n(t)$ according to this decomposition. By Hölder's inequality and the moment and regularity assumptions on ρ, $\mathbb{T}_3^n(t)$ is of size at most

$$\Delta_n^{-\frac{1}{2}} \sum_{q=1}^{Q} (\theta_n^{(q-1)}\Delta_n)\Delta_n^{2\theta^{(q)}(1-H)}, \tag{D.23}$$

which goes to 0 as $n \to \infty$ as we saw in (D.13). Similarly, thanks to the regularity property (C.1) of $\rho^{(0)}$, we further obtain $\mathbb{E}[\sup_{t\le T}|\mathbb{T}_1^n(t)|] \lesssim \Delta_n^{-1/2}\sum_{q=1}^Q(\theta_n^{(q-1)}\Delta_n)^\gamma\Delta_n^{2\theta^{(q)}(1-H)}$, and this also goes to 0 as $n\to\infty$ by our choice (D.6) of the numbers $\theta_n^{(q-1)}$. Finally,

$$\mathbb{T}_2^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{q=1}^{Q}\sum_{|\chi|=1} \partial^\chi \mu_f(\Upsilon^{n,i}) \times \left\{ \pi_{q-1}^{n,i} \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \frac{\underline{\Delta}_i^n g(s)^T \underline{\Delta}_i^n g(s)}{\Delta_n^{2H}} \mathbb{1}_{((i-\theta_n^{(q-1)})\Delta_n,(i-\theta_n^{(q)})\Delta_n)}(s)\,\mathrm{d}s \right\}^\chi,$$

where $\pi_q^{n,i} = \rho_{(i-\theta_n^{(q)})\Delta_n}(\rho^{(1)}_{(i-1)\Delta_n} - \rho^{(1)}_{(i-\theta_n^{(q)})\Delta_n})^T + (\rho^{(1)}_{(i-1)\Delta_n} - \rho^{(1)}_{(i-\theta_n^{(q)})\Delta_n})\rho^T_{(i-\theta_n^{(q)})\Delta_n}$. Define $\widetilde{\mathbb{T}}_2^n(t)$ in the same way as $\mathbb{T}_2^n(t)$ except that in the previous display, $\Upsilon^{n,i}$ is replaced by $\widetilde{\Upsilon}_{q-1}^{n,i}$, obtained from $\Upsilon^{n,i}$ by substituting $(i-\theta_n^{(q-1)})\Delta_n$ for $(i-1)\Delta_n$ everywhere. By Hölder's inequality and the regularity assumptions on ρ and σ, $\mathbb{T}_2^n(t) - \widetilde{\mathbb{T}}_2^n(t)$ is of the same size as exhibited in (D.23) and hence asymptotically negligible. Next,

$$\widetilde{\mathbb{T}}_2^n(t) = \sum_{q=1}^{Q} \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=1} \partial^\chi\mu_f(\widetilde{\Upsilon}_{q-1}^{n,i}) \biggl(\biggl\{ \Bigl(\pi_{q-1}^{n,i} - \mathbb{E}[\pi_{q-1}^{n,i} \mid \mathcal{F}^n_{i-\theta_n^{(q-1)}}] \Bigr) + \mathbb{E}[\pi_{q-1}^{n,i} \mid \mathcal{F}^n_{i-\theta_n^{(q-1)}}] \biggr\} \int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-\theta_n^{(q)})\Delta_n} \frac{\underline{\Delta}_i^n g(s)^T \underline{\Delta}_i^n g(s)}{\Delta_n^{2H}}\,\mathrm{d}s \biggr)^\chi. \tag{D.24}$$

For fixed q, the part that involves $\pi_{q-1}^{n,i} - \mathbb{E}[\pi_{q-1}^{n,i} \mid \mathcal{F}^n_{i-\theta_n^{(q-1)}}]$ is a sum where the ith summand is $\mathcal{F}^n_{i+L-1}$-measurable and has, by construction, a zero $\mathcal{F}^n_{i-\theta_n^{(q-1)}}$-conditional mean. By a martingale size estimate of the type (A.6), that part is therefore of size

$$\sum_{q=1}^{Q} \sqrt{\theta_n^{(q-1)}}(\theta_n^{(q-1)}\Delta_n)^{1/2}\Delta_n^{2\theta^{(q)}(1-H)} = \sum_{q=1}^{Q} \Delta_n^{1/2-\theta_n^{(q-1)}+2\theta^{(q)}(1-H)} \to 0$$

as $n\to\infty$ since all $\theta_n^{(q)} < \frac{1}{2}$. Clearly,

$$\mathbb{E}\Bigl[\rho^{(1),k\ell}_{(i-1)\Delta_n} - \rho^{(1),k\ell}_{(i-\theta_n^{(q-1)})\Delta_n} \Bigm| \mathcal{F}^n_{i-\theta_n^{(q-1)}}\Bigr] = \sum_{m=1}^{d} \mathbb{E}\Bigl[\int_{(i-\theta_n^{(q-1)})\Delta_n}^{(i-1)\Delta_n} \tilde{\rho}_s^{k\ell m}\,\mathrm{d}\widetilde{W}_s^m \Bigm| \mathcal{F}^n_{i-\theta_n^{(q-1)}}\Bigr] = 0.$$

Because $\rho_{(i-\theta_n^{(q)})\Delta_n}$ is $\mathcal{F}^n_{i-\theta_n^{(q)}}$-measurable, we have, in fact, $\mathbb{E}[\pi_{q-1}^{n,i} \mid \mathcal{F}^n_{i-\theta_n^{(q-1)}}] = 0$. Therefore, $\mathbb{T}_2^n(t)$ is asymptotically negligible and the proof of (D.11) is complete. □

Proof of Lemma C.4 . Recall the expressions $\mathbb{X}_1^n(t)$ and $\mathbb{X}_2^n(t)$ defined in (C.8). For a given multi-index $\chi \in \mathbb{N}_0^{(d\times L)\times(d\times L)}$, let $Q_\chi(x) = x^\chi$ for $x\in\mathbb{R}^{(d\times L)\times(d\times L)}$, which is a polynomial of degree $|\chi|$. By Taylor's theorem,

$$\mathbb{X}_1^n(t) = \Delta_n^{\frac{1}{2}} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{1}{\chi!}\partial^\chi\mu_f(\pi((i-1)\Delta_n)) \sum_{k=1}^{j}\sum_{|\chi'|=k} \frac{\Delta_n^{(j-k)(1-2H)}}{\chi'!} \times \partial^{\chi'}Q_\chi(c((i-1)\Delta_n))\Bigl\{\Upsilon^{n,i} - \pi((i-1)\Delta_n) - \Delta_n^{1-2H}c((i-1)\Delta_n)\Bigr\}^{\chi'}. \tag{D.25}$$

The key term in (D.25) is the expression in braces and we have (recall (2.6) and (1.7))

$$\begin{aligned}
&\Upsilon^{n,i} - \pi((i-1)\Delta_n) - \Delta_n^{1-2H} c((i-1)\Delta_n) \\
&\quad = (\rho\rho^T)_{(i-1)\Delta_n} \Big\{ \int_{(i-\theta_n)\Delta_n}^{(i+L-1)\Delta_n} \frac{\underline{\Delta}_i^n g(s)^T \underline{\Delta}_i^n g(s)}{\Delta_n^{2H}} \,\mathrm{d}s - (\Gamma^H_{|\ell-\ell'|})_{\ell,\ell'=1}^{L,L} \Big\} \\
&\quad = -(\rho\rho^T)_{(i-1)\Delta_n} \int_{-\infty}^{(i-\theta_n)\Delta_n} \frac{\underline{\Delta}_i^n g(s)^T \underline{\Delta}_i^n g(s)}{\Delta_n^{2H}} \,\mathrm{d}s,
\end{aligned} \tag{D.26}$$

because $\Gamma^H_{|\ell-\ell'|} = \Delta_n^{-2H} \int_{-\infty}^{\infty} \Delta_{i+\ell}^n g(s) \Delta_{i+\ell'}^n g(s)\,\mathrm{d}s$ by (B.2). The size of the last integral is $\Delta_n^{2\theta(1-H)}$ by Lemma B.1 (iii). Consequently, if we apply Hölder's inequality to (D.25), we obtain that $\mathbb{E}[\sup_{t\leq T} |\mathbb{X}_1^n(t)|] \lesssim \Delta_n^{-1/2} \sum_{j=1}^{N(H)} \sum_{k=1}^{j} \Delta_n^{(j-k)(1-2H)} \Delta_n^{k2\theta(1-H)} \lesssim \Delta_n^{-1/2+2\theta(1-H)} \to 0$ by (C.3). Using (D.26) and Assumption (CLT'), we further see that the magnitude of $\Upsilon^{n,i} - \pi((i-1)\Delta_n)$ is $\lesssim \Delta_n^{1-2H} + \Delta_n^{2\theta(1-H)}$. Thus, again by Hölder's inequality, we deduce that $\mathbb{E}[\sup_{t\leq T} |\mathbb{X}_2^n(t)|] \lesssim \Delta_n^{-1/2} (\Delta_n^{(N(H)+1)(1-2H)} + \Delta_n^{(N(H)+1)2\theta(1-H)}) \to 0$ by the definition of $N(H)$. □

Proof of Lemma C.5. The first convergence (C.9) can be shown analogously to Equation (5.3.24) in Jacod & Protter (2012) and is omitted. For (C.10), we write the left-hand side as $\sum_{j=1}^{N(H)} \mathbb{Z}_j^n(t) - \overline{\mathbb{Z}}^n(t)$ where

$$\begin{aligned}
\mathbb{Z}_j^n(t) = \Delta_n^{-\frac{1}{2}+j(1-2H)} \sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=j} \frac{1}{\chi!} \int_{(i-1)\Delta_n}^{i\Delta_n} \Big\{ &\partial^\chi \mu_f(\pi((i-1)\Delta_n)) c((i-1)\Delta_n)^\chi \\
&- \partial^\chi \mu_f(\pi(s)) c(s)^\chi \Big\} \,\mathrm{d}s,
\end{aligned}$$

$$\overline{\mathbb{Z}}^n(t) = \Delta_n^{-\frac{1}{2}} \Big(\int_0^{\theta_n\Delta_n} + \int_{([t/\Delta_n]+L-1)\Delta_n}^t \Big) \sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{1}{\chi!} \partial^\chi \mu_f(\pi(s)) \Delta_n^{j(1-2H)} c(s)^\chi \,\mathrm{d}s.$$

Using the moment assumptions on σ and ρ, since $t - ([t/\Delta_n] - L + 1)\Delta_n \leq L\Delta_n$, we readily see that $\mathbb{E}[\sup_{t\leq T} |\overline{\mathbb{Z}}^n(t)|] \lesssim \Delta_n^{-1/2}(\theta_n\Delta_n + L\Delta_n) \lesssim \Delta_n^{1/2-\theta} + \Delta_n^{1/2} \to 0$.

