ARTICLE IN PRESS

Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process☆

Xiaohu Wang [a], Weilin Xiao [b,*], Jun Yu [c,*]

[a] School of Economics, Fudan University, Shanghai, China
[b] School of Management, Zhejiang University, Hangzhou, 310058, China
[c] School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, 178903, Singapore

ARTICLE INFO

Article history:
Received 14 December 2020
Received in revised form 1 August 2021
Accepted 29 August 2021
Available online xxxx

JEL classification:
C15
C22
C32

Keywords:
Rough volatility
Fractional Ornstein–Uhlenbeck process
Hurst parameter
Long memory
Anti-persistent errors
Out-of-sample forecasting
ARFIMA

ABSTRACT

This paper proposes to model and forecast realized volatility (RV) using the fractional Ornstein–Uhlenbeck (fO–U) process with a general Hurst parameter, H. A two-stage method is introduced for estimating parameters in the fO–U process based on discrete-sampled observations. In the first stage, H is estimated based on the ratio of two second-order differences of observations from different frequencies. In the second stage, with the estimated H, the other parameters of the model are estimated by the method of moments. All estimators have closed-form expressions and are easy to implement. A large sample theory of the proposed estimators is derived. Extensive simulations show that the proposed estimators and the large-sample theory perform well in finite samples. We apply the model and the method to the logarithmic daily RV series of various financial assets. Our empirical findings suggest that H is much smaller than 1/2, indicating that the RV series have rough sample paths, and that the mean reversion parameter takes a small positive number, indicating that the RV series are stationary but have slow mean reversion. The proposed model is compared with many alternative models, including the fractional Brownian motion, ARFIMA, and HAR, in forecasting RV and logarithmic RV.

1. Introduction

With the increasing availability of high frequency financial data, daily realized volatility (RV) is readily available. As a result, modeling and forecasting daily RV have been extensively studied. For example, Andersen et al. (2001a,b) provide evidence of long-range dependence in the RV of equities and exchange rates, respectively. Andersen et al. (2003) point out that forecasts from a simple autoregressive fractionally integrated moving average (ARFIMA) model for the

☆ We are grateful to Torben Andersen (co-editor), an AE, and especially two anonymous referees for their extensive and constructive comments that help to improve this paper significantly. Our thanks also go to Federico Bandi, Markus Bibinger, Xiaohong Chen, Frank Diebold, Graham Elliott, Robert Engle, Peter Hansen, Clifford Hurvich, Ioannis Karparis, Jia Li, Oliver Linton, Tassos Magdalinos, Ulrich Müller, Katerina Petrova, Peter Phillips, Mark Podolskij, Mathieu Rosenbaum, Tian Xie, Martina Zähle and participants at various conferences and seminars for useful comments. Xiao gratefully acknowledges the financial support of the National Natural Science Foundation of China (No. 71871202). Yu gratefully acknowledges the financial support of Singapore Ministry of Education Tier 2 grant (under grant No. MOE2018-T2-2-169) and Lee Foundation. R code and data used in the paper can be downloaded from http://www.mysmu.edu/faculty/yujun/research.html.

* Corresponding authors.
E-mail addresses: wang_xh@fudan.edu.cn (X. Wang), wlxiao@zju.edu.cn (W. Xiao), yujun@smu.edu.sg (J. Yu).

https://doi.org/10.1016/j.jeconom.2021.08.001
0304-4076/

logarithmic daily RV outperforms several popular daily GARCH models and more complicated high-frequency models. Andersen and Bollerslev (1997) show how long-range dependence in volatility can arise via interactions of a large number of heterogeneous information processes. Motivated by this heterogeneity, Corsi (2009) introduces the heterogeneous autoregressive (HAR) model for the daily RV. Bollerslev et al. (2016) extend the HAR model by allowing the parameters of the HAR model to vary explicitly with the measurement error in estimating the daily RV.

More recently, Gatheral et al. (2018) call for more attention to another stylized fact of RV, namely, the roughness of the sample path. They show that the fractional Brownian motion (fBm) with the Hurst parameter $H = 0.14$ can generate a rough sample path and outperform the HAR model in forecasting logarithmic RV and RV out-of-sample. Moreover, they argue that although fBm with $H < 1/2$ is not a long-memory process, classical statistical procedures, such as the log-periodogram regression, tend to find the evidence of long memory in data generated from fBm with $H < 1/2$.[1] A growing strand of literature now supports the findings of Gatheral et al. (2018). For example, Livieri et al. (2018) report strong support for roughness using the implied volatility-based approximations to spot volatility. Bayer et al. (2016) obtain strong support for roughness via the S&P 500 volatility surface and variance swaps. Bennedsen (2017) and Bennedsen et al. (2021) document roughness in daily RVs of nearly two thousand U.S. equities.[2]

In the context of fBm, the roughness of a sample path is determined by the Hurst parameter, H, if $H < 1/2$. Let B_t^H denote an fBm that is a zero-mean Gaussian process with the covariance function

$$Cov\left(B_t^H, B_s^H\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right), \quad \forall t, s \geq 0. \tag{1.1}$$

If $H = 1/2$, B_t^H is a standard Brownian motion (Bm, denoted by W_t) that has independent increments. Whereas, when $H \neq 1/2$, the increments of B_t^H are serially correlated. The serial correlations are positive when $H \in (1/2, 1)$ and negative when $H \in (0, 1/2)$. The smaller the H is, the rougher the sample path of B_t^H.

Unfortunately, fBm is non-stationary, making it less ideal for modeling daily RV. To allow both stationarity and potential roughness of a sample path, in this paper, we consider the following discrete-time model

$$X_{i\Delta} = e^{-\kappa\Delta}X_{(i-1)\Delta} + \left(1 - e^{-\kappa\Delta}\right)\mu + \varepsilon_{i\Delta} \text{ with } \varepsilon_{i\Delta} = \sigma\int_{(i-1)\Delta}^{i\Delta} e^{-\kappa(i\Delta - s)}dB_s^H, \tag{1.2}$$

where $\{X_{i\Delta}\}_{i=0}^n$ is a set of observations on the logarithmic RV, Δ denotes the sampling interval, $\kappa \in R^+$, $\sigma \in R^+$, and $\mu \in R$ are the unknown parameters. Here, it is assumed that we have $n + 1 (n = T/\Delta)$ equally spaced observations at points $\{i\Delta\}_{i=0}^n$ over the time interval $[0, T]$ with T being the time-span of the data.[3] If $\kappa = 0$, $e^{-\kappa\Delta} = 1$ and $X_{i\Delta} - X_{(i-1)\Delta}$ is the increment of fBm, which is known as the fractional Gaussian noise. As a result, $X_{i\Delta}$ is a discrete sample from fBm. However, when $\kappa > 0$, $e^{-\kappa\Delta} < 1$ and $X_{i\Delta}$ is stationary. In this case, κ is often referred to as the mean reversion parameter. Together with H, it determines the autocorrelation function (ACF) of $X_{i\Delta}$. When $H < 0.5$ and κ takes a small positive value, which are estimates that we obtained from real data, the sample path of $X_{i\Delta}$ is rough and the ACF decays slowly.

Model (1.2) is the exact discrete-time representation of the following continuous-time fractional Ornstein–Uhlenbeck (fO–U) process

$$dX_t = \kappa\left(\mu - X_t\right)dt + \sigma dB_t^H. \tag{1.3}$$

In this paper, we propose a two-stage method to estimate all unknown parameters in Model (1.3) based on the discrete-time observations $\{X_{i\Delta}\}_{i=0}^n$. In the first stage, following Lang and Roueff (2001) and Barndorff-Nielsen et al. (2013), H is estimated based on the ratio of squared summations of the second-order differences of $X_{i\Delta}$ obtained at two different frequencies. In the second stage, estimators of the other parameters in Model (1.3) are constructed based on a set of moment conditions in which the true value of H is replaced with the estimated H obtained in the first stage. Closed-form expressions are established for all the proposed estimators, denoted by $\widehat{H}$, $\widehat{\kappa}$, $\widehat{\mu}$, and $\widehat{\sigma}$.

We then develop the large-sample theory for the proposed estimators. In particular, we consider two asymptotic schemes: (i) the in-fill scheme under which the sampling interval Δ goes to zero with a fixed time-span T; and (ii) the double scheme in which $\Delta \to 0$ and $T \to \infty$ simultaneously. Under both schemes, the consistency and asymptotic normality of $\widehat{H}$ and $\widehat{\sigma}$ are established for all $H \in (0, 1)$ and regardless of the stationarity property of the model. In addition, an explicit formula is derived for the asymptotic variance of $\widehat{H}$, which depends only on the value of H. This feature greatly facilitates statistical inference about H. Under the double scheme, the consistency and the asymptotic distributions of $\widehat{\kappa}$ and $\widehat{\mu}$ are developed. The convergence rate of $\widehat{\mu}$ is a function of H. Both the convergence rate and the asymptotic distribution of $\widehat{\kappa}$ depend crucially on H. Extensive simulations are designed to check the finite-sample performance of the proposed estimators and derived asymptotic distributions.

[1] Bennedsen et al. (2021) propose an alternative volatility model that can generate a rough sample path and has the long-memory property at the same time.

[2] Rough volatility models have also been applied in mathematical finance, such as in option pricing (Bayer et al., 2016; Garnier and Sølna, 2017), portfolio choice (Fouque and Hu, 2018), and dynamic hedging (Euch and Rosenbaum, 2018). Jaisson and Rosenbaum (2016) study microstructural foundations for the roughness.

[3] When $X_{i\Delta}$ is annualized and observed daily (weekly or monthly), then $\Delta = 1/252$ (1/52 or 1/12).

We apply Model (1.3) and the proposed estimation method in empirical studies where we model and fit the logarithmic daily RV for the S&P 500, DJIA, and NASDAQ 100. We obtain the strong evidence of $H < 1/2$ in all cases, reinforcing the finding of rough volatility documented in the literature. Also found is the evidence of κ taking a small positive number, suggesting slow mean reversion. We compare the out-of-sample performance of Model (1.3) with that of five alternative models, including fBm, ARFIMA, and HAR, in forecasting RV and logarithmic RV.

The remainder of the paper is organized as follows. Section 2 introduces the model, discusses its relationship with ARFIMA, and provides the forecasting formula with Model (1.3). Section 3 proposes a two-stage estimation approach. Section 4 establishes the asymptotic properties of the proposed estimators. In Section 5, Monte Carlo experiments are designed to check the finite-sample performance of the proposed estimators and the large-sample theory. Empirical studies are carried out in Section 6. Section 7 concludes. All proofs are collected in Appendix A. More empirical results can be found in the online supplement where we apply the proposed method to forecast the logarithmic RV, logarithmic bipower variation (BV) and logarithmic realized kernel (RK) for the S&P 500, DJIA, and NASDAQ 100. Throughout the paper, we use $\xrightarrow{p}$, $\xrightarrow{a.s.}$, $\xrightarrow{d}$, $\overset{d}{=}$ to denote convergence in probability, convergence almost surely, convergence in distribution, and equivalence in distribution, respectively.

2. Model

2.1. Preliminaries

The model with which we are concerned in this paper is given by (1.3). The stochastic differential equation in (1.3) has a unique path-wise solution as

$$X_t = e^{-\kappa t}X_0 + \left(1 - e^{-\kappa t}\right)\mu + \sigma \int_0^t e^{-\kappa(t-s)} dB_s^H, \tag{2.1}$$

where X_0 is the initial value and $\int_0^t e^{-\kappa(t-s)} dB_s^H$ exists as a path-wise Riemann–Stieltjes integral by using partial integration and exploiting that the exponential function is of finite variation. Moreover, the integral in (2.1) is almost surely continuous in t (Cheridito et al., 2003). Letting $t > s$ and $u > v$, we have

$$\mathbb{E}\left[\left(B_t^H - B_s^H\right)\left(B_u^H - B_v^H\right)\right] = \frac{1}{2}\left((s-u)^{2H} + (t-v)^{2H} - (t-u)^{2H} - (s-v)^{2H}\right), \tag{2.2}$$

which implies that $\left\{B_t^H\right\}_{t\geq 0}$ has stationary increments for any $H \in (0, 1)$.

When $H = 1/2$, $\left\{B_t^H\right\}_{t\geq 0}$ has independent increments and becomes a standard Bm. Whereas, when $H \neq 1/2$, the increments of $\left\{B_t^H\right\}_{t\geq 0}$ are serially correlated. If $H \in (1/2, 1)$, the covariances of increments of B_t^H are not summable, suggesting long-range dependence of the increment process. If $H \in (0, 1/2)$, the two-sided long-run variance of increments of B_t^H is zero, suggesting anti-persistence of the increment process. For all $\varepsilon > 0$ and $T > 0$, there exists a nonnegative random variable $G_{\varepsilon,T}$ such that $E\left(\left|G_{\varepsilon,T}\right|^p\right) < \infty$ for all $p \geq 1$, and $\left|B_t^H - B_s^H\right| \leq G_{\varepsilon,T}\,|t-s|^{H-\varepsilon}$ almost surely for all $s, t \in [0, T]$. This fact ensures that the sample path of $\left\{B_t^H\right\}_{t\geq 0}$ is (locally) Hölder continuous of order $H - \varepsilon$, which in turn suggests that X_t is (locally) Hölder continuous of order $H - \varepsilon$ (Gehringer and Li, 2020). Hence, the smaller the value of H, the rougher the sample paths of B_t^H and X_t. Cheridito et al. (2003) obtain the order of the covariance function of fO–U when the lag goes to infinity, suggesting that the decay of the covariance of fO–U is very similar to that of increments of B_t^H.[4]

With a proper initial condition, the stationarity and ergodicity of X_t are guaranteed by $\kappa > 0$ (Hu and Nualart, 2010). Whereas, if $\kappa < 0$, X_t is explosive and $X_t = O_p\left(e^{-\kappa t}\right)$ as $t \to \infty$ (Belfadli et al., 2011). When $\kappa = 0$, X_t is an fBm. In this case, $X_t = \sigma B_t^H = O_p\left(t^H\right)$ as $t \to \infty$ and hence, is nonstationary. The last equation is from the fact of $E\left[\left(B_t^H - B_0^H\right)^2\right] = t^{2H}$, which can be obtained straightforwardly from (2.2) by letting $t = u$ and $s = v = 0$.

For the case of $\kappa > 0$, set the initial condition to be

$$X_0 = \mu + \sigma \int_{-\infty}^0 e^{\kappa s} d\widetilde{B}_s^H \quad \text{with} \quad \widetilde{B}_s^H = \begin{cases} B_s^H & \text{for } s \geq 0 \\ W_{|s|}^H & \text{for } s \leq 0 \end{cases},$$

where $\widetilde{B}_s^H$ is a two-sided fBm with $W_{|s|}^H$ being an fBm independent of B_s^H. Then, X_t in (2.1) becomes a covariance stationary process with

$$\mathbb{E}\left(X_t\right) = \mu \quad \text{and} \quad Var\left(X_t\right) = \sigma^2 \kappa^{-2H} H\Gamma\left(2H\right), \tag{2.3}$$

[4] More in-depth discussions of fBm and fO–U can be found in Samorodnitsky and Taqqu (1994), Embrechts and Maejima (2002), and Cheridito et al. (2003).

where $\Gamma(\cdot)$ denotes the gamma function. Moreover, X_t can be identically represented as

$$X_t = \mu + \sigma \int_{-\infty}^{t} e^{-\kappa(t-s)} d\widetilde{B}_s^H . \tag{2.4}$$

Under a general initial condition $X_0 = O_p(1)$, X_t is asymptotically covariance stationary. In this case, the impact of the initial condition disappears as $t \to \infty$ because $e^{-\kappa t} \to 0$ as $t \to \infty$. Hence, it is easy to get $\lim_{t\to\infty} \mathbb{E}(X_t) = \mu$ and $\lim_{t\to\infty} Var(X_t) = \sigma^2 \kappa^{-2H} H\Gamma(2H)$.

2.2. Discrete-time model

When $H = 1/2$, Model (1.2) turns out to be an AR(1) model with independent Gaussian errors $\{\varepsilon_{i\Delta}\}$ (see, for example, Bergstrom, 1990). Under the in-fill asymptotic scheme where $\Delta \to 0$ with a fixed T, $e^{-\kappa\Delta} = 1 - \kappa T/n + O(n^{-2}) \to 1$, implying that $X_{i\Delta}$ is a local-to-unity process (see, for example, Phillips, 1987). Under the double asymptotic scheme where $\Delta \to 0$ and $T \to \infty$, $X_{i\Delta}$ is a mildly stationary process for it has a root moderately deviated from unity, as studied in Wang and Yu (2016).

Whereas, when $H \neq 1/2$, $\{\varepsilon_{i\Delta}\}$ is serially dependent with the covariance

$$\mathbb{E}\left(\varepsilon_{i\Delta}\varepsilon_{(i+j)\Delta}\right) = \sigma^2 H(2H-1) \int_{(i-1)\Delta}^{i\Delta} \int_{(i+j-1)\Delta}^{(i+j)\Delta} e^{-\kappa(2(i+j)\Delta - s - \tau)} |s-\tau|^{2H-2} \, ds d\tau ,$$

(see, Lemma 2.1 of Cheridito et al. (2003)). Note that, as $\Delta \to 0$,

$$\varepsilon_{i\Delta} = \sigma \int_{(i-1)\Delta}^{i\Delta} e^{-\kappa(i\Delta - s)} dB_s^H = \sigma\left(B_{i\Delta}^H - B_{(i-1)\Delta}^H\right) + O_p\left(\Delta^{1+H}\right) := \sigma v_{i\Delta} + O_p\left(\Delta^{1+H}\right),$$

where $v_{i\Delta} = B_{i\Delta}^H - B_{(i-1)\Delta}^H$. Hence, the dependence structure of $\{\varepsilon_{i\Delta}\}$ is determined by that of $\{v_{i\Delta}\}$. From (1.1), it can be proven that, for any fixed Δ, $\{v_{i\Delta}\}$ is stationary with the following autocovariance function

$$Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = \frac{1}{2}\Delta^{2H}\left\{|j+1|^{2H} - 2|j|^{2H} + |j-1|^{2H}\right\} = O\left(j^{2H-2}\right) \quad \text{as } j \to \infty.$$

If $H \in (1/2, 1)$, it has $Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) > 0$ and $\sum_{j=0}^{\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = +\infty$. In this case, both $\{v_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$ have positive serial correlations and long-range dependence. In contrast, if $H \in (0, 1/2)$, $Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) < 0$ for any $j \neq 0$ and $\sum_{j=-\infty}^{\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = 0$.[5] This property implies that $\{v_{i\Delta}\}$ is anti-persistent when $H \in (0, 1/2)$. Consequently, $\{\varepsilon_{i\Delta}\}$ has negative serial correlations that quickly decay to zero as the lag order increases.

