Estimation and Inference in Fractional Ornstein–Uhlenbeck Model with Discrete-Sampled Data*

Weilin Xiao†, Xiaohu Wang‡, Jun Yu§

October 9, 2018

Abstract

This paper proposes a two-stage method for estimating parameters in the fractional Ornstein–Uhlenbeck model based on discrete-sampled observations. In the first stage, two "diffusion" parameters (i.e. the Hurst parameter and the volatility parameter) are estimated based on the second order differences obtained at two different time scales. Their asymptotic theory is established under an in-fill asymptotic scheme. In the second stage, the two drift parameters are estimated based on the ergodic theorem. Their asymptotic theory is established under a double asymptotic scheme. Extensive simulation studies show that the proposed estimators perform well. Two empirical studies are carried out. The first empirical study, based on the daily VIX data over 2004-2017, shows that VIX is rough. The Hurst parameter is slightly but statistically significantly less than one half. It also shows that the persistence parameter is much larger than one divided by the time span, suggesting that a commonly adopted assumption in the literature can be too strong. The second empirical study, based on the daily realized volatility of S&P 500, DJIA and Nasdaq over 2011-2017, shows that the Hurst parameter is much less than one half and the persistence parameter is significantly larger than one divided by the time span.

JEL Classification: C15, C22, C32.

Keywords: Fractional Brownian motion; Second order differences; Stationary process; Ergodic estimator; Rough volatility; In-fill asymptotics; Double asymptotics

*We gratefully thank Peter C. B. Phillips, Xiaohong Chen, Graham Elliott, and participants at various conferences and seminars for useful comments.

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

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

§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

Over the last few decades, the phenomenon of long-range dependence has been found in data from hydrology, geophysics, climatology and telecommunication, economics and finance. Consequently, several time series models have been proposed to capture long-range dependence, both in discrete time and in continuous time. In economics and finance, a partial list of references include Granger and Joyeux (1980), Lo (1991), Ding et al. (1993), Cheung (1993), Baillie (1996), Baillie et al. (1996), Andersen et al. (2003) in the domain of discrete time and Comte and Renault (1996, 1998), Aït-Sahalia and Mancini (2008), Comte et al. (2012) in the domain of continuous time. Comte and Renault (1996) established the connection between the two classes of models. Typically in continuous time, to capture long-range dependence, the standard Brownian motion (W_t) is replaced with a fractional Brownian motion (B_t^H, fBm hereafter) where $H \in (0, 1)$ is the Hurst parameter. When H is greater than one half, B_t^H generates long-range dependence. Some surveys on fBm can be found in Biagini et al. (2008), Mishura (2008) and Kubilius et al. (2017).

When H is greater than or equal to one half, the smoothness of the sample path of B_t^H is the same as that of W_t. Namely both B_t^H and W_t are $1/2 - \epsilon$-Hölder continuous, for any $\epsilon > 0$. However, in a seminar study, Gatheral et al. (2018) shows that the logarithm of realized volatility is too rough for B_t^H with $H \geq 1/2$. Instead Gatheral et al. (2018) proposes to model volatility using the fractional Ornstein–Uhlenbeck (fOU hereafter) process driven by B_t^H with $H < 1/2$ and called it the rough volatility model. To see the difference between the smoothness of B_t^H with different values of H, Figure 1 plots three simulated paths from B_t^H, with $H = 0.5, 0.7, 0.15$. Also plotted is the logarithm of daily realized volatility of S&P 500. It can be seen clearly that the sample path of B_t^H with $H = 0.15$ is rougher than those with $H = 0.5, 0.7$. The real data also exhibits the feature of roughness.

Moreover, Gatheral et al. (2018) shows that when the persistence parameter (κ) in fOU is much smaller than one divided by the time span, fOU behaves locally as a fBm. Based on this assumption, Gatheral et al. (2018) documents the evidence that the logarithm of realized volatility behaves essentially as a fOU with $H \approx 0.1$ at any reasonable time scale. Moreover, Livieri et al. (2018) finds the strong support to fOU with $H < 1/2$ using implied volatility-based approximations of spot volatility while Bayer et al. (2016) found the strong support to fOU with $H < 1/2$ using the SPX volatility surface and variance swaps. Gatheral et al. (2018) and Bennedsen, et al (2017) reports the superior forecasting performance of fOU with $H < 1/2$ relative to the heterogeneous autoregressive (HAR, hereafter) model of Corsi (2009). Not surprisingly, rough volatility models have found in many applications, including option pricing (Bayer et al. (2016) and Garnier and Sølna (2017)), portfolio choice (Fouque and Hu, 2018), and dynamic hedging (Euch and Rosenbaum, 2017).

Consequently, parameter estimation of fOU has been of great interest in theory and in applied work. In particular, under the assumption of a known H and a known long term mean (set to zero), tremendous efforts have been made to estimate the persistence

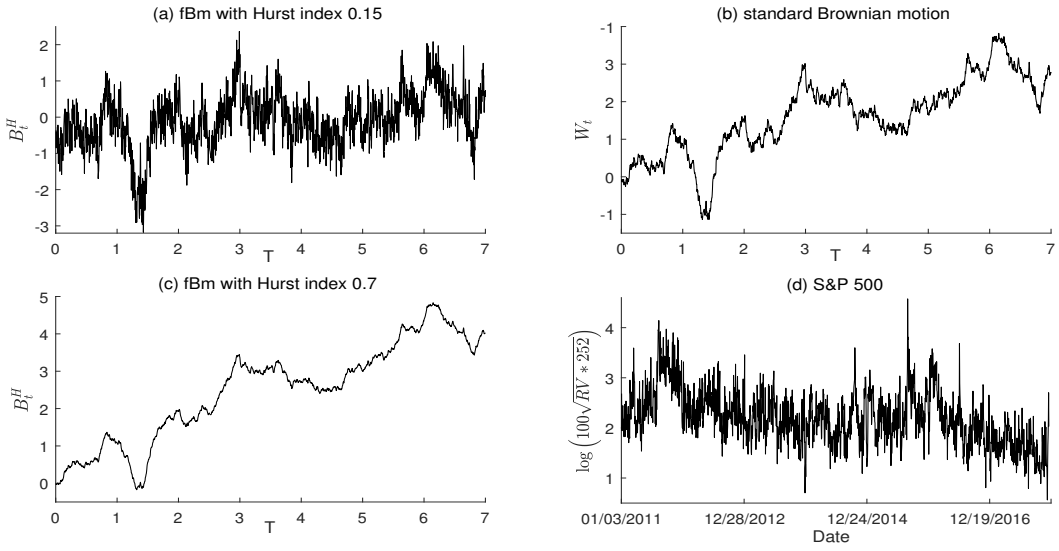

Figure 1: Plot of fBm with different Hurst parameter and the time series of log(RV) for S&P 500

parameter (κ) in fOU and to develop the asymptotic theory when a continuous record over an increasing time span is available. Alternative estimation methods include maximum likelihood estimation (MLE hereafter) (Kleptsyna and Le Breton (2002), Tudor and Viens (2007), Tanaka (2014)), least squares estimation (LSE hereafter) (Hu and Nualart (2010), Hu et al. (2018), Xiao and Yu (2018a)), minimum contrast estimation (Tanaka, 2013), and ergodic-type estimation (Hu and Nualart (2010), Xiao and Yu (2018a)). However, in practice, it is only possible to have discrete-sampled observations rather than a continuous record of observations. With discrete-sampled data, Tudor and Viens (2007) studies MLE of κ and develops its in-fill asymptotic distribution. Es-Sebaiy (2013) and Kubilius et al. (2015) study LSE of κ and develops its double asymptotic distribution. Brouste and Iacus (2013), Kubilius et al. (2015), Hu et al. (2018) study the ergodic-type estimator of κ and develops its double asymptotic distribution. Comte and Renault (1996) proposes to use the log-periodogram regression to estimate H and then estimate the other parameters by first obtaining the differentiated series based on the estimated H and second using the Bergstrom-type estimation technique.

In practice, all parameters including H are unknown and have to be estimated. Moreover, given the discussions above, it is desirable to develop an estimation method and asymptotic theory for a general value for H as H can be greater than one half or less than one half in practice. Furthermore, some important assumptions made in the literature are testable hypotheses. One example is the assumption about the value of H that H is either greater than one half or less than one half. To test the hypothesis that $H = 1/2$ (either left-sided or right-sided or two-sided), an asymptotic theory for H is needed. Another example is related to the assumption about κ that κ is much smaller than one divided by

the time span. To test such a hypothesis, an asymptotic theory for κ is needed.

In this paper, based on discrete-sampled observations, we estimate all four parameters in fOU for a complete range of Hurst parameter, namely $H \in (0,1)$. In particular, a two-stage approach is developed for parameter estimation. In the first stage, the two "diffusion" parameters (i.e. the Hurst parameter and the volatility parameter) are estimated based on the second order differences at two different time scales. Their asymptotic theory for these two parameters is established under an in-fill asymptotic scheme. In particular, the proposed asymptotic theory for H covers all $H \in (0,1)$. In the second stage, the two drift parameters are estimated based on the ergodic theorem and the plug-in method. Their asymptotic theory is established under a double asymptotic scheme. The asymptotic theory for κ critically depends on regimes for H.

The remainder of the paper is organized as follows. In Section 2, we introduce the model and our two-stage approach for parameter estimation. Asymptotic theory of the proposed estimators is developed in Section 3. Section 4 conducts Monte Carlo studies to check finite sample properties of the proposed estimators. Section 5 carries out two empirical studies using real data. Some important hypotheses are also tested in this section. Section 6 gives concluding remarks and discusses some open questions. The technical proofs are given in the Appendix.

We use the following notations throughout the paper: $\xrightarrow{p}$, $\xrightarrow{a.s.}$, $\xrightarrow{\mathcal{L}}$ and $\sim$ denote convergence in probability, convergence almost surely, convergence in distribution and asymptotic equivalence, respectively. For any matrix A, we denote its transpose by A' and its (i,j) component by A_{ij}. Moreover, let n, δ and T be the sample size, the sampling interval, and the time span of the data, respectively. In particular, we assume that $T/\delta = n$ is an integer and fOU is observed at the following discrete points in time, $t_k = k\delta$ with $k = 0, 1, 2, \ldots, n$. Obviously $T = n\delta$ is the time span. Throughout the paper, we denote a generic constant by C, which maybe context-dependent.

2 The Model and Estimators

The fOU model that we are interested in is given by the following stochastic differential equation:

$$dX_t = \kappa(\mu - X_t)\,dt + \sigma dB_t^H, \tag{2.1}$$

where $\sigma \in \mathbb{R}^+$, $\mu \in \mathbb{R}$, $\kappa \in \mathbb{R}^+$, the initial condition is set at X_0, and B_t^H, a fBm with Hurst parameter $H \in (0,1)$, is a zero mean Gaussian process with covariance

$$\mathbb{E}\left(B_t^H B_s^H\right) = R_H(s,t) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right). \tag{2.2}$$

The process B_t^H becomes the standard Brownian motion W_t when $H = 1/2$. It is negatively correlated when $0 < H < 1/2$. When $1/2 < H < 1$, it has long-range dependence in the sense that $\sum_{n=1}^{\infty} \mathbb{E}\left(B_1^H(B_{n+1}^H - B_n^H)\right) = \infty$. In this case, the positive (negative) increments are likely to be followed by positive (negative) increments. The exact discrete-time

representation of (2.1) is given by

$$X_{l\delta} = \mu + e^{-\kappa\delta}\left(X_{(l-1)\delta} - \mu\right) + \sigma \int_{(l-1)\delta}^{l\delta} e^{-\kappa(k\delta - s)} dB_s^H. \tag{2.3}$$

When $\kappa > 0$, X_t is asymptotically stationary and ergodic.