Let $j = 1, \dots, N(H)$ (in particular, everything in the following can be skipped if $H < \frac{1}{4}$) and consider, for $\chi \in \mathbb{N}_0^{(d\times L)\times(d\times L)}$, again the polynomial Q_χ introduced in proof of Lemma C.4. Using the mean-value theorem, we can write

$$\begin{aligned}
\mathbb{Z}_j^n(t) = \Delta_n^{-\frac{1}{2}+j(1-2H)} &\sum_{i=\theta_n+1}^{[t/\Delta_n]-L+1} \sum_{|\chi|=j} \frac{1}{\chi!} \int_{(i-1)\Delta_n}^{i\Delta_n} \sum_{|\chi_1+\chi_2|=1} \partial^{\chi+\chi_1} \mu_f(\zeta_{n,i}^1) \partial^{\chi_2} Q_\chi(\zeta_{n,i}^2) \\
&\times \{\pi((i-1)\Delta_n) - \pi(s)\}^{\chi_1} \{c((i-1)\Delta_n) - c(s)\}^{\chi_2} \,\mathrm{d}s
\end{aligned}$$

for some $\zeta_{n,i}^1$ and $\zeta_{n,i}^2$. By Hölder's inequality and Assumption (CLT), we deduce that $\mathbb{E}[\sup_{t\leq T} |\mathbb{Z}_j^n(t)|] \lesssim \Delta_n^{-1/2+j(1-2H)} \Delta_n^{-1} \Delta_n \Delta_n^{1/2} = \Delta_n^{j(1-2H)} \to 0$ for any $H < \frac{1}{2}$. □

E Proof of Theorem 3.2

Since φ is invertible, we can write

$$\begin{aligned}
&H = \varphi^{-1}\Big(\frac{\langle a, \Gamma^H \rangle \Pi_t}{\langle b, \Gamma^H \rangle \Pi_t} \Big) = G(\langle a, \Gamma^H \rangle \Pi_t, \langle b, \Gamma^H \rangle \Pi_t), \\
&\widetilde{H}^n = G(\langle a, \widehat{V}_t^n \rangle, \langle b, \widehat{V}_t^n \rangle) = G(\langle a, V_t^n \rangle, \langle b, V_t^n \rangle), \qquad G(x,y) = \varphi^{-1}(x/y).
\end{aligned} \tag{E.1}$$

As G is infinitely differentiable on $\mathbb{R}\times(\mathbb{R}\setminus\{0\})$, we can expand $\widetilde{H}^n$ in a Taylor sum around $(\langle a,\Gamma^H\rangle\Pi_t,\langle b,\Gamma^H\rangle\Pi_t)$ and obtain

$$\begin{aligned}\widetilde{H}^n-H&=\sum_{|\chi|=1}\partial^\chi G(\langle a,\Gamma^H\rangle\Pi_t,\langle b,\Gamma^H\rangle\Pi_t)(\langle a,V^n_t-\Gamma^H\Pi_t\rangle,\langle b,V^n_t-\Gamma^H\Pi_t\rangle)^\chi+\mathbb{H}^n,\\ \mathbb{H}^n&=\sum_{|\chi|=2}\frac{\partial^\chi G(\alpha^n)}{\chi!}(\langle a,V^n_t\Gamma^H\Pi_t\rangle,\langle b,V^n_t\Gamma^H\Pi_t\rangle)^\chi,\end{aligned}\tag{E.2}$$

where $\chi\in\mathbb{N}_0^2$ and α^n is a point between $(\langle a,V^n_t\rangle,\langle b,V^n_t\rangle)$ and $(\langle a,\Gamma^H\rangle\Pi_t,\langle b,\Gamma^H\rangle\Pi_t)$. By straightforward computations,

$$\partial^{(1,0)}G(x,y)=(\varphi^{-1})'(x/y)y^{-1}\quad\text{and}\quad\partial^{(0,1)}G(x,y)=-(\varphi^{-1})'(x/y)xy^{-2}.\tag{E.3}$$

Therefore, (E.2) becomes

$$\widetilde{H}^n-H=\frac{(\varphi^{-1})'(\varphi(H))}{\langle b,\Gamma^H\rangle\Pi_t}\{\langle a,V^n_t-\Gamma^H\Pi_t\rangle-\varphi(H)\langle b,V^n_t-\Gamma^H\Pi_t\rangle\}+\mathbb{H}^n.\tag{E.4}$$

Because $H\in(0,\frac14)$ or $a_0=b_0=0$, the term inside the braces in the last line can be written as $\{a^T-\varphi(H)b^T\}\{V^n_t-\Gamma^H\int_0^t\rho_s^2\,\mathrm{d}s-e_1\int_0^t\sigma_s^2\,\mathrm{d}s\,\Delta_n^{1-2H}\,\mathbb{1}_{[\frac14,\frac12)}(H)\}$. Moreover, by Corollary 3.1, the term $\mathbb{H}^n$ is of magnitude Δ_n and hence,

$$\begin{aligned}\Delta_n^{-\frac12}(\widetilde{H}^n-H)&=\frac{(\varphi^{-1})'(\varphi(H))}{\langle b,\Gamma^H\rangle\Pi_t}\{a^T-\varphi(H)b^T\}\\ &\qquad\times\Delta_n^{-\frac12}\Big\{V^n_t-\Gamma^H\int_0^t\rho_s^2\,\mathrm{d}s-e_1\int_0^t\sigma_s^2\,\mathrm{d}s\,\Delta_n^{1-2H}\,\mathbb{1}_{[\frac14,\frac12)}(H)\Big\}+\Delta_n^{-\frac12}\mathbb{H}^n\\ &\stackrel{\text{st}}{\longrightarrow}\frac{(\varphi^{-1})'(\varphi(H))}{\langle b,\Gamma^H\rangle\Pi_t}\{a^T-\varphi(H)b^T\}\mathcal{Z}_t\sim\mathcal{N}\Big(0,\mathrm{Var}_{H,0}\frac{\int_0^t\rho_s^4\,\mathrm{d}s}{(\int_0^t\rho_s^2\,\mathrm{d}s)^2}\Big),\end{aligned}$$

which proves (3.5).

We now turn to the convergence stated in (3.7) when $H>\frac14$. We decompose

$$\begin{aligned}V^n_{0,t}-\frac{\langle c,V^n_t\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}&=\{V^n_{0,t}-\Pi_t\}-\frac{\langle c,V^n_t-\Gamma^H\Pi_t\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}+\Pi_t\frac{\langle c,\Gamma^{\widetilde{H}^n}-\Gamma^H\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}\\ &=\{V^n_{0,t}-\Pi_t\}-\frac{c^T\{V^n_t-\Gamma^H\Pi_t\}}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}+\Pi_t\frac{\langle c,\partial_H\Gamma^{\widetilde{H}^n}\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}\{\widetilde{H}^n-H\}+\mathbb{V}^n,\\ \mathbb{V}^n&=\frac12\Pi_t\frac{\langle c,\partial_{HH}\Gamma^{\beta^n}\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}\{\widetilde{H}^n-H\}^2,\end{aligned}\tag{E.5}$$

where $\partial_{HH}\Gamma^H$ is the second derivative of $H\mapsto(\Gamma^H_0,\dots,\Gamma^H_R)$ evaluated at H and β^n is somewhere between $\widetilde{H}^n$ and H. Since $c_0\neq0$, the first two terms in the second line of (E.5) are of magnitude Δ_n^{1-2H}, while the third is of magnitude $\Delta_n^{1/2}$ by our first result (3.5). Finally, $\mathbb{V}^n$ is of magnitude Δ_n, so using Corollary 3.1, we deduce that

$$\Delta_n^{2H-1}\Big\{V^n_{0,t}-\frac{\langle c,V^n_t\rangle}{\langle c,\Gamma^{\widetilde{H}^n}\rangle}\Big\}\stackrel{\mathbb{P}}{\longrightarrow}C_t-\frac{1}{\langle c,\Gamma^H\rangle}\langle c,e_1\rangle C_t=\Big(1-\frac{c_0}{\langle c,\Gamma^H\rangle}\Big)C_t.\tag{E.6}$$

Reusing (E.4) and recalling that $a_0 = b_0 = 0$, we further have that

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}}\Big\{V_{0,t}^n - \frac{\langle c, V_t^n\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} - \Big(1 - \frac{c_0}{\langle c, \Gamma^{\widetilde{H}^n}\rangle}\Big) C_t\,\Delta_n^{1-2H}\Big\} \\
&\quad = \Delta_n^{-\frac{1}{2}}\{V_{0,t}^n - \Pi_t - C_t\,\Delta_n^{1-2H}\} - \frac{\Delta_n^{-\frac{1}{2}} c^T\{V_t^n - \Gamma^H \Pi_t - e_1 C_t\,\Delta_n^{1-2H}\}}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} \\
&\qquad + \Pi_t \frac{\langle c, \partial_H \Gamma^{\widetilde{H}^n}\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} \Delta_n^{-\frac{1}{2}}\{\widetilde{H}^n - H\} + \Delta_n^{-\frac{1}{2}} \mathbb{V}^n \\
&\quad = \Big(e_1^T - \frac{c^T}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} + \Pi_t \frac{\langle c, \partial_H \Gamma^{\widetilde{H}^n}\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle \Pi_t}\{a^T - \varphi(H) b^T\}\Big) \\
&\qquad \times \Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t - e_1 C_t\,\Delta_n^{1-2H}\} + \Delta_n^{-\frac{1}{2}}\Big(\Pi_t \frac{\langle c, \partial_H \Gamma^{\widetilde{H}^n}\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle} \mathbb{H}^n + \mathbb{V}^n\Big) \\
&\quad \stackrel{\text{st}}{\longrightarrow} \Big(e_1^T - \frac{c^T}{\langle c, \Gamma^H\rangle} + \frac{\langle c, \partial_H \Gamma^H\rangle}{\langle c, \Gamma^H\rangle} \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle}\{a^T - \varphi(H) b^T\}\Big) \mathcal{Z}_t.
\end{aligned}$$

