Research Collection School Of Economics   School of Economics

9-2019

Estimation and Inference of fractional continuous-time model with discrete-sampled data

Xiaohu WANG
Chinese University of Hong Kong

Weilin XIAO
Zhejiang University

Jun YU
Singapore Management University, yujun@smu.edu.sg

Estimation and Inference of Fractional Continuous-Time Model with Discrete-Sampled Data

Xiaohu Wang, Weilin Xiao, Jun Yu

September 2019

Estimation and Inference of Fractional Continuous-Time Model with Discrete-Sampled Data[*]

Xiaohu Wang[†]   Weilin Xiao[‡]   Jun Yu[§]

September 16, 2019

Abstract

This paper proposes a two-stage method for estimating parameters in a parametric fractional continuous-time model based on discrete-sampled observations. In the first stage, the Hurst parameter is estimated based on the ratio of two second-order differences of observations from different time scales. 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. A large sample theory of the proposed estimators is derived under either the in-fill asymptotic scheme or the double asymptotic scheme. Extensive simulations show that the proposed theory performs well in finite samples. Two empirical studies are carried out. The first, based on the daily realized volatility of equities from 2011 to 2017, shows that the Hurst parameter is much lower than 0.5, which suggests that the realized volatility is too rough for continuous-time models driven by standard Brownian motion or fractional Brownian motion with Hurst parameter larger than 0.5. The second empirical study is of the daily realized volatility of exchange rates from 1986 to 1999. The estimate of the Hurst parameter is again much lower than 0.5. Moreover, the proposed fractional continuous-time model performs better than the autoregressive fractionally integrated moving average (ARFIMA) model out-of-sample.

JEL Classification: C15, C22, C32.

Keywords: Rough Volatility; Hurst Parameter; Second-order Difference; Different Time Scales; Method of Moments; ARFIMA.

[†]Department of Economics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong. Email: xiaohu.wang@cuhk.edu.hk.

[‡]School of Management, Zhejiang University, Hangzhou, 310058, China. Email: wlxiao@zju.edu.cn.

[§]School of Economics and Lee Kong Chian School of Business, Singapore Management University, 90 Stamford Road, 178903, Singapore, Email: yujun@smu.edu.sg.

1 Introduction

In recent decades, the phenomenon of long-range dependence has been widely observed in data from hydrology, geophysics, climatology, telecommunication, and economics. In finance, Taylor (1986) and Ding et al. (1993) find that the absolute values and powers of stock returns tend to have slowly decaying autocorrelations. Following this stylized fact, many time series models are proposed to capture long-range dependence, both in discrete time and in continuous time. A partial list of references includes Granger and Joyeux (1980), Lo (1991), Ding et al. (1993), Cheung (1993), Baillie (1996), Baillie et al. (1996), and Andersen et al. (2003) in the domain of discrete time and Comte and Renault (1996, 1998), Aït-Sahalia and Mancini (2008), and Comte et al. (2012) in the domain of continuous time.

Among these models, Comte and Renault (1998) propose a continuous-time stochastic volatility model

$$dX_t = \kappa \left(\mu - X_t \right) dt + \sigma dB_t^H, \tag{1.1}$$

where $\kappa \in R^+$, $\sigma \in R^+$, and $\mu \in R$ are all constants, and B_t^H is a fractional Brownian motion (fBm) that is a zero-mean Gaussian process with covariance functions

$$Cov \left(B_t^H, B_s^H \right) = \frac{1}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right), \forall t, s \in R. \tag{1.2}$$

Although Comte and Renault (1998) impose the assumption that $H \in (1/2, 1)$, we do not impose this restriction in this paper. In general, H can take any value within the interval $(0, 1)$. Comte and Renault (1996) has established connections between the continuous-time model in (1.1) and some conventional long-memory models in the discrete-time literature.

Gatheral et al. (2018) point out that the sample path of logarithmic realized volatilities (RV) is often too rough to be fitted well by Model (1.1) with $H \in (1/2, 1)$. Instead, they propose the use of Model (1.1) with $H \in (0, 1/2)$, which they call the rough fractional stochastic volatility (RFSV) model. When examining the logarithmic RV of a DAX contract, Bund futures contract, S&P 500 index, and NASDAQ index, they document the evidence of $H \approx 0.1$ on any reasonable time scale. They also report the superior forecasting performance of the RFSV model relative to the heterogeneous autoregressive model of Corsi (2009).

A growing strand of literature now supports the findings of Gatheral et al. (2018).

For example, Bennedsen et al. (2017) document roughness in a large number of U.S. equities and reveal the superior forecasting performance of the RFSV model with the intraday volatility of the Emini S&P 500 futures contract; Livieri et al. (2018) report strong support for RFSV using the implied volatility-based approximations to spot volatility; Bayer et al. (2016) obtain strong support for RFSV via SPX volatility surface and variance swaps. The RFSV model is also applied in mathematical finance, such as in option pricing theory (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 RFSV.

To better appreciate the arguments of Gatheral et al. (2018), Figure 1 plots the sample path of the logarithmic daily RV of the S&P 500 index and three simulated sample paths generated from Model (1.1), with $H = 0.1453$, 0.5, and 0.7. To generate the sample paths of $\{X_t\}$, we set $\kappa = 1.381$, $\mu = 2.196$, and $\sigma = 0.844$, which are the estimated values when Model (1.1) is fitted to the logarithmic daily RV of the S&P 500 index using the estimation method proposed in our paper (0.1453 is the estimate of H). Figure 1 clearly shows that the sample paths of $\{X_t\}$ with $H = 0.5$ and 0.7 are much smoother than the real data. Moreover, the sample paths of $\{X_t\}$ become rougher as H decreases. When $H = 0.1453$, the level of roughness of the simulated sample-path looks very similar to that of the real data.

Despite the popularity of Model (1.1), to the best of our knowledge, statistical analysis of this model based on discrete-sampled data, including estimation and statistical inference, is limited. In this paper, we first propose a two-stage approach to the estimation of the parameters in Model (1.1). In the first stage, a novel estimator for the Hurst parameter H is introduced based on the ratio of squared summations of second-order differences of X_t obtained at different time scales. In the second stage, estimators of the other parameters in (1.1) 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.

We then develop a large sample theory for the proposed estimators. In particular, we consider two asymptotic schemes: (i) the in-fill asymptotic scheme under which the sampling interval Δ goes to zero with a fixed time-span T; and (ii) the double asymptotic scheme in which $\Delta \to 0$ and $T \to \infty$ simultaneously. Under both asymptotic schemes, the consistency and asymptotic normality of H and σ^2 are established for all $H \in (0, 1)$.

3
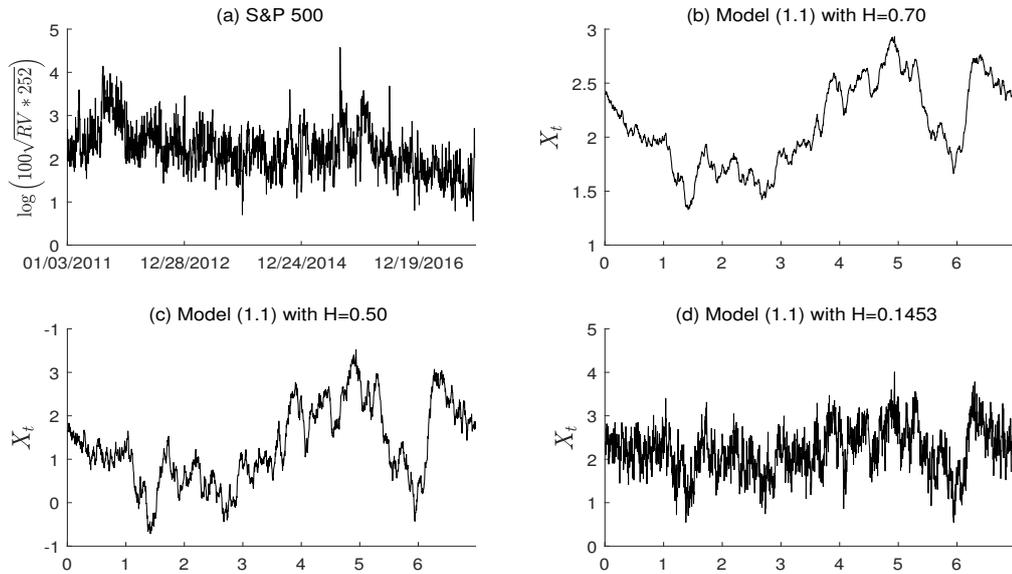
Figure 1: Time series plot of the logarithmic daily realized volatility of S&P 500 index and three simulated sample paths of $\{X_t\}$ generated by Model (1.1) with $\kappa = 1.381$, $\mu = 2.196$, $\sigma = 0.844$, and $H = 0.1453$, 0.5, and 0.7, respectively.

In addition, an explicit formula is derived for the asymptotic variance of H, which depends only on the value of H. This feature greatly facilitates statistical inference about H. Under the double asymptotic scheme, the consistency and asymptotic distributions for μ and κ are developed. The convergence rate for μ is a function of H. Both the convergence rate and the asymptotic distribution for κ depends crucially on H.

Extensive simulations demonstrate that the proposed estimators and the derived asymptotic distributions work well in finite samples. We also design an experiment to show the robustness of the proposed estimators to the microstructural noise effects. We then carry out two empirical studies. In the first, we apply the proposed two-stage estimation approach to the logarithmic daily RV for the S&P 500, the DJIA, and the Nasdaq 100. Our estimates and inference results suggest that H is statistically significantly less than 0.5. This conclusion is robust to jumps. In the second study, we apply the proposed estimation approach to the logarithmic daily RV for three spot exchange rates for the U.S. dollar, the Deutsch Mark, and the Japanese yen. Once again, we document strong evidence that $H < 0.5$ for each RV. The point estimates of H are similar to those obtained in the first empirical study and are very robust to jumps. In

4

addition, we document the superior out-of-sample performance of Model (1.1) relative to the discrete-time autoregressive fractionally integrated moving average (ARFIMA) model.

The remainder of the paper is organized as follows. Section 2 introduces the model and discusses its relationship with the discrete-time ARFIMA model. Section 3 proposes a two-stage estimation approach for the parameters in the concerned model. 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 developed large sample theory. Empirical studies are carried out in Section 6, and Section 7 presents our conclusions. All proofs are collected in the Appendix. Throughout the paper, we use $\xrightarrow{p}$, $\xrightarrow{a.s.}$, $\xrightarrow{d}$, $\overset{d}{=}$, and $\sim$ to denote convergence in probability, convergence almost surely, convergence in distribution, equivalence in distribution, and asymptotic equivalence, respectively.

2 Model and Some Preliminaries

2.1 Model

The model with which we are concerned in this paper is given by (1.1), where $\kappa \in R^+$, $\sigma \in R^+$, $\mu \in R$, and $H \in (0,1)$ are constants. The stochastic differential equation in (1.1) has a unique path-wise solution as (see, for example, Cheridito et al., 2003 in the case when $\mu = 0$)

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

where X_0 is the initial value of X_t at $t = 0$, and the stochastic integral exists as a path-wise Riemann-Stieltjes integral.

In practice, observations of X_t are available only at discrete time points, for example, at $n(:= T/\Delta)$ equally spaced points $\{i\Delta\}_{i=0}^n$, with Δ being the sampling interval and T being the time-span. When X_t is annualized and observed monthly (weekly or daily), then $\Delta = 1/12$ (1/52 or 1/252). Let $\{X_{i\Delta}\}_{i=0}^n$ denote the discrete-time observations of X_t. The exact discrete-time model of $\{X_{i\Delta}\}_{i=0}^n$ is obtained from (2.1) as

$$X_{i\Delta} = e^{-\kappa\Delta}X_{(i-1)\Delta} + \left(1 - e^{-\kappa\Delta}\right)\mu + \varepsilon_{i\Delta} \quad \text{with } \varepsilon_{i\Delta} = \sigma \int_{(i-1)\Delta}^{i\Delta} e^{-\kappa(i\Delta-s)}dB_s^H. \tag{2.2}$$

When $H = 1/2$, B_t^H becomes a standard Brownian motion and Model (2.2) turns out to be a first-order autoregressive model (AR(1)) with independent errors (see Bergstrom, 1990). Under the in-fill asymptotic scheme, which assumes $\Delta \to 0$ with a fixed T, $e^{-\kappa\Delta} \approx 1 - \kappa\Delta = 1 - \kappa T/n \to 1$. (2.2) is a local-to-unity model, as shown by Phillips (1987). In the double asymptotic scheme, which assumes $\Delta \to 0$ and $T \to \infty$, (2.2) is an AR(1) model with a root with moderate deviation from unity, as shown by Wang and Yu (2016).