In Model (2.1), there are four parameters, two "diffusion" parameters, H and σ, and two drift parameters, κ and μ.[1] Parameter κ determines the persistence in X_t. Since we assume $\kappa > 0$, X_t reverts to its long-run mean μ at the speed determined by κ. That is why sometimes κ is called the mean reversion parameter. Because of the ongoing debate about empirically relevant values for H in the volatility literature, in this paper, we set the parameter space for H to be $(0, 1)$ and use discrete-sampled observations to estimate all four parameters, H, σ, μ and κ.

In particular, we employ a two-stage estimation procedure in the same spirit as that in Phillips and Yu (2009b). However, in each stage the estimation method we will use is different from that in Phillips and Yu (2009b). In the first stage, we estimate H and σ based on the method of Barndorff-Nielsen et al. (2013). Consistent estimation of these two parameters does not require any knowledge about κ and μ. The key insight for this property is that as $\delta \to 0$, the order of magnitude of the "diffusion" term is larger than that of the drift term. Therefore, the development of asymptotic properties of estimators of H and σ only requires $\delta \to 0$. This asymptotic scheme is termed the in-fill scheme. In the second stage, following Hu and Nualart (2010) and Xiao and Yu (2018a, 2018b), we estimate κ and μ based on the ergodic theorem. As typically found in the literature, the development of asymptotic properties for drift parameters requires $T \to \infty$ as well as consistent estimation of "diffusion" parameters. Hence, a double asymptotic scheme is needed to obtain consistency and asymptotic theory for κ and μ.

2.1 Estimating the "diffusion" parameters

To explain the estimation procedure and to establish the asymptotic theory for the "diffusion" parameters, in the first stage, we make use of an important result derived in Barndorff-Nielsen et al. (2013) for the second order differences at two different time scales. According to Barndorff-Nielsen et al. (2013), for any fix T and any κ, μ and σ, as $\delta \to 0$,

$$\frac{\sum_{i=4}^{T/\delta} |X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}|^2}{\sum_{i=2}^{T/\delta} |X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}|^2} \xrightarrow{p} 2^{2H}.$$

[1] Strictly speaking, fOU is not a diffusion process as it does not have the Markov property.

Motivated by this result, we propose the following estimator for H

$$\hat{H} = \frac{1}{2}\log_2\left(\frac{\sum_{i=4}^{T/\delta}\left|X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}\right|^2}{\sum_{i=2}^{T/\delta}\left|X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}\right|^2}\right), \tag{2.4}$$

where $\log_2$ is the base-2 logarithm.

Based on the estimator of H, we can obtain a plug-in estimator for σ using the quadratic variation of the second order difference, i.e.,

$$\hat{\sigma} = \sqrt{\frac{\delta}{T\tau}\sum_{i=2}^{T/\delta}\left|X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}\right|^2}, \tag{2.5}$$

where $\tau = 4\delta^{2\hat{H}} - (2\delta)^{2\hat{H}}$.

Remark 2.1. *Our estimator of H is not only related to Barndorff-Nielsen et al. (2013) but also to Barndorff-Nielsen and Podolskij (2009). Barndorff-Nielsen and Podolskij (2009); Barndorff-Nielsen et al. (2013) also proposed estimators of the volatility parameter for Brownian semimartingale and Brownian semi-stationary processes using power, bipower, or multipower variations. In this paper, motivated by Podolskij and Wasmuth (2013), we extend the method to fOU, which lacks the semimartingale property.*

Remark 2.2. *The estimators of the Hurst coefficient and the volatility parameter can be easily extended to a general stochastic differential equation*

$$dX_t = \mu\left(t, X_t\right)dt + \sigma dG_t\,,$$

where G_t is a Gaussian process with stationary and centered increments with local behaviour being the same as that of B_t^H (see Podolskij and Wasmuth (2013) for detailed discussions about G_t) and $\mu\left(t, X_t\right)$ may not be a linear function. Under some mild conditions for $\mu\left(t, X_t\right)$, we can prove consistency of $\hat{H}$ and $\hat{\sigma}$.

Remark 2.3. *Corcuera et al. (2006) considered the asymptotic behavior of the realized power variation for the first order difference of fractional integral processes when $H \in (0, 3/4]$. The present paper removes the restriction and uses the quadratic variation of the second order difference. In fact, we conjecture that we can use the p^{th} ($p \geq 2$) power variation (or multipower variation) of the m^{th} ($m \geq 2$) order difference to estimate H for $H \in (0, 1)$. It is unknown if there is any efficiency gain in using a different order power variation or a different order difference.*

2.2 Estimating the drift parameters

We now turn to the problem of estimating the drift parameters in fOU. We first argue why it is difficult to use MLE and LSE before we propose the ergodic-type estimators.

Perhaps the leading method for estimating the drift parameters is MLE. However, the non-Markov property of fOU makes calculations of the likelihood function difficult. This is because the transition density based on discrete-sampled observations depends on the entire history. Consequently, obtaining analytical expression for MLE is not possible. Moreover, Kleptsyna and Le Breton (2002) claimed that replacing the stochastic integrals in the continuous-record MLE by their corresponding Riemann sums is problematic. To illustrate this difficulty, considering $H > 1/2$ and following the idea of Kleptsyna and Le Breton (2002), we can obtain the MLE of μ and κ based on a continuous record:

$$\begin{aligned}
\hat{\mu}_{MLE} &= \frac{\int_0^T P_H(t)dS_t \int_0^T P_H(t)dw_t^H - S_T \int_0^T (P_H(t))^2 dw_t^H}{w_T^H \int_0^T P_H(t)\, dS_t - S_T \int_0^T P_H(t)\, dw_t^H}\,, \\
\hat{\kappa}_{MLE} &= \frac{w_T^H \int_0^T P_H(t)\, dS_t - S_T \int_0^T P_H(t)\, dw_t^H}{\left(\int_0^T P_H(t)\, dw_t^H\right)^2 - w_T^H \int_0^T (P_H(t))^2\, dw_t^H}\,,
\end{aligned}$$

where

$$\begin{aligned}
&\omega_t^H = \lambda_H^{-1} t^{2-2H}\,, && \lambda_H = \frac{2H\Gamma(3-2H)\Gamma(H+\frac{1}{2})}{\Gamma(\frac{3}{2}-H)}\,, \\
&P_H(t) = \tfrac{1}{\sigma}\tfrac{d}{d\omega_t^H}\int_0^t \kappa_H(t,s)X_s ds\,, && S_t = \frac{1}{\sigma}\int_0^t \kappa_H(t,s)dX_s\,, \\
&\kappa_H(t,s) = \kappa_H^{-1} s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}\,, && \kappa_H = 2H\Gamma(\frac{3}{2}-H)\Gamma(H+\frac{1}{2})\,,
\end{aligned}$$

and $\Gamma(\cdot)$ denotes the gamma function.

Unfortunately, the process P_H depends continuously on X. The discrete-sampled observations on X do not allow one to obtain directly the discrete-sampled observations on P_H. Moreover, note that both $\hat{\mu}_{MLE}$ and $\hat{\kappa}_{MLE}$ relies on S, which is also not observable. Consequently, approximations to the continuous-record MLE are numerically challenging, if not impossible.

Similarly, for $H > 1/2$, the continuous-record LSE of μ and κ can be explicitly represented as in Xiao and Yu (2018a)

$$\begin{aligned}
\bar{\mu}_{LSE} &= \frac{(X_T - X_0)\int_0^T X_t^2 dt - \int_0^T X_t dt \left(\frac{X_T^2}{2} - \frac{X_0^2}{2} - \alpha_H \sigma^2 \int_0^T \int_0^t s^{2H-2} e^{-\kappa s} ds dt\right)}{(X_T - X_0)\int_0^T X_t dt - T\left(\frac{X_T^2}{2} - \frac{X_0^2}{2} - \alpha_H \sigma^2 \int_0^T \int_0^t s^{2H-2} e^{-\kappa s} ds dt\right)}\,, \\
\bar{\kappa}_{LSE} &= \frac{(X_T - X_0)\int_0^T X_t dt - T\left(\frac{X_T^2}{2} - \frac{X_0^2}{2} - \alpha_H \sigma^2 \int_0^T \int_0^t s^{2H-2} e^{-\kappa s} ds dt\right)}{T\int_0^T X_t^2 dt - \left(\int_0^T X_t dt\right)^2}\,,
\end{aligned}$$

where $\alpha_H = H(2H-1)$. However, $\bar{\kappa}_{LSE}$ depends on κ which is the parameter we wish to estimate. We can approximate $\bar{\mu}_{LSE}$ and $\bar{\kappa}_{LSE}$ by

$$\hat{\mu}_{LSE} = \frac{(X_{n\delta}-X_0)\sum_{l=0}^{n} X_{l\delta}^2\cdot\delta - \sum_{l=0}^{n} X_{l\delta}\cdot\delta\cdot\sum_{l=1}^{n}\left[X_{(l-1)\delta}\left(X_{l\delta}-X_{(l-1)\delta}\right)-J\left(\hat{\kappa}_{LSE}\right)\right]}{(X_{n\delta}-X_0)\sum_{l=0}^{n} X_{l\delta}\cdot\delta - n\delta\cdot\sum_{l=1}^{n}\left[X_{(l-1)\delta}\left(X_{l\delta}-X_{(l-1)\delta}\right)-J\left(\hat{\kappa}_{LSE}\right)\right]},$$

$$\hat{\kappa}_{LSE} = \frac{(X_{n\delta}-X_0)\sum_{l=0}^{n} X_{l\delta}\cdot\delta - n\delta\cdot\sum_{l=1}^{n}\left[X_{(l-1)\delta}\left(X_{l\delta}-X_{(l-1)\delta}\right)-J\left(\hat{\kappa}_{LSE}\right)\right]}{n\delta\sum_{l=0}^{n} X_{l\delta}^2\cdot\delta-\left(\sum_{l=0}^{n} X_{l\delta}\cdot\delta\right)^2},$$

with $J\left(\hat{\kappa}_{LSE}\right)=\hat{H}\left(2\hat{H}-1\right)\hat{\sigma}^2\int_{(l-1)\delta}^{l\delta}\int_0^{(l-1)\delta} e^{-\hat{\kappa}_{LSE}((l-1)\delta-s)}\left(r-s\right)^{2\hat{H}-2}drds$. Obviously, to calculate $\hat{\mu}_{LSE}$ and $\hat{\kappa}_{LSE}$, numerical methods are needed, which are cumbersome.

Motivated by Hu and Nualart (2010), Hu et al. (2018), Xiao and Yu (2018a, 2018b), the present paper considers the ergodic-type estimators. Since $\kappa>0$, X_t is stationary and ergodic. Hence, we can use functions of $\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}$ and $\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2$ to estimate κ and μ. More precisely, based on the consistent estimators of H (defined by (2.4)) and σ (defined by (2.5)), we introduce the following ergodic-type estimators for μ and κ based on discrete-sampled observations of $X_{l\delta}$, $l=0,1,\ldots,n$:

$$\hat{\mu}_s = \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\,, \tag{2.6}$$

$$\hat{\kappa}_s = \left(\frac{n\sum_{l=0}^{n} X_{l\delta}^2-\left(\sum_{l=0}^{n} X_{l\delta}\right)^2}{n^2\hat{\sigma}^2\hat{H}\Gamma\left(2\hat{H}\right)}\right)^{-\frac{1}{2\hat{H}}}. \tag{2.7}$$

Comparing them with the ergodic-type estimators of μ and κ based on a continuous record proposed in Xiao and Yu (2018a), our new ergodic-type estimators replace integrals with the corresponding Riemann sums and plug the estimated H and σ into the expressions.