It remains to normalize the left-hand side of (E.6) in order to obtain (3.7):

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}+(1-2H)}\Big\{\Big(\widehat{V}_{0,t}^n - \frac{\langle c, \widehat{V}_t^n\rangle}{\langle c, \Gamma^{\widetilde{H}^n}\rangle}\Big)\Big(1 - \frac{c_0}{\langle c, \Gamma^{\widetilde{H}^n}\rangle}\Big)^{-1} - C_t\Big\} \\
&\quad \stackrel{\text{st}}{\longrightarrow} \Big(1 - \frac{c_0}{\langle c, \Gamma^H\rangle}\Big)^{-1}\Big(e_1^T - \frac{c^T}{\langle c, \Gamma^H\rangle} + \frac{\langle c, \partial_H \Gamma^H\rangle}{\langle c, \Gamma^H\rangle} \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle}\{a^T - \varphi(H) b^T\}\Big) \mathcal{Z}_t \\
&\quad \sim \mathcal{N}\Big(0, \mathrm{Var}_C \int_0^t \rho_s^4\, \mathrm{d}s\Big).
\end{aligned}$$

Finally, we tackle (3.12). We use the mean-value theorem to decompose

$$\begin{aligned}
\Delta_n^{-\frac{1}{2}}\Big(\frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \Pi_t\Big) &= \Delta_n^{-\frac{1}{2}} \frac{\langle a, V_t^n - \Gamma^H \Pi_t\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \frac{\Pi_t}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Delta_n^{-\frac{1}{2}} \langle a, \Gamma^{\widetilde{H}^n} - \Gamma^H\rangle \\
&= \frac{a^T}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t\} - \frac{\Pi_t \langle a, \partial_H \Gamma^{\widetilde{\beta}^n}\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Delta_n^{-\frac{1}{2}}\{\widetilde{H}^n - H\},
\end{aligned} \tag{E.7}$$

where $\widetilde{\beta}^n$ is between $\widetilde{H}^n$ and H and therefore satisfies $\widetilde{\beta}^n \stackrel{\mathbb{P}}{\longrightarrow} H$. As before, because $\frac{1}{4} < H < \frac{1}{2}$ or $a_0 = b_0 = 0$, we have $V_t^n - \Gamma^H \Pi_t = V_t^n - \Gamma^H \Pi_t - C_t\,\Delta_n^{1-2H}\, \mathbb{1}_{[\frac{1}{4}, \frac{1}{2})}(H)$. Using Corollary 3.1 and our first result (3.5), we infer that $\Delta_n^{-1/2}(\langle a, V_t^n\rangle / \langle a, \Gamma^{\widetilde{H}^n}\rangle - \Pi_t)$ converges stably in distribution. Applying again the mean-value theorem, this time on the function $H \mapsto \Delta_n^{-2H}$, and recalling the identity $\Delta_n^{1-2H} \widehat{V}_{r,t}^n = V_{r,t}^n$, we further obtain

$$\begin{aligned}
\Delta_n^{-\frac{1}{2}}(\widehat{\Pi}_t^n - \Pi_t) &= \Delta_n^{-\frac{1}{2}}\Big(\frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \Pi_t\Big) + \Delta_n^{-\frac{1}{2}} \frac{\langle a, \widehat{V}_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\{\Delta_n^{1-2\widetilde{H}^n} - \Delta_n^{1-2H}\} \\
&= \Delta_n^{-\frac{1}{2}}\Big(\frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \Pi_t\Big) - 2\frac{\langle a, \widehat{V}_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Delta_n^{1-2H} (\log \Delta_n) \Delta_n^{2(H - \overline{\beta}^n)} \Delta_n^{-\frac{1}{2}}\{\widetilde{H}^n - H\}
\end{aligned}$$

for another point $\overline{\beta}^n$ between $\widetilde{H}^n$ and H. By (3.5), $\overline{\beta}^n$ converges to H at a rate of $\Delta_n^{1/2}$. Therefore, $\Delta_n^{2(H-\overline{\beta}^n)} \to 1$ as $n \to \infty$. Normalizing by $\log \Delta_n$, we conclude from (3.5) that

$$\frac{\Delta_n^{-\frac{1}{2}}}{\log \Delta_n}(\widehat{\Pi}_t^n - \Pi_t) = \frac{\Delta_n^{-\frac{1}{2}}}{\log \Delta_n}\Big(\frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \Pi_t\Big) - 2\frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\Delta_n^{2(H-\overline{\beta}^n)}\Delta_n^{-\frac{1}{2}}\{\widetilde{H}^n - H\}$$
$$\xrightarrow{\text{st}} \mathcal{N}\Big(0, 4\,\mathrm{Var}_{H,0}\int_0^t \rho_s^4\,\mathrm{d}s\Big).$$

This completes the proof of Theorem 3.2. □

F Estimators based on quadratic variation

F.1 A consistent but not asymptotically normal estimator of H

To simplify the exposition, we assume that at least one of a_0 and b_0 is zero. By symmetry, we shall consider the case where

$$a_0 \neq 0, \qquad b_0 = 0. \tag{F.1}$$

Also, again to simplify the argument and because this is not really a severe restriction from a statistical point of view, we shall assume that the true value of H satisfies

$$H \in (\tfrac{1}{4}, \tfrac{1}{2}) \setminus \mathcal{H}, \tag{F.2}$$

where $\mathcal{H}$ is the set from (2.10).

Proposition F.1. *Let $H \in (\frac{1}{4}, \frac{1}{2}) \setminus \mathcal{H}$ and suppose that $a, b \in \mathbb{R}^{1+R}$ satisfy (F.1) and are such that φ from (3.4) is invertible. Recalling that $N(H) = [1/(2-4H)]$, we further define for $j = 1, \ldots, N(H)$,*

$$\Phi_j^n = \Phi_j^n(R, a, b, \widehat{V}_t^n, \widetilde{H}^n) = \frac{(-1)^j}{j!}(\varphi^{-1})^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b, \widehat{V}_t^n\rangle^j}. \tag{F.3}$$

Then $\widetilde{H}^n$, as defined in (3.4), satisfies

$$\Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n C_t^j\Big\} \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{H,0}\,\frac{\int_0^t \rho_s^4\,\mathrm{d}s}{(\int_0^t \rho_s^2\,\mathrm{d}s)^2}\Big), \tag{F.4}$$

where $\mathrm{Var}_{H,0}$ is defined in (3.6).

For each j, the term Φ_j^n is of order $\Delta_n^{j(1-2H)}$. As a result, while $\widetilde{H}^n$ is consistent for H, it is affected by many higher-order asymptotic bias terms that depend on C_t. So our next goal is to find consistent estimators of C_t that we can use to correct H.

F.2 A consistent but not asymptotically normal estimator of C_t

With a first estimator of H at hand, we can now construct an estimator of C_t by removing the first-order limit of $\widehat{V}^n_{0,t}$, hereby replacing H by $\widetilde{H}^n$ throughout. Doing so, we have to employ an estimator of Π_t, the integrated noise volatility. To avoid even more higher-order bias terms, we need one with convergence rate $\sqrt{\Delta_n}$. One possibility is to use the estimator $\widehat{\Pi}^n_t$ from Theorem 3.2, constructed from an additional pair of weights a^0 and b^0 with $a^0_0 = b^0_0 = 0$. Note that even in the noise-free case, where $\rho = 0$, the estimator $\widehat{\Pi}^n_t$ from Theorem 3.2 converges to the desired limit 0 in probability.

Proposition F.2. *In addition to $a, b \in \mathbb{R}^{1+R}$ satisfying (F.1), choose $a^0, b^0 \in \mathbb{R}^{1+R}$ with $a^0_0 = b^0_0 = 0$ and let*

$$\widehat{P}^n_t = \frac{\langle a^0, \widehat{V}^n_t\rangle}{\langle a^0, \Gamma^{\widetilde{H}^{n,0}}\rangle}, \qquad \widetilde{H}^{n,0} = \varphi^{-1}\Big(\frac{\langle a^0, \widehat{V}^n_t\rangle}{\langle b^0, \widehat{V}^n_t\rangle}\Big). \tag{F.5}$$

Further define

$$\widetilde{C}^{n,1}_t = \Big\{\widehat{V}^n_{0,t} - \frac{\langle a, \widehat{V}^n_t\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\Big\}\Theta(\widehat{V}^n_t, \widetilde{H}^n, \widetilde{H}^{n,0})^{-1}, \tag{F.6}$$

where

$$\begin{aligned}\Theta(\widehat{V}^n_t, \widetilde{H}^n, \widetilde{H}^{n,0}) &= \Theta(R, a, b, a^0, b^0, \widehat{V}^n_t, \widetilde{H}^n, \widetilde{H}^{n,0})\\ &= 1 - \frac{a_0}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} + \frac{\widehat{P}^n_t}{\langle b, \widehat{V}^n_t\rangle}\frac{a_0\psi'(\varphi(\widetilde{H}^n))}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\end{aligned} \tag{F.7}$$

and

$$\psi(y) = \langle a, \Gamma^{\varphi^{-1}(y)}\rangle, \qquad y \in \mathbb{R}. \tag{F.8}$$

Then, under the assumptions made in Proposition F.1,

$$\Delta_n^{\frac{1}{2}-2H}\Big\{\widetilde{C}^{n,1}_t - C_t + \sum_{j=2}^{N(H)} \Psi^n_j C^j_t\Big\} \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{C,1}\int_0^t \rho_s^4\,\mathrm{d}s\Big), \tag{F.9}$$

where

$$\begin{aligned}\Psi^n_j &= \Psi^n_j(R, a, b, a^0, b^0, \widehat{V}^n_t, \widetilde{H}^n, \widetilde{H}^{n,0})\\ &= \frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a^j_0}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\frac{\widehat{P}^n_t}{\langle b, \widehat{V}^n_t\rangle^j}\Theta(\widehat{V}^n_t, \widetilde{H}^n, \widetilde{H}^{n,0})^{-1}\end{aligned} \tag{F.10}$$

for $j = 2, \dots, N(H)$ and

$$\mathrm{Var}_{C,1} = \mathrm{Var}_{C,1}(R, a, b, H) = u_1^T \mathcal{C}^H u_1, \tag{F.11}$$

$$\begin{aligned}u_1 &= \Big(e_1 - \frac{a}{\langle a, \Gamma^H\rangle} + \frac{\psi'(\varphi(H))}{\langle a, \Gamma^H\rangle\langle b, \Gamma^H\rangle}(a - \varphi(H)b)\Big)\\ &\quad \times \Big(1 - \frac{a_0}{\langle a, \Gamma^H\rangle} + \frac{\psi'(\varphi(H))}{\langle a, \Gamma^H\rangle\langle b, \Gamma^H\rangle}a_0\Big)^{-1},\end{aligned} \tag{F.12}$$

and $\mathcal{C}^H$ is the matrix in (3.2).