When $H \neq 1/2$, the increments of B_t^H, also known as fractional Gaussian noise, are serially correlated, leading to serial dependence in $\{\varepsilon_{i\Delta}\}$. From the covariance structure of B_t^H given in (1.2), it can be proven that, for any fixed Δ, the increments process $\left\{v_{i\Delta} := B_{i\Delta}^H - B_{(i-1)\Delta}^H\right\}_{i=1}^n$ is stationary with the following autocovariance function

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

It is easy to show that $Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) > 0$ and $\sum_{j=0}^{\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = +\infty$ when $H \in (1/2, 1)$. In this case, $\{\varepsilon_{i\Delta}\}$ in (2.2) has positive serial correlations and is a long-memory process, which leads to long-range dependence in $\{X_{i\Delta}\}$ (Cheridito et al., 2003). Moreover, if κ is positive and close to zero, $\{X_{i\Delta}\}$ is stationary but has a root close to unity. In this case, $\{X_{i\Delta}\}$ behaves as the cumulative sum of $\{\varepsilon_{i\Delta}\}$. The positive correlation in $\{\varepsilon_{i\Delta}\}$ makes the sample path of $\{X_{i\Delta}\}$ smooth.

In contrast, if $H \in (0, 1/2)$, then $Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) < 0$ and $\{v_{i\Delta}\}_{i=1}^n$ are anti-persistent with $\sum_{j=0}^{\infty} Cov\left(v_{i\Delta}, v_{(i+j)\Delta}\right) = 0$. In this case, $\{\varepsilon_{i\Delta}\}$ in (2.2) has negative serial correlations that quickly decay to zero as the lag order increases. Although $\{\varepsilon_{i\Delta}\}$ does not have long-range dependence, when κ is close to zero, $\{X_{i\Delta}\}$ still exhibits long-range dependence due to the feature of local-to-unity. Moreover, the negative serial correlation in $\{\varepsilon_{i\Delta}\}$ induces a rough sample path in $\{X_{i\Delta}\}$, which explains the findings of Gatheral et al. (2018) that Model (1.1) with $H \approx 0.1$ and $\kappa \approx 0$ provides a good fit to RV, which has a rough sample path and slowly decaying autocorrelations.

To better appreciate the discussion above, we simulate $\{X_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$ from Model (1.1) with various values of H. We set $\kappa = 1.3810$, $\mu = 2.1960$, $\sigma = 0.8440$, and $\Delta = 1/256$. Details of the simulation of data from Model (1.1) are given in Section 5. Figure 2 plots the autocorrelations 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 $H = 0.1453$. The Figure clearly shows that $\{\varepsilon_{i\Delta}\}$ has positive serial correlations when $H = 0.7$ and negative serial

6
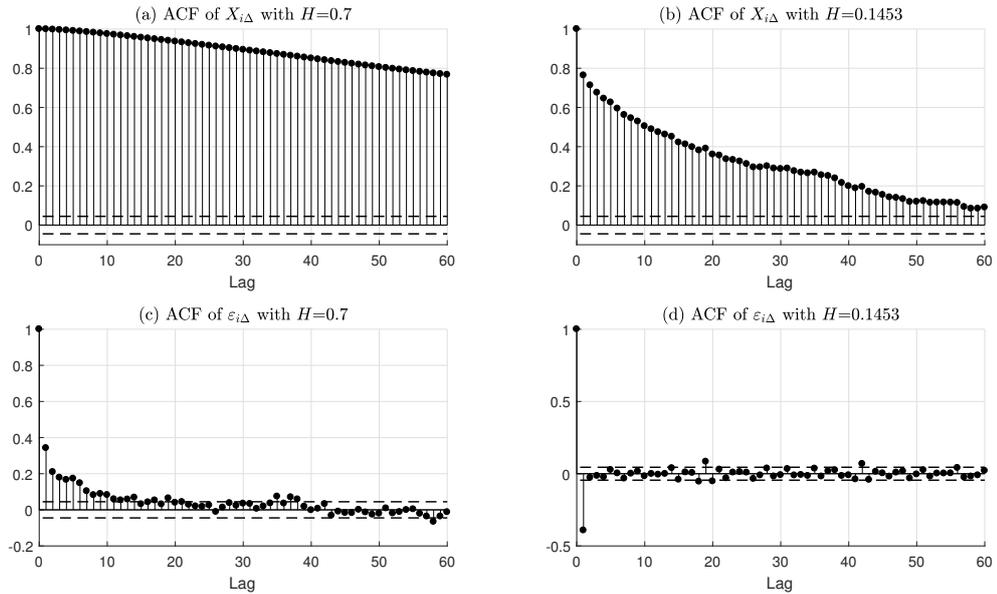
Figure 2: Autocorrelation functions of $X_{i\Delta}$ and $\varepsilon_{i\Delta}$ simulated from Model (1.1).

correlations when $H = 0.1453$. More importantly, in both cases, the process $\{X_{i\Delta}\}$ has positive and slowly decaying autocorrelations, and hence, long-range dependence.

Figure 3 plots the simulated sample paths of $\{X_{i\Delta}\}$ and $\{\varepsilon_{i\Delta}\}$ and shows that the sample paths of $\{\varepsilon_{i\Delta}\}$ are rough regardless of whether $H = 0.7$ or $H = 0.1453$. However, the sample path of $\{X_{i\Delta}\}$, which is close to that of the partial sums of $\{\varepsilon_{i\Delta}\}$, is smooth when $H = 0.7$ but remains rough when $H = 0.1453$.

When $\kappa > 0$, X_t defined in (1.1) is stationary. For simplicity, a stationary initial condition is taken as

$$X_0 = \mu + \sigma \int_{-\infty}^{0} e^{\kappa s} dB_s^H \stackrel{d}{=} N\left(\mu, \sigma^2 \kappa^{-2H} H\Gamma(2H)\right),$$

where $\Gamma(\cdot)$ denotes the gamma function, although all of the asymptotic results derived here continue to hold when X_0 is a constant or $X_0 = O_p(1)$. Under the stationary initial condition, $\{X_{i\Delta}\}_{i=1}^{n}$ is a Gaussian stationary process with

$$E(X_{i\Delta}) = \mu \quad \text{and} \quad Var(X_{i\Delta}) = \sigma^2 \kappa^{-2H} H\Gamma(2H). \tag{2.3}$$

An alternative representation of $X_{i\Delta}$ is

$$X_{i\Delta} = \mu + \sigma \int_{-\infty}^{i\Delta} e^{-\kappa(i\Delta - s)} dB_s^H. \tag{2.4}$$

7
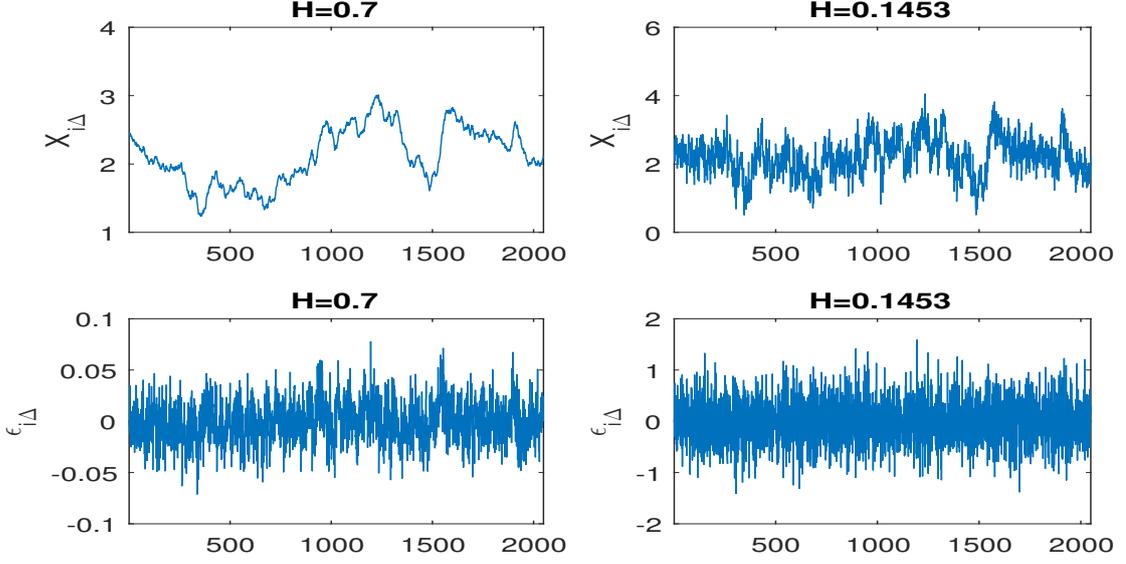
Figure 3: Time series plots of $X_{i\Delta}$ and $\varepsilon_{i\Delta}$ simulated from Model (1.1).

2.2 Relation to ARFIMA model

The continuous-time model (1.1) and its discretization (2.2) are closely related to the following stationary ARFIMA$(1, H - 1/2, 0)$ model that is widely used and extensively studied in the discrete-time literature:

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

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

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

Define $d := H - 1/2$. Because $H \in (0, 1)$, then $d \in (-1/2, 1/2)$. Together with the condition that $|\rho| < 1$, the ARFIMA model is stationary. It is well-established in the literature that the errors $\{u_{i\Delta}\}$ have a long memory when $d \in (0, 1/2)$ but are 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., $T = 1$), Davydov (1970) proves the following weak convergence result under some regular conditions: as $\Delta \to 0$,

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

8

Table 1: Mean and standard deviation (SD) of the ML estimates of d and ρ when fitting the ARFIMA$(1, d, 0)$ model by ML approach to data simulated from Model (1.1). When simulating data, we set κ =15, μ =2.8, σ =1, H =0.15, T =4, Δ =1/256. This setup implies that $d = H - 1/2 = -0.35$ and $\rho = \exp(-\kappa\Delta)$ =0.9414.

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

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 process defined in (1.1) (see also Tanaka (2013, 2015)).

The weak convergence in (2.6) may lead one to believe that the fractional continuous time model and the ARFIMA$(1, d, 0)$ model are essentially equivalent, especially when Δ is small; unfortunately, this belief is not justified. To show the difference between the fractional continuous time model and the ARFIMA model, we simulate data from Model (1.1) but fit the stationary ARFIMA model with the maximum likelihood (ML) method. When simulating the data, we set κ =15, μ =2.8, σ =1, H =0.15, T =4, and Δ =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 ML estimates 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 implied values. In fact, the ML estimates of d and ρ are very close to those obtained in the empirical study when we fit the ARFIMA model to the daily RV of exchange rates as shown in Section 6. Although not reported, decreasing the value of Δ essentially leads to no change in the mean of $\widehat{d}$ and the mean of $\widehat{\rho}$. We conjecture that there may exist critical differences between the asymptotic behavior of the two likelihoods that make the two model asymptotically non-equivalent. Here, the asymptotic equivalence is defined based on Le Cam's deficiency distance (see Le Cam (1986), Le Cam and Yang (1990)). Wang (2002) showed that the GARCH(1,1) model and the continuous-time stochastic volatility model are asymptotically non-equivalent, although the former model converges weakly to the latter model under an in-fill asymptotic scheme. Establishing such a result for the ARFIMA model and the fractional continuous-time model will be pursued in a future study.

3   A Two-Stage Estimation Approach

To estimate the parameters in (1.1) based on discrete-sampled data, it is difficult to apply the maximum likelihood method for the reason that the errors $\{\varepsilon_{t\Delta}\}$ in (2.2) have complicated dependent structure when $H \neq 1/2$.[1] In this paper, following Phillips and Yu (2009b), we proposes an alternative two-stage estimation approach, which is very easy to implement and does not require any tuning parameter.

In the first stage, motivated by Barndorff-Nielsen et al. (2013), we propose to estimate the Hurst parameter H by using

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

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 second-order differences of $\{X_{i\Delta}\}_{i=1}^n$ taken at two different time scales.[2]

Clearly, estimating H requires no information about other parameters in Model (1.1). Moreover, $\widehat{H}$ is trivial to compute from data. Section 3 develops the large sample theory of $\widehat{H}$ under two asymptotic schemes, including the in-fill asymptotic scheme (i.e., $\Delta \to 0$ with a fixed T) and the double asymptotic scheme (i.e., $\Delta \to 0$ and $T \to \infty$, simultaneously). Under both asymptotic schemes, consistency and asymptotic normality are established. The asymptotic distribution of $\widehat{H}$ depends only on the value of H itself. Hence, statistical inference of H can be done without knowing values of the other parameters in the model.

It is important to take second order differences to reduce the dependence in the data so that the central limit theory is applicable for all values of $H \in (0, 1)$. However, if

---

[1] 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. (2019) 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.

[2] While the same estimator was used in Barndorff-Nielsen et al. (2013) under a different model, Barndorff-Nielsen et al. (2013) require the process to be a Brownian semimartingale. Whereas, Model (1.1) is not a Brownian semimartingale unless $H = 1/2$. Hence, the asymptotic theory developed in Barndorff-Nielsen et al. (2013) is not applicable here.