The degree of persistence of $\{X_{i\Delta}\}$ defined in Model (1.2) is determined jointly by the autoregressive root $e^{-\kappa\Delta}$ and H. If $H = 1/2$, $\{\varepsilon_{i\Delta}\}$ is iid. In this case, the ACF of $\{X_{i\Delta}\}$ at the jth lag is $e^{-\kappa\Delta j}$ that converges to 0 at the exponential rate as $j \to \infty$ and hence summable. When $H \neq 1/2$, the convergence rate of the ACF of $\{X_{i\Delta}\}$ is $O\left(j^{2H-2}\right)$, which is solely determined by H as $j \to \infty$ and dominates the exponential rate $e^{-\kappa\Delta j}$. Clearly, only when $H \in (1/2, 1)$, can long-range dependence be found in $\{X_{i\Delta}\}$. However, if $e^{-\kappa\Delta}$ takes a value close to unity, which is ensured by κ taking a small positive value, $\{X_{i\Delta}\}$ behaves almost like the cumulative sums of $\{\varepsilon_{i\Delta}\}$. In this case, $e^{-\kappa\Delta j}$ plays an important role in determining the ACF of $\{X_{i\Delta}\}$ at small and moderate lags. As a result, even when $H \in (0, 1/2)$, the ACF of $\{X_{i\Delta}\}$ may decay slowly at small and moderate lags. However, if $H \in (1/2, 1)$, the positive serial dependence of $\{\varepsilon_{i\Delta}\}$ makes the ACF of $\{X_{i\Delta}\}$ decay "too" slowly and the sample path of $\{X_{i\Delta}\}$ "too" smooth. In practical applications, we do not know ex-ante if a slowly decaying ACF at small and moderate lag is generated by a small positive κ or by a large positive κ together with $H \in (1/2, 1)$. However, if a slowly decaying ACF and the rough feature are both important in a dataset, then a small positive κ and $H \in (0, 1/2)$ are expected. One goal of our paper is to find a reasonable method to estimate H and κ.

To better appreciate the discussion above, we simulate $\{X_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$ from Model (1.3) with various values of H. We set $\kappa = 0.2366$, $\mu = 2.4165$, $\sigma = 0.7007$, and $\Delta = 1/256$. Details about how to simulate data from Model (1.3) are given in Section 5. Fig. 1 plots simulated $\{X_{i\Delta}\}$ and the autocorrelation function of $\{X_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$, in which the left panels represent the model with $H = 0.7$ and the right panels represent the model with $H = 0.1299$ (which is the estimated H from the S&P500 reported later). This figure clearly shows that $\{\varepsilon_{i\Delta}\}$ has positive serial correlations when $H = 0.7$ and negative serial correlations when $H = 0.1299$. More importantly, as the autoregressive root $e^{-\kappa\Delta} = 0.9991$ is very close to unity, for both values of H, $\{X_{i\Delta}\}$ has positive and slowly decaying autocorrelations. The autocorrelations decay much more slowly when $H = 0.7$ than when $H = 0.1299$. Furthermore, the sample path of $\{X_{i\Delta}\}$ is smooth when $H = 0.7$, but quite rough when $H = 0.1299$.

[5] It can be shown that $\sum_{j=1}^{\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = \sum_{j=-1}^{-\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = -0.5 Var\left(v_{i\Delta}\right)$.

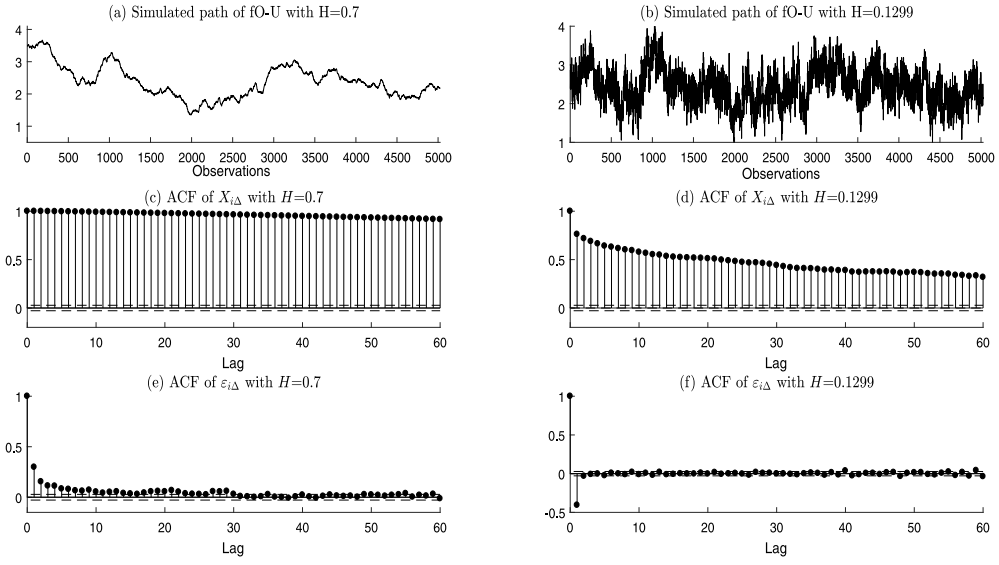

Fig. 1. Simulated sample paths of $\{X_{i\Delta}\}$, autocorrelation functions of $\{X_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$ from Model (1.3).

2.3. Relation to ARFIMA

There are close relationships and crucial differences between the models introduced in (1.2) and (1.3) and the following stationary ARFIMA$(1, d, 0)$ model with $d := H - 1/2$ that is widely used and extensively studied in the discrete-time literature:

$$\begin{aligned} &y_{i\Delta} = \mu(1-\rho) + \rho y_{(i-1)\Delta} + u_{i\Delta},\ |\rho| < 1, \\ &u_{i\Delta} = (1-L)^{-d} e_{i\Delta},\ e_{i\Delta} \sim i.i.d.(0, \sigma_e^2),\ i = 1, \ldots, n, \end{aligned} \tag{2.5}$$

where L is the lag operator with $(1-L)^{-d}$ defined as

$$(1-L)^{-d} = \sum_{j=0}^{\infty} \frac{\Gamma(j+d)}{\Gamma(d)\Gamma(j+1)} L^j.$$

As $H \in (0, 1)$, $d \in (-1/2, 1/2)$. Together with the condition that $|\rho| < 1$, the ARFIMA model is stationary. It is well-established in the literature that $\{u_{i\Delta}\}$ has long-range dependence when $d \in (0, 1/2)$ but it is anti-persistent when $d \in (-1/2, 0)$ (see, for example, Giraitis et al., 2012).

Letting $\rho = e^{-\kappa\Delta}$, $\sigma_e^2 = \frac{1-e^{-2\kappa\Delta}}{2\kappa}\sigma^2$, and $n = 1/\Delta$ (i.e., $T = 1$), using the result of Davydov (1970), we have, as $\Delta \to 0$,

$$\frac{\delta_H \Gamma(H+1/2)}{n^H} y_{\lfloor ns \rfloor} \Rightarrow X_s\,, \forall\ 0 \le s \le 1\,, \tag{2.6}$$

where $\delta_H = \sqrt{\frac{2H\Gamma(3/2-H)}{\Gamma(H+1/2)\Gamma(2-2H)}}$, $\lfloor z \rfloor$ denotes the greatest integer less than or equal to z, and $\{X_s\}$ is the fO–U process (see also Tanaka, 2013).

The weak convergence in (2.6) may lead one to believe that the fO–U model (1.3) is essentially identical to the ARFIMA$(1, d, 0)$ model. Unfortunately, this belief is not justified under popular estimation methods. To show the difference between fO–U and ARFIMA$(1, d, 0)$, we simulate data from Model (1.3) but fit the stationary ARFIMA$(1, d, 0)$ model with the frequency-domain (or Whittle) ML method of Whittle (1953). When simulating the data, we set $\kappa = 15$, $\mu = 2.8$, $\sigma = 1$, $H = 0.15$, $T = 4$, and $\Delta = 1/256$. This setup implies that $d = H - 1/2 = -0.35$ and $\rho = \exp(-\kappa\Delta) = 0.9414$. Table 1 reports the means and standard deviations (SD) of the Whittle ML estimates (MLE) of d and ρ over 200 replications. The mean of $\widehat{d}$ is very close to 0.4, whereas the mean of $\widehat{\rho}$ is very close to 0. Both values are far away from the true values. In fact, the Whittle MLE of d and ρ are very close to those obtained in the empirical studies when we fit the ARFIMA model

Table 1
Mean and standard deviation (SD) of the Whittle MLE of d and ρ when fitting the ARFIMA$(1, d, 0)$ model to data simulated from Model (1.3). When simulating data, we set $\kappa = 15$, $\mu = 2.8$, $\sigma = 1$, $H = 0.15$, $T = 4$, $\Delta = 1/256$. This setup implies that $d = H - 1/2 = -0.35$ and $\rho = \exp(-\kappa\Delta) = 0.9414$.

Mean of $\widehat{d}$	SD of $\widehat{d}$	Mean of $\widehat{\rho}$	SD of $\widehat{\rho}$
0.3954	0.0409	0.0118	0.0529

to the daily logarithmic RV of stock indices as shown in Section 6. Although not reported, the decreasing the value of Δ essentially leads to no change in the mean of $\widehat{d}$ and the mean of $\widehat{\rho}$.[6]

2.4. Out-of-sample forecast with fO–U

To forecast future X_t with the fO–U model, let F_s be the σ-algebra generated by fBm over $[0, s]$. According to Corollary 3.1 of Fink et al. (2013), the conditional expectation and conditional variance are

$$\mathbb{E}[X_t|\mathcal{F}_s] = X_s e^{-\kappa(t-s)} + \mu(1 - e^{-\kappa(t-s)}) - \frac{\kappa\mu}{\sigma}\int_0^s \Psi_H(s,t,v)dv + \frac{\kappa}{\sigma}\int_0^s \Psi_H(s,t,v)X_v dv + \frac{1}{\sigma}\int_0^s \Psi_H(s,t,v)dX_v\,, \tag{2.7}$$

$$Var[X_t|\mathcal{F}_s] = \left\|c(v)\,\mathbf{1}_{[s,t]}(v)\right\|_{H,t}^2 - \left\|\Psi_H(s,t,v)\,\mathbf{1}_{[0,s]}(v)\right\|_{H,t}^2\,, \tag{2.8}$$

where $c(v) = e^{-\kappa(t-v)}$, $\mathbf{1}_{[s,t]}(v) = 1$ if $v \in [s,t]$ and 0 otherwise,

$$\Psi_H(s,t,v) = \frac{\sin(\pi(H-1/2))}{\pi} v^{1/2-H}(s-v)^{1/2-H}\int_s^t \frac{z^{H-1/2}(z-s)^{H-1/2}}{z-v}\sigma e^{-\kappa(t-z)}dz,$$

and

$$\left\|f(v,t)\,\mathbf{1}_{[0,t]}(v)\right\|_{H,t}^2 = \frac{2\pi H(H-1/2)\int_0^t s^{1-2H}\left(\int_s^t \frac{r^{H-1/2}f(r,t)\mathbf{1}_{[0,t]}(r)}{(r-s)^{3/2-H}}dr\right)^2 ds}{\Gamma(2-2H)\sin(\pi(H-1/2))\,\Gamma^2(H-1/2)}\,. \tag{2.9}$$

By the Gaussian property of X_t, the h-step-ahead predictor of $RV_t = \exp(X_t)$ is

$$\widehat{RV}_{t+h} = \exp\left(\mathbb{E}[X_{t+h}|\mathcal{F}_t] + \frac{1}{2}Var[X_{t+h}|\mathcal{F}_t]\right)\,. \tag{2.10}$$

When only discrete-sampled data is available, say at periods $\Delta, 2\Delta, \ldots, n\Delta, (n+1)\Delta, \ldots, (n+m)\Delta$, we need to discrete the expression of $E[X_t|\mathcal{F}_s]$ and $Var[X_{t+h}|\mathcal{F}_t]$. Let $s = n\Delta$ and $t = (n+m)\Delta$. Since $\Psi_H(s,t,v)$ is a deterministic function, we can discretize $E[X_t|\mathcal{F}_s]$ using the Euler scheme:

$$\begin{aligned}\mathbb{E}[X_t|\mathcal{F}_s] &= X_s^{-\kappa(t-s)}e^{-\kappa(t-s)} + \mu(1-e^{-\kappa(t-s)}) - \frac{\kappa\mu}{\sigma}\int_0^s \Psi_H(s,t,v)dv \\ &+ \frac{\kappa}{\sigma}\int_0^s \Psi_H(s,t,v)X_v dv + \frac{1}{\sigma}\int_0^s \Psi_H(s,t,v)dX_v \\ &\approx X_s e^{-\kappa(t-s)} + \mu(1-e^{-\kappa(t-s)}) - \frac{\kappa\mu}{\sigma}\sum_{i=1}^n \Psi_H(s,t,i\Delta)\Delta \\ &+ \frac{\kappa}{\sigma}\sum_{i=1}^n \Psi_H(s,t,i\Delta)X_{i\Delta}\Delta + \frac{1}{\sigma}\sum_{i=1}^n \Psi_H(s,t,i\Delta)\left(X_{(i+1)\Delta} - X_{i\Delta}\right)\,,\end{aligned}$$

where

$$\begin{aligned}\Psi_H(s,t,v) &\approx \frac{\sin(\pi(H-1/2))}{\pi} v^{1/2-H}(n\Delta - v)^{1/2-H} \\ &\times \sum_{j=n+1}^{m+n} \frac{(j\Delta)^{H-1/2}(j\Delta - n\Delta)^{H-1/2}}{j\Delta - v}\sigma\Delta e^{-\kappa((n+m)\Delta - j\Delta)}\,.\end{aligned} \tag{2.11}$$

[6] We have also tried to use the log-periodogram method of Geweke and Porter-Hudak (1983) to fit the ARFIMA$(1, d, 0)$ to the simulated data and obtained similar results to the Whittle ML method.

To discretize the expression of $Var[X_t|\mathcal{F}_s]$, following Fink et al. (2013), for an interval $[i\Delta,(i+1)\Delta]$ where $i = 0,\ldots,n+m-1$, taking a partition $i\Delta = u_0^i \le u_1^i \le \cdots \le u_p^i = (i+1)\Delta$ with $p \in N$,[7] we have

$$\begin{aligned}
&\int_0^t s^{1-2H}\left(\int_s^t \frac{r^{H-1/2}c(r)\,\mathbf{1}_{[s,t]}(r)}{(r-s)^{3/2-H}}dr\right)^2 ds\\
&= \sum_{i=0}^{n+m+l-1}\int_{i\Delta}^{(i+1)\Delta} s^{1-2H}\left(\int_s^t \frac{r^{H-1/2}c(r)\,\mathbf{1}_{[s,t]}(r)}{(r-s)^{3/2-H}}dr\right)^2 ds\\
&\approx \sum_{i=0}^{n+m+l-1}\left[((i+1)\Delta)^{1-2H}-(i\Delta)^{1-2H}\right]\left[\sum_{j=1}^{p-1}\left[\left(u_{j+1}^i-i\Delta\right)^{H-\frac{1}{2}}-\left(u_j^i-i\Delta\right)^{H-\frac{1}{2}}\right]\right.\\
&\left.\times\left(\frac{1}{H-\frac{1}{2}}\right)^2\frac{\left(u_j^i\right)^{H-\frac{1}{2}}c\left(u_j^i\right)+\left(u_{j+1}^i\right)^{H-\frac{1}{2}}c\left(u_{j+1}^i\right)}{2}\right]^2.
\end{aligned}$$

Substituting the above result into (2.9), we obtain

$$\begin{aligned}
\left\|c(r)\,\mathbf{1}_{[s,t]}(r)\right\|_{H,t}^2 \approx{}& \frac{\pi H(2H-1)\sum_{i=0}^{n+m+l-1}\left[((i+1)\Delta)^{1-2H}-(i\Delta)^{1-2H}\right]}{4\Gamma(2-2H)\sin(\pi(H-1/2))\Gamma^2(H+1/2)}\\
&\times\left[\sum_{j=1}^{p-1}\left[\left(u_{j+1}^i-i\Delta\right)^{H-\frac{1}{2}}-\left(u_j^i-i\Delta\right)^{H-\frac{1}{2}}\right]\right.\\
&\left.\times\left[\left(u_j^i\right)^{H-\frac{1}{2}}c\left(u_j^i\right)+\left(u_{j+1}^i\right)^{H-\frac{1}{2}}c\left(u_{j+1}^i\right)\right]\right]^2. \qquad (2.12)
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\left\|\Psi_H(s,t,r)\,\mathbf{1}_{[0,s]}(r)\right\|_{H,t}^2 \approx{}& \frac{\pi H(2H-1)\sum_{i=0}^{n+m+l-1}\left[((i+1)\Delta)^{1-2H}-(i\Delta)^{1-2H}\right]}{4\Gamma(2-2H)\sin(\pi(H-1/2))\Gamma^2(H+1/2)}\\
&\times\left[\sum_{j=1}^{p-1}\left[\left(u_{j+1}^i-i\Delta\right)^{H-\frac{1}{2}}-\left(u_j^i-i\Delta\right)^{H-\frac{1}{2}}\right]\right.\\
&\left.\times\left[\left(u_j^i\right)^{H-\frac{1}{2}}\Psi_H\left(s,t,u_j^i\right)+\left(u_{j+1}^i\right)^{H-\frac{1}{2}}\Psi_H\left(s,t,u_{j+1}^i\right)\right]\right]^2,
\end{aligned}$$

where the approximation of $\Psi_H(s,t,v)$ is given by (2.11).[8]

3. A two-stage estimation approach

To estimate the parameters in (1.3) based on discrete-sampled data, it is difficult to apply the ML method to estimate all the parameters simultaneously for the reason that the errors $\{\varepsilon_{t\Delta}\}$ in (1.2) have a complicated dependence structure when $H \neq 1/2$. Following Phillips and Yu (2009b), we propose a two-stage estimation approach to estimate the parameters in Model (1.3). Our method is straightforward to implement.