Note that Ψ^n_j is of magnitude $\Delta_n^{(j-1)(1-2H)}$. Thus, just as for the initial estimator of H, the estimator $\widetilde{C}^{n,1}_t$ is consistent but has higher-order bias terms.

F.3 The first asymptotically normal estimators of H and C_t

What is different between the two initial estimators of H and C_t is that in (F.9) the bias terms only hinge on C_t, the quantity that $\widetilde{C}_t^{n,1}$ is supposed to estimate in the first place. Therefore, we can set up an iteration procedure to correct $\widetilde{C}_t^{n,1}$.

Proposition F.3. *Recall that* $N(H) = [1/(2-4H)]$ *and define*

$$\widetilde{C}_t^{n,\ell+1} = \widetilde{C}_t^{n,1} + \sum_{j=2}^{\ell+1} \Psi_j^n (\widetilde{C}_t^{n,\ell-j+2})^j, \qquad \ell \geq 0, \tag{F.13}$$

and

$$\widehat{C}_t^{n,1} = \widetilde{C}_t^{n,N(\widetilde{H}^n)}. \tag{F.14}$$

Then we have that

$$\Delta_n^{\frac{1}{2}-2H}\Big\{\widetilde{C}_t^{n,1} - C_t + \sum_{j=2}^{N(\widetilde{H}^n)} \Psi_j^n \Big(\widetilde{C}_t^{n,N(\widetilde{H}^n)-j+1}\Big)^j\Big\} = \Delta_n^{\frac{1}{2}-2H}(\widehat{C}_t^{n,1} - C_t)$$
$$\xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{C,1}\int_0^t \rho_s^4\,\mathrm{d}s\Big)$$

with the same $\mathrm{Var}_{C,1}$ *as in* (F.11)*.*

The corrected estimator $\widehat{C}_t^{n,1}$ is our first consistent and asymptotically mixed normal estimator for C_t in the setting of (F.1). With a bias-free estimator of C_t at hand, we can now proceed to correcting the initial estimator $\widetilde{H}^n$ of H.

Proposition F.4. *Recall* $\widetilde{H}^n$ *in* (3.4) *and define*

$$\widehat{H}_1^n = \widetilde{H}^n + \sum_{j=1}^{N(\widetilde{H}^n)} \Phi_j^n (\widehat{C}_t^{n,1})^j \tag{F.15}$$

with Φ_j^n *as in* (F.3)*. Then*

$$\Delta_n^{-\frac{1}{2}}(\widehat{H}_1^n - H) \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{H,1}\frac{\int_0^t \rho_s^4\,\mathrm{d}s}{(\int_0^t \rho_s^2\,\mathrm{d}s)^2}\Big),$$

where

$$\mathrm{Var}_{H,1} = \mathrm{Var}_{H,1}(R,a,b,H) = w_1^T \mathcal{C}^H w_1, \qquad w_1 = \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H\rangle}\{a - \varphi(H)b - a_0 u_1\},$$

and the vector u_1 *is exactly as in* (F.11) *and the matrix* $\mathcal{C}^H$ *as in* (3.2)*.*

F.4 A multi-step algorithm

Even though $\widehat{C}_t^{n,1}$ and $\widehat{H}_1^n$ from Propositions F.3 and F.4 are rate-optimal and asymptotically bias-free estimators of C_t and H, respectively, we can still do better: The estimator $\widehat{C}_t^{n,1}$ is based on the initial estimator $\widetilde{C}_t^{n,1}$ from (F.6), which in turn is based on the initial estimator $\widetilde{H}^n$ of H. Now that we have a better estimator of H, namely $\widehat{H}_1^n$, the idea is to use $\widehat{H}_1^n$ to construct an updated estimator, say, $\widetilde{C}_t^{n,2}$, of C_t. And with this updated estimator of C_t, we next update $\widehat{H}_1^n$ to, say, $\widehat{H}_2^n$, which we can then use to update $\widetilde{C}_t^{n,2}$ again, and so on. A related approach was used in Li et al. (2020).

Proposition F.5. *For $k = 2, \dots, m$ where $m \geq 2$ is an integer, we define iteratively*

$$\widehat{C}_t^{n,k} = \Big\{ \widehat{V}_{0,t}^n - \frac{\langle a, \widehat{V}_t^n \rangle}{\langle a, \Gamma^{\widehat{H}_{k-1}^n} \rangle} \Big\} \Big(1 - \frac{a_0}{\langle a, \Gamma^{\widehat{H}_{k-1}^n} \rangle} \Big)^{-1} \tag{F.16}$$

and

$$\widehat{H}_k^n = \widetilde{H}^n + \sum_{j=1}^{N(\widehat{H}_{k-1}^n)} \Phi_j^n (\widehat{C}_t^{n,k})^j. \tag{F.17}$$

Then

$$\Delta_n^{-\frac{1}{2}} (\widehat{H}_k^n - H) \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{H,k} \frac{\int_0^t \rho_s^4 \, \mathrm{d}s}{(\int_0^t \rho_s^2 \, \mathrm{d}s)^2}\Big), \tag{F.18}$$

$$\Delta_n^{\frac{1}{2}-2H} (\widehat{C}_t^{n,k} - C_t) \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_{C,k} \int_0^t \rho_s^4 \, \mathrm{d}s\Big), \tag{F.19}$$

where, for each $k = 2, \dots, m$,

$$\mathrm{Var}_{H,k} = \mathrm{Var}_{H,k}(R, a, b, H) = w_k^T \mathcal{C}^H w_k, \qquad \mathrm{Var}_{C,k} = \mathrm{Var}_{C,k}(R, a, b, H) = u_k^T \mathcal{C}^H u_k,$$

and

$$u_k = \Big(e_1 - \frac{a}{\langle a, \Gamma^H \rangle} + \frac{\langle a, \partial_H \Gamma^H \rangle}{\langle a, \Gamma^H \rangle} w_{k-1} \Big) \Big(1 - \frac{a_0}{\langle a, \Gamma^H \rangle} \Big)^{-1},$$
$$w_k = \frac{(\varphi^{-1})'(\varphi(H))}{\langle b, \Gamma^H \rangle} \{ a - \varphi(H) b - a_0 u_k \}.$$

Our final estimator of H is

$$\widehat{H}^n = \widehat{H}_m^n. \tag{F.20}$$

For later references, let us define

$$\mathrm{Var}_H = \mathrm{Var}_H(R, a, b, H) = \mathrm{Var}_{H,m}(R, a, b, H). \tag{F.21}$$

The next theorem exhibits our final estimators for C_t and Π_t.

Theorem F.6. *Choose $c \in \mathbb{R}^{1+R}$ and define*

$$\widehat{C}_t^n = \Big\{ \widehat{V}_{0,t}^n - \frac{\langle c, \widehat{V}_t^n \rangle}{\langle c, \Gamma^{\widehat{H}^n} \rangle} \Big\} \Big(1 - \frac{c_0}{\langle c, \Gamma^{\widehat{H}^n} \rangle} \Big)^{-1}, \tag{F.22}$$

$$\widehat{\Pi}_t^n = \Big\{ \frac{\langle a, \widehat{V}_t^n \rangle}{\langle a, \Gamma^{\widehat{H}^n} \rangle} - \frac{a_0}{\langle a, \Gamma^{\widehat{H}^n} \rangle} \widehat{C}_t^n \Big\} \Delta_n^{1-2\widehat{H}^n}. \tag{F.23}$$

Then

$$\Delta_n^{\frac{1}{2}-2H} (\widehat{C}_t^n - C_t) \xrightarrow{\text{st}} \mathcal{N}\Big(0, \mathrm{Var}_C \int_0^t \rho_s^4 \, \mathrm{d}s\Big), \tag{F.24}$$

$$\frac{\Delta_n^{-\frac{1}{2}}}{|\log \Delta_n|} (\widehat{\Pi}_t^n - \Pi_t) \xrightarrow{\text{st}} \mathcal{N}\Big(0, 4 \,\mathrm{Var}_H \int_0^t \rho_s^4 \, \mathrm{d}s\Big), \tag{F.25}$$

where

$$\begin{aligned} \mathrm{Var}_C &= \mathrm{Var}_C(R, a, b, c, H) = u^T \mathcal{C}^H u, \\ u &= \Big(e_1 - \frac{c}{\langle c, \Gamma^H \rangle} + \frac{\langle c, \partial_H \Gamma^H \rangle}{\langle c, \Gamma^H \rangle} w_m \Big) \Big(1 - \frac{c_0}{\langle c, \Gamma^H \rangle} \Big)^{-1}. \end{aligned} \tag{F.26}$$