H is known to be less than $3/4$, taking first-order differences is enough to reduce the dependence in the data. The asymptotic theory for such an estimator is developed in Phillips et al. (2019) for fractional continuous-time models with more general drift and diffusion functions.

In the second stage, we estimate the other parameters, σ, μ, κ, in Model (1.1) using the following method-of-moments estimators:

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

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

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

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, $X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta} \approx \sigma\left(B^H_{(i+2)\Delta} - 2B^H_{(i+1)\Delta} + B^H_{i\Delta}\right)$. It has been proved in Hu et al. (2019) 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 zero mean and variance $4 - 2^{2H}$ for any $H \in (0,1)$. Using the self-similarity property of B^H_t, we have

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

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).

The estimators $\widehat{\mu}$ and $\widehat{\kappa}$ are closely related to some previous studies in the continuous-time literature, where σ and H are assumed to be known and a continuous record of $\{X_t\}$ is assumed to be observed over the time interval $[0, T]$. For example, Xiao and Yu (2019a, b) have proposed the method-of-moments estimators of κ and μ with a continuous-time record of $\{X_t\}$ available, whose expressions are similar to the estimators $\widehat{\mu}$ and $\widehat{\kappa}$ in (3.3) and (3.4) with (i) the summations replaced by corresponding Riemann integrals and (ii) the estimates $\widehat{\sigma}$ and $\widehat{H}$ replaced by their true values. Pioneer studies for estimating κ with the additional condition of $\mu = 0$ are Hu and Nualart (2010) and Hu et al. (2019).

4 Asymptotic Theory

The large sample theory of $\widehat{H}$ and $\widehat{\sigma}$ defined in (3.1) and (3.2) is reported Section 3.1. 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; 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 $1/\Delta$.[3] In Section 3.2, 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 $\hat{\sigma}$ be the estimators defined in (3.1) and (3.2) for Model (1.1). For all $H \in (0,1)$, when $T\Delta \to 0$ and $n = T/\Delta \to \infty$, we have*
(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

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

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

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

---

[3]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 $\Delta \to 0$.

Remark 4.1 *It can be proved that* $\rho_j \sim O\left(j^{2H-4}\right)$ *as* $j \to \infty$. *Hence, for any* $H \in (0,1)$, *the sequence* $\{\rho_j\}_{j=1}^{\infty}$ *is square summable, ensuring that the infinite sums in* Σ_{11}, Σ_{12}, *and* Σ_{22} *are all finite. A simple proof by only using the mean value theorem for integrals suggests that, as* $j \to \infty$, *we have*

$$
2\left(4 - 2^{2H}\right)\rho_j
$$
$$
= 2H\left\{-\int_{j+1}^{j+2} x^{2H-1}dx + 3\int_{j}^{j+1} x^{2H-1}dx - 3\int_{j-1}^{j} x^{2H-1}dx + \int_{j-2}^{j-1} x^{2H-1}dx\right\}
$$
$$
= 2H\left\{-(j+1+\lambda_1)^{2H-1} + 3(j+\lambda_2)^{2H-1} - 3(j-1+\lambda_3)^{2H-1} + (j-2+\lambda_4)^{2H-1}\right\}
$$
$$
= 2H(2H-1)\left\{-\int_{j+\lambda_2}^{j+1+\lambda_1} x^{2H-2}dx + 2\int_{j-1+\lambda_3}^{j+\lambda_2} x^{2H-2}dx - \int_{j-2+\lambda_4}^{j-1+\lambda_3} x^{2H-2}dx\right\}
$$
$$
\approx 2H(2H-1)\left\{-(j+\lambda_5)^{2H-2} + 2(j-1+\lambda_6)^{2H-2} - (j-2+\lambda_7)^{2H-2}\right\}
$$
$$
= 2H(2H-1)(2H-2)\left\{-\int_{j-1+\lambda_6}^{j+\lambda_5} x^{2H-3}dx + \int_{j-2+\lambda_7}^{j-1+\lambda_6} x^{2H-3}dx\right\}
$$
$$
\approx 2H(2H-1)(2H-2)\left\{-(j-1+\lambda_8)^{2H-3} + (j-2+\lambda_9)^{2H-3}\right\}
$$
$$
= 2H(2H-1)(2H-2)(2H-3)\left\{-\int_{j-2+\lambda_9}^{j-1+\lambda_8} x^{2H-4}dx\right\} \sim O\left(j^{2H-4}\right),
$$

where the second equality and the approximate equations come from the mean value theorem for integrals, and $\{\lambda_s\}_{s=1}^{9}$ *are real numbers in the interval* $(0,3)$.

Remark 4.2 *Figure 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 over the interval* $H \in (0,1)$.

Remark 4.3 *It is worth mentioning that the asymptotic variance of* $\widehat{H}$ *only depend on* H *while the asymptotic variance of* $\widehat{\sigma}^2$ *only depends on* H *and* σ^2. *Neither depends on* κ *and* μ. *This feature greatly facilitates statistical inference about* H *and* σ^2 *because* H *and* σ^2 *can be consistently estimated when* T *is fixed but* κ *and* μ *cannot.*

Remark 4.4 *Although we have assumed* $\kappa > 0$ *in Model (1.1), the proposed estimators of* H *and* σ *still work when* $\kappa \leq 0$. *Moreover, the developed in-fill asymptotic theory still applies when* $\kappa \leq 0$. *In fact, if* $\kappa = 0$, $X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta} = \sigma\left(B_{(i+2)\Delta}^{H} - 2B_{(i+1)\Delta}^{H} + B_{i\Delta}^{H}\right)$, *making it easier to develop the asymptotic distributions of* $\widehat{H}$ *and* $\widehat{\sigma}^2$.

13
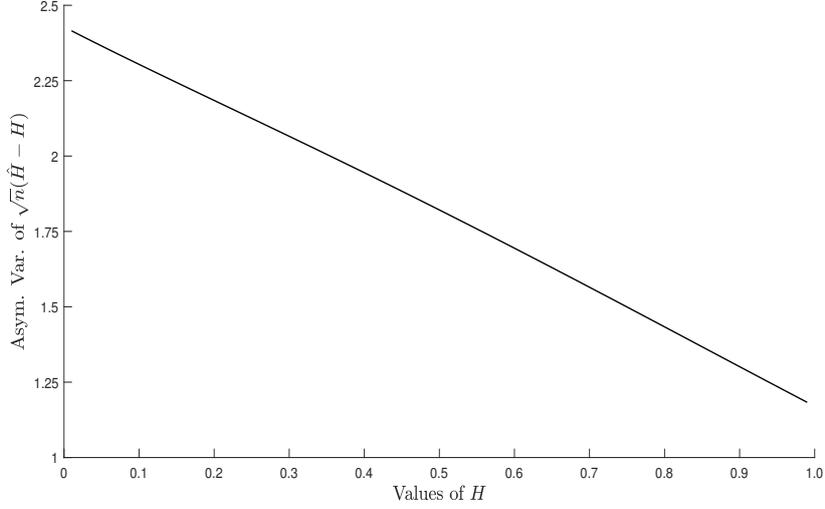
Figure 4: Asymptotic variance of $\sqrt{n}\left(\widehat{H} - H\right)$ as a function of $H \in (0, 1)$.

When $H = 1/2$, Model (1.1) becomes the Vasicek model that has been used to model interest rates in the literature. The Vasicek model enjoys the Markov property. Whereas, if $H \neq 1/2$, Model (1.1) does not have the Markov property any more. 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 $n = T/\Delta \to \infty$,*

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

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 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*

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

14

Theorem 4.4 *Let $\widehat{\kappa}$ be the estimator of κ defined in (3.4). For all $H \in (0,1)$, when $T \to \infty$ and $T\Delta \to 0$, we have $\widehat{\kappa} \overset{p}{\to} \kappa$. If, in addition,*
(a) for $H \in (0, 3/4)$, $\sqrt{T}\Delta^H \to 0$, then

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

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

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

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

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

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

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

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

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

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

15
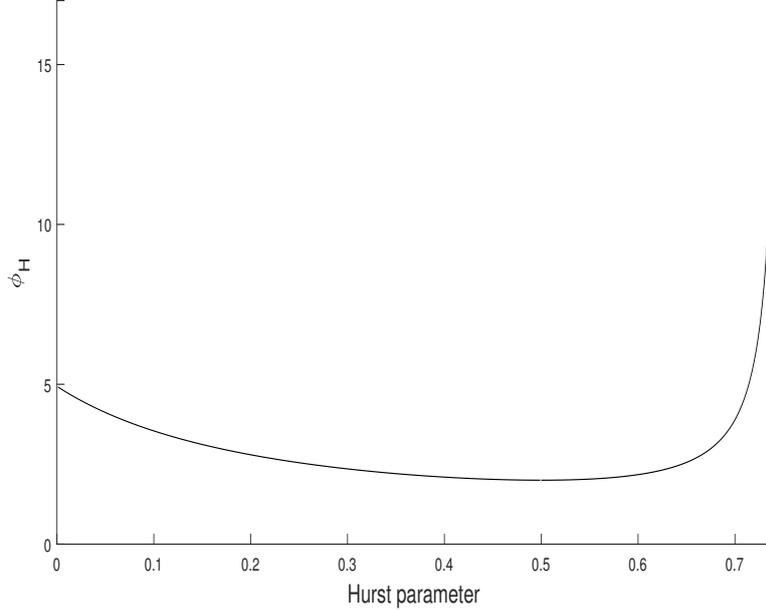
Figure 5: Plot of ϕ_H as a function of H.

Remark 4.6 *When $H = 1/2$ and is known, the double asymptotic distribution of the ML estimator of κ is known to be $N(0, 2\kappa)$; 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 4.7 *Figure 3 plots ϕ_H as a function of H. It shows that ϕ_H reaches the minimum value 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 (1.1), various values of

16

H, σ, μ and κ, and different combinations of the sampling frequency Δ and 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 (1.1).

(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 series of fractional Gaussian noise $\left\{B_{j\gamma}^H - B_{(j-1)\gamma}^H\right\}_{j=1}^{nM}$ by using fast Fourier transformation at the finer grid $\gamma := \Delta/M$.[4]

(iv) The Euler approximation of Model (1.1) over the interval $((j-1)\gamma, j\gamma)$ takes the form of

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

Starting from any pre-determined initial value X_0, the time series $\{X_{j\gamma}\}_{j=1}^{nM}$ is generated recursively based on Equation in (5.1) with the simulated fractional Gaussian noise series $\left\{B_{j\gamma}^H - B_{(j-1)\gamma}^H\right\}_{j=1}^{nM}$ obtained in Step 3. A subset of $\{X_{j\gamma}\}_{j=1}^{nM}$ is $\{X_{i\Delta}\}_{i=0}^{n}$, which gives the simulated sample path of the process X_t with the target sampling interval Δ.[5]

---

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

[5]For any target sampling interval Δ, a representation of Model (1.1) over the interval $((i-1)\Delta, i\Delta)$ is

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

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

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

which is the same as Equation (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 Equation (5.1), then the simulated data are

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

which is the same as Equation (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 the M, the smaller the approximation

17

(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 the estimator $\widehat{H}$ 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 $\kappa = 1.381$, $\mu = 2.196$, and $\sigma = 0.844$, which are the estimated values when Model (1.1) is fitted to the logarithmic daily RV of S&P 500 index using the estimation method proposed in the present paper. Simulation results are reported in Table 1, including the mean, the SD, the 2.5 percentile, and the 97.5 percentile. For the purpose of comparison, we also calculate and report (in parentheses) the SD, the 2.5 percentile and 97.5 percentile implied by the asymptotic theory given by (4.1).

Table 2 reveals several features. First, for all combinations of Δ and T, and for all values of 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, the SD, and the 95% confidence interval of $\widehat{H}$ each 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, the 2.5 percentile, and 97.5 percentile are very close to their asymptotic counterparts, which suggests that the asymptotic distribution derived in Theorem 4.1 can provide excellent approximations to finite-sample distribution.

To better show how well the derived asymptotic distribution can approximate its finite-sample counterpart, in Figure 6, we plot the histogram of the statistic $\Phi\left(\widehat{H}, H, n\right)$ defined as

$$\Phi\left(\widehat{H}, H, n\right) = \frac{2\sqrt{n}\log 2}{\sqrt{\Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}}}\left(\widehat{H} - H\right),$$

where $\Sigma_{11}, \Sigma_{22},$ and Σ_{12} are defined in Theorem 4.1, and $\widehat{H}$ comes from the first experiment. We then compare it with the density of $N(0,1)$. We consider the cases in which $n = T/\Delta$ with $T = 10$ and $\Delta = 1/256$, and $H = 0.1, 0.3, 0.5,$ and 0.7, respectively. It can be seen in Figure 6 that, in all cases, the histograms can be well approximated by

---

error generated by using Riemann sums. When Δ is already 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).