3 Asymptotic Theory

3.1 The "diffusion" parameters

This subsection develops asymptotic theory for $\hat{H}$ and $\hat{\sigma}$ under the in-fill asymptotic scheme. The following two theorems establish consistency and the asymptotic laws for $\hat{H}$ and $\hat{\sigma}$, respectively.

Theorem 3.1. *Let $H \in (0,1)$ and T be fixed. As $\delta \to 0$, we have*

$$\hat{H} \xrightarrow{p} H\,, \tag{3.1}$$

$$\hat{\sigma} \xrightarrow{p} \sigma\,. \tag{3.2}$$

Theorem 3.2. *Let $H \in (0,1)$ and T be fixed. As $\delta \to 0$, we have*

$$\frac{\sigma_H}{\sqrt{\delta}}\left(\hat{H} - H\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)\,, \tag{3.3}$$

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma} - \sigma\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{4}\Gamma_{22}\right)\,, \tag{3.4}$$

where

$$\sigma_H = \frac{\sqrt{\delta \log(64) F_\delta^n} \sum_{i=2}^{T/\delta-2} \left|X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}\right|^2}{\sqrt{(1, -F_\delta^n)\,\Gamma\,(1, -F_\delta^n)' \sum_{i=2}^{T/\delta-2} \left|X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}\right|^4}}\,,$$

$$F_\delta^n = \frac{\sum_{i=4}^{T/\delta} |X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}|^2}{\sum_{i=2}^{T/\delta} |X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}|^2}\,,$$

and $\Gamma = (\Gamma_{ij})_{1\le i,j\le 2}$ are given by

$$\Gamma_{11} = 2 + 2^{2-4H} \sum_{k=1}^{\infty} \left|\rho_{k+2} - 4\rho_{k+1} + 6\rho_k - 4\rho_{|k-1|} + \rho_{|k-2|}\right|^2\,,$$

$$\Gamma_{12} = \Gamma_{21} = 2^{1-2H}(\rho_1 - 1) + 2^{2-2H} \sum_{k=0}^{\infty} \left|\rho_{k+2} - 2\rho_{k+1} + \rho_k\right|^2\,,$$

$$\Gamma_{22} = 2 + 4 \sum_{k=1}^{\infty} \rho_k^2\,,$$

with

$$\rho_k = \frac{1}{2\left(4 - 2^{2H}\right)}\left[-(k+2)^{2H} + 4(k+1)^{2H} - 6k^{2H} + 4|k-1|^{2H} - |k-2|^{2H}\right]\,.$$

Remark 3.1. *Using the formula $(1+u)^\alpha = 1 + \sum_{k=1}^{\infty} \frac{\alpha(\alpha-1)\times\cdots\times(\alpha-k+1)}{k!}\, u^k$ for $-1 < u < 1$,*

we can rewrite ρ_k *for any* $k \geq 3$ *as*

$$\begin{aligned}
\rho_k &= \frac{1}{(4-2^{2H})}\frac{1}{2}\left[-|k-2|^{2H}+4\,|k-1|^{2H}-6k^{2H}+4\,(k+1)^{2H}-(k+2)^{2H}\right] \\
&= \frac{1}{(4-2^{2H})}\frac{k^{2H}}{2}\left[-\left(1-\frac{2}{k}\right)^{2H}+4\left(1-\frac{1}{k}\right)^{2H}-6+4\left(1+\frac{1}{k}\right)^{2H}-\left(1+\frac{2}{k}\right)^{2H}\right] \\
&= \frac{1}{(4-2^{2H})}\frac{k^{2H}}{2}\sum_{m=1}^{\infty}\frac{2H(2H-1)\times\cdots\times(2H-m+1)}{m!}\left(-(-2)^m+4(-1)^m+4-2^m\right)k^{-m} \\
&= \frac{1}{(4-2^{2H})}k^{2H}\sum_{m=1}^{\infty}\frac{2H(2H-1)\times\cdots\times(2H-2m+1)}{(2m)!}\left(4-4^m\right)k^{-2m}\,.
\end{aligned}$$

Note that the sign of $2H(2H-1)\times\cdots\times(2H-2m+1)$ *is the same as that of* $2H-1$ *and*

$$\begin{aligned}
|2H(2H-1)\times\cdots\times(2H-2m+1)| &= 2H|2H-1|(2-2H)\times\cdots\times(2m-1-2H) \\
&\leq 2(2m-1)!\,.
\end{aligned}$$

Using the inequality $(1-H)\log 4 < 4-2^{2H}$*, we have, for any* $k \geq 3$*,*

$$\begin{aligned}
|\rho_k| &\leq \frac{k^{2H}}{4-2^{2H}}\sum_{m=1}^{\infty}\frac{2H|2H-1|(2-2H)\times\cdots\times(2m-1-2H)}{(2m)!}\left(4^m-4\right)k^{-2m} \\
&\leq \frac{k^{2H}}{\log 4}\sum_{m=2}^{\infty}\frac{4^m-4}{m}\left(\frac{1}{k^2}\right)^m = \frac{k^{2H}}{\log 4}\left(4\log\left(1-\frac{1}{k^2}\right)-\log\left(1-\frac{4}{k^2}\right)\right) \\
&\leq \frac{243}{20\log 4}\,k^{2H-4}\,.
\end{aligned}$$

For the last inequality we have used $\log(1-u) = -\sum_{k=1}^{\infty}\frac{u^k}{k}$ *if* $0 \leq u < 1$ *and* $4\log(1-u)-\log(1-4u) \leq \frac{243}{20}u^2$ *if* $0 \leq u \leq \frac{1}{6}$*. Moreover, a standard calculation implies*

$$\rho_1 = \left(7-3^{2H}+4\times 2^{2H}\right)/\left(2\left(4-2^{2H}\right)\right),$$

and

$$\rho_2 = \left(4-4^{2H}+4\times 3^{2H}-6\times 2^{2H}\right)/\left(2\left(4-2^{2H}\right)\right).$$

Based on the discussion above, for any $k \geq 1$*, we have* $|\rho_k| \sim k^{2H-4}$*, which implies* $\sum_{k=1}^{\infty}\rho_k^2 < \infty$*. Figure 2 plots* $\sum_{k=1}^{\infty}\rho_k^2$ *as a function of* H*, which decreases monotonically in* H*. So the asymptotic variance of* $\hat{\sigma}$ *decreases as* H *increases.*

Remark 3.2. *From Barndorff-Nielsen et al. (2013), we know that* $X_{i\delta}-2X_{(i-1)\delta}+X_{(i-2)\delta} = O_p(\delta^H)$*. Consequently,*

$$\begin{aligned}
X_{i\delta}-2X_{(i-2)\delta}+X_{(i-4)\delta} &= \left(X_{i\delta}-2X_{(i-1)\delta}+X_{(i-2)\delta}\right)+2\left(X_{(i-1)\delta}-2X_{(i-2)\delta}+X_{(i-3)\delta}\right) \\
&\quad +\left(X_{(i-2)\delta}-2X_{(i-3)\delta}+X_{(i-4)\delta}\right) \\
&= O_p(\delta^H).
\end{aligned}$$

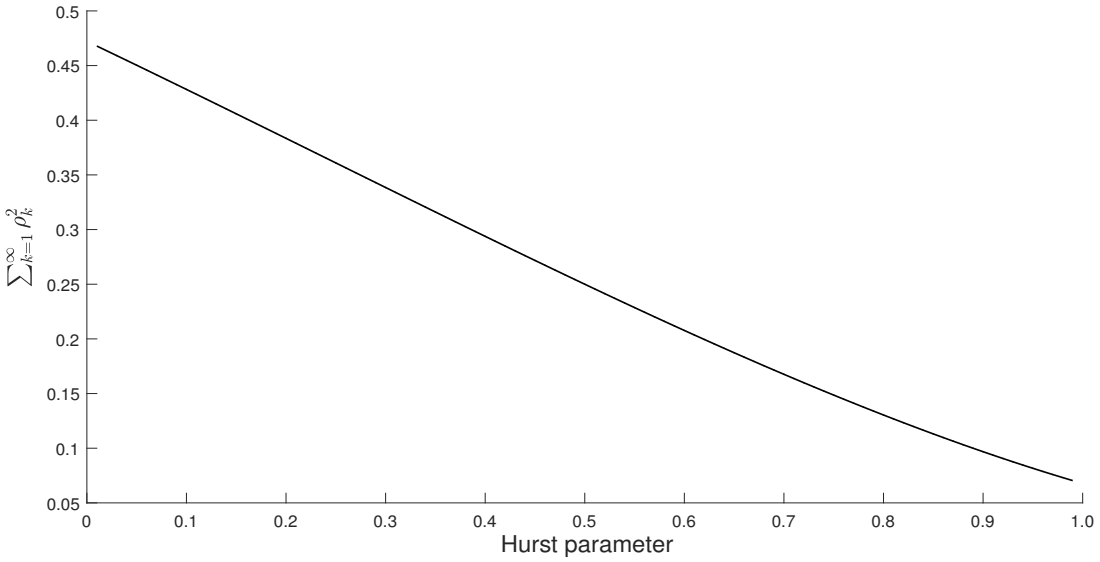

Figure 2: Plot of $\sum_{k=1}^{\infty} \rho_k^2$ as a function of H

Remark 3.3. *From the results in Remark 3.1 and Remark 3.2, we obtain*

$$\sigma_H \sim \frac{\sqrt{\delta}\sqrt{\frac{\frac{T}{\delta}O_p(\delta^{2H})}{\frac{T}{\delta}O_p(\delta^{2H})}\frac{T}{\delta}O_p(\delta^{2H})}}{\sqrt{\left(\frac{\frac{T}{\delta}O_p(\delta^{2H})}{\frac{T}{\delta}O_p(\delta^{2H})}\right)^2\frac{T}{\delta}O_p(\delta^{4H})}} \sim \sqrt{T}.$$

Hence, the convergence rate of $\hat{H}$ is $\sqrt{T/\delta}$.

Remark 3.4. *The development of asymptotic theory for the "diffusion" parameters only requires $\delta \to 0$, a feature shared by the two-stage method of Phillips and Yu (2009b). This is because as $\delta \to 0$, the "diffusion" term in both models dominates the drift term, making the drift term irrelevant in the development of asymptotic theory for the "diffusion" parameters. Since our model is driven by a fBm but the model of Phillips and Yu (2009b) is driven by a standard Brownian motion, the two sets of asymptotic distributions are different.*

Remark 3.5. *When $H = 1/2$, we can easily obtain $\rho_0 = 1$, $\rho_1 = -1/2$ and $\rho_k = 0$ for $k \geq 2$. Consequently, a standard calculation yields $\Gamma_{11} = 135/2$, $\Gamma_{12} = \Gamma_{21} = 7$ and $\Gamma_{22} = 3$. The limit distribution in (3.4) becomes $\frac{\sqrt{T}}{\sqrt{\delta}}(\hat{\sigma} - \sigma) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{3\sigma^2}{4}\right)$ which is the standard result.*

3.2 The drift parameters

We now consider the asymptotic theory for $\hat{\mu}_s$ and $\hat{\kappa}_s$ under the double asymptotic scheme. We first deal with consistency of $\hat{\mu}_s$ and $\hat{\kappa}_s$.

Theorem 3.3. *Let $H \in (0,1)$ and $X_0 = o_p\left(\sqrt{n\delta}\right)$. Assume that $\delta \to 0$ and $T \to \infty$. Then we have*

$$\hat{\mu}_s \xrightarrow{p} \mu\,, \tag{3.5}$$

$$\hat{\kappa}_s \xrightarrow{p} \kappa\,. \tag{3.6}$$

To develop the asymptotic distributions for $\hat{\mu}_s$ and $\hat{\kappa}_s$, we first state the following technical lemma for the ergodic estimators of μ and κ based on a continuous record of observations. We omit the proof of the lemma as it follows directly from Xiao and Yu (2018a, 2018b).