F.5 Proofs

Proof of Proposition F.1. Starting from (E.1), we expand

$$\Delta_n^{-\frac{1}{2}}(\widetilde{H}^n - H) = -\sum_{j=1}^{N(H)} \sum_{|\chi|=j} \frac{\partial^\chi G(\langle a, V_t^n\rangle, \langle b, V_t^n\rangle)}{\chi!}(-1)^j$$
$$\times \Delta_n^{-\frac{1}{2}}(\langle a, V_t^n - \Gamma^H \Pi_t\rangle, \langle b, V_t^n - \Gamma^H \Pi_t\rangle)^\chi - \mathbb{I}^n, \tag{F.27}$$
$$\mathbb{I}^n = \sum_{|\chi|=N(H)+1} \frac{\partial^\chi G(\overline{\alpha}^n)}{\chi!}(-1)^{|\chi|}\Delta_n^{-\frac{1}{2}}(\langle a, V_t^n - \Gamma^H \Pi_t\rangle, \langle b, V_t^n - \Gamma^H \Pi_t\rangle)^\chi,$$

where $\chi \in \mathbb{N}_0^2$ and $\overline{\alpha}^n$ is a point between $(\langle a, \Gamma^H\rangle \Pi_t, \langle b, \Gamma^H\rangle \Pi_t)$ and $(\langle a, V_t^n\rangle, \langle b, V_t^n\rangle)$. In contrast to the proof of (3.5), we expanded $\widetilde{H}^n$ around $(\langle a, \Gamma^H\rangle \Pi_t, \langle b, \Gamma^H\rangle \Pi_t)$ and not $(\langle a, V_t^n\rangle, \langle b, V_t^n\rangle)$. We consider the terms where $\chi = (j, 0)$ for some $j = 1, \ldots, N(H)$ and where $\chi = (0, 1)$ separately. In the first case, we have $\partial^\chi G(x, y) = (\varphi^{-1})^{(j)}(x/y)y^{-j}$ for all $\chi = (j, 0)$ and $j \geq 1$; in the second case, $\partial^\chi G(x, y)$ was computed in (E.3). With that in mind, and recalling (F.3), we have that

$$\Delta_n^{-\frac{1}{2}}\left(\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n C_t^j\right)$$
$$= (\varphi^{-1})'(\varphi(\widetilde{H}^n))\frac{a^T - \varphi(\widetilde{H}^n)b^T}{\langle b, V_t^n\rangle}\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t - e_1 C_t\, \Delta_n^{1-2H}\}$$
$$+ \sum_{j=2}^{N(H)} \frac{(-1)^{j+1}}{j!}(\varphi^{-1})^{(j)}(\varphi(\widetilde{H}^n))\frac{1}{\langle b, V_t^n\rangle^j}\Delta_n^{-\frac{1}{2}}\left\{\langle a, V_t^n - \Gamma^H \Pi_t\rangle^j - a_0^j C_t^j \Delta_n^{j(1-2H)}\right\}$$
$$- \sum_{j=2}^{N(H)} \sum_{\chi \neq (j,0)} \frac{\partial^\chi G(\langle a, V_t^n\rangle, \langle b, V_t^n\rangle)}{\chi!}(-1)^j \Delta_n^{-\frac{1}{2}}(\langle a, V_t^n - \Gamma^H \Pi_t\rangle, \langle b, V_t^n - \Gamma^H \Pi_t\rangle)^\chi$$
$$- \mathbb{I}^n. \tag{F.28}$$

By Corollary 3.1, one can see that $\Delta_n^{-1/2}\{\langle a, V_t^n - \Gamma^H \Pi_t\rangle^j - a_0^j C_t^j \Delta_n^{j(1-2H)}\}$ is of magnitude $\Delta_n^{(j-1)/2}$. Thus, the second term on the right-hand side of (F.28) is asymptotically negligible. And so are the third term in (F.28) and $\mathbb{I}^n$: For any $\chi = (j - i, i) \in \mathbb{N}_0^2$, Corollary 3.1 and assumption (F.1) imply that $\Delta_n^{-1/2}(\langle a, V_t^n - \Gamma^H \Pi_t\rangle, \langle b, V_t^n - \Gamma^H \Pi_t\rangle)^\chi$ is of magnitude $\Delta_n^{(j-i)(1-2H)+i/2-1/2}$ and therefore asymptotically negligible as soon as $i \geq 1$ and $j - i \geq 1$. Similarly, $\mathbb{I}^n$ is of magnitude at most $\Delta_n^{(N(H)+1)(1-2H)-1/2}$, which goes to 0 by the definition of $N(H)$. Altogether, we obtain by Corollary 3.1 that

$$\Delta_n^{-\frac{1}{2}}\left(\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n C_t^j\right) \xrightarrow{\text{st}} (\varphi^{-1})'(\varphi(H))\frac{a^T - \varphi(H)b^T}{\langle b, \Gamma^H\rangle \Pi_t}\mathcal{Z}_t,$$

which concludes the proof. □

Proof of Proposition F.2. We start similarly to the proof of (3.8) and decompose

$$V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} = \{V_{0,t}^n - \Pi_t\} - \frac{\langle a, V_t^n - \Gamma^H \Pi_t\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} + \frac{\Pi_t \langle a, \Gamma^{\widetilde{H}^n} - \Gamma^H\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}. \tag{F.29}$$

We further analyze the last term in the above display and write

$$\langle a,\Gamma^{\widetilde{H}^n}\rangle = K(\langle a,\widehat{V}^n_t\rangle,\langle b,\widehat{V}^n_t\rangle) = K(\langle a,V^n_t\rangle,\langle b,V^n_t\rangle),$$
$$\langle a,\Gamma^{H}\rangle = K(\langle a,\Gamma^H\rangle,\langle b,\Gamma^H\rangle) = K(\langle a,\Gamma^H\rangle\Pi_t,\langle b,\Gamma^H\rangle\Pi_t),$$

where $K(x,y)=\psi(x/y)$ and ψ is the function from (F.8). We now expand $\langle a,\Gamma^H\rangle$ in a Taylor sum around the point $(\langle a,V^n_t\rangle,\langle b,V^n_t\rangle)$ up to order $N(H)$, singling out the two first-order derivatives as well as the derivatives $\partial^{(j,0)}$: noting that $\partial^{(j,0)}K(x,y)=\psi^{(j)}(x/y)y^{-j}$ for $j\geq 1$ and $\partial^{(0,1)}K(x,y)=-\psi'(x/y)xy^{-2}$, we have that

$$\begin{aligned}\langle a,\Gamma^{\widetilde{H}^n}-\Gamma^H\rangle &= \psi'(\varphi(\widetilde{H}^n))\frac{1}{\langle b,V^n_t\rangle}\Big(\langle a,V^n_t-\Gamma^H\Pi_t\rangle-\varphi(\widetilde{H}^n)\langle b,V^n_t-\Gamma^H\Pi_t\rangle\Big)\\ &\quad-\sum_{j=2}^{N(H)}\frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{1}{\langle b,V^n_t\rangle^j}\langle a,V^n_t-\Gamma^H\Pi_t\rangle^j-\mathbb{J}^n,\end{aligned}\tag{F.30}$$

where

$$\begin{aligned}\mathbb{J}^n &= \sum_{j=2}^{N(H)}\sum_{\chi\neq(j,0)}\frac{\partial^\chi K(\langle a,V^n_t\rangle,\langle b,V^n_t\rangle)}{\chi!}(-1)^j\Delta_n^{-\frac12}(\langle a,V^n_t-\Gamma^H\Pi_t\rangle,\langle b,V^n_t-\Gamma^H\Pi_t\rangle)^\chi\\ &\quad+\sum_{|\chi|=N(H)+1}\frac{\partial^\chi K(\widetilde{\alpha}^n)}{\chi!}(-1)^{|\chi|}\Delta_n^{-\frac12}(\langle a,V^n_t-\Gamma^H\Pi_t\rangle,\langle b,V^n_t-\Gamma^H\Pi_t\rangle)^\chi\end{aligned}$$

and $\widetilde{\alpha}^n$ is between $(\langle a,\Gamma^H\rangle\Pi_t,\langle b,\Gamma^H\rangle\Pi_t)$ and $(\langle a,V^n_t\rangle,\langle b,V^n_t\rangle)$. Using (F.29) for the first and (F.30) for the second equality, we find that

$$\begin{aligned}&\Delta_n^{-\frac12}\bigg(\Big\{V^n_{0,t}-\frac{\langle a,V^n_t\rangle}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\Big\}-\Big(1-\frac{a_0}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}+\frac{\Pi_t\psi'(\varphi(\widetilde{H}^n))}{\langle a,\Gamma^{\widetilde{H}^n}\rangle\langle b,V^n_t\rangle}a_0\Big)C_t\,\Delta_n^{1-2H}\\ &\qquad+\frac{\Pi_t}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\sum_{j=2}^{N(H)}\frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b,V^n_t\rangle^j}C_t^j\,\Delta_n^{j(1-2H)}\bigg)\\ &\quad=\Delta_n^{-\frac12}\{V^n_{0,t}-\Pi_t-C_t\Delta_n^{1-2H}\}-\frac{1}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\Delta_n^{-\frac12}\{\langle a,V^n_t-\Gamma^H\Pi_t\rangle-a_0C_t\,\Delta_n^{1-2H}\}\\ &\qquad+\frac{\Pi_t}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\Delta_n^{-\frac12}\Big\{\langle a,\Gamma^{\widetilde{H}^n}-\Gamma^H\rangle+\sum_{j=1}^{N(H)}\frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b,V^n_t\rangle^j}C_t^j\,\Delta_n^{j(1-2H)}\Big\}\\ &\quad=\Big\{e_1^T-\frac{a^T}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}+\frac{\Pi_t\psi'(\varphi(\widetilde{H}^n))}{\langle a,\Gamma^{\widetilde{H}^n}\rangle\langle b,V^n_t\rangle}(a^T-\varphi(\widetilde{H}^n)b^T)\Big\}\Delta_n^{-\frac12}\{V^n_t-\Gamma^H\Pi_t-e_1C_t\Delta_n^{1-2H}\}\\ &\qquad-\frac{\Pi_t}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\sum_{j=2}^{N(H)}\frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{1}{\langle b,V^n_t\rangle^j}\Delta_n^{-\frac12}\{\langle a,V^n_t-\Gamma^H\Pi_t\rangle^j-a_0^jC_t^j\Delta_n^{j(1-2H)}\}\\ &\qquad-\frac{\Pi_t}{\langle a,\Gamma^{\widetilde{H}^n}\rangle}\Delta_n^{-\frac12}\mathbb{J}^n.\end{aligned}\tag{F.31}$$

For the exact same reasons as explained after (F.27), the term involving $\mathbb{J}^n$ is asymptotically negligible: $(\langle a,V^n_t-\Gamma^H\Pi_t\rangle,\langle b,V^n_t-\Gamma^H\Pi_t\rangle)^\chi$ is of magnitude $\Delta_n^{(j-i)(1-2H)+i/2}\leq\Delta_n^{3/2-2H}$ if $|\chi|=2,\dots,N(H)$ and $\chi\neq(j,0)$, and it is of magnitude $\leq\Delta_n^{(N(H)+1)(1-2H)}$

if $|\chi| = N(H)+1$; in both cases, the exponent is strictly bigger than $\frac{1}{2}$. Moreover, by Corollary 3.1, $\Delta_n^{-j(1-2H)}(\langle a, V_t^n\rangle - \langle a, \Gamma^H\rangle \Pi_t)^j \xrightarrow{\mathbb{P}} a_0^j C_t^j$, which implies that the second term on the right-hand side of (F.31) is of magnitude $\Delta_n^{(j-1)1/2}$ for $j = 2, \dots, N(H)$. Thus, by Corollary 3.1, the left-hand side of (F.31) converges stably in law to

$$\mathcal{Z}_t' = \Big\{ e_1^T - \frac{a^T}{\langle a, \Gamma^H\rangle} + \frac{\Pi_t}{\langle a, \Gamma^H\rangle} \frac{\psi'(\varphi(H))}{\langle b, \Gamma^H\rangle \Pi_t} (a^T - \varphi(H) b^T) \Big\} \mathcal{Z}_t. \tag{F.32}$$