Table 2: Finite sample properties of the estimator $\widehat{H}$ defined in (3.1). Values reported in parentheses are calculated based on the asymptotic theory given in (4.1). In the simulations, we set $M = 8$, and fix $\kappa = 1.381$, $\mu = 2.196$, and $\sigma = 0.844$, which are the estimated values when Model (1.1) is fitted to the logarithmic daily RV of S&P 500 index using the estimation method proposed in the present paper.

Value of H		0.1	0.2	0.3	0.5	0.7	0.8	0.9
Panel A: $T = 4$								
	Mean	0.0984	0.1982	0.2981	0.4979	0.6977	0.7976	0.8974
	SD	0.0470	0.0460	0.0449	0.0423	0.0392	0.0375	0.0357
		(0.0474)	(0.0461)	(0.0449)	(0.0421)	(0.0390)	(0.0374)	(0.0356)
$\Delta = \frac{1}{256}$	2.5%	0.0057	0.1071	0.2092	0.4142	0.6190	0.7211	0.8258
		(0.0055)	(0.1085)	(0.2115)	(0.4173)	(0.6227)	(0.7251)	(0.8270)
	97.5%	0.1892	0.2867	0.3834	0.5785	0.7732	0.8707	0.9670
		(0.1944)	(0.2914)	(0.3884)	(0.5826)	(0.7772)	(0.8748)	(0.9729)
	Mean	0.0995	0.1994	0.2993	0.4992	0.6991	0.7990	0.8989
	SD	0.0334	0.0326	0.0319	0.0301	0.0280	0.0269	0.0256
		(0.0335)	(0.0326)	(0.0317)	(0.0298)	(0.0276)	(0.0264)	(0.0252)
$\Delta = \frac{1}{512}$	2.5%	0.0339	0.1355	0.2367	0.4394	0.6432	0.7456	0.8478
		(0.0331)	(0.1353)	(0.2374)	(0.4415)	(0.6453)	(0.7470)	(0.8484)
	97.5%	0.1648	0.2632	0.3613	0.5582	0.7542	0.8515	0.9483
		(0.1668)	(0.2646)	(0.3625)	(0.5584)	(0.7546)	(0.8529)	(0.9515)
Panel B: $T = 16$								
	Mean	0.0994	0.1994	0.2993	0.4991	0.6989	0.7988	0.8986
	SD	0.0239	0.0232	0.0225	0.0211	0.0196	0.0187	0.0179
		(0.0237)	(0.0230)	(0.0224)	(0.0210)	(0.0195)	(0.0187)	(0.0178)
$\Delta = \frac{1}{256}$	2.5%	0.0521	0.1534	0.2544	0.4570	0.6598	0.7616	0.8636
		(0.0527)	(0.1542)	(0.2557)	(0.4586)	(0.6613)	(0.7625)	(0.8635)
	97.5%	0.1453	0.2441	0.3427	0.5400	0.7368	0.8351	0.9334
		(0.1472)	(0.2457)	(0.3442)	(0.5413)	(0.7386)	(0.8374)	(0.9364)
	Mean	0.0998	0.1998	0.2998	0.4997	0.6997	0.7997	0.8996
	SD	0.0166	0.0161	0.0157	0.0148	0.0137	0.0131	0.0125
		(0.0167)	(0.0163)	(0.0158)	(0.0149)	(0.0138)	(0.0132)	(0.0126)
$\Delta = \frac{1}{512}$	2.5%	0.0671	0.1681	0.2692	0.4706	0.6727	0.7739	0.8750
		(0.0665)	(0.1676)	(0.2687)	(0.4707)	(0.6726)	(0.7735)	(0.8742)
	97.5%	0.1325	0.2312	0.3304	0.5290	0.7270	0.8255	0.9241
		(0.1334)	(0.2323)	(0.3312)	(0.5292)	(0.7273)	(0.8264)	(0.9257)
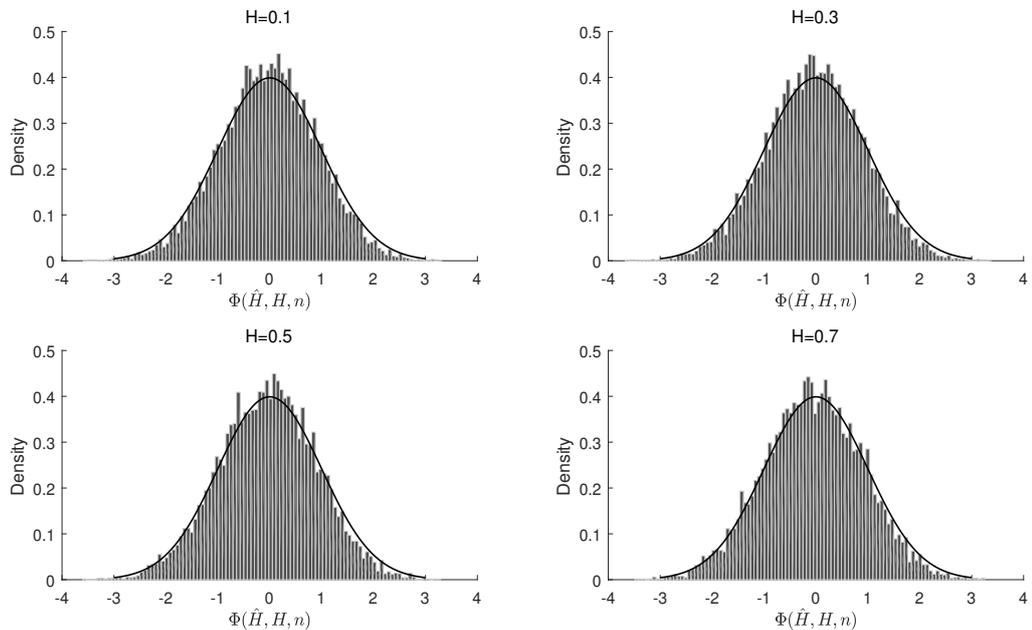
Figure 6: Histograms of $\Phi\left(\widehat{H}, H, n\right)$ with $n = T/\Delta$, $T = 10$, $\Delta = 1/256$, and $H = 0.1, 0.3, 0.5, 0.7$, respectively. Superimposed in the solid line is the density of $N(0, 1)$.

the density of $N(0, 1)$, which suggests that the derived asymptotic distribution works well in finite samples.

In the second experiment, we set $\sigma = 1$, $\mu = 2.8$, $\kappa = 5$, $T = 16$, $\Delta = 1/256$, and $M = 8$ 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, comparison with the results in Panel B of Table 2 shows the estimation of H is not sensitive to the change in the values of σ, μ, or κ, which indicates that the good properties of $\widehat{H}$ reported in Table 2 continue to hold for various values of σ, μ, and κ. Second, the parameters σ and μ can be accurately estimated. The means and medians are always close to their respective true parameter values, and the SDs are small. 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 Figure 2, which shows that the asymptotic variance of $\widehat{\sigma}$ is a decreasing function of H. Furthermore, as the value of H changes, 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 H has a negative effect on the precision of $\widehat{\mu}$.

Table 3: Finite sample properties of the estimates of the parameters (H, σ, μ, κ) with various values of H, $T = 16$ and $\Delta = 1/256$

	H	σ	μ	κ	::::	H	σ	μ	κ
True value	.10	1.00	2.80	5.00		.30	1.00	2.80	5.00
Mean	.0994	1.0063	2.7999	5.6397		.2993	1.0049	2.7997	5.1998
Median	.0997	.9981	2.7997	4.7395		.2993	.9978	2.7993	5.0043
SD	.0239	.1324	.0173	4.2997		.0225	.1284	.0288	1.8281
2.5%	.0521	.7689	2.7664	.1671		.2544	.7711	2.744	2.1814
97.5%	.1453	1.2853	2.8339	16.1090		.3427	1.2749	2.8562	9.2864
True value	.50	1.00	2.80	5.00		.70	1.00	2.80	5.00
Mean	.4991	1.0037	2.7995	5.2745		.6989	1.0029	2.7995	5.6529
Median	.4994	.9969	2.7990	5.1378		.6988	.9932	2.7986	5.5571
SD	.0211	.1269	.0499	1.3391		.0196	.1317	.0865	1.2643
2.5%	.4570	.7758	2.7038	2.9904		.6598	.7674	2.6326	3.4720
97.5%	.5400	1.2723	2.8966	8.2454		.7368	1.2832	2.9694	8.4214

Third, the parameter κ can be estimated with less precision. The SDs are comparatively large, and the bias and skewness in $\widehat{\kappa}$ are noticeable. The difficulties in estimating κ have been well studied for continuous-time models driven by standard Brownian motion (i.e., $H = 1/2$); see, for example, Phillips and Yu (2005, 2009a) and Wang et al. (2011). Tang and Chen (2009) and Yu (2012) derive analytical expressions to approximate the bias in the MLE 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 the true value of κ but also on the value of 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 the value of H.

In the third experiment, we fix H to 0.45 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 $\mu = 2.8$, $\kappa = 5$, and σ varies from 0.3 to 2. The simulation results confirm the prediction from the asymptotic theories developed here that a change in σ should have no effect on the 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 $\sigma = 0.3$, $\kappa = 5$, and μ varies from 0.5 to 2. As predicted by the asymptotic theories, the estimation results of H, σ,

Table 4: Estimates of (H, σ, μ, κ) when $H = 0.45$, $T = 16$, $\Delta = 1/256$, and (σ, μ, κ) take various values.

	H	σ	μ	κ	::::	H	σ	μ	κ
Panel A: σ varies									
True value	.45	.30	2.80	5.00		.45	.50	2.80	5.00
Mean	.4492	.3012	2.7999	5.2383		.4492	.5020	2.7998	5.2383
SD	.0215	.0381	.0131	1.4123		.0215	.0635	.0218	1.4123
True value	.45	1.00	2.80	5.00		.45	2.00	2.80	5.00
Mean	.4492	1.0040	2.7996	5.2383		.4492	2.0081	2.7992	5.2383
SD	.0215	.1271	.0436	1.4123		.0215	.2541	.0871	1.4123
Panel B: μ varies									
True value	.45	.30	.50	5.00		.45	.30	1.00	5.00
Mean	.4492	.3012	.4999	5.2383		.4492	.3012	.9999	5.2383
SD	.0215	.0381	.0131	1.4123		.0215	.0381	.0131	1.4123
True value	.45	0.30	1.50	5.00		.45	.30	2.00	5.00
Mean	.4492	.3012	1.4999	5.2383		.4492	.3012	1.9999	5.2383
SD	.0215	.0381	.0131	1.4123		.0215	.0381	.0131	1.4123
Panel C: κ varies									
True value	.45	1.00	2.80	1.00		.45	1.00	2.80	3.00
Mean	.4494	1.0045	2.7978	1.2595		.4494	1.0045	2.7974	3.2483
SD	.0215	.1271	.2123	.5612		.0215	.1271	.0724	1.0246
True value	.45	1.00	2.80	5.00		.45	1.00	2.80	10.00
Mean	.4492	1.0040	2.7996	5.2383		.4485	1.0008	2.7998	10.1683
SD	.0215	.1271	.0436	1.4123		.0215	.1267	.0218	2.2102

and κ and the SD of $\widehat{\mu}$ all remain the same when the value of μ is changed.

Panel C of Table 4 reports the simulation results when $\sigma = 1$, $\mu = 2.8$, and κ varies from 1 to 10. It shows that the results of the estimation of H and σ are insensitive to the change in κ, whereas when κ is increased from 1 to 10, 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 theories.

To see the effects of Δ and M on the estimates, we design the fourth experiment by fixing T and the four parameters (H, σ, μ, and κ), but varying the value of Δ from $1/256$ to $1/2048$ and the value of M from 16 to 32. Note that M is chosen to control

Table 5: Performance of the estimators when $T = 16$ and (Δ, M) vary.