In the first stage, following Lang and Roueff (2001) and Barndorff-Nielsen et al. (2013), we propose to estimate the Hurst parameter H by using the change-of-frequency (COF) estimator based on the second-order differences of X_t:

$$\widehat{H} = \frac{1}{2}\log_2\left(\frac{\sum_{i=1}^{n-4}\left(X_{(i+4)\Delta}-2X_{(i+2)\Delta}+X_{i\Delta}\right)^2}{\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}\right), \qquad (3.1)$$

where $\log_2(\cdot)$ is the base-2 logarithm, $\left\{X_{(i+4)\Delta}-2X_{(i+2)\Delta}+X_{i\Delta}\right\}_{i=1}^{n-4}$ and $\left\{X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right\}_{i=1}^{n-2}$ are the second-order differences of $\{X_{i\Delta}\}_{i=1}^n$ taken at two different frequencies.

The square summation of the second-order differences is known to be related to the smoothness parameter of a stochastic process. Lang and Roueff (2001) consider the COF estimator for Gaussian processes with stationary increments.

[7] In the empirical studies, we set $p = 100$.

[8] To forecast future RV with fBm, using Eq. (5.2) in Gatheral et al. (2018), we have

$$\widehat{RV}_{t+h} = \exp\left(\widehat{B}_{t+h}^H + 2cv^2h^{2H}\right), \qquad (2.13)$$

where $\widehat{B}_{t+h}^H$ is the h-step-ahead predictor of B_t^H, c and v^2 are defined in Section 5.2 of Gatheral et al. (2018).

Barndorff-Nielsen et al. (2013) consider the COF estimator for Brownian semi-stationary (BSS) processes when $H \in (0, 1/2) \cup (1/2, 3/4)$ while Corcuera et al. (2013) consider a modified COF estimator for BSS processes that is applicable when $H \in (3/4, 1)$. All these studies develop the in-fill asymptotic distribution of the COF estimator. Moreover, Bennedsen et al. (2019) introduce bootstrap procedures to improve the finite-sample performance of the COF-based hypothesis testing.

Gaussian processes with stationary increments and BSS processes include many interesting processes, including the fO–U model with $\kappa > 0$, as special cases (see, Corollary 4.3 and Remark 4.4 in Barndorff-Nielsen and Basse-O'Connor (2011)). However, the fO–U model with $\kappa < 0$ is not a special case of the Gaussian processes with stationary increments and BSS processes. While focusing our attention on the simple fO–U model, we extend the COF estimator to the explosive fO–U model in which $\kappa < 0$ and develop both the in-fill asymptotic and the double asymptotic distributions for all $H \in (0, 1)$.

Besides the COF estimator, many other estimators for H have been proposed and studied in the literature. Comprehensive surveys of these estimators can be found in Gneiting et al. (2012), and Bardet (2018). While the COF estimator may not be statistically efficient, we consider it for the following reasons. First, it is easy to implement and performs well in finite samples. Second, we can develop its asymptotic distribution for all values of H and κ. Third, its asymptotic variance has an analytical expression, which depends only on H. Hence, statistical inference of H can be made without knowing the values of the other parameters in the model.

In the second stage, assuming Model (1.3) is stationary, we estimate the other parameters, σ, μ, κ using the following method-of-moments estimators:

$$\widehat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n-2} \left(X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}\right)^2}{n\left(4 - 2^{2\widehat{H}}\right)\Delta^{2\widehat{H}}}}, \tag{3.2}$$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_{i\Delta}, \tag{3.3}$$

$$\widehat{\kappa} = \left(\frac{n\sum_{i=1}^{n} X_{i\Delta}^2 - \left(\sum_{i=1}^{n} X_{i\Delta}\right)^2}{n^2\widehat{\sigma}^2\widehat{H}\,\Gamma\left(2\widehat{H}\right)}\right)^{-1/(2\widehat{H})}. \tag{3.4}$$

Note that $\widehat{\sigma}$ depends on $\widehat{H}$ obtained in the first stage and $\widehat{\kappa}$ depends on both $\widehat{\sigma}$ and $\widehat{H}$.

The estimators in the second stage are based on a set of moment conditions. When Δ is small, $X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta} = \sigma\left(B^H_{(i+2)\Delta} - 2B^H_{(i+1)\Delta} + B^H_{i\Delta}\right) + O_p\left(\Delta^{1+H}\right)$. It is well studied in the literature that, when $\Delta = 1$, the process $\left\{B^H_{i+2} - 2B^H_{i+1} + B^H_i\right\}_{i=1}^n$ is stationary and ergodic with mean zero and variance $4 - 2^{2H}$ for any $H \in (0, 1)$. Using the self-similarity property of B^H_t, we have

$$Var\left(X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}\right) = \sigma^2\left(4 - 2^{2H}\right)\Delta^{2H} + o\left(\Delta^{2H}\right),$$

which justifies the estimator of σ^2 given in (3.2). The estimators $\widehat{\mu}$ and $\widehat{\kappa}$ come from the expressions of the unconditional mean and variance of $X_{i\Delta}$ given in (2.3).

Our estimators $\widehat{\mu}$ and $\widehat{\kappa}$ can be regarded as the discrete-time version of the ergodic-type estimators of κ and μ of Xiao and Yu (2019a,b). However, since Xiao and Yu (2019a,b) assume that σ^2 and H are known and a continuous record of $\{X_t\}$ is observed, we have to modify their estimators by (i) replacing σ and H with $\widehat{\sigma}$ and $\widehat{H}$; (ii) replacing the Riemann integral with the Riemann sum.

There is somewhat related but separate literature that focuses on estimating κ in the fO–U process. Almost all studies in this literature assume that H is known. When σ and H are known and a continuous record of X_t is available over the time interval $[0, T]$, Kleptsyna and Le Breton (2002) and Tanaka et al. (2020) obtain expressions for the exact MLE of κ which involve stochastic integrals. Replacing these stochastic integrals by corresponding Riemann sums calculated from discrete-time observations $\{X_{i\Delta}\}$, Tudor and Viens (2007) introduce an approximate MLE of κ with discrete-sampled data. However, the approximate MLE is challenging to implement, and its limiting distribution is unknown. Moreover, when σ and H are unknown, how to obtain an approximate MLE from discrete-sampled data remains as an unsolved problem. Bishwal (2011) proposes a minimum contrast (MC) estimator of κ. Hu and Nualart (2010) and Hu et al. (2019) develop the least-squares (LS) estimator of κ and the ergodic type estimator of κ. Tanaka (2013, 2015) and Xiao and Yu (2019a,b) extend these estimators from the stationary case with $\kappa > 0$ to the non-stationary case with $\kappa \leq 0$. When discrete-time observations of $\{X_t\}$ are available, the MC, LS, and ergodic type estimators have been studied by Ludeña (2004), Es-Sebaiy (2013) and Hu et al. (2019), respectively. A critical difference from these studies is that our study does not assume H is known when estimating κ.

For practical applications, it is important to assume that H is unknown when estimating other parameters. Moreover, it is important to estimate parameters based on a discrete record of data. It is because we estimate H and other parameters in the model based on a discrete record of data, we can examine the performance of fO–U model in empirical studies.

The asymptotic distribution of LS estimator of κ has been derived in the O–U model (H is assumed to be 1/2). For example, assuming $\kappa > 0$, Tang and Chen (2009) obtain the long-span and double asymptotic distributions of MLE of κ. Assuming $\kappa < 0$, Wang and Yu (2016) obtain the double asymptotic distribution of the LS estimator of κ. We extend the asymptotics of κ from the O–U model to the fO–U model.

4. Asymptotic theory

The large-sample theory of $\widehat{H}$ and $\widehat{\sigma}$ defined in (3.1) and (3.2) is developed in Section 4.1. We first show that $\widehat{H}$ and $\widehat{\sigma}$ are consistent as long as $T\Delta \to 0$ and $n = T/\Delta \to \infty$, a condition that is satisfied under either (i) the in-fill asymptotic scheme where $\Delta \to 0$ with a fixed T[9]; or (ii) the double asymptotic scheme where $\Delta \to 0$ and $T \to \infty$ simultaneously with T diverging at a lower rate than that of $1/\Delta$.[10] In Section 4.2, we show that $T \to \infty$ is a necessary condition for the consistency of $\widehat{\mu}$ and $\widehat{\kappa}$ defined in (3.3) and (3.4) and report the double asymptotic theory of $\widehat{\mu}$ and $\widehat{\kappa}$.

4.1. Asymptotic theory of $\widehat{H}$ and $\widehat{\sigma}$

Theorem 4.1. *Let $\widehat{H}$ and $\widehat{\sigma}$ be the estimators defined by (3.1) and (3.2) for Model (1.3). For all $H \in (0, 1)$, when $T\Delta \to 0$ and $n = T/\Delta \to \infty$, it has*

(a) $\widehat{H} \xrightarrow{p} H$ and

$$\sqrt{n}\left(\widehat{H} - H\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}}{(2\log 2)^2}\right). \tag{4.1}$$

(b) $\widehat{\sigma} \xrightarrow{p} \sigma$ and

$$\frac{\sqrt{n}}{\log(1/\Delta)}\left(\widehat{\sigma} - \sigma\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}}{(2\log 2)^2}\sigma^2\right), \tag{4.2}$$

where

$$\Sigma_{11} = 2 + 2^{2-4H}\sum_{j=1}^{\infty}\left(\rho_{j+2} + 4\rho_{j+1} + 6\rho_j + 4\rho_{|j-1|} + \rho_{|j-2|}\right)^2, \tag{4.3}$$

$$\Sigma_{12} = 2^{1-2H}\left(4(\rho_1 + 1)^2 + 2\sum_{j=0}^{\infty}\left(\rho_{j+2} + 2\rho_{j+1} + \rho_j\right)^2\right), \tag{4.4}$$

$$\Sigma_{22} = 2 + 4\sum_{j=1}^{\infty}\rho_j^2, \tag{4.5}$$

with

$$\rho_j = \frac{1}{2\left(4 - 2^{2H}\right)}\left(-|j+2|^{2H} + 4|j+1|^{2H} - 6|j|^{2H} + 4|j-1|^{2H} - |j-2|^{2H}\right). \tag{4.6}$$

Remark 4.1. The asymptotics of $\widehat{H}$ and $\widehat{\sigma}$ apply to all $H \in (0, 1)$. Moreover, they apply to all κ, including $\kappa > 0$, $\kappa = 0$, and $\kappa < 0$.

Remark 4.2. It can be proved that $\rho_j = O\left(j^{2H-4}\right)$ as $j \to \infty$. Hence, for any $H \in (0, 1)$, the sequence $\left\{\rho_j\right\}_{j=1}^{\infty}$ is square summable, ensuring that the infinite sums in Σ_{11}, Σ_{12}, and Σ_{22} are all finite. Using the mean value theorem for integrals repeatedly, we have, as $j \to \infty$,

$$\begin{aligned}
&2\left(4 - 2^{2H}\right)\rho_j \\
=&2H\left\{-(j+1+\lambda_1)^{2H-1} + 3(j+\lambda_2)^{2H-1} - 3(j-1+\lambda_3)^{2H-1} + (j-2+\lambda_4)^{2H-1}\right\} \\
=&O\left(j^{2H-4}\right),
\end{aligned}$$

where $\{\lambda_s\}_{s=1}^{6}$ are real numbers in the interval $(0, 3)$.

Remark 4.3. Fig. 2 plots the values of the asymptotic variance of $\sqrt{n}\left(\widehat{H} - H\right)$ for $H \in (0, 1)$. It shows that the asymptotic variance of $\sqrt{n}\left(\widehat{H} - H\right)$ is a decreasing function in H over the interval $H \in (0, 1)$.

[9] The relevance and superiority of the in-fill asymptotic theory relative to the long-span asymptotic theory have been well documented in econometrics, even when the available frequency is daily, weekly or monthly; see, Yu (2014) and Zhou and Yu (2015), Jiang et al. (2018, 2021) for further details.

[10] The consistency of $\widehat{H}$ only requires $\Delta \to 0$. In other words, even when T diverges faster than $1/\Delta$, violating the condition $T\Delta \to 0$, $\widehat{H}$ is still consistent as long as $\Delta \to 0$.

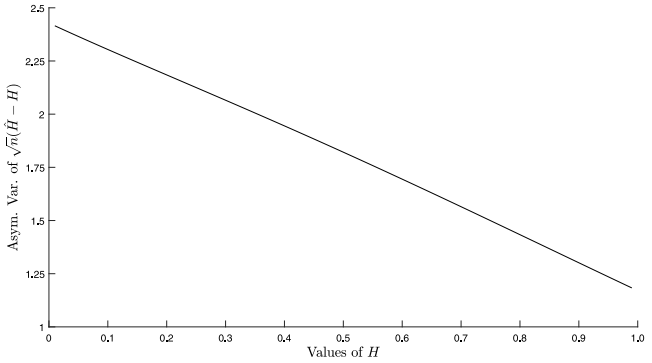

Fig. 2. Asymptotic variance of $\sqrt{n}\left(\widehat{H}-H\right)$ as a function of $H \in (0, 1)$.

Remark 4.4. It is worth mentioning that the asymptotic variance of $\widehat{H}$ only depends on H while the asymptotic variance of $\widehat{\sigma}$ only depends on H and σ. This feature greatly facilitates statistical inference about H and σ because H and σ can be consistently estimated when T is fixed but κ and μ cannot. If an alternative estimator of H is used, as long as the rate of convergence remains to be $\sqrt{n}$, $\sqrt{n}\left(\widehat{\sigma}-\sigma\right)/(\sigma \log(\Delta))$ should have the same limiting distribution.

When $H = 1/2$, Model (1.3) becomes the O–U model which enjoys the Markov property. Whereas, if $H \neq 1/2$, Model (1.3) does not have the Markov property anymore. To facilitate the test of the hypothesis $H = 1/2$, Corollary 4.2 gives the value of the asymptotic variance of $\sqrt{n}\left(\widehat{H}-1/2\right)$. Putting $H = 1/2$ into the formulae given in Theorem 4.1, we get that $\rho_0 = 1$, $\rho_1 = -1/2$, $\rho_j = 0$ for $j \geq 2$, $\Sigma_{11} = 7/2$, $\Sigma_{12} = 3/2$, and $\Sigma_{22} = 3$, and then Corollary 4.2 is obtained directly and reported below.

Corollary 4.2. *When $H = 1/2$, we have, as $T\Delta \to 0$ and $n = T/\Delta \to \infty$,*

$$\sqrt{n}\left(\widehat{H}-1/2\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{7}{8(\log 2)^2}\right).$$

4.2. Asymptotic theory of $\widehat{\mu}$ and $\widehat{\kappa}$

To develop the asymptotic theory of $\widehat{\mu}$ and $\widehat{\kappa}$ defined in (3.3) and (3.4), we need the double asymptotic scheme where $T \to \infty$ and $\Delta \to 0$. We may also need a condition to govern the relative divergence/convergence rates of T and Δ.

Theorem 4.3. *Let $\widehat{\mu}$ be the estimator of μ defined in (3.3) for Model (1.3) with $\kappa > 0$. For all $H \in (0, 1)$, when $T \to \infty$ and $\Delta \to 0$, we have $\widehat{\mu} \xrightarrow{p} \mu$. If, in addition, $T^{1-H}\Delta^H \to 0$, then*

$$T^{1-H}\left(\widehat{\mu}-\mu\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2/\kappa^2\right). \tag{4.7}$$

Theorem 4.4. *Let $\widehat{\kappa}$ be the estimator of κ defined in (3.4) for Model (1.3) with $\kappa > 0$. For all $H \in (0, 1)$, when $T \to \infty$ and $T\Delta \to 0$, we have $\widehat{\kappa} \xrightarrow{p} \kappa$. If, in addition,*

(a) for $H \in (0, 3/4)$, $\sqrt{T}\Delta^H \to 0$, then

$$\sqrt{T}\left(\widehat{\kappa}-\kappa\right) \xrightarrow{d} \mathcal{N}\left(0, \kappa\phi_H\right), \tag{4.8}$$

with

$$\phi_H = \begin{cases} \frac{1}{4H^2}\left[(4H-1) + \frac{2\Gamma(2-4H)\Gamma(4H)}{\Gamma(2H)\Gamma(1-2H)}\right] & \text{if } H \in (0, \frac{1}{2}) \\ \frac{4H-1}{4H^2}\left[1 + \frac{\Gamma(3-4H)\Gamma(4H-1)}{\Gamma(2-2H)\Gamma(2H)}\right] & \text{if } H \in [\frac{1}{2}, \frac{3}{4}) \end{cases};$$

(b) for $H = 3/4$, $\sqrt{T}\Delta^H/\log(T) \to 0$, then

$$\frac{\sqrt{T}}{\log(T)}\left(\widehat{\kappa}-\kappa\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{16\kappa}{9\pi}\right);$$

(c) for $H \in (3/4, 1)$, $T^{2-2H}\Delta^H \to 0$, then

$$T^{2-2H}\left(\widehat{\kappa}-\kappa\right) \xrightarrow{d} \frac{-\kappa^{2H-1}}{H\Gamma(2H+1)}R,$$

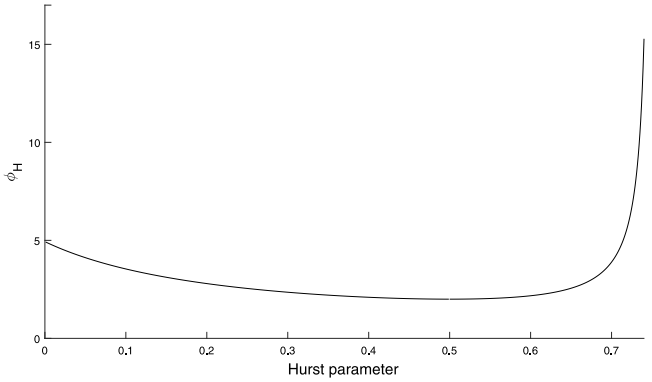

Fig. 3. Plot of ϕ_H as a function of H.

where R is the Rosenblatt random variable whose characteristic function is given by

$$c(s) = \exp\left(\frac{1}{2}\sum_{k=2}^{\infty}\left(2\sqrt{-1}s\psi(H)\right)^k \frac{a_k}{k}\right),$$

with $\psi(H) = \sqrt{H(H-1/2)}$ *and*

$$a_k = \int_0^1\int_0^1\cdots\int_0^1 |x_1 - x_2|^{H-1}\cdots|x_{k-1} - x_k|^{H-1}|x_k - x_1|^{H-1}\,dx_1\cdots dx_k.$$

Remark 4.5. If an alternative estimator of H is used, as long as the rate of convergence remains to be $\sqrt{n}$ and the condition that governs the relative divergence/convergence rates of T and Δ is imposed, we expect that Theorems 4.3 and 4.4 continue to hold.