Next, we replace Π_t in the first two lines of (F.31) by $\Delta_n^{1-2H} \widehat{P}_t^n$, where $\widehat{P}_t^n$ was introduced in (F.5). The resulting difference is given by

$$\frac{\Delta_n^{-\frac{1}{2}} \{\Delta_n^{1-2H} \widehat{P}_t^n - \Pi_t\}}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \sum_{j=1}^{N(H)} \frac{(-1)^j}{j!} \psi^{(j)}(\varphi(\widetilde{H}^n)) \frac{a_0^j}{\langle b, V_t^n\rangle^j} C_t^j \, \Delta_n^{j(1-2H)}. \tag{F.33}$$

By the proof of Theorem 3.2 (see (E.7) in particular), $\Delta_n^{-1/2}\{\Delta_n^{1-2H} \widehat{P}_t^n - \Pi_t\}$ converges stably in distribution. As a consequence, the expression in the previous display converges to 0 in probability as $n \to \infty$. By (F.7), (F.10) and (F.32), it follows that

$$\begin{aligned} &\Delta_n^{-\frac{1}{2}+(1-2H)} \Big(\Big\{ V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Big\} \Delta_n^{2H-1} \Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})^{-1} - C_t + \sum_{j=2}^{N(H)} \Psi_j^n C_t^j \Big) \\ &\quad = \Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})^{-1} \Delta_n^{-\frac{1}{2}} \Big(\Big\{ V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \Big\} - \Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0}) C_t \, \Delta_n^{1-2H} \\ &\qquad + \frac{\Delta_n^{1-2H} \widehat{P}_t^n}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \sum_{j=2}^{N(H)} \frac{(-1)^j}{j!} \psi^{(j)}(\varphi(\widetilde{H}^n) \frac{a_0^j}{\langle b, V_t^n\rangle^j} C_t^j \, \Delta_n^{j(1-2H)} \Big) \\ &\quad \xrightarrow{\text{st}} \Big(1 - \frac{a_0}{\langle a, \Gamma^H\rangle} + \frac{\psi'(\varphi(H))}{\langle a, \Gamma^H\rangle \langle b, \Gamma^H\rangle} a_0 \Big)^{-1} \mathcal{Z}_t' \sim \mathcal{N}\Big(0, \mathrm{Var}_{C,1} \int_0^t \rho_s^4 \, \mathrm{d}s\Big). \end{aligned} \tag{F.34}$$

The CLT stated in (F.9) is proved. □

Proof of Proposition F.3. We first prove by induction that for $\ell = 0, \dots, N(H)-2$, the difference $\widetilde{C}_t^{n,\ell+1} - C_t$ converges in probability with a convergence rate of $\Delta_n^{(1+\ell)(1-2H)}$. If $\ell = 0$, then $\widetilde{C}_t^{n,\ell+1} = \widetilde{C}_t^{n,1}$, so by (F.34),

$$\Delta_n^{2H-1}(\widetilde{C}_t^{n,1} - C_t) \xrightarrow{\mathbb{P}} -\frac{\psi^{(2)}(\varphi(H)) a_0^2}{2\langle a, \Gamma^H\rangle \langle b, \Gamma^H\rangle^2} \Big(1 - \frac{a_0}{\langle a, \Gamma^H\rangle} + \frac{\psi'(\varphi(H))}{\langle a, \Gamma^H\rangle \langle b, \Gamma^H\rangle} a_0 \Big)^{-1} C_t^2.$$

Suppose now that $\widetilde{C}_t^{n,\ell+1} - C_t$ converges at a rate of $\Delta_n^{(1+\ell)(1-2H)}$ for $\ell = 0, \dots, \ell'-1$. Decomposing

$$\widetilde{C}_t^{n,\ell'+1} - C_t = \Big\{ \widetilde{C}_t^{n,1} - C_t + \sum_{j=2}^{\ell'+1} \Psi_j^n C_t^j \Big\} + \sum_{j=2}^{\ell'+1} \Psi_j^n \Big\{ (\widetilde{C}_t^{n,\ell'-j+2})^j - C_t^j \Big\}, \tag{F.35}$$

we note that the first term on the right-hand side converges at a rate of $\Delta_n^{(\ell'+1)(1-2H)}$ by (F.34). The second term can be rewritten as

$$\sum_{j=2}^{\ell'+1} \Psi_j^n \Big\{ (\widetilde{C}_t^{n,\ell'-j+2})^j - C_t^j \Big\} = \sum_{j=2}^{\ell'+1} \sum_{m=1}^{j} \frac{j!}{(j-m)!} C_t^{j-m} \Big\{ \Psi_j^n (\widetilde{C}_t^{n,\ell'-j+2} - C_t)^m \Big\}. \tag{F.36}$$

By assumption, $\widetilde{C}_t^{n,(\ell'-j+1)+1} - C_t$ is of size $\Delta_n^{(\ell'-j+2)(1-2H)}$. Moreover, from (F.10), the product $\Psi_j^n \Delta_n^{(1-j)(1-2H)}$ converges in probability. Thus, Ψ_j^n is of magnitude $\Delta_n^{(j-1)(1-2H)}$ and we conclude that $\Psi_j^n(\widetilde{C}_t^{n,\ell'-j+2} - C_t)^m$ is of magnitude $\Delta_n^{(j-1+m(\ell'-j+2))(1-2H)} \leq \Delta_n^{(\ell'+1)(1-2H)}$. Altogether, $\widetilde{C}_t^{n,\ell'+1} - C_t$ is of magnitude $\Delta_n^{(\ell'+1)(1-2H)}$.

We can now complete the proof of the proposition. By a similar decomposition to (F.35) with $\ell' = N(H) - 1$,

$$\widetilde{C}_t^{n,N(H)} - C_t = \Big\{\widetilde{C}_t^{n,1} - C_t + \sum_{j=2}^{N(H)} \Psi_j^n C_t^j\Big\} + \sum_{j=2}^{N(H)} \Psi_j^n \Big\{(\widetilde{C}_t^{n,N(H)-j+1})^j - C_t^j\Big\}. \tag{F.37}$$

We know that $\widetilde{C}_t^{n,N(H)-j+1} - C_t$ is of magnitude $\Delta_n^{(N(H)-j+1)(1-2H)}$. Therefore, proceeding exactly as in (F.36), we see that the right-hand side of (F.37) times $\Delta_n^{1/2-2H}$ is of size $\Delta_n^{(N(H)+1)(1-2H)-1/2}$ which goes to 0 as $n \to \infty$ since the exponent is positive by the definition of $N(H)$. So $\Delta_n^{1/2-2H}\{\widetilde{C}_t^{n,N(H)} - C_t\}$ converges stably to the same distribution as $\Delta_n^{1/2-2H}\{\widetilde{C}_t^{n,1} - C_t\}$ does. Finally, $\Delta_n^{-1/2}\{\widetilde{C}_t^{n,N(\widetilde{H}^n)} - C_t\} = \Delta_n^{-1/2}\{\widetilde{C}_t^{n,N(H)} - C_t\} + \Delta_n^{-1/2}\{\widetilde{C}_t^{n,N(\widetilde{H}^n)} - \widetilde{C}_t^{n,N(H)}\}$. Since $\widetilde{H}^n$ is a consistent estimator for H and $H \notin \mathcal{H}$, for small enough $\varepsilon > 0$ (such that the event $\{|\widetilde{H}^n - H| \leq \varepsilon\} \subseteq \{N(\widetilde{H}^n) = N(H)\}$),

$$\mathbb{P}(\Delta_n^{-\frac{1}{2}}|\widetilde{C}_t^{n,N(\widetilde{H}^n)} - \widetilde{C}_t^{n,N(H)}| > \varepsilon) \leq \mathbb{P}(|\widetilde{H}^n - H| > \varepsilon) \to 0. \tag{F.38}$$

Thus, the CLT of $\widetilde{C}_t^{n,N(H)} - C_t$ is not affected when $N(H)$ is replaced by $N(\widetilde{H}^n)$. □

Proof of Proposition F.4. We first decompose

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n (\widehat{C}_t^{n,1})^j\Big\} \\
&\quad= \Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n C_t^j\Big\} + \Phi_1^n \Delta_n^{-\frac{1}{2}}\{\widetilde{C}_t^{n,N(H)} - C_t\} \\
&\qquad+ \sum_{j=2}^{N(H)} \Phi_j^n \Delta_n^{-\frac{1}{2}}\Big\{(\widetilde{C}_t^{n,N(H)})^j - C_t^j\Big\} + \sum_{j=1}^{N(H)} \Phi_j^n \Delta_n^{-\frac{1}{2}}\Big\{(\widetilde{C}_t^{n,N(\widetilde{H}^n)})^j - (\widetilde{C}_t^{n,N(H)})^j\Big\} \\
&\quad= \Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n C_t^j\Big\} + \Phi_1^n \Delta_n^{-\frac{1}{2}}\Big\{\widetilde{C}_t^{n,1} - C_t + \sum_{k=2}^{N(H)} \Psi_k^n C_t^k\Big\} + \mathbb{I}_1^n,
\end{aligned} \tag{F.39}$$

where

$$\begin{aligned}
\mathbb{I}_1^n &= \sum_{k=2}^{N(H)} \Phi_1^n \Psi_k^n \Delta_n^{-\frac{1}{2}}\Big\{(\widetilde{C}_t^{n,N(H)-k+1})^k - C_t^k\Big\} + \sum_{j=2}^{N(H)} \Phi_j^n \Delta_n^{-\frac{1}{2}}\Big\{(\widetilde{C}_t^{n,N(H)})^j - C_t^j\Big\} \\
&\quad+ \sum_{j=1}^{N(H)} \Phi_j^n \Delta_n^{-\frac{1}{2}}\Big\{(\widetilde{C}_t^{n,N(\widetilde{H}^n)})^j - (\widetilde{C}_t^{n,N(H)})^j\Big\}.
\end{aligned}$$

By the proof of Proposition F.3 and the mean-value theorem, $(\widetilde{C}_t^{n,N(H)-k+1})^k - C_t^k$ is of size $\Delta_n^{(N(H)-k+1)(1-2H)}$ and $(\widetilde{C}_t^{n,N(H)})^j - C_t^j$ is of size $\Delta_n^{2H-1/2}$. Furthermore, from (F.3), we see that $\Phi_j^n \Delta_n^{-j(1-2H)}$ converges in probability. Hence, $\Phi_j^n\{(\widetilde{C}_t^{n,N(H)})^j - C_t^j\}$ is of size $\Delta_n^{1/2+(j-1)(1-2H)}$. Also, Ψ_k^n is of size $\Delta_n^{(k-1)(1-2H)}$, so $\Phi_1^n\Psi_k^n\Delta_n^{-1/2}\{(\widetilde{C}_t^{n,N(H)-k+1})^k - C_t^k\}$ is of size $\Delta_n^{(N(H)+1)(1-2H)-1/2}$. Recall also that $\Delta_n^{-1/2}\{\widetilde{C}_t^{n,N(\widetilde{H}^n)} - \widetilde{C}_t^{n,N(H)}\}$ is negligible by the last part of the proof of Proposition F.3. Altogether, $\mathbb{I}_1^n$ is asymptotically negligible.