Lemma 3.1. *Let $\hat{\mu}_{TS}$ and $\hat{\kappa}_{TS}$ be the ergodic estimators of μ and κ based on a continuous record of observations over $[0,T]$, that is,*

$$\hat{\mu}_{TS} = \frac{1}{T}\int_0^T X_t dt\,, \tag{3.7}$$

$$\hat{\kappa}_{TS} = \left(\frac{T\int_0^T X_t^2 dt - \left(\int_0^T X_t dt\right)^2}{T^2\sigma^2 H\Gamma(2H)}\right)^{-\frac{1}{2H}}. \tag{3.8}$$

Assuming H and σ are known, then for $H \in (0,1)$, $X_0 = o_p(T^H)$, we have, as $T \to \infty$,

$$T^{1-H}\left(\hat{\mu}_{TS} - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{\kappa^2}\right). \tag{3.9}$$

Moreover, as $T \to \infty$. Then the following convergence results hold true.

(i) *If $H \in (0, 3/4)$ and $X_0 = o_p(\sqrt{T})$, then*

$$\sqrt{T}\left(\hat{\kappa}_{TS} - \kappa\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \kappa\phi_H\right), \tag{3.10}$$

where

$$\phi_H = \begin{cases} \frac{1}{4H^2}\left[(4H-1) + \frac{2\Gamma(2-4H)\Gamma(4H)}{\Gamma(2H)\Gamma(1-2H)}\right], & \text{when } H \in (0, \frac{1}{2}), \\ \frac{4H-1}{4H^2}\left[1 + \frac{\Gamma(3-4H)\Gamma(4H-1)}{\Gamma(2-2H)\Gamma(2H)}\right], & \text{when } H \in [\frac{1}{2}, \frac{3}{4}). \end{cases} \tag{3.11}$$

(ii) *If $H = 3/4$ and $X_0 = o_p\left(\sqrt{T\log(T)}\right)$, then*

$$\frac{\sqrt{T}}{\log(T)}\left(\hat{\kappa}_{TS} - \kappa\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{16\kappa}{9\pi}\right).$$

(iii) *If $H \in (3/4, 1)$ and $X_0 = o_p(T^{2H-1})$, then*

$$T^{2-2H}\left(\hat{\kappa}_{TS} - \kappa\right) \xrightarrow{\mathcal{L}} \frac{-\kappa^{2H-1}}{H\Gamma(2H+1)} R\,,$$

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

$$c(s) = \exp\left(\frac{1}{2}\sum_{k=2}^{\infty} \left(2is\sigma(H)\right)^k \frac{a_k}{k}\right), \tag{3.12}$$

with $i = \sqrt{-1}$, $\sigma(H) = \sqrt{H(H-1/2)}$ and

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

The following theorem derives the asymptotic laws for $\hat{\mu}_s$ and $\hat{\kappa}_s$ which mimic those of $\hat{\mu}_{TS}$ and $\hat{\kappa}_{TS}$.

Theorem 3.4. *Assume that $X_0 = o_p(T^H)$, $\delta \to 0$, $T \to \infty$ and $n\delta^p \to 0$ for some $p \in (1, 2 \wedge (1 + \frac{H}{1-H}))$. Then, we have*

$$T^{1-H}\left(\hat{\mu}_s - \mu\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\sigma^2}{\kappa^2}\right). \tag{3.13}$$

Moreover, as $n \to \infty$, the following convergence results hold true.

(i) *Let $X_0 = o_p(\sqrt{T})$ and when $H \in (0, 3/4)$, $n\delta^p \to 0$ for some $p \in \left(1, \frac{3+2H}{1+2H} \wedge (1+2H)\right)$, we have*

$$\sqrt{T}\left(\hat{\kappa}_s - \kappa\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \kappa\phi_H\right), \tag{3.14}$$

where ϕ_H is defined by (3.11).

(ii) *Let $X_0 = o_p\left(\sqrt{T\log(T)}\right)$ and when $H = 3/4$, $\frac{n\delta^p}{\log(T)} \to 0$ for some $p \in \left(1, \frac{9}{5}\right)$, we have*

$$\frac{\sqrt{T}}{\log(T)}\left(\hat{\kappa}_s - \kappa\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{16\kappa}{9\pi}\right).$$

(iii) *Let $X_0 = o_p\left(T^{2H-1}\right)$ and when $H \in (3/4, 1)$, $n\delta^p \to 0$ for some $p \in \left(1, \frac{3-H}{2-H}\right)$, we have*

$$T^{2-2H}\left(\hat{\kappa}_s - \kappa\right) \xrightarrow{\mathcal{L}} \frac{-\kappa^{2H-1}}{H\Gamma(2H+1)} R\,,$$

where R is the Rosenblatt random variable.

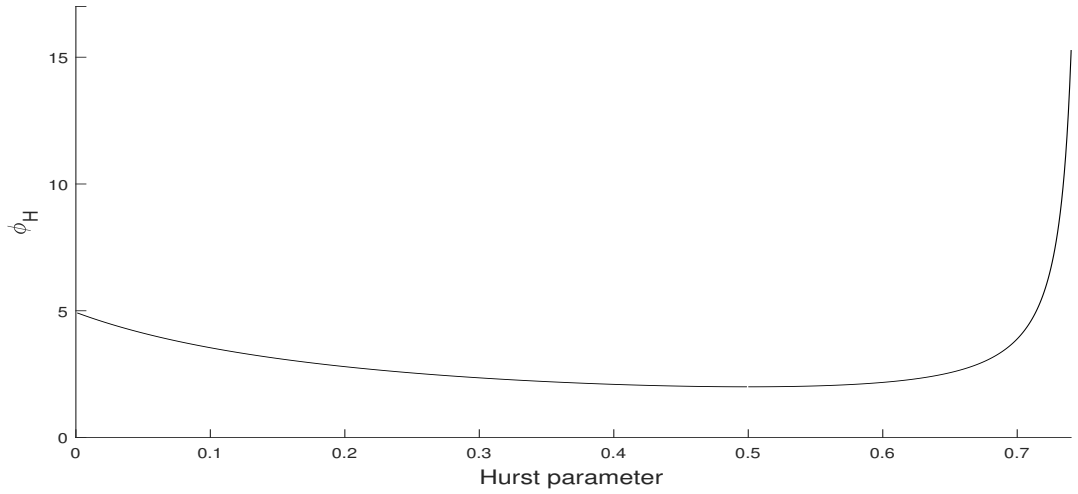

Figure 3: Plot of ϕ_H as a function of H.

Remark 3.6. *The asymptotic theory of $\hat{\kappa}_s$ depends on rrgions where the true value of H lies in. In particular, when $H \in (0, 3/4)$, the asymptotic variance of $\hat{\kappa}_s$ depends on H via ϕ_H. Figure 3 plots ϕ_H as a function of H. It is important to point out that ϕ_H reaches its minimum (equals two) at $H = 1/2$, and monotonically decreases over $(0, 1/2)$ but monotonically increases over $(1/2, 3/4)$. When $H = 1/2$, the asymptotic distribution of $\hat{\kappa}_s$ is $\mathcal{N}(0, 2\kappa)$, which is the same as that of MLE of κ when $H = 1/2$ and is known; see Theorem 3.2.2. in Tang et al. (2009). This result is very interesting as it suggests that the ergodic-type estimate of κ attains the Cramer-Rao lower bound asymptotically. Another interesting feature that can be observed is that the asymptotic variance of $\hat{\kappa}_s$ increases as H increase when $H \in (1/2, 3/4)$ and tends to infinity as H approaches $3/4$. Hence ϕ_H has a singularity at $H = 3/4$.*

4 Simulation Studies

In this section, we study the finite sample properties of the proposed two-stage estimators via data simulated from fOU with different parameter values for H, σ, μ and κ and different values for δ and T. Several experiments are designed.

Once data are simulated, the estimators are straightforward to obtain as they all have closed-form expressions. The main difficulty is to simulate a path from a fBm. In the literature, there are several methods to deal with the problem of simulating fBm (Coeurjolly, 2000). The present paper employs Paxson's algorithm (Paxson, 1997). In particular, we first generate fractional Gaussian noises based on Paxson's method by fast Fourier transformation at a grid that is finer than the target grid. Then, we obtain a path of fBm based on the result that fBm is a partial sum of fractional Gaussian noises.

Based on simulated fBm at the fine grid, the Euler discretization of fOU allows us to obtain simulated data from fOU at the fine grid. Data used for estimation are then extracted from the simulated fOU at the target (coarse) grid. For better understanding of our method, we describe all the steps for data simulation and parameter estimation as follows.

(i) Set the targeted sampling interval to δ, the time span to T, and the initial value to X_0. Hence, the target number of observations for parameter estimation is $n = T/\delta$;

(ii) Set values for four parameters H, μ, κ, σ;

(iii) Choose a fine grid $\{0, \delta/M, 2\delta/M, \ldots, \delta, (M+1)\delta/M, \ldots, 2\delta, \ldots, n\delta\}$ with $M > 1$ so that the corresponding sampling interval is δ/M (denoted by γ) which is smaller than δ. Use the fast Fourier transform to generate fBm B_t^H based on Paxson's method at the fine grid;

(iv) Obtain $\left\{X_{(i+1)\gamma}\right\}$ recursively from

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

where $i = 0, 1, \ldots, M, M+1, \ldots, 2M, \ldots, nM$; Retain only $\{X_{i\delta}\}_{i=0}^n$ for parameter estimation;

(v) Calculate $\hat{H}$ using (2.4);

(vi) Calculate $\hat{\sigma}$ using (2.5);

(vii) Calculate $\hat{\mu}_s$ using (2.6);

(viii) Calculate $\hat{\kappa}_s$ using (2.7).

In total we design three experiments. In all experiments, we set the number of replications to 10,000. In the first experiment, we set the time span $T = 16$, $\delta = 1/256$, $M = 8$, but vary H, σ, μ and κ. Note that $\delta = 1/256$ corresponds approximately to the daily frequency for annualized data. Mean, median, 2.5 percentile, 97.5 percentile, and standard deviation (SD hereafter) for each of the four estimators are computed across 10,000 replications.

Table 1 reports simulation results for fixed values for σ, μ and κ, but varied values for H (from 0.1 to 0.7). All parameters, except κ, can be accurately estimated. The means and medians are always close to their respective true parameter values and the SDs are small. Judged by larger SDs, κ is estimated with a less precision. Moreover, there is a noticeable upward bias and skewness in $\hat{\kappa}_s$ in each case. This finding is not surprising as Phillips and Yu (2005, 2009a) documented the upward bias in the MLE of κ when $H = 1/2$. Yu (2012) obtained an analytical expression to approximate the bias in the MLE of κ when $H = 1/2$. However, when H is unknown, the simulation results suggest that the bias depends on both κ and H. Moreover, as the value of H changes,

Table 1: Performance of the two-stage estimates for different H with $T = 16$ 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	.0995	1.0064	2.8000	5.6398		.2993	1.0050	2.7997	5.1998
Median	.0997	.9981	2.8000	4.7396		.2994	.9978	2.7993	5.0043
SD	.0239	.1324	.0173	4.2998		.0226	.1285	.0289	1.8282
2.5%	.0522	.7689	2.7664	.1672		.2544	.7711	2.7441	2.1814
97.5%	.1453	1.2854	2.8339	16.1090		.3428	1.2750	2.8563	9.2865
True value	.50	1.00	2.80	5.00		.70	1.00	2.80	5.00
Mean	.4992	1.0037	2.8000	5.2745		.6990	1.0029	2.7996	5.6530
Median	.4995	.9969	2.7991	5.1378		.6989	.9933	2.7987	5.5572
SD	.0212	.1270	.0500	1.3392		.0196	.1218	.0865	1.2643
2.5%	.4570	.7758	2.7039	2.9904		.6598	.7674	2.6327	3.4721
97.5%	.5401	1.2723	2.8967	8.2455		.7368	1.2832	2.9695	8.4214

the SD of $\hat{\sigma}$ decreases, as predicted by the asymptotic theory given by (3.4); see Remark 3.1. Furthermore, as the value of H changes, the SD of $\widehat{\mu}_s$ increases. This because the asymptotic theory given by (3.13) suggests that the rate of convergence is T^{1-H} and hence H negatively affects the precision of $\widehat{\mu}_s$.