	$\Delta = 1/256$				::::	$\Delta = 1/2048$			
	H	σ	μ	κ		H	σ	μ	κ
True value	.45	1.00	2.80	5.00		.45	1.00	2.80	5.00
Panel A: $M = 16$									
Mean	.4492	1.0037	2.8003	5.2301		.4498	1.0006	2.8004	5.2402
Median	.4493	.9953	2.7999	5.0811		.4500	.9995	2.8006	5.1518
S.Dev.	.0215	.1271	.0429	1.4179		.0076	.0601	.0434	1.0321
2.5%	.4074	.7795	2.7172	2.8915		.4352	.8888	2.7165	3.5075
97.5%	.4912	1.2752	2.8862	8.4212		.4644	1.1217	2.8865	7.5137
Panel B: $M = 32$									
Mean	.4494	1.0040	2.8007	5.2406		.4500	1.0010	2.7992	5.2232
Median	.4493	.9952	2.8004	5.1082		.4500	.9993	2.7992	5.1324
S.Dev.	.0213	.1262	.0432	1.4070		.0076	.0601	.0436	1.0189
2.5%	.4075	.7815	2.7181	2.9167		.4353	.8899	2.7136	3.4769
97.5%	.4905	1.2716	2.8864	8.4084		.4649	1.1259	2.8834	7.4461

the discretization errors generated by applying the Euler discretization when simulating data. If Δ is already small enough, the Euler discretization error is negligible; hence, the choice of M does not materially change the simulation results. In contrast, the value of Δ may affect the performance of the estimators. Here, $1/256$ corresponds roughly to daily observations, whereas $1/2048$ corresponds to hourly observations. The asymptotic theory given in Theorem 4.1 shows that the estimators $\widehat{H}$ and $\widehat{\sigma}$ are consistent under the scheme of $\Delta \to 0$. Hence, it is expected that, when Δ changes from $1/256$ to $1/2048$, the performances of $\widehat{H}$ and $\widehat{\sigma}$ should improve. The consistency of $\widehat{\mu}$ and $\widehat{\kappa}$ requires $T \to \infty$. Hence, the change in Δ may have only a limited effect on $\widehat{\mu}$ and $\widehat{\kappa}$. Each of these predictions is supported by the simulation results reported in Table 5.

Table 6 presents the simulation results when $\Delta = 1/256$ and $M = 16$, but the time-span T varies from 8 to 16. The Table shows that the performance of all estimators improves as T increases. In particular, the bias in $\widehat{\kappa}$ is reduced by approximately half when the value of T is doubled. The SDs for $\widehat{H}$, $\widehat{\sigma}$, and $\widehat{\kappa}$ are reduced by a factor of $\sqrt{2}$, whereas the SD for $\widehat{\mu}$ is reduced by a factor of $2^{0.55} \approx 1.46$ when the value of T is doubled. Again, these findings are consistent with the prediction of our asymptotic theory.

Table 6: Performance of the estimators when $\Delta = 1/256$, $M = 16$, but T varies from 8 to 16.

	$T = 8$				::::	$T = 16$			
	H	σ	μ	κ		H	σ	μ	κ
True value	.45	1.00	2.80	5.00		.45	1.00	2.80	5.00
Mean	.4491	1.01149	2.7994	5.5351		.4491	1.0036	2.8002	5.2301
Median	.4490	.9928	2.7987	5.2522		.4493	.9953	2.7999	5.0810
SD	.0306	.1834	.0634	2.1019		.0214	.1270	.0429	1.4179
2.5%	.3897	.7016	2.6780	2.3040		.4074	.7795	2.7172	2.8915
97.5%	.5091	1.4168	2.9219	10.4051		.4912	1.2752	2.8862	8.4212

Thus far, we have assumed that X_t is observed without measurement errors. In the empirical applications considered below, X_t corresponds to daily integrated volatility, which is estimated by the daily RV obtained from high-frequency data. The presence of microstructural noise in ultra high-frequency data indicates that non-negligible estimation errors are present in the daily RV. In the last experiment, we assume that discrete observations of logarithmic daily integrated volatility are generated from Model (1.1) (denoted as X_t^*), where we set $T = 16$, $\Delta = 1/256$, $M = 8$, $\kappa = 5$, $\mu = 2.8$, and $\sigma = 1$, but vary $H = 0.1, 0.2, ..., 0.9$. We add an estimation error $v_t \sim N(0, 10^{-3})$ to the generated series to produce the RV X_t (i.e., $X_t = X_t^* + v_t$). Such a choice is consistent with the empirical results in the literature regarding the size of estimation errors relative to the size of the daily RV (see, for example, Bollerslev et al., 2016).[6] Table 7 reports the estimates of H by using data with and without measurement errors. It can be seen that the empirically relevant measurement errors do not affect the accuracy in the estimates of H, especially when $H \leq 0.7$.

6 Empirical Studies

This section includes two empirical studies. In the first empirical study, we fit Model (1.1) to three logarithmic daily RV series for equities. We apply the proposed estimation method and the new asymptotic theory to test the null hypothesis of $H = 0.5$. In the

---

[6]Bollerslev et al. (2016) report the median realized quarticity of 0.000175 and the median RV of 0.6295 in the S&P500. The ratio between the two quantities is 2.78×10^{-4}, which is slightly less than the ratio between 10^{-3} and 2.8 used in our simulation.

Table 7: Performance of $\widehat{H}$ when data has measurement errors with variance 10^{-3}. The experiment sets $T = 16$, $\Delta = 1/256$, $M = 8$, $\kappa = 5$, $\mu = 2.8$, $\sigma = 1$.

	$H = 0.1$	$H = 0.2$	$H = 0.3$	$H = 0.4$	$H = 0.5$	$H = 0.6$	$H = 0.7$	$H = 0.8$	$H = 0.9$
Panel A: Estimation results without measurement errors									
Mean	0.0993	0.1993	0.2992	0.3992	0.4991	0.5990	0.6990	0.7988	0.8987
Median	0.0992	0.1993	0.2992	0.3991	0.4993	0.5992	0.6991	0.7991	0.8988
SD	0.0238	0.0232	0.0225	0.0218	0.0211	0.0203	0.0195	0.0186	0.0177
Panel B: Estimation results with measurement errors									
Mean	0.0993	0.1992	0.2992	0.3991	0.4988	0.5979	0.6943	0.7781	0.7875
Median	0.0993	0.1993	0.2992	0.3992	0.4991	0.5980	0.6945	0.7782	0.7876
SD	0.0238	0.0232	0.0225	0.0218	0.0211	0.0203	0.0195	0.0189	0.0190

second study, we mainly compare the out-of-sample performance of Model (1.1) relative to an ARFIMA$(1, d, 0)$ model for forecasting three daily RV series for exchange rates.

6.1 RV for equities

We now fit Model (1.1) to three logarithmic daily RV series for the S&P 500, DJIA, and Nasdaq 100. The three RV series are obtained from the Ox-Mann realized library and based on 5-minute returns.[7] The sample period is from 01/03/2011 to 12/04/2017. Figure 7 plots three time series of $\log(100\sqrt{RV \times 252})$ which is the logarithmic annualized RV.

Table 8 reports three sets of estimation results, including the point estimates and the 95% confidence intervals for all four parameters. The confidence intervals are obtained from our asymptotic theory. In all cases, the estimated H is much less than 0.5, ranging between .0946 for DJIA to .2550 for Nasdaq 100. The point estimates of H are similar to those used by Gatheral et al. (2018). The 95% confidence intervals suggest that we have strong evidence against the null hypothesis of $H = 0.5$. Hence, each RV series is better modeled by RFSV. This finding once again supports the results found by Bayer et al. (2016), Gatheral et al. (2018), and Livieri et al. (2018). We also find clear evidence against the model used by Comte and Renault (1998), where it is assumed that $H \in (1/2, 1)$.

Model (1.1) is a Gaussian process that does not allow for jumps in volatility. To check

---

[7]The data are obtained from https://realized.oxford-man.ox.ac.uk/.
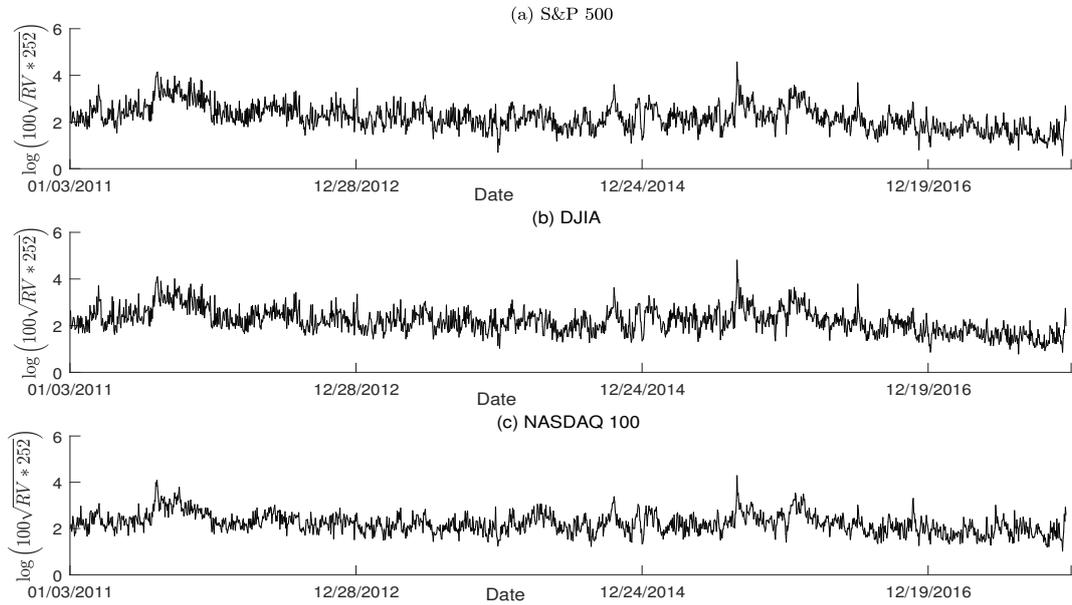
Figure 7: Time series plot of log(RV) for S&P 500, DJIA, Nasdaq 100

Table 8: Empirical results for logarithmic RV of S&P 500, DJIA, Nasdaq 100

Name	H	σ	μ	κ
S&P 500	.1453	.8440	2.1960	1.3810
	(.0738, .2166)	(.8331, .8549)	(1.9665, 2.4253)	(1.2829, 1.4790)
DJIA	.0946	.6788	2.2019	.2382
	(.0200, .1672)	(.6698, .6877)	(1.2318, 3.1718)	(.1946, .2816)
Nasdaq 100	.2550	1.2849	2.2220	14.8874
	(.1861, .3238)	(1.2688, 1.3008)	(2.1819, 2.2621)	(14.5993, 15.1754)

Table 9: Empirical results for logarithmic RV of S&P 500, DJIA, NASDAQ 100 with 7 largest observations removed.

Name	H	σ	μ	κ
S&P 500	.1518	.8706	2.1883	1.9407
	(.08041, .2232)	(.8594, .8818)	(2.0171, 2.3594)	(1.8251, 2.0562)
DJIA	.0965	.6807	2.1942	.3269
	(.0238, .1692)	(.6717, .6897)	(1.4803, 2.9080)	(.2760, .3778)
Nasdaq 100	.2577	1.2973	2.2152	16.9296
	(.1887, .3266)	(1.2811, 1.3135)	(2.1793, 2.2511)	(16.6225, 17.2367)

Table 10: Empirical results for logarithmic RV of S&P 500, DJIA, NASDAQ 100 with 14 largest observations removed.

Name	H	σ	μ	κ
S&P 500	.1592	.9009	2.1817	2.5822
	(.0878, .2305)	(.8893, .9126)	(2.0463, 2.3172)	(2.4498, 2.7146)
DJIA	.1061	.7089	2.1877	.6106
	(.0335, .1787)	(.6996, .7182)	(1.7807, 2.5947)	(.5418, .6793)
Nasdaq 100	.2536	1.2675	2.2101	17.3052
	(.1843, .3227)	(1.2516, 1.2833)	(2.1759, 2.2441)	(16.9929, 17.6173)

the robustness of our empirical results against the potential jumps in volatility over the period, we remove the 7 and 14 largest observations from each series. These choices of jump intensity correspond to 1 jump and 2 jumps per annum and are empirically reasonable. Tables 9 and 10 report three sets of estimation results, including the point estimates and 95% confidence intervals for all four parameters, with the jumps removed. The empirical results are very similar to those reported in Table 8, which suggests that the results, including the estimated H, are robust to jumps.