Now, recalling (F.6), (F.7) and (F.10), we decompose

$$\begin{aligned}
&\Delta_n^{\frac{1}{2}-2H}\Big\{\widetilde{C}_t^{n,1} - C_t + \sum_{k=2}^{N(H)} \Psi_k^n C_t^k\Big\} \\
&= \frac{\Delta_n^{\frac{1}{2}-2H}}{\Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})}\bigg\{\Big\{V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\Big\}\Delta_n^{2H-1} - \Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})C_t \\
&\quad + \frac{\Delta_n^{1-2H}\widehat{P}_t^n}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\sum_{j=2}^{N(H)} \frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b, V_t^n\rangle^j} C_t^j\, \Delta_n^{(j-1)(1-2H)}\bigg\} \\
&= \frac{\Delta_n^{-\frac{1}{2}}}{\Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})}\bigg\{V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} - \Big(1 - \frac{a_0}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} + \frac{\Pi_t a_0 \psi'(\varphi(\widetilde{H}^n))}{\langle a, \Gamma^{\widetilde{H}^n}\rangle\langle b, V_t^n\rangle}\Big)C_t\Delta_n^{1-2H} \\
&\quad + \frac{\Pi_t}{\langle a, \Gamma^{\widetilde{H}^n}\rangle}\sum_{j=2}^{N(H)} \frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b, V_t^n\rangle^j} C_t^j\, \Delta_n^{j(1-2H)}\bigg\} \\
&\quad + \frac{\Delta_n^{-\frac{1}{2}}\{\Delta_n^{1-2H}\widehat{P}_t^n - \Pi_t\}}{\Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})\langle a, \Gamma^{\widetilde{H}^n}\rangle}\sum_{j=1}^{N(H)} \frac{(-1)^j}{j!}\psi^{(j)}(\varphi(\widetilde{H}^n))\frac{a_0^j}{\langle b, V_t^n\rangle^j} C_t^j\, \Delta_n^{j(1-2H)}.
\end{aligned} \tag{F.40}$$

The last term is asymptotically negligible as already seen in the discussion following (F.33), while the first term on the right-hand side of (F.40) was analyzed in the (F.31). Combining this with (F.28), we continue the computations started in (F.39):

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n (\widehat{C}_t^{n,1})^j\Big\} \\
&= \bigg\{(\varphi^{-1})'(\varphi(\widetilde{H}^n))\frac{a^T - \varphi(\widetilde{H}^n)b^T}{\langle b, V_t^n\rangle} + \frac{\Phi_1^n \Delta_n^{2H-1}}{\Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})}\Big\{e_1^T - \frac{a^T}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} \\
&\quad + \frac{\Pi_t \psi'(\varphi(\widetilde{H}^n))}{\langle a, \Gamma^{\widetilde{H}^n}\rangle\langle b, V_t^n\rangle}(a^T - \varphi(\widetilde{H}^n)b^T)\Big\}\bigg\}\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t - e_1 C_t \Delta_n^{1-2H}\} + \widehat{\mathbb{I}}_1^n \\
&= w_1(\widetilde{H}^n, \widetilde{H}^{n,0}, V_t^n)\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H \Pi_t - e_1 C_t \Delta_n^{1-2H}\} + \widehat{\mathbb{I}}_1^n,
\end{aligned} \tag{F.41}$$

where

$$\begin{aligned}
w_1(\widetilde{H}^n, \widetilde{H}^{n,0}, V_t^n) &= \frac{(\varphi^{-1})'(\varphi(\widetilde{H}^n))}{\langle b, V_t^n\rangle}\Big\{a^T - \varphi(\widetilde{H}^n)b^T - a_0 u_1(\widetilde{H}^n, \widetilde{H}^{n,0}, V_t^n)\Big\}, \\
u_1(\widetilde{H}^n, \widetilde{H}^{n,0}, V_t^n) &= \Theta(\widehat{V}_t^n, \widetilde{H}^n, \widetilde{H}^{n,0})^{-1} \\
&\qquad \times \Big(e_1^T - \frac{a^T}{\langle a, \Gamma^{\widetilde{H}^n}\rangle} + \frac{\Pi_t \psi'(\varphi(\widetilde{H}^n))}{\langle b, V_t^n\rangle\langle a, \Gamma^{\widetilde{H}^n}\rangle}(a^T - \varphi(\widetilde{H}^n)b^T)\Big).
\end{aligned} \tag{F.42}$$

In $\widehat{\mathbb{I}}_1^n$, we have incorporated the last three terms on the right-hand side of (F.28), the last two terms on the right-hand side of (F.31), the last expression in (F.40) as well as $\mathbb{I}_1^n$ from

(F.39). By the discussions following these equations, we know that $\widehat{\mathbb{I}}_1^n$ is asymptotically negligible. Therefore, we obtain

$$\Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(H)} \Phi_j^n(\widehat{C}_t^{n,1})^j\Big\} \xrightarrow{\text{st}} \frac{w_1^T}{\Pi_t}\mathcal{Z}_t \sim \mathcal{N}\Big(0, \mathrm{Var}_{H,1}\frac{\int_0^t \rho_s^4\,\mathrm{d}s}{(\int_0^t \rho_s^2\,\mathrm{d}s)^2}\Big).$$

To conclude, it remains to observe that this CLT is not affected when $N(H)$ is replaced by $N(\widetilde{H}^n)$ because $H \notin \mathcal{H}$; cf. the argument used to show (F.38). □

Proof of Proposition F.5. For $k = 2, \dots, m$, define

$$\begin{aligned} &u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) \\ &\quad = \Big\{e_1^T - \frac{a^T}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle} + \frac{\Pi_t\langle a, \partial_H\Gamma^H}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\rangle w_{k-1}(\widehat{H}_{k-2}^n, \widetilde{H}^n, V_t^n)\Big\}\Big(1 - \frac{a_0}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Big)^{-1}, \\ &w_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) = \frac{(\varphi^{-1})'(\varphi(\widetilde{H}^n))}{\langle b, V_t^n\rangle}\Big\{a^T - \varphi(\widetilde{H}^n)b^T - a_0 u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)\Big\}. \end{aligned}$$

In the definition of $u_2(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)$, the term $w_1(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)$ is replaced by the term $w_1(\widetilde{H}^n, \widetilde{H}^{n,0}, V_t^n)$ from (F.42). By induction over k, we are going to show for all $k = 1, \dots, m$ that

$$\Delta_n^{-\frac{1}{2}}(\widehat{H}_k^n - H) = w_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H\Pi_t - e_1C_t\Delta_n^{1-2H}\} + \widehat{\mathbb{I}}_k^n \tag{F.43}$$

for some asymptotically negligible expression $\widehat{\mathbb{I}}_k^n$ and that

$$u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) \xrightarrow{\mathbb{P}} u_k^T, \qquad w_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) \xrightarrow{\mathbb{P}} \frac{w_k^T}{\Pi_t}, \tag{F.44}$$

where, for $k = 1$, we take the expressions in (F.42) instead. For $k = 1$, (F.43) was already shown in (F.41), and (F.44) is obvious, so we may consider $k \geq 2$ now and assume (F.43) and (F.44) for $k - 1$. In particular,

$$\Delta_n^{-\frac{1}{2}}\{\widehat{H}_{k-1}^n - H\} \xrightarrow{\text{st}} \frac{w_{k-1}^T}{\Pi_t}\mathcal{Z}_t \sim \mathcal{N}\Big(0, \mathrm{Var}_{H,k-1}\frac{\int_0^t \rho_s^4\,\mathrm{d}s}{(\int_0^t \rho_s^2\,\mathrm{d}s)^2}\Big).$$