Table 2 reports simulation results for fixed values for H, μ and κ, but varied values for σ (from 0.3 to 2). As the estimated value for H in our first empirical study based on VIX is around 0.45, we fix $H = 0.45$. The estimation results for H and κ stay the same when σ changes its value. This is expected and consistent with our asymptotic theory. As σ increases, both the SD of $\hat{\sigma}$ and the SD of $\hat{\mu}_s$ increase, consistent with the prediction of asymptotic theory given by (3.3) and (3.13), respectively.

Table 3 reports simulation results for fixed values for H, σ and κ, but varied values for μ (from 0.5 to 2). As expected, the estimation results for H, σ, κ stay the same when μ changes its value. When μ changes, both the mean and the median of $\hat{\mu}_s$ change but the SD of $\hat{\mu}_s$ stays the same.

Table 4 reports simulation results for fixed values for H, σ and μ, but varied values for κ (from 1.0 to 10.0). The estimation results for H, σ are robust to the change in κ, consistent with the prediction of the asymptotic theory. However, as κ increases, the SD of $\hat{\mu}_s$ decreases, consistent with (3.13). As κ increases, the SD of $\hat{\kappa}_s$ increases, consistent with (3.14).

To see the influence of the sampling interval and M, we design the second experiment by fixing T and all four parameters, but varying values of δ from 1/256 to 1/2048

Table 2: Performance of the two-stage estimates for different σ with $T = 16$ and $\delta = 1/256$

	H	σ	μ	κ	::::	H	σ	μ	κ
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
Median	.4496	.2994	2.7997	5.0970		.4496	.4990	2.7996	5.0970
SD	.0215	.0381	.0131	1.4123		.0215	.0635	.0218	1.4123
2.5%	.4063	.2322	2.7749	2.8321		.4063	.3870	2.7581	2.8321
97.5%	.4910	.3811	2.8252	8.3794		.4910	.6351	2.8420	8.3794
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
Median	.4496	.9981	2.7991	5.0970		.4496	1.9962	2.7983	5.0970
SD	.0215	.1271	.0436	1.4123		.0215	.2541	.0871	1.4123
2.5%	.4063	.7740	2.7162	2.8321		.4063	1.5481	2.6324	2.8321
97.5%	.4910	1.2703	2.8841	8.3794		.4910	2.5405	2.9682	8.3794

Table 3: Performance of the two-stage estimates for different μ with $T = 16$ and $\delta = 1/256$

	H	σ	μ	κ	::::	H	σ	μ	κ
True value	.45	.30	.50	5.00		.45	.30	1.00	5.00
Mean	.4492	.3012	.4999	5.2383		.4492	.3012	.9999	5.2383
Median	.4496	.2994	.4997	5.0970		.4496	.2994	.9997	5.0970
SD	.0215	.0381	.0131	1.4123		.0215	.0381	.0131	1.4123
2.5%	.4063	.2322	.4749	2.8321		.4063	.2322	.9749	2.8321
97.5%	.4910	.3811	.5252	8.3794		.4910	.3811	1.0252	8.3794
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
Median	.4496	.2994	1.4997	5.0970		.4496	.2994	1.9997	5.0970
SD	.0215	.0381	.0131	1.4123		.0215	.0381	.0131	1.4123
2.5%	.4063	.2322	1.4749	2.8321		.4063	.2322	1.9749	2.8321
97.5%	.4910	.3811	1.5252	8.3794		.4910	.3811	2.0252	8.3794

Table 4: Performance of the two-stage estimators for different κ with $T = 16$ and $\delta = 1/256$

	H	σ	μ	κ	::::	H	σ	μ	κ
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
Median	.4498	.9986	2.7953	1.1626		.4498	.9987	2.7986	3.1337
SD	.0215	.1271	.2123	.5612		.0215	.1271	.0724	1.0246
2.5%	.4065	.7746	2.3913	.4454		.4065	.7746	2.6608	1.5773
97.5%	.4913	1.2707	3.2109	2.5969		.4912	1.2707	2.9392	5.5608
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
Median	.4496	.9981	2.7991	5.0970		.4489	.9950	2.7996	10.0188
SD	.0215	.1271	.0436	1.4123		.0215	.1267	.0218	2.2102
2.5%	.4063	.7740	2.7162	2.8321		.4065	.7713	2.7580	6.2109
97.5%	.4910	1.2703	2.8841	8.3794		.4903	1.2667	2.8420	14.8804

and values of M from 16 to 32. Here, 1/256 corresponds roughly to daily observations while 1/2048 corresponds to hourly observations. Table 5 reports simulation results across 10,000 simulated replications. It is clear that increasing M from 16 to 32 does not materially change the performance of the proposed estimators. Note that M is used to control the discretization bias in the Euler discretization when simulating data. Our simulation results suggest that the discretization bias is very small. However, when δ changes from 1/256 to 1/2048, the frequency of data becomes higher and the performance of $\hat{H}$ and $\hat{\sigma}$ improves, as predicted by the asymptotic theory given in (3.3) and (3.4).

In the last experiment, to evaluate the influence of time span, we fix the value of $\delta = 1/256$, $M = 16$ and the four parameters, but vary T from 10 to 20. It can be seen that as T increases, the performance of all four estimators improves. In particular, the bias in $\hat{\kappa}_s$ approximately reduces by half when T doubles its value. All the standard deviations reduce by a factor of $\sqrt{2}$ or so when T doubles its value, consistent with the prediction of our asymptotic theory.

5 Empirical Studies

In this section, we carry out two empirical studies, applying the proposed estimation method and the new asymptotic theory for making statistical inference to volatility data in the risk-neutral measure and in the physical measure, respectively. In the first study, we fit fOU to daily VIX from CBOE. In the second study, we fit fOU to three daily

Table 5: Performance of the two-stage estimators for different δ and M with $T = 16$

	$\delta = 1/256$				::::	$\delta = 1/2048$			
	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

realized volatility series.

5.1 VIX

VIX, the CBOE volatility index, attempts to track the volatility of S&P 500. As it is based on the implied volatility in the 30-day options of the S&P 500, a model of VIX can be regarded as a model in the risk-neutral measure. VIX was employed in our empirical study simply because it is widely accepted as the premier measure of stock market volatility and investor sentiment. The daily data utilized in our empirical investigation are extracted from the CBOE spanning from 01/02/2004 through 12/29/2017.[2] In total there are $3,524$ trading days. Hence, we can set $\delta = 1/252$, and $T = 14$ in estimation. Moreover, for the sake of comparison, we divide the whole sample into two subsamples and fit fOU to each subsample. The first subsample is from 01/02/2004 to 12/30/2011 (about $2,015$ data points with $T = 8$) while the second subsample is from 01/03/2012 to 12/29/2017 (about $1,509$ data points with $T = 6$). Figure 4 plots the time series of log(VIX) in the whole

[2]The data were obtained from http://www.cboe.com/products/vix-index-volatility/vix-options-and-futures/vix-index/vix-historical-data. In 2003, the methodology for calculating VIX was revised. The new method is model-free, and uses European style options. Moreover, this new VIX is able to incorporate information from the volatility smile by using a wider range of strike prices.

Table 6: Performance of the two-stage estimators for different T

	$T = 10$				::::	$T = 20$			
	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	.4487	.8139	2.7996	5.3850		.4491	.8148	2.8002	5.1764
Median	.4488	.8035	2.7994	5.1698		.4490	.8048	2.8001	5.0558
SD	.0273	.1312	.0447	1.8490		.0192	.0919	.0311	1.2522
2.5%	.3949	.5866	2.7116	2.4410		.4113	.6447	2.7408	3.0662
97.5%	.5020	1.0989	2.8884	9.6494		.4868	1.0046	2.8619	7.9525

Table 7: Empirical results for VIX index

Period	H	σ	μ	κ	$1/T$
2004–2017	.4496	.3707	1.2325	2.7434	.0714
	(.4443, .4547)	(.3600, .3814)	(1.1706, 1.2945)	(2.6656, 2.8213)	
2004–2011	.4315	.3254	1.2842	1.6194	.1250
	(.4216, .4413)	(.3129, .3380)	(1.1635, 1.4050)	(1.5399, 1.6990)	
2012–2017	.4710	.4336	1.1635	10.9451	.1667
	(.4601, .4819)	(.4145, .4527)	(1.1334, 1.1936)	(10.7085, 11.1818)	

sample.

Three sets of estimation results, including the point estimates and 95% confidence intervals for all four parameters (2.5% and 97.5% percentiles in brackets), and the value for $1/T$, are reported in Table 7. Several important empirical results are found. First, in all cases, the estimated H is around 0.45, less than 0.5. The 95% confidence intervals suggest that we have strong evidence against the null hypothesis of $H = 0.5$. Hence, VIX is better modeled by a rough volatility model. This finding supports results found in Bayer et al. (2016), Gatheral et al. (2018) and Livieri et al. (2018). Second, the estimated κ is not very small. It is always much larger than $1/T$, suggesting that we have strong evidence against the hypothesis of $\kappa \ll 1/T$. Therefore, we cannot use Proposition 3.1 of Gatheral et al. (2018) to argue that fOU behaves locally as a fBm.

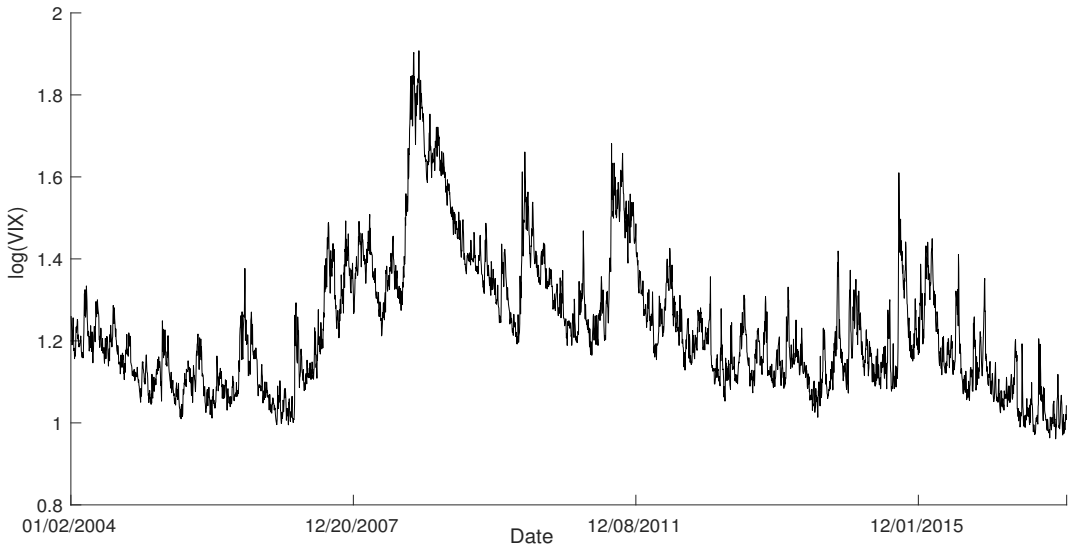

Figure 4: Time series plot of log(VIX)

5.2 Realized volatility

We now fit fOU to the logarithm of three realized volatility (RV hereafter) series for S&P 500, DJIA, and Nasdaq 100. A model of RV can be regarded as a model in the physical measure. The three RV series are obtained from Ox-Mann realized library and based on 5-minute returns.[3] The sample period is from 01/03/2011 to 12/04/2017. Figure 5 plots three time series of $\sqrt{100 \times \log(RV \times 252)}$.