Remark 4.6. Note that ϕ_H in Part (a) of Theorem 4.4 is continuous at $H = 1/2$. Using the formula $\Gamma(z+1) = z\Gamma(z)$, ϕ_H for $H \in (0, 1/2)$ can be rewritten as

$$\phi_H = \frac{1}{4H^2}\left[(4H-1) + \frac{\Gamma(3-4H)\,\Gamma(4H)}{\Gamma(2H)\,\Gamma(2-2H)}\right].$$

Hence, when $H \to 1/2$ from the left side of $1/2$, we have

$$\lim_{H\nearrow 1/2}\phi_H = \left[1 + \frac{\Gamma(1)\,\Gamma(2)}{\Gamma(1)\,\Gamma(1)}\right] = 2.$$

If $H = 1/2$, $\phi_H = 2$. Hence, ϕ_H is continuous at $H = 1/2$.

Remark 4.7. When $H = 1/2$ and is known, the double asymptotic distribution of the MLE of κ is known to be $\mathcal{N}(0, 2\kappa)$; see, for example, Tang and Chen (2009). Since $\phi_H = 2$ when $H = 1/2$, our method-of-moments estimator $\widehat{\kappa}$ has the same limiting distribution as the MLE in this case. Therefore, $\widehat{\kappa}$ is asymptotically efficient when $H = 1/2$.

Remark 4.8. Fig. 3 plots ϕ_H as a function of H that reaches the minimum at $H = 1/2$. Over the interval $(0, 1/2]$, ϕ_H is decreasing in H. Whereas, over the interval $[1/2, 3/4)$, ϕ_H monotonically increases to $+\infty$ as $H \to 3/4$. This feature suggests that the convergence rate of $\widehat{\kappa} - \kappa$ should be lower than $1/\sqrt{T}$ when $H = 3/4$. Part (b) of Theorem 4.4 shows that the convergence rate of $\widehat{\kappa} - \kappa$ is $\log(T)/\sqrt{T}$ when $H = 3/4$.

5. Monte Carlo studies

This section checks the finite-sample performance of the proposed estimators and the developed asymptotic theory with data simulated from Model (1.3), various values of H, σ, μ and κ, and different combinations of the sampling frequency Δ and the time span T. The data simulation and parameter estimation steps are summarized as follows:

(i) Set values for parameters H, μ, κ, σ, in Model (1.3).

(ii) Choose the values of Δ and T, and hence, the number of observations for parameter estimation $n = T/\Delta$.

(iii) For any given Δ, choose the value of $M > 1$ to get a finer grid

$$\{0, \Delta/M, 2\Delta/M, \ldots, \Delta; (M+1)\Delta/M, (M+2)\Delta/M, \ldots, 2\Delta; \ldots, n\Delta\}.$$

Then, generate a series of fractional Gaussian noise $\left\{B^H_{j\gamma} - B^H_{(j-1)\gamma}\right\}_{j=1}^{nM}$ by using fast Fourier transformation at the finer grid $\gamma := \Delta/M$.[11]

(iv) The Euler approximation of Model (1.3) over the interval $((j-1)\gamma, j\gamma)$ takes the form of

$$X_{j\gamma} = X_{(j-1)\gamma} + \kappa\left(\mu - X_{(j-1)\gamma}\right)\gamma + \sigma\left(B^H_{j\gamma} - B^H_{(j-1)\gamma}\right). \tag{5.1}$$

Starting from any pre-determined initial value X_0, the time series $\left\{X_{j\gamma}\right\}_{j=1}^{nM}$ is generated recursively based on Eq. (5.1) with the simulated fractional Gaussian noise series $\left\{B^H_{j\gamma} - B^H_{(j-1)\gamma}\right\}_{j=1}^{nM}$ obtained in Step 3. A subset of $\left\{X_{j\gamma}\right\}_{j=1}^{nM}$ is $\{X_{i\Delta}\}_{i=0}^{n}$, which gives the simulated sample path of the process X_t with the target sampling interval Δ.[12]

(v) Using the simulated data $\{X_{i\Delta}\}_{i=0}^{n}$, estimate H, μ, κ, and σ based on the estimators defined in (3.1), (3.2), (3.3), and (3.4), respectively.

(vi) Replicate the above procedure 10,000 times.

In the first experiment, we investigate the finite-sample properties of $\widehat{H}$ defined by (3.1) under various combinations of the sampling frequency Δ and the time span T. We let the true value of H vary from 0.1 to 0.9, and set $\kappa = 0.2366$, $\mu = 2.4165$, and $\sigma = 0.7007$, which are the estimated values when Model (1.3) is fitted to the logarithmic daily RV of S&P 500 index. Simulation results are reported in Table 2, including the mean and SD. For comparison, we also report, in parentheses, the SD implied by the asymptotic theory given by (4.1).

Table 2 reveals several features. First, for all combinations of Δ, T, and H, the estimator $\widehat{H}$ always has a very small bias and a small SD. This suggests that H can be accurately estimated by $\widehat{H}$. Second, the bias and the SD become smaller when the sampling interval Δ decreases, or the time span T increases. This finding supports the asymptotic theory of $\widehat{H}$ given by (4.1). Third, the finite-sample SD is very close to the asymptotic counterpart, suggesting that the asymptotic distribution derived in Theorem 4.1 provides excellent approximations to finite-sample distribution. Since our interest in this paper is to model logarithmic daily RV, $\Delta = 1/256$ is more relevant.

In the second experiment, we set $\kappa = 0.2366$, $\mu = 2.4165$, $\sigma = 0.7007$, $T = 16$, $\Delta = 1/256$, and let H vary from 0.1 to 0.7. Table 3 reports the estimation results of each parameter (H, σ, μ, and κ) and reveals several features. First, σ and μ can always be accurately estimated with negligible biases and small SDs. When the value of H increases from 0.1 to 0.7, the SD of $\widehat{\sigma}$ decreases, as predicted by the asymptotic theory given by (4.2) and by Fig. 2, which shows that the asymptotic variance of $\widehat{\sigma}$ is a decreasing function of H.[13] Furthermore, as H increases, the SD of $\widehat{\mu}$ increases. This observation is also supported by the asymptotic theory given in (4.7), which shows that the convergence rate of $\widehat{\mu}$ is T^{1-H}, hence larger values of H worsen the precision of $\widehat{\mu}$. Second, the parameter κ can be estimated with less precision. The SDs are comparatively large, and the bias in $\widehat{\kappa}$ are noticeable. Also, the finite-sample SDs are very different from the asymptotic SDs. The difficulties in estimating κ have been well studied for continuous-time models driven by standard Brownian motion; see, Phillips and Yu (2005, 2009a) and Wang et al. (2011). Tang and Chen (2009) and Yu (2012) derive

[11] Details of the use of fast Fourier transformation to generate a series of fractional Gaussian noise can be found in Paxson (1997). Other methods for simulating fBm can be seen in a survey paper by Coeurjolly (2000).

[12] For any target sampling interval Δ, a representation of Model (1.3) over the interval $((i-1)\Delta, i\Delta)$ is

$$X_{i\Delta} = X_{(i-1)\Delta} + \kappa\mu\Delta - \kappa\int_{(i-1)\Delta}^{i\Delta} X_t dt + \sigma\left(B^H_{i\Delta} - B^H_{(i-1)\Delta}\right). \tag{5.2}$$

If we let $\gamma = \Delta$ (i.e. $M = 1$), Eq. (5.1) for simulating the data becomes

$$X_{i\Delta} = X_{(i-1)\Delta} + \kappa\mu\Delta - \kappa X_{(i-1)\Delta}\Delta + \sigma\left(B^H_{i\Delta} - B^H_{(i-1)\Delta}\right),$$

which is the same as Eq. (5.2) but with the integral $\int_{(i-1)\Delta}^{i\Delta} X_t dt$ replaced by $X_{(i-1)\Delta}\Delta$. If we choose an $M > 1$, by dividing the interval $((i-1)\Delta, i\Delta)$ into M equally-spaced subintervals as $\cup_{j=(i-1)M+1}^{iM}((j-1)\gamma, j\gamma]$ and simulating data based on Eq. (5.1), then the simulated data are

$$X_{i\Delta} = X_{(i-1)\Delta} + \kappa\mu\Delta - \kappa\sum_{j=(i-1)M+1}^{iM} X_{(j-1)\gamma}\gamma + \sigma\left(B^H_{i\Delta} - B^H_{(i-1)\Delta}\right),$$

which is the same as Eq. (5.2) but with the integral replaced by the corresponding Riemann sum, i.e., $\int_{(i-1)\Delta}^{i\Delta} X_t dt \approx \sum_{j=(i-1)M+1}^{iM} X_{(j-1)\gamma}\gamma$. Clearly, the larger M is, the smaller the approximation error generated by using Riemann sums. When Δ is small, the approximation error can be ignored even when M is a relatively small number. Our idea is the same as the in-fill technique used in Elerian et al. (2001).

[13] While the asymptotic variance of $\widehat{\sigma}$ does not strictly monotonically decrease in H, when we increase the number of replications to 100,000, the strict monotonicity is found.

Table 2
Finite sample properties of $\widehat{H}$. Values reported in parentheses correspond to the asymptotic theory in (4.1). We set $M = 8$, and fix $\kappa = 0.2366$, $\mu = 2.4165$, and $\sigma = 0.7007$.

Value of H		0.1	0.2	0.3	0.5	0.7	0.8	0.9
Panel A: $T = 4$								
$\Delta = \frac{1}{256}$	Mean	0.0984	0.1983	0.2982	0.4982	0.6981	0.7981	0.8981
	SD	0.0470 (0.0474)	0.0460 (0.0461)	0.0449 (0.0449)	0.0423 (0.0421)	0.0392 (0.0390)	0.0375 (0.0374)	0.0357 (0.0356)
$\Delta = \frac{1}{512}$	Mean	0.0995	0.1995	0.2994	0.4993	0.6992	0.7991	0.8991
	SD	0.0334 (0.0335)	0.0326 (0.0326)	0.0319 (0.0317)	0.0301 (0.0298)	0.0280 (0.0276)	0.0269 (0.0264)	0.0256 (0.0252)
Panel B: $T = 16$								
$\Delta = \frac{1}{256}$	Mean	0.0995	0.1995	0.2994	0.4994	0.6993	0.7993	0.8993
	SD	0.0239 (0.0237)	0.0232 (0.0230)	0.0225 (0.0224)	0.0211 (0.0210)	0.0196 (0.0195)	0.0187 (0.0187)	0.0179 (0.0178)
$\Delta = \frac{1}{512}$	Mean	0.0998	0.1998	0.2998	0.4998	0.6998	0.7998	0.8998
	SD	0.0166 (0.0167)	0.0161 (0.0163)	0.0157 (0.0158)	0.0147 (0.0149)	0.0137 (0.0138)	0.0131 (0.0132)	0.0125 (0.0126)

Table 3
Finite sample properties of the estimates of the parameters (H, σ, μ, κ) with various values of H, $T = 16$ and $\Delta = 1/256$. Values reported in parentheses are the asymptotic SDs implied by (4.1) for $\widehat{H}$, (4.2) for $\widehat{\sigma}$, (4.7) for $\widehat{\mu}$, and (4.8) for $\widehat{\kappa}$.

	H	σ	μ	κ	::::	H	σ	μ	κ
True value	0.1000	0.7007	2.4165	0.2366		0.3000	0.7007	2.4165	0.2366
Mean	0.0995	0.7046	2.4305	0.5847		0.2994	0.7040	2.4266	0.4975
SD	0.0239 (0.0237)	0.0926 (0.0921)	0.2315 (0.2442)	0.8977 (0.2288)		0.0225 (0.0224)	0.0900 (0.0872)	0.3813 (0.4252)	0.4089 (0.1865)
True value	0.5000	0.7007	2.4165	0.2366		0.7000	0.7007	2.4165	0.2366
Mean	0.4994	0.7037	2.4242	0.5395		0.6993	0.7040	2.4229	0.6120
SD	0.0211 (0.0210)	0.0890 (0.0819)	0.6284 (0.7403)	0.3455 (0.1719)		0.0196 (0.0195)	0.0925 (0.0759)	1.0390 (1.2890)	0.3467 (0.2398)

analytical expressions to approximate the bias in the LS of κ when $H = 1/2$. Our simulation results show that the bias in estimating κ continues to exist for continuous-time models driven by fBm and depends not only on κ but also on H in a nonlinear fashion. This finding is supported by the asymptotic theory given in Theorem 4.4, which shows that both the convergence rate and the asymptotic variance of $\widehat{\kappa}$ depend crucially on H.

In the third experiment, we fix H to 0.1299 (which is the estimated H for the S&P500 reported later) and allow the other parameters (σ, μ, and κ) to take various values to determine how a change in one parameter affects the estimates of the other parameters. Panel A of Table 4 reports the simulation results when $\kappa = 0.2366$, $\mu = 2.4165$, and σ varies from 0.3 to 2. The simulation results confirm the developed asymptotic theory that an increase in σ should not affect estimation of H and κ, but it should increase the variance of $\widehat{\sigma}$ and $\widehat{\mu}$.

Panel B of Table 4 reports the simulation results when $\kappa = 0.2366$, $\sigma = 0.7007$, and μ varies from 1 to 4. As predicted by the asymptotic theory, the estimation results of H, σ, and κ and the SD of $\widehat{\mu}$ all remain the same when the value of μ changes.

Panel C of Table 4 reports the simulation results when $\mu = 2.4165$, $\sigma = 0.7007$, and κ varies from 0.1 to 12.5. It shows that the estimation results for H and σ are insensitive to the change in κ, whereas when the value of κ increases, the SD of $\widehat{\mu}$ decreases and the SD of $\widehat{\kappa}$ increases. Again, these findings are consistent with the suggestions of the developed asymptotic theory.

6. Empirical studies

This section reports empirical studies where we assume the logarithmic daily RV of equities follows Model (1.2) and hence, under the in-fill scheme, Model (1.3). We apply the proposed model, the estimation method, and the new asymptotic theory to each logarithmic daily RV series and test the hypothesis of $H = 0.5$. We also compare the out-of-sample performance of Model (1.3) relative to five competing models, namely the random walk (RW), AR(1), HAR of Corsi (2009), ARFIMA$(1, d, 0)$, and fBm in forecasting RV. When forecasting future RV, we replace the underlying parameters in each candidate model with their estimates. Then, we evaluate the forecasting performance of the competing models

Table 4
Estimates of (H, σ, μ, κ) when $H = 0.1299$, $T = 16$, $\Delta = 1/256$, and (σ, μ, κ) take various values. Values reported in parentheses are the asymptotic SDs implied by (4.1) for $\widehat{H}$, (4.2) for $\widehat{\sigma}$, (4.7) for $\widehat{\mu}$, and (4.8) for $\widehat{\kappa}$.

	H	σ	μ	κ	::::	H	σ	μ	κ
Panel A: σ varies									
True value	0.1299	0.3000	2.4165	0.2366		0.1299	0.5000	2.4165	0.2366
Mean	0.1294	0.3016	2.4302	0.5292		0.1294	0.5027	2.4300	0.5312
SD	0.0237	0.0394	0.1068	0.7056		0.0237	0.0657	0.1780	0.7062
	(0.0235)	(0.0391)	(0.1136)	(0.2199)		(0.0235)	(0.0652)	(0.1893)	(0.2199)
True value	0.1299	1.0000	2.4165	0.2366		0.1299	2.0000	2.4165	0.2366
Mean	0.1294	1.0055	2.4293	0.5319		0.1294	2.0110	2.4281	0.5321
SD	0.0237	0.1315	0.3561	0.7061		0.0237	0.2631	0.7123	0.7059
	(0.0235)	(0.1304)	(0.3786)	(0.2199)		(0.0235)	(0.2609)	(0.7573)	(0.2199)
Panel B: μ varies									
True value	0.1299	0.7007	1.0000	0.2366		0.1299	0.7007	2.0000	0.2366
Mean	0.1294	0.7045	1.0049	0.5321		0.1294	0.7045	2.0108	0.5318
SD	0.0237	0.0922	0.2495	0.7060		0.0237	0.0922	0.2495	0.7062
	(0.0235)	(0.0914)	(0.2653)	(0.2199)		(0.0235)	(0.0914)	(0.2653)	(0.2199)
True value	0.1299	0.7007	3.0000	0.2366		0.1299	0.7007	4.0000	0.2366
Mean	0.1294	0.7045	3.0166	0.5314		0.1294	0.7045	4.0225	0.5307
SD	0.0237	0.0922	0.2495	0.7062		0.0237	0.0922	0.2495	0.7061
	(0.0235)	(0.0914)	(0.2653)	(0.2199)		(0.0235)	(0.0914)	(0.2653)	(0.2199)
Panel C: κ varies									
True value	0.1299	0.7007	2.4165	0.1000		0.1299	0.7007	2.4165	0.5000
Mean	0.1294	0.7045	2.6585	0.4307		0.1294	0.7046	2.4160	0.7737
SD	0.0237	0.0921	0.4840	0.6292		0.0237	0.0922	0.1236	0.8866
	(0.0235)	(0.0914)	(0.6278)	(0.1430)		(0.0235)	(0.0914)	(0.1255)	(0.3198)
True value	0.1299	0.7007	2.4165	2.5000		0.1299	0.7007	2.4165	12.5000
Mean	0.1294	0.7048	2.4163	2.8762		0.1289	0.7044	2.4164	12.7455
SD	0.0237	0.0922	0.0253	2.2122		0.0237	0.0922	0.0058	6.2194
	(0.0235)	(0.0914)	(0.0251)	(0.7151)		(0.0235)	(0.0914)	(0.0050)	(1.5990)

Table 5
Empirical results for $\log(100\sqrt{RV \times 252})$ of S&P 500, DJIA, and NASDAQ 100.