It is straightforward to see that

$$u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) \xrightarrow{\mathbb{P}} u_k^T \qquad \text{and} \qquad w_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n) \xrightarrow{\mathbb{P}} \frac{w_k^T}{\Pi_t}, \tag{F.45}$$

so we can proceed to showing (F.43) for k. Expanding $\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle$ around H and using the induction hypothesis, we can find β_{k-1}^n between $\widehat{H}_{k-1}^n$ and H such that

$$\begin{aligned}
&\Delta_n^{-\frac{1}{2}}\Big\{V_{0,t}^n - \frac{\langle a, V_t^n\rangle}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle} - \Big(1 - \frac{a_0}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Big)C_t\Delta_n^{1-2H}\Big\} \\
&= \Delta_n^{-\frac{1}{2}}\{V_{0,t}^n - \Pi_t - C_t\Delta_n^{1-2H}\} - \frac{1}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Delta_n^{\frac{1}{2}}\{\langle a, V_t^n - \Gamma^H\Pi_t\rangle - a_0C_t\Delta_n^{1-2H}\} \\
&\quad + \frac{\Pi_t}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Delta_n^{-\frac{1}{2}}\langle a, \Gamma^{\widehat{H}_{k-1}^n} - \Gamma^H\rangle \\
&= \Delta_n^{-\frac{1}{2}}\{V_{0,t}^n - \Pi_t - C_t\Delta_n^{1-2H}\} - \frac{1}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Delta_n^{\frac{1}{2}}\{\langle a, V_t^n - \Gamma^H\Pi_t\rangle - a_0C_t\Delta_n^{1-2H}\} \\
&\quad + \frac{\Pi_t\langle a, \partial_H\Gamma^H\rangle}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Big\{w_{k-1}(\widehat{H}_{k-2}^n, \widetilde{H}^n, V_t^n)\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H\Pi_t - e_1C_t\Delta_n^{1-2H}\} + \widehat{\mathbb{I}}_{k-1}^n\Big\} \\
&\quad + \frac{1}{2!}\frac{\Pi_t}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\langle a, \partial_{HH}\Gamma^{\xi^n}\rangle\Delta_n^{-\frac{1}{2}}\{\widehat{H}_{k-1}^n - H\}^2 \\
&= u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)\Big(1 - \frac{a_0}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Big)\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H\Pi_t - e_1C_t\Delta_n^{1-2H}\} + \mathbb{J}_k^n,
\end{aligned} \tag{F.46}$$

where $\mathbb{J}_k^n = \Pi_t\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle^{-1}(\langle a, \partial_H\Gamma^H\rangle\widehat{\mathbb{I}}_{k-1}^n + \frac{1}{2}\langle a, \partial_{HH}\Gamma^{\beta_{k-1}^n}\rangle\Delta_n^{-1/2}\{\widehat{H}_{k-1}^n - H\}^2)$. Because $\widehat{I}_{k-1}^n$ is negligible by the induction hypothesis and $\widehat{H}_{k-1}^n - H$ is of size $\Delta_n^{1/2}$, we see that $\mathbb{J}_k^n \xrightarrow{\mathbb{P}} 0$. Recalling (F.16), we infer from (F.46) and (F.45) that $\Delta_n^{1/2-2H}(\widehat{C}_t^{n,k} - C_t) \xrightarrow{\text{st}} u_k^T\mathcal{Z}_t$, which is (F.19). Now recall the definitions (F.16) and (F.17). Using (F.28) and the formula $\Phi_1^n = -\Delta_n^{1-2H}(\varphi^{-1})(\varphi(\widetilde{H}^n))a_0/\langle b, V_t^n\rangle$ for the second equality and (F.46) for the third, we obtain

$$\begin{aligned}
\Delta_n^{-\frac{1}{2}}\{\widehat{H}_k^n - H\} &= \Delta_n^{-\frac{1}{2}}\Big\{\widetilde{H}^n - H + \sum_{j=1}^{N(\widehat{H}_{k-1}^n)}\Phi_j^nC_t^j\Big\} + \Phi_1^n\Delta_n^{-\frac{1}{2}}\{\widehat{C}_t^{n,k} - C_t\} \\
&\quad + \sum_{j=2}^{N(\widehat{H}_{k-1}^n)}\Phi_j^n\Delta_n^{-\frac{1}{2}}\{(\widehat{C}_t^{n,k})^j - C_t^j\} \\
&= \frac{(\varphi^{-1})'(\varphi(\widetilde{H}^n))}{\langle b, V_t^n\rangle}\Big\{a^T - \varphi(\widetilde{H}^n)b^T - a_0u_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)\Big\} \\
&\quad \times \Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H\Pi_t - e_1C_t\Delta_n^{1-2H}\} + \mathbb{I}_k^n \\
&= w_k(\widehat{H}_{k-1}^n, \widetilde{H}^n, V_t^n)\Delta_n^{-\frac{1}{2}}\{V_t^n - \Gamma^H\Pi_t - e_1C_t\Delta_n^{1-2H}\} + \widehat{\mathbb{I}}_k^n.
\end{aligned} \tag{F.47}$$

In the last line, $\widehat{\mathbb{I}}_k^n$ contains the last three terms on the right-hand side of (F.28) and

$$\mathbb{J}_k^n\Big(1 - \frac{a_0}{\langle a, \Gamma^{\widehat{H}_{k-1}^n}\rangle}\Big)^{-1} + \sum_{j=2}^{N(H)}\Phi_j^n\Delta_n^{-\frac{1}{2}}\{(\widehat{C}_t^{n,k})^j - C_t^j\} + \Delta_n^{-\frac{1}{2}}\sum_{j=N(H)\wedge N(\widehat{H}_{k-1}^n)}^{N(H)\vee N(\widehat{H}_{k-1}^n)}\Phi_j^n(\widehat{C}_t^{n,k})^j.$$

The term $\Phi_j^n\Delta_n^{-1/2}\{(\widehat{C}_t^{n,k})^j - C_t^j\}$ is of size $\Delta_n^{(j-1)(1-2H)}$ because Φ_j^n is of size $\Delta_n^{j(1-2H)}$. Also, the last sum goes to 0 in probability by a similar argument to (F.38). Therefore, $\widehat{\mathbb{I}}_k^n$

is asymptotically negligible. This together with (F.47) implies (F.43) and our induction argument is complete. From (F.43), we immediately obtain (F.18). □

Proof of Theorem F.6. The proof of (F.24) and (F.25) is similar to that of (F.19) and (3.12) in Theorem F.5 and 3.2, respectively. □

G Choosing the tuning parameters

We fix the number of iterations in the multi-step algorithm of Section F.4 at $m = 50$. In fact, for an overwhelming majority of estimates obtained in the simulation and the empirical analysis of Section 5, a precision of 10^{-5} was attained after fewer than 50 steps. We further make the choice $R = 60$, which corresponds to considering quadratic variations with time lags up to one minute. In order to tune the remaining parameters, we want to choose the vectors $a, b, c \in \mathbb{R}^{1+R}$ in such a way that $(a, b, c) \mapsto \mathrm{Var}_C(R, a, b, c, H)$ is as small as possible. Due to the complexity of how Var_C depends on a, b and c, we were not able to find (and doubt there is) an analytical expression for the minimizers. In addition, Var_C depends on H, which is unknown. Pretending we knew H for the moment and $H \in (\frac{1}{4}, \frac{1}{2})$, in order to resolve the first issue, we choose

$$a = c = \frac{\Gamma^H - \langle \Gamma^H, b\rangle b}{\|\Gamma^H - \langle \Gamma^H, b\rangle b\|}, \qquad b = \frac{\partial_H \Gamma^H}{\|\partial_H \Gamma^H\|} \tag{G.1}$$

as initial values (This is a heuristic choice: with these vectors, $\langle c, \partial_H \Gamma^H\rangle = 0$ in (F.26) and $\langle u, \Gamma^H\rangle = 0$. Consequently, if $\mathcal{C}^H_{0,1}$ and $\mathcal{C}^H_{0,2}$ denote the two zeroth-order terms in (3.2), then $u^T \mathcal{C}^H_{0,2} u = 0$ and, in $u^T \mathcal{C}^H_{0,1} u = \sum_{i,j=0}^R u_i u_j (\mathcal{C}^H_{0,1})_{ij}$, the part of the sum where $i = j$ is 0.) Then we run the R function `fminsearch()` from the package `pracma` to find (local) minimizers $a(H)$, $b(H)$ and $c(H)$ of $(a, b, c) \mapsto \mathrm{Var}_C(R, a, b, c, H)$ from (F.26). Similarly, we obtain $a^0(H)$, $b^0(H)$ and $c^0(H)$ as minimizers of $(a, b, c) \mapsto \mathrm{Var}_C(R, a, b, c, H)$ from (3.9) by taking the same initial weights b and c from (G.1) for b^0 and c^0 and by choosing a^0 as the vector obtained by substituting 0 for the first component of a from (G.1). As H is unknown in this process, we simply take the minimizers at $H_0 = 0.35$, that is,

$$a^0 = a^0(0.35), \qquad b^0 = b^0(0.35), \qquad c^0 = c^0(0.35), \tag{G.2}$$

$$a = a(0.35), \qquad b = b(0.35), \qquad c = c(0.35). \tag{G.3}$$

One could, of course, plug in a consistent estimator of H (e.g., $\widehat{H}^n$, computed for some initial choice of a, b and c), determine the minimizing vectors, use them to construct an update of $\widehat{H}^n$, and repeat this procedure. However, such an adaptive scheme of constructing $\widehat{H}^n$ makes the weight vectors dependent on the latest estimator of H and therefore changes its asymptotic variance in every step. Unfortunately, we see no way of keeping track of those changes, in particular because we do not know the precise form of how $a(H)$, $b(H)$ and $c(H)$ depend on H. It turns out that the variances $\mathrm{Var}_C(R, a^0(H_0), b^0(H_0), c^0(H_0), H)$ and $\mathrm{Var}_C(R, a(H_0), b(H_0), c(H_0), H)$ at other values of H based on the choice $H_0 = 0.35$ turn out to be reasonably close to the H-dependent minimal values $\mathrm{Var}_C(R, a^0(H), b^0(H), c^0(H), H)$ and $\mathrm{Var}_C(R, a(H), b(H), c(H), H)$, respectively (no more than 2.1% larger in the former case; no more than 8.1% larger in the latter case for all H in (4.1) except for $H = 0.45$, where the variance based on (G.3) is 2.6 times larger).

References

Chong, C. (2020*a*), 'High-frequency analysis of parabolic stochastic PDEs', *The Annals of Statistics*, **48**, 1143–1167.

Chong, C. (2020*b*), High-frequency analysis of parabolic stochastic PDEs with multiplicative noise. arXiv:1908.04145.

Chong, C. (2020*c*), 'Supplement to "High-frequency analysis of parabolic stochastic PDEs"', *The Annals of Statistics Supplementary Material*.

Jacod, J. & Protter, P. (2012), *Discretization of Processes*, Vol. 67 of *Stochastic Modelling and Applied Probability*, Springer, Heidelberg.

Li, Z. M., Laeven, R. J. A. & Vellekoop, M. H. (2020), 'Dependent microstructure noise and integrated volatility estimation from high-frequency data', *Journal of Econometrics*, **215**, 536–558.

Nualart, D. (2006), *The Malliavin calculus and related topics*, Probability and its Applications (New York), second edn, Springer-Verlag, Berlin.

Protter, P. E. (2005), *Stochastic Integration and Differential Equations*, Vol. 21 of *Stochastic Modelling and Applied Probability*, Springer-Verlag, Berlin. Second edition. Version 2.1, Corrected third printing.