6.2 RV for exchange rates

We now fit Model (1.1) to three well-known RV series for spot exchange rates for the U.S. dollar, the Deutsch Mark, and the Japanese yen over the period from December 1, 1986 to June 30, 1999. The same data are used by Andersen et al. (2003). Figure 8 plots three time series, each of which with 3045 observations. The roughness in the data is once again obvious from the plot. Because our model is closely related to the
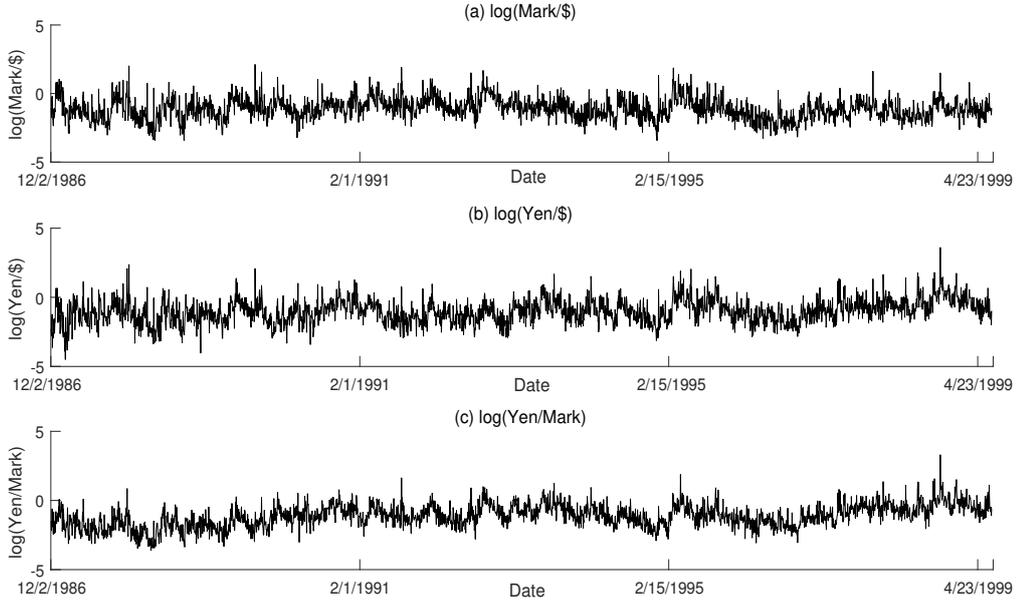
Figure 8: Time series plot of log(RV) for the U.S. dollar, the Deutschmark and the Japanese Yen over the period from December 1, 1986 to June 30, 1999

discrete time ARFIMA model, which has very good out-of-sample performance among alternative univariate models, as shown in Andersen et al. (2003), we focus our attention on the forecasting performance between the proposed model and the ARFIMA model.

To forecast future logarithmic RV, following Gatheral et al. (2018), we assume that $\kappa = 0$ in Model (1.1). When H is known, the h-step-ahead forecasting formula based on the history $(-\infty, t]$ in a continuous record, derived by Nuzman and Poor (2000), is given by[8]

$$\mathbb{E}\left(X_{t+h}|\mathcal{F}_t\right) = \frac{\cos\left(H\pi\right)}{\pi} h^{H+1/2} \int_{-\infty}^{t} \frac{X_s}{(t-s+h)(t-s)^{H+1/2}} ds. \qquad (6.2)$$

When H is unknown, we replace it with the estimate $\widehat{H}$ defined in (3.1). It has been noted in Remark 4.4 that the estimate of H is consistent even when $\kappa = 0$.

---

[8]When a truncated discrete record is available at $s = 1, ..., t$, we must modify the forecasting formula to

$$\mathbb{E}\left(X_{t+h}|\mathcal{F}_t\right) = \frac{\cos\left(H\pi\right)}{\pi} h^{H+1/2} \frac{\sum_{s=1}^{t} \frac{X_s}{(t-s+1+h)(t-s+1)^{H+1/2}}}{\sum_{s=1}^{t} (t-s+1+h)^{-1}(t-s+1)^{-H+1/2}}. \qquad (6.1)$$

To forecast with the following ARFIMA$(1, d, 0)$ model,

$$X_t = \mu + \rho \left(X_{t-1} - \mu\right) + \varepsilon_t, \ (1 - L)^d \varepsilon_t = e_t, \ e_t \sim i.i.d.\mathcal{N}(0, \sigma_e^2), \ d \in (-1/2, 1/2), \ (6.3)$$

if μ, ρ, and d are known, one may use the forecasting formula given by Hosking (1981), as follows

$$\mathbb{E}\left(X_{t+h}|\mathcal{F}_t\right) = -\sum_{s=1}^{\infty} \pi_s \widehat{X}_{t+h-s},$$

where

$$\pi_s = \frac{(s - d - 2)!}{(s - 1)!(-d - 1)!} \left\{1 - \rho - \frac{(1 + d)}{s}\right\}, \quad (6.4)$$

and $\widehat{X}_{t+h-s} = \mathbb{E}\left(X_{t+h-s}|\mathcal{F}_t\right)$ when $h - s > 0$ and $\widehat{X}_{t+h-s} = X_{t+h-s}$ when $h - s \leq 0$.[9] When μ, ρ, and d are unknown, we estimate them with the ML method.

Following Andersen et al. (2003), we initially divide the entire sample period into two periods. The first period is from December 1, 1986 to December 1, 1996, which contains 2449 observations, and the second period is from December 2, 1996 to June 30, 1999, which contains 596 days for out-of-sample evaluation. On each day in the second period, 1-day-ahead and h-day-ahead (with $h = 2$, 5, 10) forecasts are obtained. Each forecast is based on the MLE of μ, ρ, and d for the ARFIMA model and the proposed estimate of H for Model (1.1). The estimates are obtained on the basis of an expanding window as more observations become available. The last date for which a 10-day-ahead forecast is carried out is June 20, 1999. The data from between December 1, 1986 and June 20, 1999 are used to obtain estimates in both models. The last date for which a 1-day-ahead forecast is carried out is June 29, 1999. The data from between December 1, 1986 and June 29, 1999 are used to obtain the estimates in both models.

To gain an idea about the point estimates in both models, Table 11 reports the estimated H in Model (1.1) and the estimated μ, ρ, and d in the ARFIMA model based on the full sample of each volatility series. In all cases, the estimated value of d is very close to 0.401, which is the value used by Andersen et al. (2003) to forecast RV. The estimated value of H is close to those obtained in the first empirical study and very close to 0.14, which is the value used by Gatheral et al. (2018).

---

[9] When implementing the forecasting formula, if $s \geq 100$, noting that $\frac{(s-d-2)!}{(s-1)!(-d-1)!} \left\{1 - \rho - \frac{(1+d)}{s}\right\} \sim \frac{(1-\rho)}{(-d-1)!} s^{-d-1}$, we simply use $\frac{(1-\rho)}{(-d-1)!} s^{-d-1}$ to approximate π_s.

Table 11: Estimate of H for the fractional continuous time model and estimate of μ, ρ and d for the ARFIMA model

	Model (1.1)	ARFIMA		
	H	μ	ρ	d
Mark/$	0.1543	-1.1357	-0.0630	0.4155
Mark/Yen	0.1307	-1.1184	-0.1008	0.4663
Yen/$	0.1711	-1.0279	-0.0379	0.4330

Table 12: One-step-ahead forecasting results in the Mark/$ RV

	intercept	slope of ARFIMA	slope of our model	R^2
ARFIMA	-0.0608	0.9335	-	0.2801
Our model	-0.2395	-	0.7926	0.2855
Joint	-0.2916	-0.2516	1.003	0.2857

To compare the out-of-sample performances of these two competing models, following Mincer and Zarnowitz (1969) and Andersen et al. (2003), we project the logarithmic RV on a constant and the two model forecasts (first individually and then jointly). Table 12 reports the results for the 1-day-ahead forecasts for Mark/$, including the estimated intercept, the estimated slopes, and R^2 in each regression. We find the results very interesting. First, in single forecasting regressions, our model generates a slightly higher value for R^2 than the ARFIMA model. Note that the ARFIMA model can generate very accurate volatility forecasts according to Andersen et al. (2003). The joint forecasting regression indicates that the ARFIMA forecasts have a negative relationship with the true logarithmic RV after controlling for the forecasts obtained from our model.

Table 13 reports the results for multiple-day-ahead forecasts. To save space, we report R^2 only in single forecasting regressions. In all cases, our model generates higher values for R^2 than the ARFIMA model.

Table 14 reports the R^2 results in single forecasting regressions for the other two exchange rates for 1-day-ahead up to 10-day-ahead forecasts. The general conclusions about the superiority of the fractional continuous time model hold true, except for the

Table 13: Multi-step-ahead forecasting results in the Mark/$ RV

	2-day	5-day	10-day
ARFIMA	.4364	.4792	.5041
Our model	.4640	.5159	.5452

Table 14: Forecasting results in the Mark/Yen RV and in the Yen/$ RV

		1-day	2-day	5-day	10-day
Mark/Yen	ARFIMA	.4219	.5284	.5581	.5661
Mark/Yen	Our model	.4253	.5434	.5821	.5929
Yen/$	ARFIMA	.3704	.5046	.5340	.5355
Yen/$	Our model	.3685	.5273	.5707	.5740

Table 15: Estimate of H for the fractional continuous time model and estimate of μ, ρ and d for the ARFIMA model with 14 and 28 largest observations removed

	Model (1.1)	ARFIMA		
	H	μ	ρ	d
Mark/$ with 14 observations removed	0.1532	-1.1464	-0.0689	0.4165
Mark/$ with 28 observations removed	0.1585	-1.1557	-0.0652	0.4138
Mark/Yen with 14 observations removed	0.1400	-1.1352	-0.1006	0.4636
Mark/Yen with 14 observations removed	0.1391	-1.1459	-0.1027	0.4610
Yen/$ with 14 observations removed	0.1713	-1.0432	-0.0437	0.4323
Yen/$ with 14 observations removed	0.1690	-1.0540	-0.0448	0.4297

1-day-ahead forecast of Yen/$ RV.

Model (1.1) is a Gaussian process that does not allow for jumps. To check the robustness of our empirical results against the potential jumps over the period, we remove the 14 and 28 largest observations from each series. Table 15 reports three sets of estimation results, including the point estimates and 95% confidence intervals for all four parameters. The empirical results are very similar to those reported in Tables 11, which suggests that the results, especially the estimated H, are robust to jumps.

To gain a deeper understanding of why the continuous time model and the ARFIMA model show differences in performance in the out-of-sample exercise, Figure 9 compares the values of the weight functions used in these two models for one-step-ahead forecast (i.e., $h = 1$). The weights used for the ARFIMA model are $\{-\pi_s\}$ given in (6.4), and s denotes the lag length. To facilitate the comparison, we rewrite the forecasting formula used by the continuous time model as

$$E\left(X_{t+h}|\mathcal{F}_t\right) = \sum_{s=1}^{t} \theta_s X_{t+1-s} \quad \text{with } \theta_s = \frac{\cos\left(H\pi\right)}{(s+h)(s)^{H+1/2}\pi} h^{H+1/2} / \sum_{s=1}^{t}(s+h)^{-1}(s)^{-H+1/2}.$$

Let $t = 3045$, which is the sample size of the Mark/$ RV data. Let $H = 0.1543$, which corresponds to the estimated H in our proposed model based on the full sample of

31
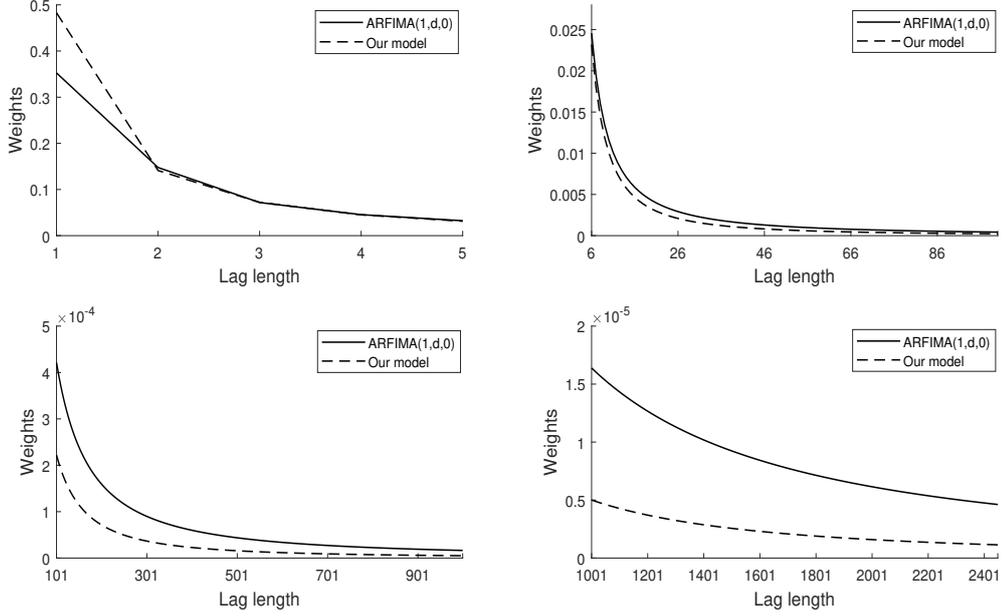
Figure 9: Weights of ARFIMA$(1, d, 0)$ and our model for forecasting Mark/$

Mark/$ RV. Let $\rho = -0.063$ and $d = 0.4155$, which correspond to the ML estimates of ρ and d in the stationary ARFIMA model based on the full sample of Mark/$ RV. Figure 9 reports the values of $\{-\pi_s\}$ and $\{\theta_s\}$ for $s = 1, ..., 2401$. It shows that $\{-\pi_s\}$ and $\{\theta_s\}$ are positive and that they decay to zero at the same rate. More importantly, it can be seen that $\theta_1 > -\pi_1$ and $\theta_s < -\pi_s$ for $s > 1$, which means that, relative to the ARFIMA model, our model assigns a higher weight to the most recent observation and lower weights to earlier observations.