Name	H	σ	μ	κ	T_n
S&P 500	.1299	.7007	2.4165	.2366	3.3027
	(.0876, .1722)	(.6953, .7061)	(1.9869, 2.8460)	(.2122, .2609)	
DJIA	.0829	.5770	2.4296	.0154	3.1366
	(.0399, .1258)	(.5726, .5815)	(−2.2950, 7.1542)	(.0087, .0220)	
NASDAQ 100	.2269	1.0482	2.5166	2.9182	2.4605
	(.1858, .2678)	(1.0403, 1.0559)	(2.4468, 2.5862)	(2.8411, 2.9952)	

based on the root mean squared error (RMSE), the Mincer–Zarnowitz-R^2 (MZ-R^2) of Mincer and Zarnowitz (1969).[14] We also check the statistical significance of the forecasts from these competing models using the Diebold–Mariano (DM) test of Diebold and Mariano (1995), and the model confidence set (MCS) of Hansen et al. (2011). When comparing the nested models, we use the Clark and West (CW) test of Clark and West (2007). In the online supplement, we apply our method to the logarithmic daily bipower variation (BV) and the logarithmic daily realized kernel (RK) for the equities. We show that the empirical results reported in the main text continue to hold.

We first fit Model (1.3) to three logarithmic daily RV series for the S&P 500, DJIA, and NASDAQ 100. The three RV series are obtained from the Oxford-Man realized library and based on 5-minute returns.[15] The sample period is from January 3, 2000 to December 31, 2019. Fig. 4 plots three time series of $\log\left(100\sqrt{RV \times 252}\right)$ which is the logarithmic annualized RV.

Table 5 reports the estimation results for the fO–U model, including the point estimates and the 95% confidence intervals (CIs) for all four parameters, and the CUSUM statistic T_n. The CIs are obtained from our asymptotic theory with the asymptotic variances obtained by the plug-in method. In all cases, the estimated H is much less than 0.5, ranging

[14] As in Andersen et al. (2003), we project the actual RV data on a constant and each of the competing forecasted series. The MZ-R^2 is the coefficient of determination of the projection that captures the correlation between the actual and forecast series, and hence, is often used to compare the performance of competing forecasts.

[15] The data are obtained from https://realized.oxford-man.ox.ac.uk/.

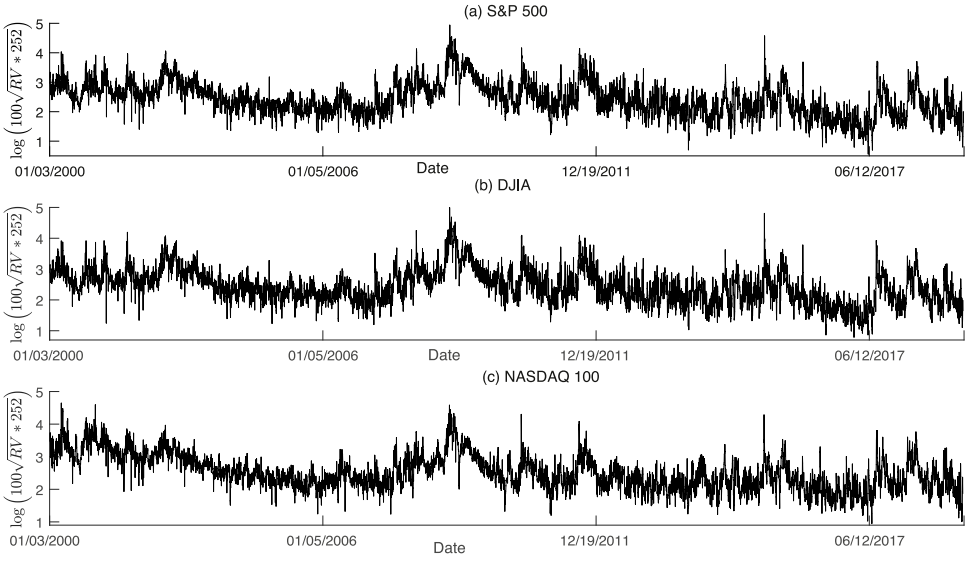

Fig. 4. Time series plots of $\log(100\sqrt{RV \times 252})$ for the three equity series.

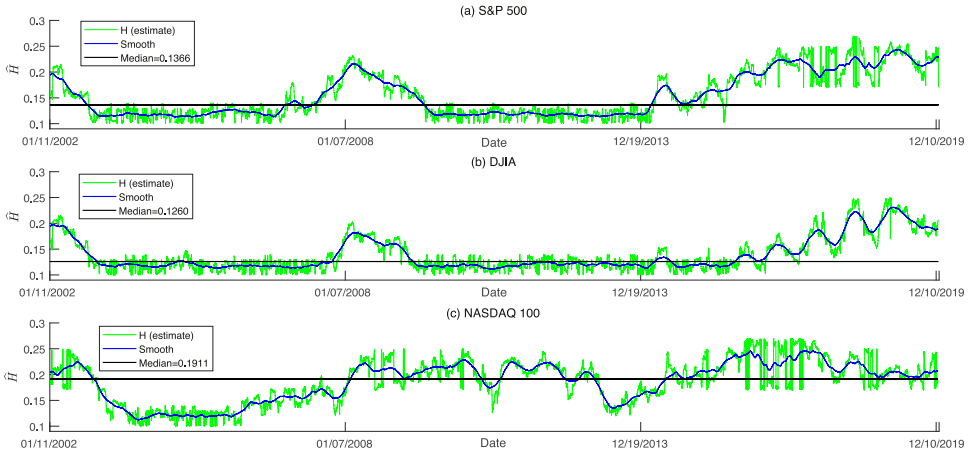

Fig. 5. Rolling-window estimates of H for $\log(100\sqrt{RV \times 252})$. The smoothed version is a simple moving average filter using 100 values on both sides.

between 0.0829 for DJIA to 0.2269 for NASDAQ 100. The 95% CIs suggest that we have strong evidence against $H \geq 1/2$. In all cases, the estimated κ is positive but very close to zero, but the 95% confidence intervals of κ exclude zero. Hence, each RV series is better modeled by fO–U with $H < 1/2$ than by fO–U with $H > 1/2$, fBm and O–U.

To test if there are structural breaks in H, we employ the CUSUM test, T_n, of Bibinger (2020) that converges to the Kolmogorov–Smirnov limit law for any $H \in (0, 1)$. The CUSUM test T_n rejects the hypothesis of a constant H.[16] To show the degree of instability in $\widehat{H}$, following Bennedsen et al. (2021), we estimate H based on rolling windows, each of which has 504 observations. Fig. 5 plots the rolling-window estimates of H, a smoothed version of the estimates, and the median value of the estimates. Although $\widehat{H}$ varies over time, their values always lie in the interval $[0.1, 0.25]$. The results are qualitatively the same for different window sizes or for expanding windows.

We now compare performance of fO–U, fBm, ARFIMA, HAR, AR(1), and RW in forecasting RV. To evaluate the out-of-sample forecasting performance, we split the sample period into two periods. The first period is between January 3, 2000 and December 31, 2018 and the second period is between January 2, 2019 and December 31, 2019. On each day in the second period, h-day-ahead (with $h = 1, \ldots, 10$) forecasts of RV are obtained from six competing models, each of which is estimated using data from January 3, 2000 to the day before.

16 The critical value at the 5% level is $1.3581/\sqrt{n}$. When $n = 5071$, it is 0.0192.

Table 6
The ratio of RMSE of different models and that of HAR for h-day-ahead-forecast of RV.

h	1	2	3	4	5	6	7	8	9	10
Panel A: S&P 500										
RW	1.0231	1.0733	1.1629	1.1940	1.2526	1.2745	1.2986	1.2782	1.2742	1.2956
AR(1)	1.0103	1.0331	1.0724	1.0362	1.0663	1.0521	1.0472	1.0000	1.0274	1.0803
HAR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ARFIMA	0.9000	0.9125	0.9367	0.9232	0.9317	0.9279	0.9312	0.9098	0.9068	0.9197
fBm	0.8974	0.9078	0.9299	0.9147	0.9213	0.9158	0.9175	0.8947	0.8885	0.9015
fO–U	**0.8897**	**0.9054**	**0.9276**	**0.9126**	**0.9151**	**0.9118**	**0.9096**	**0.8891**	**0.8848**	**0.8960**
Panel B: DJIA										
RW	1.0379	1.1272	1.2902	1.3729	1.4416	1.4356	1.4978	1.4470	1.4283	1.4177
AR(1)	1.0352	1.0050	1.1000	1.0855	1.1005	1.0378	1.1143	1.0064	1.0594	1.1024
HAR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ARFIMA	0.9106	0.9077	0.9585	0.9549	0.9673	0.9422	0.9753	0.9364	0.9385	0.9458
fBm	0.9024	0.9002	0.9488	0.9454	0.9579	0.9333	0.9619	0.9280	0.9283	0.9317
fO–U	**0.8970**	**0.8953**	**0.9439**	**0.9430**	**0.9533**	**0.9289**	**0.9596**	**0.9237**	**0.9242**	**0.9297**
Panel C: NASDAQ 100										
RW	1.0360	1.0843	1.1704	1.2293	1.2878	1.3250	1.3912	1.3625	1.3511	1.3717
AR(1)	1.0000	1.0752	1.1256	1.1157	1.1237	1.1083	1.1381	1.0438	1.0725	1.1151
HAR	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
ARFIMA	0.9041	0.9339	0.9574	0.9541	0.9616	0.9563	0.9833	0.9462	0.9485	0.9679
fBm	0.8993	0.9248	0.9507	0.9476	0.9531	0.9479	0.9749	0.9382	0.9313	0.9434
fO–U	**0.8897**	**0.9157**	**0.9350**	**0.9323**	**0.9382**	**0.9333**	**0.9582**	**0.9223**	**0.9218**	**0.9358**

Table 7
MZ-R^2 for h-day-ahead-forecast of RV in different models.

h	1	2	3	4	5	6	7	8	9	10
Panel A: S&P 500										
RW	0.3569	0.2732	0.2020	0.1382	0.1095	0.0795	0.0595	0.0332	0.0184	0.0137
AR(1)	0.3935	0.2990	0.2038	0.1570	0.1125	0.0822	0.0603	0.0411	0.0255	0.0153
HAR	0.3941	0.3002	0.2077	0.1593	0.1145	0.0829	0.0633	0.0421	0.0255	0.0162
ARFIMA	0.4444	0.3310	0.2218	0.1634	0.1152	0.0898	0.0687	0.0495	0.0312	0.0184
fBm	0.4447	0.3311	0.2227	0.1685	0.1157	0.0926	0.0805	0.0816	0.0414	0.0198
fO–U	**0.4520**	**0.3353**	**0.2274**	**0.1709**	**0.1237**	**0.0965**	**0.0851**	**0.0869**	**0.0477**	**0.0208**
Panel B: DJIA										
RW	0.2908	0.1991	0.1674	0.1493	0.1292	0.0846	0.0824	0.0465	0.0251	0.0140
AR(1)	0.3307	0.2805	0.1708	0.1565	0.1325	0.1105	0.0848	0.0674	0.0369	0.0161
HAR	0.3317	0.2817	0.1756	0.1617	0.1353	0.1115	0.0878	0.0679	0.0380	0.0168
ARFIMA	0.3722	0.2837	0.1916	0.1662	0.1362	0.1184	0.0926	0.0742	0.0419	0.0185
fBm	0.3770	0.2838	0.1951	0.1689	0.1377	0.1380	0.0945	0.1206	0.0514	0.0210
fO–U	**0.3847**	**0.2908**	**0.2019**	**0.1760**	**0.1430**	**0.1471**	**0.0954**	**0.1344**	**0.0683**	**0.0325**
Panel C: NASDAQ 100										
RW	0.2218	0.1660	0.1509	0.1223	0.0950	0.0848	0.0697	0.0469	0.0196	0.0123
AR(1)	0.2828	0.1888	0.1383	0.1246	0.0979	0.0894	0.0717	0.0599	0.0272	0.0141
HAR	0.2834	0.1898	0.1399	0.1255	0.1048	0.0928	0.0759	0.0746	0.0338	0.0147
ARFIMA	0.3236	0.2221	0.1720	0.1443	0.1110	0.0929	0.0820	0.0856	0.0423	0.0165
fBm	0.3268	0.2281	0.1794	0.1549	0.1224	0.1055	0.0871	0.1030	0.0460	0.0215
fO–U	**0.3363**	**0.2405**	**0.1945**	**0.1703**	**0.1357**	**0.1183**	**0.0936**	**0.1085**	**0.0517**	**0.0217**

Tables 6–7 report the RMSE and the MZ-R^2 of six competing models for h-day-ahead-forecast of RV with $h = 1, 2, \ldots, 10$ with the best result highlighted in boldface for each h. For RMSE, we report the ratio of RMSE of each candidate model and that of HAR. Most importantly, regardless of the RV series and forecasting horizon, fO–U always performs the best, followed by fBm and then by ARFIMA. ARFIMA model always outperforms the RW, AR(1), HAR models in both criteria and by a wide margin. This result confirms the finding of Andersen et al. (2003) that the ARFIMA model can generate accurate volatility forecasts.

To test if forecasts from fO–U are statistically significantly different from those of competing models, Table 8 reports the DM test based on the squared forecast errors and the p-value (in parenthesis) with the benchmark being fO–U (boldface means statistically significant at the 10% level). To save space, we only report the results for HAR, ARFIMA and fBm in Table 8. In terms of DM, fO–U is always statistically different from HAR, ARFIMA and fBm at the 10% level, regardless of the RV series and forecasting horizon.

To find out which models contain the best model in the set of competing models, we implement the MCS of Hansen et al. (2011). The MCS determines which models can be considered to be statistically superior, and at what level of

Table 8
DM statistic with fO–U being the benchmark model.

h	1	2	3	4	5	6	7	8	9	10
Panel A: S&P 500										
HAR	−3.7063	−3.3613	−3.0482	−3.3732	−5.7227	−3.5121	−5.3147	−4.2277	−5.2174	−4.3701
	(0.0001)	**(0.0003)**	**(0.0011)**	**(0.0003)**	**(0.0000)**	**(0.0002)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
ARFIMA	−2.8461	−2.9371	−2.3965	−2.5667	−3.2590	−3.1858	−3.1162	−4.0784	−4.2191	−4.3414
	(0.0022)	**(0.0016)**	**(0.0082)**	**(0.0051)**	**(0.0005)**	**(0.0007)**	**(0.0009)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
fBm	−2.5026	−1.7870	−1.9006	−1.5298	−3.1347	−2.7829	−2.6281	−2.3105	−3.0311	−4.0768
	(0.0061)	**(0.0369)**	**(0.0286)**	**(0.0630)**	**(0.0008)**	**(0.0026)**	**(0.0042)**	**(0.0104)**	**(0.0012)**	**(0.0000)**
Panel B: DJIA										
HAR	−3.5431	−4.2664	−3.5632	−3.1413	−3.5105	−3.5138	−3.1507	−3.6108	−3.3440	−3.5700
	(0.0001)	**(0.0000)**	**(0.0001)**	**(0.0008)**	**(0.0002)**	**(0.0002)**	**(0.0008)**	**(0.0001)**	**(0.0004)**	**(0.0001)**
ARFIMA	−3.2895	−3.6539	−2.9467	−2.7869	−2.5570	−3.2736	−2.7536	−2.9251	−3.1646	−3.4715
	(0.0005)	**(0.0001)**	**(0.0016)**	**(0.0026)**	**(0.0052)**	**(0.0005)**	**(0.0029)**	**(0.0017)**	**(0.0007)**	**(0.0002)**
fBm	−1.8430	−2.3946	−1.8642	−2.3838	−1.9719	−2.1959	−1.7686	−1.5442	−2.6065	−2.6816
	(0.0326)	**(0.0083)**	**(0.0311)**	**(0.0085)**	**(0.0243)**	**(0.0140)**	**(0.0384)**	**(0.0612)**	**(0.0045)**	**(0.0036)**
Panel C: NASDAQ 100										
HAR	−3.5700	−4.5170	−4.3910	−3.3634	−3.4575	−3.6766	−3.3347	−3.1904	−3.1106	−4.0507
	(0.0001)	**(0.0000)**	**(0.0000)**	**(0.0003)**	**(0.0002)**	**(0.0001)**	**(0.0004)**	**(0.0007)**	**(0.0009)**	**(0.0000)**
ARFIMA	−3.4715	−3.3687	−3.7547	−3.1701	−3.4524	−3.1080	−3.2860	−2.7206	−2.9136	−3.4443
	(0.0002)	**(0.0003)**	**(0.0000)**	**(0.0007)**	**(0.0002)**	**(0.0009)**	**(0.0005)**	**(0.0032)**	**(0.0017)**	**(0.0002)**
fBm	−2.6816	−3.3310	−1.9921	−2.4964	−2.4277	−2.9629	−1.5113	−2.3930	−2.6990	−1.7309
	(0.0036)	**(0.0004)**	**(0.0231)**	**(0.0062)**	**(0.0075)**	**(0.0015)**	**(0.0653)**	**(0.0083)**	**(0.0034)**	**(0.0417)**

Table 9
p-values of MSC to compare forecasts of RV from six competing models.

h	1	2	3	4	5	6	7	8	9	10
Panel A: S&P 500										
RW	0.0005	0.0030	0.0050	0.0060	0.0015	0.0025	0.0005	0.0015	0.0020	0.0010
AR(1)	0.0015	0.0075	0.0085	0.0195	0.0025	0.0030	0.0015	0.0020	0.0025	0.0015
HAR	0.0065	0.0120	0.0185	0.0245	0.0030	0.0180	0.0035	0.0075	0.0030	0.0022
ARFIMA	0.0085	**0.1105**	0.0320	0.0470	**0.1060**	0.0654	0.0895	0.0325	0.0212	0.0050
fBm	**0.1030**	**0.1835**	**0.2325**	**0.3450**	**0.1550**	**0.3245**	**0.1015**	**0.1320**	**0.1022**	**0.1125**
fO–U	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Panel B: DJIA										
RW	0.0000	0.0000	0.0000	0.0020	0.0010	0.0000	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0000	0.0000	0.0005	0.0005	0.0025	0.0000	0.0000	0.0005	0.0001	0.0003
HAR	0.0000	0.0010	0.0025	0.0014	0.0095	0.0015	0.0005	0.0005	0.0025	0.0085
ARFIMA	0.0765	0.0060	0.0030	0.0217	0.0239	0.0130	0.0270	0.0235	0.0225	0.0225
fBm	**0.1205**	**0.1175**	0.0523	**0.1535**	**0.1870**	**0.1105**	**0.2105**	**0.1980**	**0.1245**	**0.1185**
fO–U	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Panel C: NASDAQ 100										
RW	0.0020	0.0000	0.0000	0.0020	0.0010	0.0005	0.0000	0.0000	0.0000	0.0000
AR(1)	0.0035	0.0005	0.0005	0.0010	0.0060	0.0025	0.0010	0.0005	0.0025	0.0005
HAR	0.0060	0.0025	0.0010	0.0120	0.0085	0.0045	0.0065	0.0240	0.0030	0.0025
ARFIMA	0.0070	0.0035	0.0090	0.0125	0.1458	0.0070	0.0295	0.0045	0.0195	0.0275
fBm	**0.1470**	**0.1245**	**0.1465**	**0.2170**	**0.2260**	**0.1730**	**0.1530**	**0.1515**	**0.1495**	**0.1585**
fO–U	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

significance. Table 9 reports the p-value of the semi-quadratic statistic obtained from 2000 bootstrap iterations with a block length of 12. Values in boldface denote that the model belongs to the confidence set of the best models. HAR, AR(1), and RW are always rejected regardless of the RV series and forecasting horizon. In all but two cases, ARFIMA is also rejected. In one case, fBm is rejected. Most importantly, in no case, fO–U is rejected.