Three sets of estimation results, including the point estimates and 95% confidence intervals for all four parameters, and the value for $1/T$, are reported in Table 8. Several important empirical results are found. First, in all cases, the estimated H is much smaller than 0.5, ranging between .0946 for DJIA to .2550 for Nasdaq 100. The point estimates of H are very close to what was used in Gatheral et al (2018). The 95% confidence intervals suggest that we have strong evidence against the null hypothesis of $H = 0.5$. Hence, each of the RV series is better modeled by a rough volatility model. This finding once again supports results found in Bayer et al. (2016), Gatheral et al. (2018) and Livieri et al. (2018). Second, the estimated κ is always larger than $1/T$, suggesting that we have strong evidence against the hypothesis of $\kappa \ll 1/T$. Once again, we cannot use Proposition 3.1 of Gatheral et al. (2018).

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

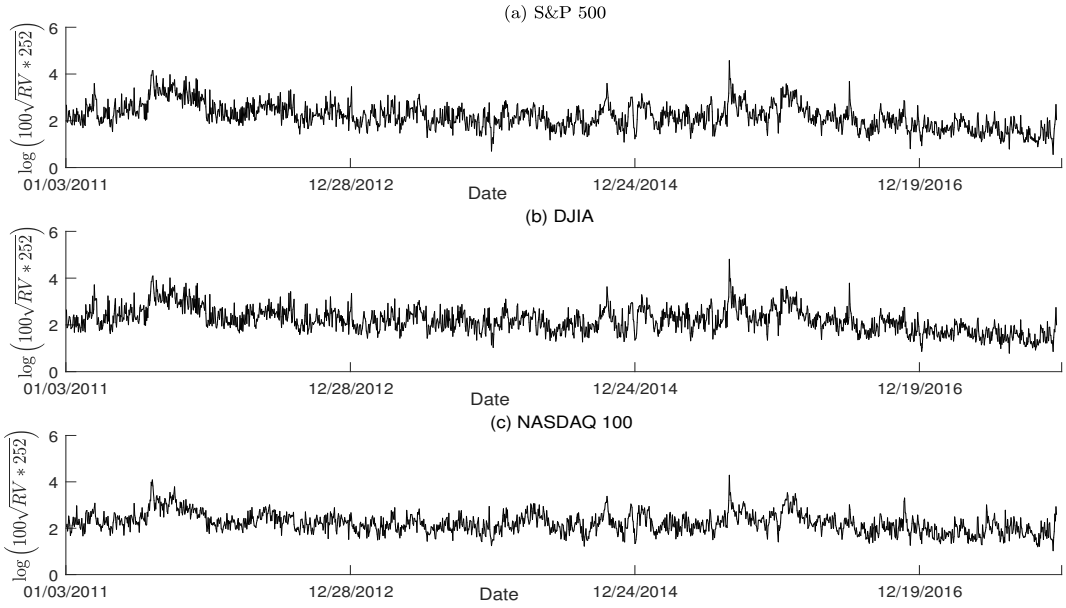

Figure 5: Time series plot of log(RV) for S&P 500, DJIA, Nasdaq 100

Table 8: Empirical results for realized volatility of S&P 500, DJIA, Nasdaq 100

Name	H	σ	μ	κ	$1/T$
S&P 500	.1453	.8440	2.1960	1.3810	.1445
	(.1295, .1610)	(.8062, .8818)	(1.9666, 2.4253)	(1.2830, 1.4791)	
DJIA	.0946	.6788	2.2019	.2382	.1445
	(.0768, .1124)	(.6481, .7096)	(1.2318, 3.1719)	(.1947, .2816)	
Nasdaq 100	.2550	1.2849	2.2220	14.8874	.1445
	(.2432, .2668)	(1.2289, 1.3408)	(2.1819, 2.2621)	(14.5993, 15.1755)	

6 Conclusion

Over the past two decades, it has been a general consensus that volatility of financial asset displays the long-range dependence (see, for example, Comte and Renault (1998)). In the continuous time setting, long-range dependence can be modeled with the help of fBm when the Hurst parameter is greater than one half. However, Gatheral et al. (2018) shows that the logarithm of realized volatility behaves essentially as an fBm with the Hurst parameter taking a value of around 0.1 at any reasonable time scale. Using at-the-money options on the S&P500 index with short maturity, Livieri et al. (2018) further confirms that volatility is rough.

This present paper makes contributions to the literature by proposing a novel estimation method for all the parameters in the fOU model based on discrete-sampled observations when the parameter space for the Hurst parameter is $(0, 1)$. We have also developed asymptotic theory for the proposed estimators which facilitate statistical inference. By employing the property of different order of magnitude between the "diffusion" term and the drift term, the proposed method estimates the "diffusion" parameters in the first stage and the drift parameters in the second stage. In the first stage, using the asymptotic theory for the second order differences of fOU, we construct a two-time-scale estimator of H and an estimator of σ. In the second stage, by employing the property of stationarity and ergodicity, we propose ergodic-type estimators of κ and μ. Consistency and asymptotic laws for these estimators are established. Simulations suggest that our two-stage estimators perform well in finite samples. The method is applied to two empirical examples, VIX and realized volatility of S&P 500, DJIA, Nasdaq 100. The empirical studies show that the volatility is rough, reinforcing the findings in Gatheral, et al. (2018). However, we find the evidence that κ is larger than one divided by the time span, making Proposition 3.1 of Gatheral et al. (2018) not applicable.

This study also suggests several important directions for future research. First, while our estimators are consistent and easy to use, they may not be asymptotically efficient. Finding an asymptotically more efficient estimation technique and obtaining the asymptotic relative inefficiency of our two-stage estimators would be of great interest. Second, the fOU model considered in the present paper does not have any jump, even though fOU with $H < 1/2$ can be rough. Although removal of a few jumps from 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 fOU with jumps is important and we leave it for further research. Third, this paper assumes that the Hurst parameter does not change over time. This assumption can be too restrictive. How to test if H changes its value in the sample and how to model time-varying H are some important questions to ask. Finally, when fitting fOU to RV series, we assume RV measures integrated volatility without measurement errors. This assumption is clearly too strong. How robust the empirical results to measurement errors in RV should be examined and will be explored in future research.

APPENDIX

A.1. Proof of Theorem 3.1

It is well known that fOU is defined as the unique solution of the following Langevin equation

$$X_t = X_0 + \kappa \int_0^t (\mu - X_s)\, ds + \sigma B_t^H \,. \tag{A.1}$$

Consequently, a standard but tedious calculation shows that Assumptions 1 and 2 of Barndorff-Nielsen et al. (2013) hold true.

For $k = 1, 2$ and $t \geq 0$, let $\tau_k = 4R_{ks} - R_{2ks}$ with $R_t = \mathbb{E}\left[\left|B_{t+s}^H - B_s^H\right|^2\right]$. Then, using Lemma 1 of Barndorff-Nielsen et al. (2013), we have

$$\frac{\delta}{\tau_2} \sum_{i=4}^{T/\delta} \left|X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}\right|^2 \xrightarrow{p} \int_0^T \sigma^2 ds \;=\; \sigma^2 T \,, \tag{A.2}$$

$$\frac{\delta}{\tau_1} \sum_{i=2}^{T/\delta} \left|X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}\right|^2 \xrightarrow{p} \int_0^T \sigma^2 ds \;=\; \sigma^2 T \,. \tag{A.3}$$

Combining (A.2) with (A.3), we deduce that

$$\frac{\sum_{i=4}^{T/\delta} |X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}|^2}{\sum_{i=2}^{T/\delta} |X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}|^2} \xrightarrow{p} \frac{\tau_2}{\tau_1} \,. \tag{A.4}$$

Using the polarization identities, we have

$$\begin{aligned} \frac{\tau_2}{\tau_1} &= \frac{4\mathbb{E}\left[|B_{s+2\delta}^H - B_s^H|^2\right] - \mathbb{E}\left[|B_{s+4\delta}^H - B_s^H|^2\right]}{4\mathbb{E}\left[|B_{s+\delta}^H - B_s^H|^2\right] - \mathbb{E}\left[|B_{s+2\delta}^H - B_s^H|^2\right]} = \frac{4 \cdot 2^{2H}\delta^{2H} - (4\delta)^{2H}}{4 \cdot \delta^{2H} - (2\delta)^{2H}} \\ &= \frac{4 \cdot 2^{2H} - 2^{2H}2^{2H}}{4 - 2^{2H}} = 2^{2H} \,. \end{aligned} \tag{A.5}$$

Using (A.4) and (A.5), we have

$$\frac{\sum_{i=4}^{T/\delta} |X_{i\delta} - 2X_{(i-2)\delta} + X_{(i-4)\delta}|^2}{\sum_{i=2}^{T/\delta} |X_{i\delta} - 2X_{(i-1)\delta} + X_{(i-2)\delta}|^2} \xrightarrow{p} 2^{2H} \,. \tag{A.6}$$

From the continuous mapping theorem and (A.6), we obtain (3.1). Similarly, using the continuous mapping theorem, (3.1) and (A.3), we obtain (3.2).

A.2. Proof of Theorem 3.2

An elementary but tedious calculation shows that conditions of Theorem 2 and (30) in Barndorff-Nielsen et al. (2013) hold true. Then the result of (3.3) follows directly from (41) in Section 4.3 of Barndorff-Nielsen et al. (2013) and is omitted.

Now, we are left with (3.4). For the sake of convenience, we define

$$\begin{aligned}\tilde{\sigma}^2 &= \frac{\delta}{T\left(4\cdot\delta^{2H}-(2\delta)^{2H}\right)}\sum_{i=2}^{T/\delta}\left|X_{i\delta}-2X_{(i-1)\delta}+X_{(i-2)\delta}\right|^2 \\ &= \frac{\delta}{T\cdot\tau_1}\sum_{i=2}^{T/\delta}\left|X_{i\delta}-2X_{(i-1)\delta}+X_{(i-2)\delta}\right|^2 . \end{aligned} \tag{A.7}$$

Note that

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma}^2-\sigma^2\right)=\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)+\frac{\sqrt{T}}{\sqrt{\delta}}\left(\tilde{\sigma}^2-\sigma^2\right) . \tag{A.8}$$

By Taylor's theorem we have

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)=\frac{\partial\tilde{\sigma}^2}{\partial H}\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{H}-H\right)+\frac{\partial\tilde{\sigma}^2}{\partial H}L\left(\hat{H},X_i\right)\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{H}-H\right) ,$$

for some reminder function $L\left(\hat{H},X_i\right)$ which converges to zero as $\hat{H}\xrightarrow{p}H$. Therefore, by the fact $\hat{H}-H=o_p(\sqrt{\delta})$ (see Eq. (46) or Eq. (51) in Barndorff-Nielsen et al. (2013)) and the assumption of a fixed T, we obtain

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)\xrightarrow{p}0 . \tag{A.9}$$

Moreover, from Theorem 2 in Barndorff-Nielsen et al. (2013) and the properties of stable convergence, we can easily obtain

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\tilde{\sigma}^2-\sigma^2\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\sigma^4\left(2+4\sum_{k=1}^{\infty}\rho_k^2\right)\right) . \tag{A.10}$$

Combining (A.8), (A.9), (A.10) with Slutsky's theorem, we obtain

$$\frac{\sqrt{T}}{\sqrt{\delta}}\left(\hat{\sigma}^2-\sigma^2\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\sigma^4\left(2+4\sum_{k=1}^{\infty}\rho_k^2\right)\right) .$$

By the delta method, we obtain (3.4).