7 Conclusions

Over the past two decades, the general consensus is that the volatility of financial assets displays long-range dependence. In the continuous-time setting, long-range dependence can be modeled with the help of fBm. Gatheral et al. (2018) show that the logarithmic RV behaves essentially as an fBm, and the Hurst parameter takes a value of around 0.1 at any reasonable time scale. Using at-the-money options on the S&P 500 index with short maturity, Livieri et al. (2018) further confirm that volatility is rough.

This study contributes to the literature by proposing a novel estimation method for all parameters in a fractional continuous-time model based on discrete-sampled observations when the parameter space for the Hurst parameter is $(0, 1)$. In the first stage, the Hurst parameter is estimated based on the ratio of two second-order differences of observations obtained at various time scales. 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 developed the asymptotic theory for the proposed estimators that facilitates statistical inference.

Simulations suggest that our two-stage estimators perform well in finite samples. The method is applied to two empirical examples, the logarithmic RV of the S&P 500, DJIA, and Nasdaq 100 and the logarithmic RV of Mark/$, Mark/Yen, and Yen/$ exchange rates. Empirical studies show that the volatility is rough, which reinforces the findings of Gatheral et al. (2018) in all six series. This empirical finding is robust to jumps. We also compare the out-of-sample forecasting performance of the fractional continuous-time model with the stationary ARFIMA model and find evidence of the superior performance of the fractional continuous-time model.

This study also suggests several important directions for future research. First, although our estimators are consistent and easy to use, they may not be asymptotically efficient. The development of an asymptotically more efficient estimation technique and determination of the asymptotic relative inefficiency of our two-stage estimators hold great interest. Second, the model considered in this paper has no jumps, even though the proposed model with $H < 1/2$ is rough. Although the removal of a few jumps from the data cannot change the feature of roughness, jumps may have implications for the magnitude of parameter estimates. Extending the estimate method and asymptotic theory to cover fractional continuous-time models with jumps is important, and we leave it for future studies. Third, this paper assumes that the Hurst parameter does not change over time. This assumption can be too restrictive. How to test whether the value of H changes in the sample and how to model time-varying H values are some important questions to ask. Fourth, in this paper we fit the fractional continuous time model to RV series. By doing so, we assume that the RV measures integrated volatility without measurement error. This assumption is clearly too strong. The degree of robustness of the empirical results to measurement errors in RV will be explored in a future study. In addition, the robustness of the roughness in other time-series data such

as interest rates is pursued by Phillips et al. (2019). Finally, our results indicate that the fractional continuous time model and the discrete time ARFIMA model may not be asymptotically equivalent, even though the discrete time ARFIMA model converges weakly to the fractional continuous time model under the in-fill asymptotic scheme. Such an asymptotic non-equivalence should be established.

APPENDIX

Lemma 7.1 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 $E(y_i) = 0$ and $Var(y_i) = 4 - 2^{2H}$, and has autocorrelation functions as, for $j = 0, 1, 2, \ldots$,

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

(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 $E(y_{i,*}) = 0$ and $Var(y_{i,*}) = 2^{2H}(4 - 2^{2H})$, 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 $\xi_{i,*} = y_{i,*}^2 - E\left(y_{i,*}^2\right)$ and $\xi_i = y_i^2 - E\left(y_i^2\right)$, for $i = 0, 1, 2, \ldots$. The bivariate process $\left\{ \begin{pmatrix} \xi_{i,*} & \xi_i \end{pmatrix}' \right\}$ is a weakly stationary process with mean zero and autocovariance matrices as, for $j = 0, 1, 2, \ldots$,

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

Lemma 7.2 Let $B^H(t)$ be the same fBm as in Lemma 7.1 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

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

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

It can be proved that, as $n \to \infty$,

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

where Γ_j are the covariance matrices defined in Lemma 7.1, and the long-run covariance matrix in the limiting distribution is well-defined.

Proof of Lemma 7.1: The results in Parts (a)-(b) are obtained straightforwardly based on the definition of fBm and its covariance structure given in (1.2). Details are tedious and omitted here for simplicity.

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

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

where the third equality comes form the Issei' theorem (Isserlis, 1918) for computing higher-order moments of 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 $E\left(\xi_{i+j,*}\xi_{i,*}\right) = 2\left(4 - 2^{2H}\right)^2 2^{4H} \rho_{j,*}^2$.

We now derive the expressions of $E\left(\xi_{i+j}\xi_{i,*}\right)$ and $E\left(\xi_{i+j,*}\xi_i\right)$ and show that they only depends on j, not i. For any $i = 0, 1, 2, \ldots,$ it can be seen that

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

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 the Isserlis' theorem (Isserlis, 1918) again, we have

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

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

In Remark 4.1, we have proved that $\rho_j \sim O\left(j^{2H-4}\right)$ as $j \to \infty$. Therefore, the sequence of covariance matrices $\{\Gamma_j\}$ 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 7.2: From the self-similarity property of fBm, it can be obtained that, for any $i = 0, 1, 2, \ldots$,

$$\frac{B_{(i+4)\Delta}^H - 2B_{(i+2)\Delta}^H + B_{i\Delta}^H}{\Delta^H} \stackrel{d}{=} B^H\left(i+4\right) - 2B^H\left(i+2\right) + B^H\left(i\right)$$

and

$$\frac{B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H}{\Delta^H} \stackrel{d}{=} B^H\left(i+2\right) - 2B^H\left(i+1\right) + B^H\left(i\right).$$

As a result, we have $\left\{\left(\eta_{i,*} \quad \eta_i\right)'\right\}_{i=1}^{n-4} \stackrel{d}{=} \left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}_{i=1}^{n-4}$, where $\left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}$ is the weakly stationary bivariate process defined in Part (c) of Lemma 7.1. Then, applying the conventional central limit theorem for stationary vector process to the process $\left\{\left(\xi_{i,*} \quad \xi_i\right)'\right\}$ gives the asymptotic normal distribution reported in the theorem.

Proof of Theorem 4.1: (a) From Equation (2.2), we have, as $\Delta \to 0$,

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

and

$$X_{(i+2)\Delta} - 2X_{(i+1)\Delta} + X_{i\Delta}$$

$$= \left(X_{(i+2)\Delta} - X_{(i+1)\Delta}\right) - \left(X_{(i+1)\Delta} - X_{i\Delta}\right)$$

$$= \left(e^{-\kappa\Delta} - 1\right)\left(X_{(i+1)\Delta} - X_{i\Delta}\right) + \sigma \left(\int_{(i+1)\Delta}^{(i+2)\Delta} e^{-\kappa[(i+2)\Delta - s]} dB_s^H - \int_{i\Delta}^{(i+1)\Delta} e^{-\kappa[(i+1)\Delta - s]} dB_s^H\right)$$

$$= O_p\left(\Delta^{1+H}\right) + \sigma \left(\int_{(i+1)\Delta}^{(i+2)\Delta} \{1 + O\left(\Delta\right)\} dB_s^H - \int_{i\Delta}^{(i+1)\Delta} \{1 + O\left(\Delta\right)\} dB_s^H\right)$$

$$= \sigma \left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right) + O_p\left(\Delta^{1+H}\right).$$

Therefore, by using the results in Lemma 7.2, we have, as long as $\Delta \to 0$,

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

$$= \frac{1}{n\Delta^{2H}} \sum_{i=1}^{n-2} \left\{\left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right)^2 + O_p\left(\Delta^{1+2H}\right)\right\}$$

$$= \frac{1}{n} \sum_{i=1}^{n-2} \left(\frac{B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H}{\Delta^H}\right)^2 + \frac{1}{n} \sum_{i=1}^{n-2} O_p\left(\Delta\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n-2} \left[\eta_i + \left(4 - 2^{2H}\right)\right] + O_p\left(\Delta\right) \xrightarrow{p} 4 - 2^{2H}, \tag{A.1}$$

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

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

$$= \frac{1}{\sqrt{n}\Delta^{2H}} \sum_{i=1}^{n-2} \left\{\left(B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H\right)^2 - \left(4 - 2^{2H}\right)\Delta^{2H} + O_p\left(\Delta^{1+2H}\right)\right\}$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-2} \left[\left(\frac{B_{(i+2)\Delta}^H - 2B_{(i+1)\Delta}^H + B_{i\Delta}^H}{\Delta^H}\right)^2 - \left(4 - 2^{2H}\right)\right] + \frac{1}{n} \sum_{i=1}^{n-2} O_p\left(\sqrt{T\Delta}\right)$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-2} \eta_i + o_p\left(1\right) \xrightarrow{d} \mathcal{N}\left(0, 2\left(4 - 2^{2H}\right)^2 \left(\rho_0^2 + 2\sum_{j=1}^{\infty} \rho_j^2\right)\right), \tag{A.2}$$

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 7.2 again, we have, as long as $\Delta \to 0$,

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

$$= \frac{1}{n} \sum_{i=1}^{n-4} \left[\eta_{i,*} + 2^{2H}\left(4 - 2^{2H}\right)\right] + O_p(\Delta) \xrightarrow{p} 2^{2H}\left(4 - 2^{2H}\right), \qquad (A.3)$$

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

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

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n-4} \eta_{i,*} + o_p(1) \xrightarrow{d} \mathcal{N}\left(0, 2^{1+4H}\left(4 - 2^{2H}\right)^2\left(\rho_{0,*}^2 + 2\sum_{j=1}^{\infty}\rho_{j,*}^2\right)\right), \qquad (A.4)$$

the asymptotic variance in which has an identical representation as

$$2^{1+4H}\left(4 - 2^{2H}\right)^2\left(\rho_{0,*}^2 + 2\sum_{j=1}^{\infty}\rho_{j,*}^2\right)$$

$$= 2^{1+4H}\left(4 - 2^{2H}\right)^2\left(1 + 2^{1-4H}\sum_{j=1}^{\infty}\left[\left(\rho_{j+2} + 4\rho_{j+1} + 6\rho_j + 4\rho_{|j-1|} + \rho_{|j-2|}\right)^2\right]\right)$$

$$= 2^{4H}\left(4 - 2^{2H}\right)^2 \Sigma_{11},$$

where the first equation comes from the relationship between $\rho_{j,*}$ and ρ_j given in Lemma 7.1, and Σ_{11} is defined in (4.3).

Then, based on (A.3) and (A.1), 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 7.2,

$$
\lim_{n\to\infty} Cov\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n-4}\eta_{i,*}, \frac{1}{\sqrt{n}}\sum_{i=1}^{n-2}\eta_i\right)
$$

$$
= 2\left(4-2^{2H}\right)^2\left(\left(\rho_2+2\rho_1+\rho_0\right)^2 + \sum_{j=1}^{\infty}\left[\left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2 + \left(\rho_j+2\rho_{|j-1|}+\rho_{|j-2|}\right)^2\right]\right)
$$

$$
= 2\left(4-2^{2H}\right)^2\left(4\left(\rho_0+\rho_1\right)^2 + 2\sum_{j=0}^{\infty}\left(\rho_{j+2}+2\rho_{j+1}+\rho_j\right)^2\right) = 2^{2H}\left(4-2^{2H}\right)^2\Sigma_{12},
$$

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

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

where Σ_{12} is defined as in (4.4). Then, together with the results given in (A.1), (A.2) and (A.4), the asymptotic distribution of $2^{2\widehat{H}}-2^{2H}$ is obtained as $\Delta\to 0$ and $T\Delta\to 0$:

$$
\sqrt{n}\left(2^{2\widehat{H}}-2^{2H}\right)
$$

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

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

$$
\overset{d}{\to} \frac{\mathcal{N}\left(0, 2^{4H}\left(4-2^{2H}\right)^2\left[\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}\right]\right)}{4-2^{2H}} \overset{d}{=} \mathcal{N}\left(0, 2^{4H}\left[\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}\right]\right).
$$

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 $T\Delta\to 0$,

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

The proof is completed.