There are nested models in the set of competing models. For example, fO–U nests fBm, AR(1) and RW. To compare the forecasting performance across nested models, we calculate MSFE-adjusted statistics and p-values of the CW test of Clark and West (2007) that account for estimation errors. Table 10 reports the CW test and the p-value (in parenthesis) with the encompassing model being fO–U (boldface means statistically significant at the 10% level). According to CW, the performance of fO–U is always statistically significantly better than RW, AR(1), and fBm, regardless of the RV series and forecasting horizon.

Table 10
CW test with fO–U being the encompassing model.

h	1	2	3	4	5	6	7	8	9	10
Panel A: S&P 500										
RW	5.8431	8.2638	11.1017	12.3055	13.3093	13.7838	14.3113	14.5086	14.6219	14.7257
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
AR(1)	4.9955	5.1822	4.9254	4.8855	6.0645	5.1925	5.7541	5.0158	6.2011	6.7644
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
fBm	2.6120	1.5814	1.6575	1.3153	4.6266	2.6107	4.5480	4.1035	5.1021	4.3290
	(0.0045)	**(0.0569)**	**(0.0487)**	**(0.0942)**	**(0.0000)**	**(0.0045)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
Panel B: DJIA										
RW	6.5436	10.7224	13.3905	14.6893	15.5466	15.9252	16.3186	16.4104	16.1193	15.8740
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
AR(1)	5.0058	8.7172	7.3277	7.6079	7.5306	6.8555	8.3416	8.1724	9.3732	9.1739
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
fBm	1.9344	4.8768	4.8277	3.0862	4.5777	3.2953	3.0694	2.9148	3.0860	2.6142
	(0.0265)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0010)**	**(0.0005)**	**(0.0011)**	**(0.0018)**	**(0.0010)**	**(0.0045)**
Panel C: NASDAQ 100										
RW	5.1952	7.5749	10.1456	11.8752	13.2111	14.0357	14.7021	15.0329	14.8542	14.7266
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
AR(1)	4.4821	5.5555	5.1159	4.5121	4.6783	5.0579	4.6578	4.9708	4.8524	5.5991
	(0.0000)	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**
fBm	3.2709	3.8405	4.1662	3.6594	3.8851	3.5977	3.7466	3.0263	3.1796	2.2354
	(0.0005)	**(0.0001)**	**(0.0000)**	**(0.0001)**	**(0.0001)**	**(0.0002)**	**(0.0001)**	**(0.0012)**	**(0.0007)**	**(0.0127)**

7. Conclusions

Over the past two decades, the consensus is that the volatility of financial assets displays long-range dependence. More recently, a new feature named rough volatility has been documented. In our continuous-time fO–U model, the ACF of RV is jointly determined by H and κ at small and moderate lags but solely determined by H when the lag goes to infinity. Moreover, the smoothness in RV is determined by H. When $H < 1/2$, the sample path of fO–U is rough. Although the fO–U model with $H < 1/2$ and κ taking a small positive value does not lead long-range dependence asymptotically, it can generate an ACF that decays slowly at small and moderate lags due to the near unit root feature in the discretized model.

Our study contributes to the literature by developing an estimation method for all parameters in the fO–U model based on discrete-sampled observations when the parameter space for H is $(0, 1)$. In the first stage, H is estimated based on the ratio of two second-order differences of observations at two frequencies and hence, determined by the smoothness of the sample path. In the second stage, the other parameters are estimated by the method of moments. All estimators have closed-form expressions and are easy to obtain. We also develop the asymptotic distributions for the proposed estimators that facilitate statistical inference.

Simulations suggest that our two-stage estimators perform well in finite samples. The method is applied to the logarithmic RV of three equity indices. Empirical studies show that the logarithmic RV is better modeled by fO–U with $H < 1/2$ and a small positive mean-reversion parameter κ. We compare the out-of-sample performance of fO–U with those of fBm, ARFIMA, RW, AR(1) and HAR in forecasting RV. All statistics considered suggest that fO–U outperforms the other models for the forecasting horizons considered, and hence, raise the benchmark for volatility forecast. In the online appendix, the method is applied to the logarithmic BP and RK of the three equity indices. All the empirical results continue to hold.

There are several ways to extend our study. First, the estimation method proposed in the present paper is by no means efficient. How to conduct the full-likelihood-based inference for the fO–U model from discrete-time observations $\{X_{t\Delta}\}_{t=1}^{n}$ and how well the model performs in forecasting RV are among interesting questions. We plan to pursue this line of research in future work.

Second, our model is not the only one in the literature that disentangles the short-term from the long-term behavior. Gneiting and Schlather (2004) propose a general Cauchy process that allows for both types of behavior. Bennedsen et al. (2021) show how the Cauchy process can decouple the short-term and long-term behavior of volatility. Unlike fO–U, the Cauchy processes are not self-similar nor BSS. Hence, conceptually they are more complicated than the fO–U model. Another example is the non-Gaussian O–U model with infinite activity jumps. It would be interesting to examine the forecasting performance of fO–U relative to these alternative specifications.

Third, perhaps a better model for logarithmic RV is fO–U with jumps. It is important to develop an estimation method of H and other parameters and establish asymptotic theory for the fO–U model with jumps.

Appendix A

Lemma A.1. *Let $B_t^H = B^H(t)$ be an fBm with the Hurst parameter $H \in (0,1)$ and $t \in [0,\infty)$.*

(a) Define $y_i = B^H(i+2) - 2B^H(i+1) + B^H(i)$ for $i = 0,1,2,\ldots$. The process $\{y_i\}$ is a Gaussian stationary process with $\mathbb{E}(y_i) = 0$ and $Var(y_i) = 4 - 2^{2H}$, and has autocorrelation functions as, for $j = 0,1,2,\ldots$,

$$\rho_j = \frac{1}{2\left(4-2^{2H}\right)}\left\{-|j+2|^{2H} + 4|j+1|^{2H} - 6|j|^{2H} + 4|j-1|^{2H} - |j-2|^{2H}\right\};$$

(b) Define $y_{i,*} = B^H(i+4) - 2B^H(i+2) + B^H(i)$ for $i = 0,1,2,\ldots$. The process $\{y_{i,*}\}$ is a Gaussian stationary process with $\mathbb{E}(y_{i,*}) = 0$ and $Var(y_{i,*}) = 2^{2H}\left(4-2^{2H}\right)$, and has autocorrelation functions as, for $j = 0,1,2,\ldots$,

$$\rho_{j,*} = 2^{-2H}\left(\rho_{j+2} + 4\rho_{j+1} + 6\rho_j + 4\rho_{|j-1|} + \rho_{|j-2|}\right);$$

(c) Define $\xi_{i,*} = y_{i,*}^2 - \mathbb{E}(y_{i,*}^2)$ and $\xi_i = y_i^2 - \mathbb{E}(y_i^2)$, for $i = 0,1,2,\ldots$. The bivariate process $\left\{(\xi_{i,*} \quad \xi_i)'\right\}$ is a weakly stationary process with mean zero and autocovariance matrices as, for $j = 0,1,2,\ldots$,

$$\Gamma_j = \mathbb{E}\left(\begin{pmatrix}\xi_{i+j,*}\\ \xi_{i+j}\end{pmatrix}(\xi_{i,*} \quad \xi_i)\right) = 2\left(4-2^{2H}\right)^2\begin{pmatrix}2^{4H}\rho_{j,*}^2 & \left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2\\ \left(\rho_j + 2\rho_{|j-1|} + \rho_{|j-2|}\right)^2 & \rho_j^2\end{pmatrix}.$$

Lemma A.2. *Let $B^H(t)$ be the same fBm as in Lemma A.1 with $t \in [0,T]$, where T is the time span. Suppose $B^H(t)$ are observed at discrete-time points with sampling interval Δ, denoted by $\left\{B_{i\Delta}^H = B^H(i\Delta)\right\}_{i=0}^n$, where $n = \lfloor T/\Delta \rfloor$ is the number of observations. Define*

$$\eta_{i,*} = \left(\frac{B^H_{(i+4)\Delta} - 2B^H_{(i+2)\Delta} + B^H_{i\Delta}}{\Delta^H}\right)^2 - 2^{2H}\left(4-2^{2H}\right) \quad \text{for } i = 0,1,2,\ldots,n-4,$$

$$\eta_i = \left(\frac{B^H_{(i+2)\Delta} - 2B^H_{(i+1)\Delta} + B^H_{i\Delta}}{\Delta^H}\right)^2 - \left(4-2^{2H}\right) \quad \text{for } i = 0,1,2,\ldots,n-2.$$

It has, as $n \to \infty$,

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=0}^{n-4}\eta_{i,*}\\ \sum_{i=0}^{n-2}\eta_i\end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j'\right)\right),$$

where Γ_j are the covariance matrices defined in Lemma A.1, and the long-run covariance matrix in the limiting distribution is well-defined.

Proof of Lemma A.1. The results in Parts (a)–(b) can be obtained straightforwardly based on the definition of fBm and its covariance structure given in (1.1). Details are tedious and omitted here for simplicity.

For Part (c), let us first prove that $\left\{\xi_i = y_i^2 - \mathbb{E}(y_i^2)\right\}_{i=0}^{\infty}$ is a stationary process. From the stationarity of $\{y_i\}$, it can be obtained that $\mathbb{E}(\xi_i) = 0$. Then, we have, for $j = 0,1,2,\ldots$,

$$\begin{aligned} Cov(\xi_{i+j},\xi_i) &= \mathbb{E}(\xi_{i+j}\xi_i) = \mathbb{E}(y_{i+j}^2 y_i^2) - \mathbb{E}(y_{i+j}^2)\mathbb{E}(y_i^2)\\ &= Var(y_{i+j})\,Var(y_i) + 2\left[Cov(y_{i+j},y_i)\right]^2 - \mathbb{E}(y_{i+j}^2)\mathbb{E}(y_i^2)\\ &= 2\left(4-2^{2H}\right)^2\rho_j^2 \end{aligned}$$

where the third equality comes from Isserlis' theorem (Isserlis, 1918) for computing high-order moments of the multivariate normal distribution, and the last equation is from the stationarity properties of $\{y_i\}$ given in Part (a).

Taking the same procedure above with the stationarity properties of $\{y_{i,*}\}$ shown in Part (b) gives a proof of $\{\xi_{i,*}\}$ being a stationary process with mean zero and $\mathbb{E}(\xi_{i+j,*}\xi_{i,*}) = 2\left(4-2^{2H}\right)^2 2^{4H}\rho_{j,*}^2$.

We now derive the expressions of $\mathbb{E}(\xi_{i+j}\xi_{i,*})$ and $\mathbb{E}(\xi_{i+j,*}\xi_i)$ and show that they only depend on j, not i. For any

$i = 0, 1, 2, \ldots$, it can be seen that

$$\begin{aligned} y_{i,*} &= \left[B^H(i+4) - 2B^H(i+3) + B^H(i+2)\right] \\ &\quad + 2\left[B^H(i+3) - 2B^H(i+2) + B^H(i+1)\right] + \left[B^H(i+2) - 2B^H(i+1) + B^H(i)\right] \\ &= y_{i+2} + 2y_{i+1} + y_i. \end{aligned}$$

Hence, for any $j = 0, 1, 2, \ldots$,

$$Cov\left(y_{i+j}, y_{i,*}\right) = Cov\left(y_{i+j}, y_{i+2} + 2y_{i+1} + y_i\right) = \left(4 - 2^{2H}\right)\left(\rho_{|j-2|} + 2\rho_{|j-1|} + \rho_j\right).$$

Then, by using Isserlis' theorem (Isserlis, 1918) again, we have

$$\begin{aligned} \mathbb{E}\left(\xi_{i+j}\xi_{i,*}\right) &= \mathbb{E}\left(y_{i+j}^2 y_{i,*}^2\right) - \mathbb{E}\left(y_{i+j}^2\right)\mathbb{E}\left(y_{i,*}^2\right) \\ &= Var\left(y_{i+j}\right) Var\left(y_{i,*}\right) + 2\left[Cov\left(y_{i+j}, y_{i,*}\right)\right]^2 - \mathbb{E}\left(y_{i+j}^2\right)\mathbb{E}\left(y_{i,*}^2\right) \\ &= 2\left(4 - 2^{2H}\right)^2\left(\rho_{|j-2|} + 2\rho_{|j-1|} + \rho_j\right)^2. \end{aligned}$$

Similarly, it can be proved that $\mathbb{E}\left(\xi_{i+j,*}\xi_i\right) = 2\left(4 - 2^{2H}\right)^2\left(\rho_{j+2} + 2\rho_{j+1} + \rho_j\right)^2$. Then, the covariance matrices $\left\{\Gamma_j\right\}$ are obtained.

In Remark 4.2, we have proved that $\rho_j = O\left(j^{2H-4}\right)$ as $j \to \infty$. Therefore, the sequence of covariance matrices $\left\{\Gamma_j\right\}$ is absolutely summable, and the long-run covariance matrix $\Gamma_0 + \sum_{j=1}^{\infty}\left(\Gamma_j + \Gamma_j'\right)$ is well-defined. Hence, the bivariate process $\left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}$ is weakly stationary.

Proof of Lemma A.2. For the case where $T = 1$ and $m = \lfloor 1/\Delta \rfloor$, the asymptotic normality of $\frac{1}{\sqrt{m}}\left(\sum_{i=0}^{m-4}\eta_{i,*} \quad \sum_{i=0}^{m-2}\eta_i\right)'$ as $m \to \infty$ can be obtained straightforwardly from the results in Section 2.4 of Bégyn (2007) and also from Coeurjolly (2001). Note that, from the self-similarity property of fBm, for any $i = 0, 1, \ldots$, we have

$$\frac{B^H_{(i+4)\Delta} - 2B^H_{(i+2)\Delta} + B^H_{i\Delta}}{\Delta^H} \overset{d}{=} B^H(i+4) - 2B^H(i+2) + B^H(i) := y_{i,*}, \tag{A.1}$$

and

$$\frac{B^H_{(i+2)\Delta} - 2B^H_{(i+1)\Delta} + B^H_{i\Delta}}{\Delta^H} \overset{d}{=} B^H(i+2) - 2B^H(i+1) + B^H(i) := y_i. \tag{A.2}$$

Hence, for any value of m,

$$\frac{1}{\sqrt{m}}\begin{pmatrix}\sum_{i=0}^{m-4}\eta_{i,*} \\ \sum_{i=0}^{m-2}\eta_i\end{pmatrix} \overset{d}{=} \frac{1}{\sqrt{m}}\begin{pmatrix}\sum_{i=0}^{m-4}\left[y_{i,*}^2 - 2^{2H}\left(4 - 2^{2H}\right)\right] \\ \sum_{i=0}^{m-2}\left[y_i^2 - \left(4 - 2^{2H}\right)\right]\end{pmatrix} = \frac{1}{\sqrt{m}}\begin{pmatrix}\sum_{i=0}^{m-4}\xi_{i,*} \\ \sum_{i=0}^{m-2}\xi_i\end{pmatrix}$$

where $\xi_{i,*}$ and ξ_i are the same variables defined in Lemma A.1. Therefore, we get the asymptotic normality of $\frac{1}{\sqrt{m}}\left(\sum_{i=0}^{m-4}\xi_{i,*} \quad \sum_{i=0}^{m-2}\xi_i\right)'$ as $m \to \infty$. It is important to note that the variables $\xi_{i,*}$ and ξ_i are independent of the sampling frequency Δ.

For general values of T, we define $n = \lfloor T/\Delta \rfloor$. By the self-similarity property of fBm, we get

$$\frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=0}^{n-4}\eta_{i,*} \\ \sum_{i=0}^{n-2}\eta_i\end{pmatrix} \overset{d}{=} \frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=0}^{n-4}\left[y_{i,*}^2 - 2^{2H}\left(4 - 2^{2H}\right)\right] \\ \sum_{i=0}^{n-2}\left[y_i^2 - \left(4 - 2^{2H}\right)\right]\end{pmatrix} = \frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=0}^{n-4}\xi_{i,*} \\ \sum_{i=0}^{n-2}\xi_i\end{pmatrix}.$$

Therefore, as long as $n \to \infty$, we get the asymptotic normality of $\frac{1}{\sqrt{n}}\left(\sum_{i=0}^{n-4}\xi_{i,*} \quad \sum_{i=0}^{n-2}\xi_i\right)'$ and the asymptotic normality of $\frac{1}{\sqrt{n}}\left(\sum_{i=0}^{n-4}\eta_{i,*} \quad \sum_{i=0}^{n-2}\eta_i\right)'$. Based on the fact that $\left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}$ is a weakly stationary vector process, the corresponding asymptotic covariance is just the long-run variance of the process $\left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}$, as given in the theorem.