A.1. Proof of Theorem 3.3

In order to avoid the integration with respect to the fBm for $0 < H < \frac{1}{2}$, we write the solution of (2.1) as

$$X_t = \mu + (X_0 - \mu)\, e^{-\kappa t} + \sigma \left(B_t^H - \kappa \int_0^t B_s^H e^{-\kappa(t-s)} ds \right) . \tag{A.11}$$

Using (A.11), for any $s, t \in [0, T]$ and $s < t$, we can obtain

$$\begin{aligned} |X_t - X_s| &\leq |X_0 - \mu| \left| e^{-\kappa t} - e^{-\kappa s} \right| + \sigma \left| B_t^H - B_s^H \right| + \sigma\kappa \int_0^t B_s^H e^{-\kappa(t-s)} ds \\ &\quad + \sigma\kappa \int_0^t B_v^H e^{-\kappa(s-v)} ds \\ &\leq Y_1 |t-s| + Y_2 |t-s|^{H-\epsilon} + Y_3 |t-s| , \end{aligned} \tag{A.12}$$

where $Y_1 := \kappa e^{-\kappa s} |X_0 - \mu|$, $Y_2 := \sigma \sup_{t \neq s \in [0,T]} \frac{\left| B_t^H - B_s^H \right|}{|t-s|^{H-\epsilon}}$ and $Y_3 := 2\sigma\kappa \sup_{0 \leq s \leq T} \left| B_s^H \right|$. Moreover, from Remark 2.3 in Azmoodeh and Viitasaari (2015) and the self-similarity property of fBm, for all $p \geq 1$ and a constant C, we have

$$\mathbb{E}[Y_1] \leq C , \qquad \mathbb{E}[Y_2] \leq C (n\delta)^{\varepsilon p} , \qquad \mathbb{E}[Y_3] \leq C (n\delta)^{Hp} , \tag{A.13}$$

for an arbitrarily small positive variable ε.

Now, we consider (3.5). Obviously, we can write

$$\hat{\mu}_s - \mu = \left(\hat{\mu}_s - \frac{1}{T} \int_0^T X_t dt \right) + \left(\frac{1}{T} \int_0^T X_t dt - \mu \right) , \tag{A.14}$$

where first term is the discretization error and the second term is the "ergodic theorem" term. Using a similar argument as Xiao and Yu (2018a, b), we have

$$\frac{1}{T} \int_0^T X_t dt \overset{a.s.}{\to} \mu , \tag{A.15}$$

for $H \in (0, 1)$. Let ϵ be an arbitrarily small positive variable. Applying Markov's inequality for $\eta > 0$, $q > 1$ yields

$$\mathbb{P} \left(\left| \frac{1}{n} \sum_{l=0}^{n} X_{l\delta} - \frac{1}{T} \int_0^T X_t dt \right| > \eta \right) \leq \eta^{-q} \mathbb{E} \left| \frac{1}{n} \sum_{l=0}^{n} X_{l\delta} - \frac{1}{T} \int_0^T X_t dt \right|^q . \tag{A.16}$$

Moreover, applying Minkowski's inequality and using (A.12) and (A.13), we obtain

$$\begin{aligned}
\mathbb{E}\left|\frac{1}{n}\sum_{l=0}^{n} X_{l\delta} - \frac{1}{T}\int_0^T X_t dt\right|^q &= \frac{1}{(n\delta)^q}\mathbb{E}\left|\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}(X_{l\delta} - X_t)\,dt\right|^q \\
&\leq \frac{1}{(n\delta)^q}\left(\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left(\mathbb{E}\left|X_{l\delta} - X_t\right|^q\right)^{\frac{1}{q}} dt\right)^q \\
&\leq \frac{1}{(n\delta)^q}\Bigg(\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}|t - l\delta|\left(\mathbb{E}\left|Y_1\right|^q\right)^{\frac{1}{q}} dt \\
&\quad + \sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}|t - l\delta|^{H-\epsilon}\left(\mathbb{E}\left|Y_2\right|^q\right)^{\frac{1}{q}} dt \\
&\quad + \sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}|t - l\delta|\left(\mathbb{E}\left|Y_3\right|^q\right)^{\frac{1}{q}} dt\Bigg)^q \\
&\leq C\delta^q + C\delta^{(H-\epsilon)q}(n\delta)^{q\epsilon} + C\delta^q(n\delta)^{qH}\,.
\end{aligned}$$

Plugging the above inequality to (A.16), we get

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{l=0}^{n} X_{l\delta} - \frac{1}{T}\int_0^T X_t dt\right| > \eta\right) \leq C\eta^{-q}\left(\delta^q + \delta^{(H-\epsilon)q}(n\delta)^{q\epsilon} + \delta^q(n\delta)^{qH}\right). \quad \text{(A.17)}$$

Obviously, if the right-hand side of the above inequality is sumable with respect to n, then $\left|\frac{1}{n}\sum_{l=0}^{n} X_{l\delta} - \frac{1}{T}\int_0^T X_t dt\right| \overset{a.s.}{\to} 0$ by the Borel-Cantelli Lemma. In fact, the right-hand side of (A.17) can be written as

$$C\eta^{-q}\left(\delta^q + \delta^{(H-\epsilon)q}(n\delta)^{q\epsilon} + \delta^q(n\delta)^{qH}\right) = Cn^{-1-\lambda}\left[(n\delta^{\alpha_1})^{\gamma_1} + (n\delta^{\alpha_2})^{\gamma_2} + (n\delta^{\alpha_3})^{\gamma_3}\right],$$

where $\alpha_1 = \frac{q}{1+\lambda}$, $\gamma_1 = 1+\lambda$, $\alpha_2 = \frac{qH}{1+\lambda+q\epsilon}$, $\gamma_2 = 1+\lambda+q\epsilon$, $\alpha_3 = \frac{q+qH}{1+\lambda+qH}$, $\gamma_1 = 1+\lambda+qH$.

Note that the positive variables ϵ and λ can be arbitrarily small and q can be arbitrarily large. In this way, we have $\alpha_1 \in (1,+\infty)$, $\alpha_2 \in (1,+\infty)$ and $\alpha_3 \in (1, 1+\frac{1}{H})$. Hence, if $n\delta^p \to 0$ for some $p \in (1, 1+\frac{1}{H})$, then using (A.17) and the Borel-Cantelli Lemma, we obtain

$$\hat{\mu}_s - \frac{1}{T}\int_0^T X_t dt \overset{a.s.}{\to} 0\,. \quad \text{(A.18)}$$

Consequently, combining (A.14), (A.15), (A.18) and the assumption $X_0 = o_p(n)$, we obtain (3.5).

To prove (3.6), we can write

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 - \left(\mu^2 + \sigma^2\kappa^{-2H}H\Gamma(2H)\right) \\
&= \left[\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 - \frac{1}{T}\int_0^T X_t^2 dt\right] + \left[\frac{1}{T}\int_0^T X_t^2 dt - \left(\mu^2 + \sigma^2\kappa^{-2H}H\Gamma(2H)\right)\right]. \quad \text{(A.19)}
\end{aligned}$$

Similar to Xiao and Yu (2018a, 2018b), for $H \in (0,1)$, we can easily obtain

$$\frac{1}{T}\int_0^T X_t^2 dt \xrightarrow{a.s.} \mu^2 + \sigma^2 \kappa^{-2H} H\Gamma(2H)\,. \tag{A.20}$$

For $\eta > 0$ and $q > 1$, by applying Markov's inequality, Minkowski's inequality, Hölder's inequality, the fact $\mathbb{E}\,|X_t|^p \le C$ for all $t > 0$ and $p > 1$, (A.12) and (A.13) yields

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{T}\int_0^T X_t^2 dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2\right| > \eta\right) &\le \eta^{-q}\mathbb{E}\left|\frac{1}{T}\int_0^T X_t^2 dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2\right|^q \\
&= \frac{\eta^{-q}}{(n\delta)^q}\mathbb{E}\left|\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left(X_t^2 - X_{l\delta}^2\right) dt\right|^q \\
&= \frac{\eta^{-q}}{(n\delta)^q}\mathbb{E}\left|\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left(X_t - X_{l\delta}\right)\left(X_t + X_{l\delta}\right) dt\right|^q \\
&\le \frac{\eta^{-q}}{(n\delta)^q}\left(\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left(\mathbb{E}\left(\left|X_{l\delta} - X_t\right|\left|X_{l\delta} + X_t\right|\right)^q\right)^{\frac{1}{q}} dt\right)^q \\
&\le \frac{\eta^{-q}}{(n\delta)^q}\Bigg(\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t - l\delta\right|\left(\mathbb{E}\left|Y_1\left|X_t + X_{l\delta}\right|\right|^q\right)^{\frac{1}{q}} dt \\
&\quad + \sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t - l\delta\right|^{H-\epsilon}\left(\mathbb{E}\left|Y_2\left|X_t + X_{l\delta}\right|\right|^q\right)^{\frac{1}{q}} dt \\
&\quad + \sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t - l\delta\right|\left(\mathbb{E}\left|Y_3\left|X_t + X_{l\delta}\right|\right|^q\right)^{\frac{1}{q}} dt\Bigg)^q \\
&\le C\delta^q + C\delta^{(H-\epsilon)q}\left(n\delta\right)^{q\epsilon} + C\delta^q\left(n\delta\right)^{qH}\,.
\end{aligned}$$

Similarly, if $n\delta^p \to 0$ for some $p \in (1, 1 + \frac{1}{H})$, then using the above result and the Borel-Cantelli Lemma, we obtain

$$\frac{1}{T}\int_0^T X_t^2 dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 \xrightarrow{a.s.} 0\,. \tag{A.21}$$

From (A.19)-(A.21) and the assumption $X_0 = o_p(n)$, we can see

$$\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 \xrightarrow{a.s.} \mu^2 + \sigma^2\kappa^{-2H}H\Gamma(2H)\,. \tag{A.22}$$

Combining (3.1), (3.2), (3.5), (A.22) with the continuous mapping theorem, we obtain the desired result of (3.6).