(b) Based on the result that $\sqrt{n}\left(\widehat{H}-H\right)=O_p\left(1\right)$ as $\Delta\to0$ and $T\Delta\to0$, we have

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

$$=\exp\left\{2\sqrt{n}\left(\widehat{H}-H\right)\frac{2\sqrt{\Delta}\log\left(\sqrt{\Delta}\right)}{\sqrt{T}}\right\}\overset{p}{\to}1,$$

where the last limit is due to $\sqrt{\Delta}\log\left(\sqrt{\Delta}\right)\to0$ as $\Delta\to0$. Together with the limiting result given in (A.1), the consistency of $\widehat{\sigma}^2$ is obtained under the condition of $\Delta\to0$ and $T\Delta\to0$:

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

To derive the asymptotic distribution of $\widehat{\sigma}^2$, we first prove that, as $\Delta\to0$ and $T\Delta\to0$,

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

and

$$\frac{\sqrt{n}}{\log\left(\Delta\right)}\left(\Delta^{2\widehat{H}-2H}-1\right)=2\sqrt{n}\left(\widehat{H}-H\right)+o_p\left(1\right)\overset{d}{\to}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{\left\{\log\left(2\right)\right\}^2}\right). \quad\text{(A.5)}$$

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

$$\widehat{\sigma}^2-\sigma^2$$

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

$$=\frac{O_p\left(1\right)}{O_p\left(\sqrt{n}\right)}+\frac{\left(n-2\right)\left(4-2^{2H}\right)\sigma^2}{n\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}-\sigma^2$$

$$=\frac{\sigma^2}{\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}\left\{\frac{n-2}{n}\left(4-2^{2H}\right)-\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}\right\}+O_p\left(\frac{1}{\sqrt{n}}\right)$$

$$=\frac{\sigma^2}{\left(4-2^{2\widehat{H}}\right)\Delta^{2\widehat{H}-2H}}\left\{\left(2^{2\widehat{H}}-2^{2H}\right)-\left(4-2^{2\widehat{H}}\right)\left(\Delta^{2\widehat{H}-2H}-1\right)+O\left(\frac{1}{n}\right)\right\}+O_p\left(\frac{1}{\sqrt{n}}\right),$$

40

where the second equation is from the result in (A.1). 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

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

where the last limit comes from the asymptotic result proved in (A.5).

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

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

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

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

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

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

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 $\Delta \to 0$,

$$\frac{1}{n} \sum_{i=1}^{n} X_{i\Delta}^2 = \frac{1}{T} \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} X_{i\Delta}^2 dt + \frac{X_T^2 - X_0^2}{n}$$

$$= \frac{1}{T} \sum_{i=0}^{n-1} \int_{i\Delta}^{(i+1)\Delta} \left(X_t^2 + O_p \left(\Delta^H \right) \right) dt + \frac{X_T^2 - X_0^2}{n}$$

$$= \frac{1}{T} \int_0^T X_t^2 dt + O_p \left(\Delta^H \right) + O_p \left(1/n \right)$$

$$\xrightarrow{p} E \left(X_t^2 \right)$$

$$= \sigma^2 \kappa^{-2H} H \Gamma \left(2H \right) + \mu^2,$$

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

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

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

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

and

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

where $\Gamma \left(\cdot \right)$, $\exp \left\{ \cdot \right\}$, and $\log \left\{ \cdot \right\}$ are continuous functions. The consistency of $\widehat{\kappa}$ is proved.

To derive the asymptotic distribution of $\sqrt{T} \left(\widehat{\kappa} - \kappa \right)$ 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

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

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

$$- \kappa^{-2H} \left\{ \widehat{\sigma}^2 \widehat{H} \Gamma \left(2\widehat{H} \right) - \sigma^2 H \Gamma \left(2H \right) \right\}. \tag{A.6}$$

42

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

$$
\begin{aligned}
&\widehat{\sigma}^2 \widehat{H} \Gamma\left(2\widehat{H}\right) - \sigma^2 H \Gamma\left(2H\right) \\
&= \left(\widehat{\sigma}^2 - \sigma^2\right) \widehat{H} \Gamma\left(2\widehat{H}\right) + \sigma^2 \left(\widehat{H} - H\right) \Gamma\left(2\widehat{H}\right) - \sigma^2 H \left[\Gamma\left(2\widehat{H}\right) - \Gamma\left(2H\right)\right] \\
&= O_p\left(\frac{\log\left(\Delta\right)}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right) + O_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned} \tag{A.7}
$$

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

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

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

Define

$$
\widehat{\kappa}_{HN} = \left(\frac{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2}{\sigma^2 H \Gamma\left(2H\right)}\right)^{-1/(2H)}. \tag{A.8}
$$

Theorem 3.3 of Xiao and Yu (2019a) and Theorem 3.1 of Xiao and Yu (2019b) have proved, for the cases where $H \in [1/2, 3/4)$ and $H \in (0, 1/2)$ respectively, that, as $T \to \infty$,

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

where ϕ_H is defined as in Theorem 4.4 in the current paper. As a result, we have

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

and, as $T \to \infty$,

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

43

Then, for the first term in (A.6), it is obtained that, as $T \to \infty$ and $\sqrt{T}\Delta^H \to 0$,

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

$$= \sqrt{T}\left\{\frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2 - \sigma^2\kappa^{-2H}H\Gamma\left(2H\right) + O_p\left(\Delta^H\right) + O_p\left(\frac{1}{n}\right)\right\}$$

$$\xrightarrow{d} \sigma^2 H\Gamma\left(2H\right)\cdot\left(-2H\kappa^{-2H-1}\right)\cdot\mathcal{N}\left(0,\kappa\phi_H\right). \tag{A.9}$$

Now, putting (A.7) and (A.9) in Equation (A.6), we have, as $T \to \infty$, $T\Delta \to 0$, and $\sqrt{T}\Delta^H \to 0$,

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

and

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

Note that the first-order Taylor expansion of $\widehat{\kappa}^{-2\widehat{H}}$ at the point $\widehat{\kappa} = \kappa$ 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 the point $\widehat{H} = H$. Finally, we have, as $T \to \infty$, $T\Delta \to 0$, and $\sqrt{T}\Delta^H \to 0$,

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

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.8), Xiao and Yu (2019a) have proved that, when $H = 3/4$,

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

44

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

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

where R denotes the Rosenblatt random variable. 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.

References

Aït-Sahalia, Y. and Mancini, T. S. (2008). Out of sample forecasts of quadratic variation. *Journal of Econometrics*, 147(1):17–33.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2): 579–625.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73(1):5–59.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1): 3–30.

Barndorff-Nielsen, O. E., Corcuera, J. M., and 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*, volume 33, pages 69–96. Springer Science & Business Media.

Bayer, C., Friz, P., and Gatheral, J. (2016). Pricing under rough volatility. *Quantitative Finance*, 16(6):887–904.

Bennedsen, M., Lunde, A., and Pakkanen, M.S., (2017). Decoupling the short- and long-term behavior of stochastic volatility. Working Paper, CREATES, Aarhus University.

Bergstrom, A.R., (1990). Continuous time econometric modelling. Oxford University Press.

Bollerslev, T., Patton, A. J., and Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting, *Journal of Econometrics*, 192(1): 1–18.

Cheridito, P., Kawaguchi, H., Maejima, M. (2003). Fractional Ornstein-Uhlenbeck processes. *Electronic Journal of Probability*, 8(3): 1-14.

Cheung, Y. W. (1993). Long memory in foreign-exchange rates. *Journal of Business and Economic Statistics*, 11(1):93–101.

Cheung, Y.-W., Diebold, F. X. (1994). On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean. *Journal of Econometrics*, 62(2): 301–316.

Coeurjolly, J. (2000). Simulation and identification of the fractional Brownian motion: a bibliographical and comparative study. *Journal of statistical software*, 5(7):1–53.

Comte, F., Coutin, L. and Renault, E. (2012). Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378.

Comte, F. and Renault, E. (1996). Long memory continuous time models. *Journal of Econometrics*, 73(1):101–149.

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2): 174–196.

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1): 83–106.

Elerian, O., Chib, S. and N. Shephard, Likelihood Inference for Discretely Observed Nonlinear Diffusions, *Econometrica*, 69(4), 959–993.

Euch, O. E. and Rosenbaum, M. (2018). Perfect hedging in rough Heston models. *Annals of Applied Probability*, 28(6): 3813–3856.

Fouque, J. P. and Hu, R. (2018). Optimal portfolio under fast mean–reverting fractional stochastic environment. *SIAM Journal on Financial Mathematics*, 9(2):564–601.

Garnier, J. and Sølna, K. (2017). Correction to Black–Scholes formula due to fractional stochastic volatility. *SIAM Journal on Financial Mathematics*, 8(1):560–588.

Gatheral, J., Jaisson, T. and Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance*, 18(6):933–949.

Giraitis L., Koul H., Surgailis D. (2012). *Large Sample Inference for Long memory Processes*, Imperial College Press.

Granger, C. W. J. and Joyeux, R. (1980). An introduction to long–memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1): 15–29.

Hosking, J. (1981). Fractional differencing. *Biometrika*, 68(1): 165–176.

Hu, Y. and Nualart, D. (2010). Parameter estimation for fractional Ornstein–Uhlenbeck processes. *Statistics and Probability Letters*, 80(11):1030–1038.

Hu, Y., Nualart, D., and Zhou, H. (2019). Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. *Statistical Inference for Stochastic Processes*, 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(1/2): 134-139.

Jaisson, T., Rosenbaum, M. (2016). Rough fractional diffusions as scaling limits of nearly unstable heavy tailed Hawkes processes. *Annals of Applied Probability*, 26(5): 2860-2882.

Kleptsyna, M. and Le Breton, A. (2002). Statistical analysis of the fractional Ornstein–Uhlenbeck type process. *Statistical Inference for stochastic processes*, 5(3):229–248.

Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin.

Le Cam, L. and Yang, G. (1990). *Asymptotics in Statistics. Some Basic Concepts*. Springer, New York.

Livieri, G., Mouti, S., Pallavicini, A. and Rosenbaum, M. (2018). Rough volatility: evidence from option prices. *IISE Transactions*, 50(9):767–776.

Lo, A. (1991). Long-term memory in stock market prices. *Econometrica*, 59(5):1279–313.

Mincer, J. A., Zarnowitz, V. (1969). The evaluation of economic forecasts. *In Economic Forecasts and Expectations, ed. by J. Mincer. New York: National Bureau of Economic Research*, 3–46.

Nuzman, C. J., Poor, H. V. (2000). Linear estimation of self-similar processes via Lamperti's transformation. *Journal of Applied Probability*, 37(2): 429–452.

Paxson, V. (1997). Fast, approximate synthesis of fractional gaussian noise for generating self-similar network traffic. *ACM SIGCOMM Computer Communication Review*, 27(5):5–18.

Phillips, P. C. B., Wang, X., Xiao, W. and Yu, J. (2019). Nonparametric estimation of fractional continuous-time model for interest rates. Work in progress, Singapore Management University

47

Phillips, P. C. B. and Yu, J. (2005). Jackknifing bond option prices. *Review of Financial Studies*, 18(2): 707–742.

Phillips, P. C. B. and Yu, J. (2009a). Simulation-based estimation of contingent–claims prices. *Review of Financial Studies*, 22(9): 3669–3705.

Phillips, P. C. B. and Yu, J. (2009b). A two-stage realized volatility approach to estimation of diffusion processes with discrete data. *Journal of Econometrics*, 150(2):139–150.

Tanaka, K. (2013). Distributions of the maximum likelihood and minimum contrast estimators associated with the fractional Ornstein–Uhlenbeck process. *Statistical Inference for Stochastic Processes*, 16(3):173–192.

Tanaka, K. (2015). Maximum likelihood estimation for the non-ergodic fractional Ornstein–Uhlenbeck process. *Statistical Inference for Stochastic Processes*, 18(3):315–332.

Tanaka, K., Xiao, W., Yu, J. (2019). Maximum likelihood estimation for the fractional Vasicek model. *SMU Economics and Statistics Working Paper Series, Paper No. 08-2019.*

Tang, C. Y., Chen, S. X. (2009). Parameter estimation and bias correction for diffusion processes. *Journal of Econometrics*, 149(1):65–81.

Taylor, S. J. (1986). *Modelling financial time series.* John Wiley & Sons, New York.

Tudor, C. and Viens, F. (2007). Statistical aspects of the fractional stochastic calculus. *Annals of Statistics*, 35(3):1183–1212.

Xiao, W. and Yu, J. (2019a). Asymptotic theory for estimating the drift parameters in the fractional Vasicek model. *Econometric Theory*, 35(1): 198–231.

Xiao, W. and Yu, J. (2019b). Asymptotic theory for rough fractional Vasicek models. *Economics Letters*, 177: 26-29.

Wang, Y. (2002). Asymptotic nonequivalence of GARCH models and diffusions, *Annals of Statistics*, 30(3), 754–783.

Wang, X., Phillips, P. C. B., Yu, J. (2011). Bias in estimating multivariate and univariate diffusions. *Journal of Econometrics*, 161(2): 228–245.

Wang, X., Yu, J. (2016). Double asymptotics for explosive continuous time models. *Journal of Econometrics*, 193(1): 35–53.

Yu, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics*, 169(1):114–122.