Proof of Theorem 4.1. (a) From Eq. (1.2), we have, as $\Delta \to 0$,

$$\begin{aligned} X_{(i+1)\Delta} - X_{i\Delta} &= \left(e^{-\kappa\Delta} - 1\right)\left(X_{i\Delta} - \mu\right) + \sigma\int_{i\Delta}^{(i+1)\Delta} e^{-\kappa[(i+1)\Delta - s]}dB_s^H \\ &= O_p(\Delta) + \sigma\int_{i\Delta}^{(i+1)\Delta}\{1 + O(\Delta)\}\,dB_s^H \\ &= \sigma\left(B^H_{(i+1)\Delta} - B^H_{i\Delta}\right) + O_p(\Delta) = O_p\left(\Delta^H\right), \end{aligned}$$

and

$$\begin{aligned}
&X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta} \\
&= \left(X_{(i+2)\Delta} - X_{(i+1)\Delta}\right) - \left(X_{(i+1)\Delta} - X_{i\Delta}\right) \\
&= \left(e^{-\kappa\Delta} - 1\right)\left(X_{(i+1)\Delta} - X_{i\Delta}\right) + \sigma\left(\int_{(i+1)\Delta}^{(i+2)\Delta} e^{-\kappa[(i+2)\Delta - s]} dB_s^H - \int_{i\Delta}^{(i+1)\Delta} e^{-\kappa[(i+1)\Delta - s]} dB_s^H\right) \\
&= O_p\left(\Delta^{1+H}\right) + \sigma\left(\int_{(i+1)\Delta}^{(i+2)\Delta} \{1 + O(\Delta)\} dB_s^H - \int_{i\Delta}^{(i+1)\Delta} \{1 + O(\Delta)\} dB_s^H\right) \\
&= \sigma\left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right) + O_p\left(\Delta^{1+H}\right).
\end{aligned}$$

Therefore, by using the results in Lemma A.2, we have, as $\Delta \to 0$,

$$\begin{aligned}
&\frac{\sigma^{-2}}{n\Delta^{2H}} \sum_{i=1}^{n-2} \left(X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}\right)^2 \\
&= \frac{1}{n\Delta^{2H}} \sum_{i=1}^{n-2} \left\{\left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right)^2 + O_p\left(\Delta^{1+2H}\right)\right\} \\
&= \frac{1}{n} \sum_{i=1}^{n-2} \left(\frac{B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H}{\Delta^H}\right)^2 + \frac{1}{n} \sum_{i=1}^{n-2} O_p(\Delta) \\
&= \frac{1}{n} \sum_{i=1}^{n-2} \left[\eta_i + \left(4 - 2^{2H}\right)\right] + O_p(\Delta) \xrightarrow{p} 4 - 2^{2H},
\end{aligned} \tag{A.3}$$

and, as $\Delta \to 0$ and $T\Delta \to 0$,

$$\begin{aligned}
&\frac{\sigma^{-2}}{\sqrt{n}\Delta^{2H}} \sum_{i=1}^{n-2} \left\{\left(X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}\right)^2 - \sigma^2\left(4 - 2^{2H}\right)\Delta^{2H}\right\} \\
&= \frac{1}{\sqrt{n}\Delta^{2H}} \sum_{i=1}^{n-2} \left\{\left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right)^2 - \left(4 - 2^{2H}\right)\Delta^{2H} + O_p\left(\Delta^{1+2H}\right)\right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-2} \left[\left(\frac{B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H}{\Delta^H}\right)^2 - \left(4 - 2^{2H}\right)\right] + \frac{1}{n} \sum_{i=1}^{n-2} O_p\left(\sqrt{T\Delta}\right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-2} \eta_i + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, 2\left(4 - 2^{2H}\right)^2 \left(\rho_0^2 + 2\sum_{j=1}^{\infty} \rho_j^2\right)\right),
\end{aligned} \tag{A.4}$$

where the asymptotic variance can be equivalently represented as $\left(4 - 2^{2H}\right)^2 \Sigma_{22}$ with Σ_{22} defined as in (4.5).

Similarly, using the results in Lemma A.2 again, we have, as $\Delta \to 0$,

$$\begin{aligned}
&\frac{\sigma^{-2}}{n\Delta^{2H}} \sum_{i=1}^{n-4} \left(X_{(i+4)\Delta} - 2X_{(i+2)\Delta} + X_{i\Delta}\right)^2 \\
&= \frac{1}{n} \sum_{i=1}^{n-4} \left[\eta_{i,*} + 2^{2H}\left(4 - 2^{2H}\right)\right] + O_p(\Delta) \xrightarrow{p} 2^{2H}\left(4 - 2^{2H}\right),
\end{aligned} \tag{A.5}$$

and, as $\Delta \to 0$ and $T\Delta \to 0$,

$$\begin{aligned}
&\frac{\sigma^{-2}}{\sqrt{n}\Delta^{2H}} \sum_{i=1}^{n-4} \left\{\left(X_{(i+4)\Delta} - 2X_{(i+2)\Delta} + X_{i\Delta}\right)^2 - \sigma^2 2^{2H}\left(4 - 2^{2H}\right)\Delta^{2H}\right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-4} \eta_{i,*} + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, 2^{1+4H}\left(4 - 2^{2H}\right)^2 \left(\rho_{0,*}^2 + 2\sum_{j=1}^{\infty} \rho_{j,*}^2\right)\right).
\end{aligned} \tag{A.6}$$

The asymptotic variance has an identical representation as

$$\begin{aligned}
&2^{1+4H}\left(4-2^{2H}\right)^2\left(\rho_{0,*}^2+2\sum_{j=1}^{\infty}\rho_{j,*}^2\right)\\
&=2^{1+4H}\left(4-2^{2H}\right)^2\left(1+2^{1-4H}\sum_{j=1}^{\infty}\left[\left(\rho_{j+2}+4\rho_{j+1}+6\rho_j+4\rho_{|j-1|}+\rho_{|j-2|}\right)^2\right]\right)\\
&=2^{4H}\left(4-2^{2H}\right)^2\Sigma_{11},
\end{aligned}$$

where the first equation comes from the relationship between $\rho_{j,*}$ and ρ_j given in Lemma A.1, and Σ_{11} is defined in (4.3).

Then, based on (A.3) and (A.5), the consistency of $2^{2\widehat{H}}$ is achieved as long as $\Delta\to 0$:

$$2^{2\widehat{H}}=\frac{\frac{\sigma^{-2}}{n\Delta^{2H}}\sum_{i=1}^{n-4}\left(X_{(i+4)\Delta}-2X_{(i+2)\Delta}+X_{i\Delta}\right)^2}{\frac{\sigma^{-2}}{n\Delta^{2H}}\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}\xrightarrow{p}\frac{2^{2H}\left(4-2^{2H}\right)}{4-2^{2H}}=2^{2H}.$$

With the continuity of $\log_2(\cdot)$, the consistency of $\widehat{H}=\frac{1}{2}\log_2\left(2^{2\widehat{H}}\right)$ is obtained straightforwardly.

To derive the asymptotic distribution, we first note that, from Lemma A.2,

$$\begin{aligned}
&\lim_{n\to\infty}Cov\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n-4}\eta_{i,*},\frac{1}{\sqrt{n}}\sum_{i=1}^{n-2}\eta_i\right)\\
&=2\left(4-2^{2H}\right)^2\left((\rho_2+2\rho_1+\rho_0)^2+\sum_{j=1}^{\infty}\left[\left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2+\left(\rho_j+2\rho_{|j-1|}+\rho_{|j-2|}\right)^2\right]\right)\\
&=2\left(4-2^{2H}\right)^2\left(4(\rho_0+\rho_1)^2+2\sum_{j=0}^{\infty}\left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2\right)=2^{2H}\left(4-2^{2H}\right)^2\Sigma_{12},
\end{aligned}$$

which leads to the asymptotic result that, as $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n-4}\eta_{i,*}-2^{2H}\frac{1}{\sqrt{n}}\sum_{i=1}^{n-2}\eta_i\xrightarrow{d}\mathcal{N}\left(0,2^{4H}\left(4-2^{2H}\right)^2\left[\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}\right]\right),$$

where Σ_{12} is defined as in (4.4). Then, together with the results given in (A.3), (A.4) and (A.6), the asymptotic distribution of $2^{2\widehat{H}}-2^{2H}$ is obtained as $\Delta\to 0$ and $T\Delta\to 0$:

$$\begin{aligned}
&\sqrt{n}\left(2^{2\widehat{H}}-2^{2H}\right)\\
&=\frac{\frac{\sigma^{-2}}{\sqrt{n}\Delta^{2H}}\left\{\sum_{i=1}^{n-4}\left(X_{(i+4)\Delta}-2X_{(i+2)\Delta}+X_{i\Delta}\right)^2-2^{2H}\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2\right\}}{\frac{\sigma^{-2}}{n\Delta^{2H}}\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}\\
&=\frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n-4}\eta_{i,*}-2^{2H}\frac{1}{\sqrt{n}}\sum_{i=1}^{n-2}\eta_i-\frac{1}{\sqrt{n}}2^{1+2H}\left(4-2^{2H}\right)}{\frac{\sigma^{-2}}{n\Delta^{2H}}\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}\\
&\xrightarrow{d}\frac{\mathcal{N}\left(0,2^{4H}\left(4-2^{2H}\right)^2\left[\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}\right]\right)}{4-2^{2H}}\overset{d}{=}\mathcal{N}\left(0,2^{4H}\left[\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}\right]\right).
\end{aligned}$$

Note that $2^{2\widehat{H}}=2^{2H}+2\log(2)\cdot 2^{2\widetilde{H}}\left(\widehat{H}-H\right)$, where $\widetilde{H}$ lies between H and $\widehat{H}$. Therefore, as $\Delta\to 0$ and $T\Delta\to 0$,

$$\sqrt{n}\left(\widehat{H}-H\right)=\frac{\sqrt{n}\left(2^{2\widehat{H}}-2^{2H}\right)}{2^{2\widetilde{H}}2\log(2)}\xrightarrow{d}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{\{2\log(2)\}^2}\right).$$

The proof is completed.

(b) Based on the result that $\sqrt{n}\left(\widehat{H}-H\right)=O_p(1)$ as $\Delta\to 0$ and $T\Delta\to 0$, we have

$$\begin{aligned}\Delta^{2\widehat{H}-2H}&=\exp\left\{2\left(\widehat{H}-H\right)\log(\Delta)\right\}=\exp\left\{2\sqrt{n}\left(\widehat{H}-H\right)\frac{\log(\Delta)}{\sqrt{n}}\right\}\\&=\exp\left\{2\sqrt{n}\left(\widehat{H}-H\right)\frac{2\sqrt{\Delta}\log\left(\sqrt{\Delta}\right)}{\sqrt{T}}\right\}\xrightarrow{p}1,\end{aligned}$$

where the last limit is due to $\frac{\log(\Delta)}{\sqrt{n}}\to 0$ as $\Delta\to 0$ and $n\to\infty$. Together with the limiting result given in (A.3), the consistency of $\widehat{\sigma}^2$ is obtained under the condition of $\Delta\to 0$ and $T\Delta\to 0$:

$$\widehat{\sigma}^2=\frac{\frac{\sigma^{-2}}{n\Delta^{2H}}\sum_{i=1}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}{\sigma^{-2}\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}\xrightarrow{p}\frac{4-2^{2H}}{\sigma^{-2}\left(4-2^{2H}\right)}=\sigma^2.$$

To derive the asymptotic distribution of $\widehat{\sigma}^2$, we first prove that, as $\Delta\to 0$ and $T\Delta\to 0$,

$$\Delta^{2\widehat{H}-2H}-1=\exp\left\{2\sqrt{n}\left(\widehat{H}-H\right)\frac{\log(\Delta)}{\sqrt{n}}\right\}-1=2\sqrt{n}\left(\widehat{H}-H\right)\frac{\log(\Delta)}{\sqrt{n}}+o_p\left(\frac{\log(\Delta)}{\sqrt{n}}\right),$$

and

$$\frac{\sqrt{n}}{\log(\Delta)}\left(\Delta^{2\widehat{H}-2H}-1\right)=2\sqrt{n}\left(\widehat{H}-H\right)+o_p(1)\xrightarrow{d}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{\{\log(2)\}^2}\right).\tag{A.7}$$

Then, from the representation of $\widehat{\sigma}^2$ given in (3.2), we have

$$\begin{aligned}&\widehat{\sigma}^2-\sigma^2\\&=\frac{\frac{\sigma^{-2}}{\sqrt{n}\Delta^{2H}}\sum_{i=1}^{n-2}\left\{\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2-\sigma^2\left(4-2^{2H}\right)\Delta^{2H}\right\}}{\sqrt{n}\sigma^{-2}\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}+\frac{\sigma^2(n-2)\left(4-2^{2H}\right)}{n\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}-\sigma^2\\&=\frac{O_p(1)}{O_p\left(\sqrt{n}\right)}+\frac{(n-2)\left(4-2^{2H}\right)\sigma^2}{n\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}-\sigma^2\\&=\frac{\sigma^2}{\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}\left\{\frac{n-2}{n}\left(4-2^{2H}\right)-\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}\right\}+O_p\left(\frac{1}{\sqrt{n}}\right)\\&=\frac{\sigma^2}{\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}\left\{\left(2^{2\widehat{H}}-2^{2H}\right)-\left(4-2^{2\widehat{H}}\right)\left(\Delta^{2\widehat{H}-2H}-1\right)+O\left(\frac{1}{n}\right)\right\}+O_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

where the second equation is from the result in (A.3). Note that $2^{2\widehat{H}}-2^{2H}=O_p\left(1/\sqrt{n}\right)$ and $\Delta^{2\widehat{H}-2H}\xrightarrow{p}1$. Therefore, as $\Delta\to 0$ and $T\Delta\to 0$, we have

$$\begin{aligned}\frac{\sqrt{n}}{\log(\Delta)}\left(\widehat{\sigma}^2-\sigma^2\right)&=-\frac{\sqrt{n}}{\log(\Delta)}\frac{\left(\Delta^{2\widehat{H}-2H}-1\right)\sigma^2}{\Delta^{2\widehat{H}-2H}}+O_p\left(\frac{1}{\log(\Delta)}\right)\\&\xrightarrow{d}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{\{\log(2)\}^2}\sigma^4\right),\end{aligned}$$

where the last limit comes from the asymptotic result proved in (A.7).

Proof of Theorem 4.3. Starting from the definition of $\widehat{\mu}$ given in (3.3), we have, as $\Delta\to 0$,

$$\begin{aligned}\widehat{\mu}&=\frac{1}{n}\sum_{i=1}^{n}X_{i\Delta}=\frac{1}{T}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}X_{i\Delta}dt+\frac{X_T-X_0}{n}\\&=\frac{1}{T}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta}\left(X_t+O_p\left(\Delta^H\right)\right)dt+\frac{X_T-X_0}{n}=\frac{1}{T}\int_0^T X_t dt+O_p\left(\Delta^H\right)+O_p\left(\frac{1}{n}\right).\end{aligned}$$

Therefore, as $T\to\infty$ and $\Delta\to 0$,

$$\widehat{\mu}=\frac{1}{T}\int_0^T X_t dt+o_p(1)\xrightarrow{p}\mathbb{E}(X_t)=\mu,$$

where the last limit comes from the ergodicity of the process $\{X_t\}$ when $\kappa>0$ (see Xiao and Yu (2019a,b)).

To derive the limiting distribution, first notice that, according to Theorem 3.3 of Xiao and Yu (2019a) and Theorem 3.1 of Xiao and Yu (2019b), as $T \to \infty$,

$$T^{1-H}\left(\frac{1}{T}\int_0^T X_t dt - \mu\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{\sigma^2}{\kappa^2}\right),$$

for the cases where $H \in [1/2, 1)$ and $H \in (0, 1/2)$, respectively. Consequently, when $T \to \infty$, $\Delta \to 0$, and $T^{1-H}\Delta^H \to 0$, we have

$$\begin{aligned} T^{1-H}(\widehat{\mu} - \mu) &= T^{1-H}\left(\frac{1}{T}\int_0^T X_t dt - \mu\right) + O_p\left(T^{1-H}\Delta^H\right) + O_p\left(T^{1-H}\Delta^H \frac{\Delta^{1-H}}{T}\right) \\ &= T^{1-H}\left(\frac{1}{T}\int_0^T X_t dt - \mu\right) + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2/\kappa^2\right). \end{aligned}$$

Proof of Theorem 4.4. We first prove the consistency of $\widehat{\kappa}$ for all $H \in (0, 1)$ under the condition of $T \to \infty$ and $T\Delta \to 0$. From the definition of $\widehat{\kappa}$ given in (3.4), we have

$$\widehat{\kappa}^{-2\widehat{H}} = \frac{\frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2}{\widehat{\sigma}^2 \widehat{H}\Gamma\left(2\widehat{H}\right)}.$$

Note that, as $T \to \infty$ and $\Delta \to 0$,

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 &= \frac{1}{T}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} X_{i\Delta}^2 dt + \frac{X_T^2 - X_0^2}{n} \\ &= \frac{1}{T}\sum_{i=0}^{n-1}\int_{i\Delta}^{(i+1)\Delta} \left(X_t^2 + O_p\left(\Delta^H\right)\right) dt + \frac{X_T^2 - X_0^2}{n} \\ &= \frac{1}{T}\int_0^T X_t^2 dt + O_p\left(\Delta^H\right) + O_p(1/n) \\ &\xrightarrow{p} \mathbb{E}\left(X_t^2\right) = \sigma^2\kappa^{-2H}H\Gamma(2H) + \mu^2, \end{aligned}$$

where the limit has been proved in Xiao and Yu (2019a,b) for $H \in [1/2, 1)$ and $H \in (0, 1/2)$, respectively. With the limit of $\frac{1}{n}\sum_{i=1}^n X_{i\Delta}$ obtained in the proof of Theorem 4.3, it is obtained that

$$\frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2 \xrightarrow{p} \mathbb{E}\left(X_t^2\right) - \mu^2 = \sigma^2\kappa^{-2H}H\Gamma(2H).$$

The consistency of $\widehat{\sigma}^2$ and $\widehat{H}$ has been proved in Theorem 4.1 under the condition of $T\Delta \to 0$. As a result, we have, when $T \to \infty$ and $T\Delta \to 0$,

$$\widehat{\kappa}^{-2\widehat{H}} = \frac{\frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2}{\widehat{\sigma}^2 \widehat{H}\Gamma\left(2\widehat{H}\right)} \xrightarrow{p} \frac{\sigma^2\kappa^{-2H}H\Gamma(2H)}{\sigma^2 H\Gamma(2H)} = \kappa^{-2H},$$

and

$$\widehat{\kappa} = \exp\left\{-\frac{1}{2\widehat{H}}\log\left\{\widehat{\kappa}^{-2\widehat{H}}\right\}\right\} \xrightarrow{p} \exp\left\{-\frac{1}{2H}\log\left\{\kappa^{-2H}\right\}\right\} = \kappa.$$

To derive the asymptotic distribution of $\sqrt{T}(\widehat{\kappa} - \kappa)$ as shown in Part (a) of the theorem, we will first find the asymptotic distribution of $\sqrt{T}\left(\widehat{\kappa}^{-2\widehat{H}} - \kappa^{-2H}\right)$. Notice that

$$\begin{aligned} \widehat{\sigma}^2\widehat{H}\Gamma\left(2\widehat{H}\right)\left(\widehat{\kappa}^{-2\widehat{H}} - \kappa^{-2H}\right) &= \frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2 - \kappa^{-2H}\widehat{\sigma}^2\widehat{H}\Gamma\left(2\widehat{H}\right) \\ &= \frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2 - \sigma^2\kappa^{-2H}H\Gamma(2H) \\ &\quad - \kappa^{-2H}\left\{\widehat{\sigma}^2\widehat{H}\Gamma\left(2\widehat{H}\right) - \sigma^2 H\Gamma(2H)\right\}. \end{aligned} \tag{A.8}$$

From the asymptotic theory of $\widehat{\sigma}^2$ and $\widehat{H}$ provided in Theorem 4.1, we have, as $T\Delta \to 0$,

$$\begin{aligned}
&\widehat{\sigma}^2\widehat{H}\Gamma\left(2\widehat{H}\right) - \sigma^2 H\Gamma(2H) \\
&= \left(\widehat{\sigma}^2 - \sigma^2\right)\widehat{H}\Gamma\left(2\widehat{H}\right) + \sigma^2\left(\widehat{H} - H\right)\Gamma\left(2\widehat{H}\right) - \sigma^2 H\left[\Gamma\left(2\widehat{H}\right) - \Gamma(2H)\right] \\
&= O_p\left(\frac{\log(\Delta)}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(A.9)}
\end{aligned}$$

The order of the term $\Gamma\left(2\widehat{H}\right) - \Gamma(2H)$ is from the Taylor expansion as

$$\Gamma\left(2\widehat{H}\right) - \Gamma(2H) = \Gamma'\left(2\widetilde{H}\right)2\left(\widehat{H} - H\right),$$

where $\widetilde{H}$ takes values between $\widehat{H}$ and H and $\Gamma'(\cdot)$ is finite over the interval $(0, 4)$.