A.1. Proof of Theorem 3.4

From (A.14), we have

$$(n\delta)^{1-H}\left(\hat{\mu}_s-\mu\right)=(n\delta)^{1-H}\left(\hat{\mu}_s-\frac{1}{T}\int_0^T X_t dt\right)+(n\delta)^{1-H}\left(\frac{1}{T}\int_0^T X_t dt-\mu\right). \quad \text{(A.23)}$$

From Lemma 3.1, we can see that

$$(n\delta)^{1-H}\left(\frac{1}{T}\int_0^T X_t dt-\mu\right)\xrightarrow{\mathcal{L}}\mathcal{N}\left(0,\frac{\sigma^2}{\kappa^2}\right). \quad \text{(A.24)}$$

Moreover, using (A.12) and (A.13), we obtain

$$\begin{aligned}
\mathbb{E}\left[(n\delta)^{1-H}\left|\frac{1}{n}\sum_{l=0}^{n}X_{l\delta}-\frac{1}{T}\int_0^T X_t dt\right|\right] &= \frac{1}{(n\delta)^H}\mathbb{E}\left|\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left(X_{l\delta}-X_t\right)dt\right| \\
&= \frac{1}{(n\delta)^H}\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\mathbb{E}\left|X_{l\delta}-X_t\right|dt \\
&\leq \frac{1}{(n\delta)^H}\Bigg(\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t-l\delta\right|\left(\mathbb{E}\left|Y_1\right|\right)dt \\
&\quad +\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t-l\delta\right|^{H-\epsilon}\left(\mathbb{E}\left|Y_2\right|\right)dt \\
&\quad +\sum_{l=0}^{n}\int_{l\delta}^{(l+1)\delta}\left|t-l\delta\right|\left(\mathbb{E}\left|Y_3\right|\right)dt\Bigg) \\
&\leq C\left[\left(n\delta^{1+\frac{1}{1-H}}\right)^{1-H}+\left(n\delta^{1+\frac{H-\epsilon}{1+\epsilon-H}}\right)^{1+\epsilon-H}+n\delta^2\right].
\end{aligned}$$

Consequently, if $H\in(0,\frac{1}{2})$, we choose $n\delta^{1+\frac{H}{1-H}}\to 0$. Otherwise, if $H\in\left[\frac{1}{2},1\right)$, we choose $n\delta^2\to 0$. Under these conditions, we can obtain

$$(n\delta)^{1-H}\left(\hat{\mu}_s-\frac{1}{T}\int_0^T X_t dt\right)\xrightarrow{p}0. \quad \text{(A.25)}$$

Using (A.23), (A.24), (A.25) and Slutsky's theorem, we obtain (3.13).

In what follows, we show the asymptotic law of $\hat{\kappa}_s$. For simplicity, let

$$\begin{aligned}
\omega_n &= \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 - \left(\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\right)^2, \\
\psi_n &= \frac{1}{T}\int_0^T X_t^2 dt - \left(\frac{1}{T}\int_0^T X_t dt\right)^2, \\
f(x) &= \sqrt{x}\mathbf{1}_{\{0<H<\frac{3}{4}\}} + \frac{\sqrt{x}}{\log(x)}\mathbf{1}_{\{H=\frac{3}{4}\}} + x^{2-2H}\mathbf{1}_{\{\frac{3}{4}<H<1\}}, \\
g_n(y) &= -\frac{1}{2H}\frac{\omega_n-\psi_n}{\hat{\sigma}^2\hat{H}\Gamma(2\hat{H})}\left(y\frac{\omega_n-\psi_n}{\hat{\sigma}^2\hat{H}\Gamma(2\hat{H})} + \hat{\kappa}_{TS}^{-2H}\right)^{-\frac{1}{2H}-1}.
\end{aligned}$$

Then, a simple calculation shows that

$$\begin{aligned}
f(n\delta)\left(\omega_n-\psi_n\right) &= f(n\delta)\left(\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 - \frac{1}{T}\int_0^T X_t^2 dt\right) + \left[\left(\frac{1}{T}\int_0^T X_t dt + \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\right)\right. \\
&\quad \left. f(n\delta)\left(\frac{1}{T}\int_0^T X_t dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\right)\right]. \qquad \text{(A.26)}
\end{aligned}$$

Now, we consider three cases: (i) $0<H<\frac{3}{4}$; (ii) $H=\frac{3}{4}$; (iii) $\frac{3}{4}<H<1$.

We first consider the case of $0<H<\frac{3}{4}$. In this situation, $f(n\delta)=\sqrt{n\delta}$ and we can write (A.26) as

$$\begin{aligned}
\sqrt{n\delta}\left(\omega_n-\psi_n\right) &= \sqrt{n\delta}\left(\frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2 - \frac{1}{T}\int_0^T X_t^2 dt\right) + \left[\left(\frac{1}{T}\int_0^T X_t dt + \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\right)\right. \\
&\quad \left. \sqrt{n\delta}\left(\frac{1}{T}\int_0^T X_t dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}\right)\right]. \qquad \text{(A.27)}
\end{aligned}$$

Using similar arguments as (A.18) and (A.21), for an arbitrarily small positive variable ϵ, we have

$$\begin{aligned}
\mathbb{E}\left[\sqrt{n\delta}\left|\frac{1}{n}\sum_{l=0}^{n} X_{l\delta} - \frac{1}{T}\int_0^T X_t dt\right|\right] &\le C\left[\left(n\delta^3\right)^{\frac{1}{2}} + \left(n\delta^{\frac{\frac{1}{2}+H}{\frac{1}{2}+\epsilon}}\right)^{\frac{1}{2}+\epsilon} + \left(n\delta^{1+\frac{1}{\frac{1}{2}+H}}\right)^{\frac{1}{2}+H}\right], \\
\mathbb{E}\left[\sqrt{n\delta}\left|\frac{1}{T}\int_0^T X_t^2 dt - \frac{1}{n}\sum_{l=0}^{n} X_{l\delta}^2\right|\right] &\le C\left[\left(n\delta^3\right)^{\frac{1}{2}} + \left(n\delta^{\frac{\frac{1}{2}+H}{\frac{1}{2}+\epsilon}}\right)^{\frac{1}{2}+\epsilon} + \left(n\delta^{1+\frac{1}{\frac{1}{2}+H}}\right)^{\frac{1}{2}+H}\right].
\end{aligned}$$

Combining the above inequalities with (A.27), we obtain

$$\sqrt{n\delta}\left(\omega_n-\psi_n\right) \xrightarrow{p} 0, \qquad \text{(A.28)}$$

with $n\delta^p \to 0$ for some $p \in \left(1, \frac{3+2H}{1+2H} \wedge (1+2H)\right)$.

Similarly, when $H = \frac{3}{4}$, we have

$$\frac{\sqrt{n\delta}}{\log(n\delta)} (\omega_n - \psi_n) \xrightarrow{p} 0 \,. \tag{A.29}$$

with $n\delta^p \to 0$ for some $p \in \left(1, \frac{9}{5}\right)$.

When $\frac{3}{4} < H < 1$, a simple calculation shows that

$$(n\delta)^{2-2H} (\omega_n - \psi_n) \xrightarrow{p} 0 \,, \tag{A.30}$$

with $n\delta^p \to 0$ for some $p \in \left(1, \frac{3-H}{2-H}\right)$. Moreover, from Eq. (46) and Eq. (51) in Barndorff-Nielsen et al. (2013), we can see

$$\hat{H} - H = o_p(\sqrt{\delta}) \,, \quad \hat{\sigma}^2 - \sigma^2 = o_p(\sqrt{\delta}) \,. \tag{A.31}$$

Note that for $x > 0$ and any $\epsilon \to 0$, the series expansion of the gamma function $\Gamma(x+\epsilon)$ can be written as

$$\Gamma(x+\epsilon) = \Gamma(x) + \Gamma(x)\left(\psi^{(0)}(x)\right)\epsilon + \Gamma(x)\left(\left(\psi^{(0)}(x)\right)^2 + \psi^{(1)}(x)\right)\frac{\epsilon^2}{2} + O(\epsilon^3) \,,$$

where $\psi^{(n)}(x) = \frac{d^{n+1}}{dx^{n+1}} \log(\Gamma(x))$ for $n = 0, 1$.

Hence, we have

$$\Gamma\left(2\hat{H}\right) = \Gamma\left(2H + o_p\left(\sqrt{\delta}\right)\right) = \Gamma(2H) + o_p\left(\sqrt{\delta}\right) . \tag{A.32}$$

Using (A.31) and (A.32), we can rewrite $g_n(y)$ as

$$g_n(y) = -\frac{1}{2H} \frac{\omega_n - \psi_n}{\sigma^2 H\Gamma(2H) + o_p\left(\sqrt{\delta}\right)} \left(y \frac{\omega_n - \psi_n}{\sigma^2 H\Gamma(2H) + o_p\left(\sqrt{\delta}\right)} + \hat{\kappa}_{TS}^{-2H} \right)^{-\frac{1}{2H}-1} . \tag{A.33}$$

Using (A.26), (A.28)-(A.30) and (A.31), we have

$$\lim_{n\to\infty} \int_0^1 f(n\delta) g_n(y) dy = 0 \,, \tag{A.34}$$

almost surely by the dominated convergence theorem.

On the other hand, a standard calculation yields

$$\begin{aligned} f(n\delta)(\hat{\kappa}_s - \kappa) &= f(n\delta) \left[\left(\frac{\omega_n - \psi_n}{\hat{\sigma}^2 \hat{H}\Gamma(2\hat{H})} + \hat{\kappa}_{TS}^{-2H} \right)^{-\frac{1}{2H}} - \kappa \right] \\ &= f(n\delta) \left[\hat{\kappa}_{TS} - \kappa + \int_0^1 g_n(y) dy \right] \\ &= f(T)(\hat{\kappa}_{TS} - \kappa) + f(n\delta) \int_0^1 g_n(y) dy \,. \end{aligned} \tag{A.35}$$

From (A.34) and (A.35), we can see that $f(n\delta)(\hat{\kappa}_s - \kappa)$ converges in law to the same random variable as $f(T)(\hat{\kappa}_{TS} - \kappa)$ when T tends to infinity. By Lemma 3.1 and Slutsky's theorem, we finish the proof.

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.

Azmoodeh, E. and Viitasaari, L. (2015). Parameter estimation based on discrete observations of fractional Ornstein–Uhlenbeck process of the second kind. *Statistical Inference for Stochastic Processes*, 18(3):205–227.

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.

Barndorff-Nielsen, O. E. and Podolskij, M. (2009). Multipower variation for Brownian semistationary processes. *Bernoulli*, 17(4):1159–1194.

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.

Biagini, F., Hu, Y., Øksendal, B., and Zhang, T. (2008). *Stochastic calculus for fractional Brownian motion and applications*. Springer.

Brouste, A. and Iacus, S. M. (2013). Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package. *Computational Statistics*, 28(4):1529–1547.

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

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.

Corcuera, J. M., Nualart, D., and Woerner, J. H. C. (2006). Power variation of some integral fractional processes. *Bernoulli*, 12(4):713–735.

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.

Es-Sebaiy, K. (2013). Berry–Esséen bounds for the least squares estimator for discretely observed fractional Ornstein–Uhlenbeck processes. *Statistics and Probability Letters*, 83(10):2372–2385.

Euch, O. E. and Rosenbaum, M. (2017). Perfect hedging in rough Heston models. *arXiv preprint arXiv:1703.05049*.

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.

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.

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. (2018). Parameter estimation for fractional Ornstein–Uhlenbeck processes of general Hurst parameter. *to appear in Statistical Inference for Stochastic Processes*, https://doi.org/10.1007/s11203–017–9168–2.

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.

Kubilius, K., Mishura, Y., and Ralchenko, K. (2017). *Drift parameter estimation in diffusion and fractional diffusion models*. Springer.

Kubilius, K., Mishura, Y., Ralchenko, K. and Seleznjev, O. (2015). Consistency of the drift parameter estimator for the discretized fractional Ornstein–Uhlenbeck process with Hurst index $H \in (0, 1/2)$. *Electronic Journal of Statistics*, 9(2):1799–1825.

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.

Mishura, Y. (2008). *Stochastic calculus for fractional Brownian motion and related processes*. Springer.

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. 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.

Podolskij, M. and Wasmuth, K. (2013). Goodness-of-fit testing for fractional diffusions. *Statistical Inference for Stochastic Processes*, 16(2):147–159.

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. (2014). Distributions of quadratic functionals of the fractional Brownian motion based on a martingale approximation. *Econometric Theory*, 30(5):1078–1109.

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

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. (2018a). Asymptotic theory for estimating the drift parameters in the fractional Vasicek model. *Econometric Theory*, forthcoming.

Xiao, W. and Yu, J. (2018b). Asymptotic theory for rough fractional Vasicek models. *Singapore Management University, School of Economics, Paper No. 2018-007.*

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