Define

$$\widehat{\kappa}_{HN} = \left(\frac{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2}{\sigma^2 H\Gamma(2H)}\right)^{-1/(2H)}. \qquad \text{(A.10)}$$

According to Theorem 3.3 of Xiao and Yu (2019a) and Theorem 3.1 of Xiao and Yu (2019b), when $H \in [1/2, 3/4)$ and $H \in (0, 1/2)$, as $T \to \infty$,

$$\sqrt{T}\left(\widehat{\kappa}_{HN} - \kappa\right) \xrightarrow{d} \mathcal{N}\left(0, \kappa\phi_H\right).$$

As a result, we have

$$\begin{aligned}
\frac{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2}{\sigma^2 H\Gamma(2H)} &= \left(\widehat{\kappa}_{HN}\right)^{-2H} \\
&= \kappa^{-2H} - 2H\kappa^{-2H-1}\left(\widehat{\kappa}_{HN} - \kappa\right) + O_p\left(\left(\widehat{\kappa}_{HN} - \kappa\right)^2\right),
\end{aligned}$$

and, as $T \to \infty$,

$$\begin{aligned}
&\sqrt{T}\left\{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2 - \sigma^2\kappa^{-2H} H\Gamma(2H)\right\} \\
&= \sqrt{T}\sigma^2 H\Gamma(2H)\left\{\left(\widehat{\kappa}_{HN}\right)^{-2H} - \kappa^{-2H}\right\} \\
&= \sqrt{T}\sigma^2 H\Gamma(2H)\left\{-2H\kappa^{-2H-1}\left(\widehat{\kappa}_{HN} - \kappa\right) + O_p\left(\left(\widehat{\kappa}_{HN} - \kappa\right)^2\right)\right\} \\
&\xrightarrow{d} \sigma^2 H\Gamma(2H)\left(-2H\kappa^{-2H-1}\right)\mathcal{N}\left(0, \kappa\phi_H\right).
\end{aligned}$$

Then, for the first term in (A.8), it is obtained that, as $T \to \infty$ and $\sqrt{T}\Delta^H \to 0$,

$$\begin{aligned}
&\sqrt{T}\left\{\frac{1}{n}\sum_{i=1}^n X_{i\Delta}^2 - \left(\frac{1}{n}\sum_{i=1}^n X_{i\Delta}\right)^2 - \sigma^2\kappa^{-2H} H\Gamma(2H)\right\} \\
&= \sqrt{T}\left\{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2 - \sigma^2\kappa^{-2H} H\Gamma(2H) + O_p\left(\Delta^H\right) + O_p\left(\frac{1}{n}\right)\right\} \\
&\xrightarrow{d} \sigma^2 H\Gamma(2H)\left(-2H\kappa^{-2H-1}\right)\mathcal{N}\left(0, \kappa\phi_H\right). \qquad \text{(A.11)}
\end{aligned}$$

Now, putting (A.9) and (A.11) in Eq. (A.8), we have, as $T \to \infty$, $T\Delta \to 0$, and $\sqrt{T}\Delta^H \to 0$,

$$\sqrt{T}\widehat{\sigma}^2\widehat{H}\Gamma\left(2\widehat{H}\right)\left(\widehat{\kappa}^{-2\widehat{H}} - \kappa^{-2H}\right) \xrightarrow{d} \sigma^2 H\Gamma(2H)\left(-2H\kappa^{-2H-1}\right)\mathcal{N}\left(0, \kappa\phi_H\right),$$

and

$$\sqrt{T}\left(\widehat{\kappa}^{-2\widehat{H}} - \kappa^{-2H}\right) \xrightarrow{d} \left(-2H\kappa^{-2H-1}\right)\mathcal{N}\left(0, \kappa\phi_H\right).$$

Note that the first-order Taylor expansion of $\widehat{\kappa}^{-2\widehat{H}}$ at $\widehat{\kappa} = \kappa$ takes the form of

$$\widehat{\kappa}^{-2\widehat{H}} = \kappa^{-2\widehat{H}} - 2\widehat{H}\widetilde{\kappa}^{-2\widehat{H}-1}\left(\widehat{\kappa} - \kappa\right),$$

where $\widetilde{\kappa}$ lies between $\widehat{\kappa}$ and κ. As a result, we have

$$-2\widehat{H}\widetilde{\kappa}^{-2\widehat{H}-1}\left(\widehat{\kappa}-\kappa\right)=\widehat{\kappa}^{-2\widehat{H}}-\kappa^{-2\widehat{H}}=\left(\widehat{\kappa}^{-2\widehat{H}}-\kappa^{-2H}\right)-\left(\kappa^{-2\widehat{H}}-\kappa^{-2H}\right)$$
$$=\left(\widehat{\kappa}^{-2\widehat{H}}-\kappa^{-2H}\right)+2\log\left(\kappa\right)\kappa^{-2H}\left(\widehat{H}-H\right)+O_p\left(\left(\widehat{H}-H\right)^2\right),$$

where the third equation comes from the first-order Taylor expansion of $\kappa^{-2\widehat{H}}$ at $\widehat{H}=H$. Finally, we have, as $T\to\infty$, $T\Delta\to 0$, and $\sqrt{T}\Delta^H\to 0$,

$$-2\widehat{H}\widetilde{\kappa}^{-2\widehat{H}-1}\sqrt{T}\left(\widehat{\kappa}-\kappa\right)=\sqrt{T}\left(\widehat{\kappa}^{-2\widehat{H}}-\kappa^{-2H}\right)+O_p\left(\sqrt{\Delta}\right)\xrightarrow{d}\left(-2H\kappa^{-2H-1}\right)\mathcal{N}\left(0,\kappa\phi_H\right),$$

thereby,

$$\sqrt{T}\left(\widehat{\kappa}-\kappa\right)\xrightarrow{d}\mathcal{N}\left(0,\kappa\phi_H\right),$$

which gives the asymptotic distribution shown in Part (a) of the theorem.

For $\widehat{\kappa}_{HN}$ given in (A.10), Xiao and Yu (2019a) have proved that, when $H=3/4$,

$$\frac{\sqrt{T}}{\log(T)}\left(\widehat{\kappa}-\kappa\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{16\kappa}{9\pi}\right),\text{ as }\ T\to\infty,$$

and, when $H\in(3/4,1)$,

$$T^{2-2H}\left(\widehat{\kappa}-\kappa\right)\xrightarrow{d}\frac{-\kappa^{2H-1}}{H\Gamma(2H+1)}R,\text{ as }\ T\to\infty,$$

where R denotes the Rosenblatt random variable with $\mathbb{E}\left(R^2\right)=2H^2(2H-1)/(4H-3)$. Using these results and taking the same procedure above for the proof of Part (a) of the theorem will give the asymptotic distributions in Part (b)–(c) of the theorem, respectively. The proof of the theorem is completed.

Appendix B. Supplementary empirical results

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2021.08.001.

References

Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. J. Finance 52 (3), 975–1005.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Ebens, H., 2001a. The distribution of realized stock return volatility. J. Financ. Econ. 61 (1), 43–76.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001b. The distribution of realized exchange rate volatility. J. Amer. Statist. Assoc. 96 (453), 42–55.
Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71 (2), 579–625.
Bardet, J.M., 2018. Theoretical and numerical comparisons of the parameter estimator of the fractional Brownian motion. In: Mathematical Structures and Applications: In Honor of Mahouton Norbert Hounkonnou. pp. 153–173.
Barndorff-Nielsen, O.E., Basse-O'Connor, A., 2011. Quasi Ornstein–Uhlenbeck processes. Bernoulli 17 (3), 916–941.
Barndorff-Nielsen, O.E., Corcuera, J.M., Podolskij, M., 2013. Limit theorems for functionals of higher order differences of Brownian semi-stationary processes. In: Prokhorov and Contemporary Probability Theory: In Honor of Yuri V. Prokhorov, Vol. 33. Springer Science & Business Media, pp. 69–96.
Bayer, C., Friz, P., Gatheral, J., 2016. Pricing under rough volatility. Quant. Finance 16 (6), 887–904.
Bégyn, A., 2007. Asymptotic expansion and central limit theorem for quadratic variations of Gaussian processes. Bernoulli 13 (3), 712–753.
Belfadli, R., Es-Sebaiy, K., Ouknine, Y., 2011. Parameter estimation for fractional Ornstein–Uhlenbeck processes: non-ergodic case. Front. Sci. Eng. 1, 1–16.
Bennedsen, M., 2017. Rough Continuous-Time Processes: Theory and Applications (Ph.D. Thesis). Aarhus University.
Bennedsen, M., Hounyo, U., Lunde, A., Pakkanen, M.S., 2019. The local fractional bootstrap. Scand. J. Stat. 46 (1), 329–359.
Bennedsen, M., Lunde, A., Pakkanen, M.S., 2021. Decoupling the short-and long-term behavior of stochastic volatility. J. Financ. Econom. forthcoming.
Bergstrom, A.R., 1990. Continuous Time Econometric Modelling. Oxford University Press.
Bibinger, M., 2020. Cusum tests for changes in the Hurst exponent and volatility of fractional Brownian motion. Stat. Probab. Lett. 161, 108725.
Bishwal, J.P., 2011. Minimum contrast estimation in fractional Ornstein–Uhlenbeck process: Continuous and discrete sampling. Fract. Calc. Appl. Anal. 14 (3), 375–410.
Bollerslev, T., Patton, A.J., Quaedvlieg, R., 2016. Exploiting the errors: A simple approach for improved volatility forecasting. J. Econometrics 192 (1), 1–18.
Cheridito, P., Kawaguchi, H., Maejima, M., 2003. Fractional Ornstein–Uhlenbeck processes. Electron. J. Probab. 8 (3), 1–14.
Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. J. Econometrics 138, 291–311.
Coeurjolly, J.F., 2000. Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. J. Stat. Softw. 5 (7), 1–53.
Coeurjolly, J.F., 2001. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. Stat. Inference Stoch. Process. 4 (2), 199–227.
Corcuera, J.M., Hedevang, E., Pakkanen, M.S., Podolskij, M., 2013. Asymptotic theory for Brownian semi-stationary processes with application to turbulence. Stochastic Process. Appl. 123 (7), 2552–2574.
Corsi, F., 2009. A simple approximate long-memory model of realized volatility. J. Financ. Econom. 7 (2), 174–196.

Davydov, Y.A., 1970. The invariance principle for stationary processes. Theory Probab. Appl. 15 (3), 487–498.
Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. J. Bus. Econ. Stat. 13, 253–265.
Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69 (4), 959–993.
Embrechts, P., Maejima, M., 2002. Selfsimilar Processes. Princeton University Press, Princeton, NJ.
Es-Sebaiy, K., 2013. Berry–Esséen bounds for the least squares estimator for discretely observed fractional Ornstein–Uhlenbeck processes. Statist. Probab. Lett. 83 (10), 2372–2385.
Euch, O.E., Rosenbaum, M., 2018. Perfect hedging in rough Heston models. Ann. Probab. 28 (6), 3813–3856.
Fink, H., Klüppelberg, C., Zähle, M., 2013. Conditional distributions of processes related to fractional Brownian motion. J. Appl. Probab. 50 (1), 166–183.
Fouque, J.P., Hu, R., 2018. Optimal portfolio under fast mean–reverting fractional stochastic environment. SIAM J. Financ. Math. 9 (2), 564–601.
Garnier, J., Sølna, K., 2017. Correction to Black–Scholes formula due to fractional stochastic volatility. SIAM J. Financ. Math. 8 (1), 560–588.
Gatheral, J., Jaisson, T., Rosenbaum, M., 2018. Volatility is rough. Quant. Finance 18 (6), 933–949.
Gehringer, J., Li, M., 2020. Functional limit theorems for the fractional Ornstein–Uhlenbeck process. J. Theoret. Probab. forthcoming.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Series Anal. 4 (4), 221–238.
Giraitis, L., Koul, H., Surgailis, D., 2012. Large Sample Inference for Long Memory Processes. Imperial College Press.
Gneiting, T., Schlather, M., 2004. Stochastic models that separate fractal dimension and the Hurst effect. SIAM Rev. 46 (2), 269–282.
Gneiting, T., Ševčíková, H., Percival, D.B., 2012. Estimators of fractal dimension: Assessing the roughness of time series and spatial data. Statist. Sci. 27 (2), 247–277.
Hansen, P.R., Lunde, A., Nason, J.M., 2011. The model confidence set. Econometrica 79 (2), 453–497.
Hu, Y., Nualart, D., 2010. Parameter estimation for fractional Ornstein–Uhlenbeck processes. Stat. Probab. Lett. 80 (11), 1030–1038.
Hu, Y., Nualart, D., Zhou, H., 2019. Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. Stat. Inference Stoch. Process. 22 (1), 111–142.
Isserlis, L., 1918. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. Biometrika 12, 134–139.
Jaisson, T., Rosenbaum, M., 2016. Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. Ann. Appl. Probab. 26 (5), 2860–2882.
Jiang, L., Wang, X., Yu, J., 2018. New distribution theory for the estimation of structural break point in mean. J. Econometrics 205 (1), 156–176.
Jiang, L., Wang, X., Yu, J., 2021. In-fill asymptotic theory for structural break point in autoregressions. Econometric Rev. 40, 359–386.
Kleptsyna, M., Le Breton, A., 2002. Statistical analysis of the fractional Ornstein–Uhlenbeck type process. Stat. Inference Stoch. Process. 5 (3), 229–248.
Lang, G., Roueff, F., 2001. Semi-parametric estimation of the Hölder exponent of a stationary Gaussian process with minimax rates. Stat. Inference Stoch. Process. 4 (3), 283–306.
Livieri, G., Mouti, S., Pallavicini, A., Rosenbaum, M., 2018. Rough volatility: evidence from option prices. IISE Trans. 50 (9), 767–776.
Ludeña, C., 2004. Minimum contrast estimation for fractional diffusions. Scand. J. Stat. 31 (4), 613–628.
Mincer, J.A., Zarnowitz, V., 1969. The evaluation of economic forecasts. In: Mincer, J. (Ed.), Economic Forecasts and Expectations. National Bureau of Economic Research, New York, pp. 3–46.
Paxson, V., 1997. Fast, approximate synthesis of fractional Gaussian noise for generating self-similar network traffic. Comput. Commun. Rev. 27 (5), 5–18.
Phillips, P.C.B., 1987. Towards a unified asymptotic theory for autoregression. Biometrika 74 (3), 535–547.
Phillips, P.C.B., Yu, J., 2005. Jackknifing bond option prices. Rev. Financ. Stud. 18 (2), 707–742.
Phillips, P.C.B., Yu, J., 2009a. Simulation-based estimation of contingent–claims prices. Rev. Financ. Stud. 22 (9), 3669–3705.
Phillips, P.C.B., Yu, J., 2009b. A two-stage realized volatility approach to estimation of diffusion processes with discrete data. J. Econom. 150 (2), 139–150.
Samorodnitsky, G., Taqqu, M., 1994. Non-Gaussian Stable Processes: Stochastic Models with Infinite Variance. Chapman and Hall/CRC Press.
Tanaka, K., 2013. Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. Stat. Inference Stoch. Process. 16 (3), 173–192.
Tanaka, K., 2015. Maximum likelihood estimation for the non-ergodic fractional Ornstein–Uhlenbeck process. Stat. Inference Stoch. Process. 18 (3), 315–332.
Tanaka, K., Xiao, W., Yu, J., 2020. Maximum likelihood estimation for the fractional Vasicek model. Econometrics 8 (32), 1–28.
Tang, C.Y., Chen, S.X., 2009. Parameter estimation and bias correction for diffusion processes. J. Econometrics 149 (1), 65–81.
Tudor, C., Viens, F., 2007. Statistical aspects of the fractional stochastic calculus. Ann. Statist. 35 (3), 1183–1212.
Wang, X., Phillips, P.C.B., Yu, J., 2011. Bias in estimating multivariate and univariate diffusions. J. Econom. 161 (2), 228–245.
Wang, X., Yu, J., 2016. Double asymptotics for explosive continuous-time models. J. Econom. 193 (1), 35–53.
Whittle, P., 1953. Estimation and information in stationary time series. Ark. Mat. 2 (5), 423–434.
Xiao, W., Yu, J., 2019a. Asymptotic theory for estimating the drift parameters in the fractional Vasicek model. Econom. Theory 35 (1), 198–231.
Xiao, W., Yu, J., 2019b. Asymptotic theory for rough fractional Vasicek models. Econ. Lett. 177, 26–29.
Yu, J., 2012. Bias in the estimation of the mean reversion parameter in continuous-time models. J. Econom. 169 (1), 114–122.
Yu, J., 2014. Econometric analysis of continuous time models: A survey of Peter Phillips's work and some new results. Econom. Theory 30 (4), 737–774.
Zhou, Q., Yu, J., 2015. Asymptotic theory for linear diffusions under alternative sampling schemes. Econom. Lett. 128, 1–5.