Maximum Likelihood Estimation of Fractional Ornstein-Uhlenbeck Process with Discretely Sampled Data*

Xiaohu Wang, Weilin Xiao, Jun Yu, Chen Zhang

March 26, 2025

Abstract

This paper first derives two analytic formulae for the autocovariance of the discretely sampled fractional Ornstein-Uhlenbeck (fOU) process. Utilizing the analytic formulae, two main applications are demonstrated: (i) investigation of the accuracy of the likelihood approximation by the Whittle method; (ii) the optimal forecasts with fOU based on discretely sampled data. The finite sample performance of the Whittle method and the derived analytic formula motivate us to introduce a feasible exact maximum likelihood (ML) method to estimate the fOU process. The long-span asymptotic theory of the ML estimator is established, where the convergence rate is a smooth function of the Hurst parameter (i.e., H) and the limiting distribution is always Gaussian, facilitating statistical inference. The asymptotic theory is different from that of some existing estimators studied in the literature, which are discontinuous at $H = 3/4$ and involve non-standard limiting distributions. The simulation results indicate that the ML method provides more accurate parameter estimates than all the existing methods, and the proposed optimal forecast formula offers a more precise forecast than the existing formula. The fOU process is applied to fit daily realized volatility (RV) and daily trading volume series. When forecasting RVs, it is found that the forecasts generated using the optimal forecast formula together with the ML estimates outperform those generated from all possible combinations of alternative estimation methods and alternative forecast formula.

JEL Classification: C15, C22, C32.

Keywords: Fractional Ornstein-Uhlenbeck process; Hurst parameter; Out-of-sample forecast; Maximum likelihood; Whittle likelihood; Composite likelihood.

*Xiaohu Wang, School of Economics, Fudan University, Shanghai, China. Email: wang_xh@fudan.edu.cn. Weilin Xiao, School of Management, Zhejiang University, Hangzhou, 310058, China. Email: wlxiao@zju.edu.cn. Jun Yu and Chen Zhang, Faculty of Business Administration, University of Macau, Macao, China. Email: junyu@um.edu.mo and chenzhang@um.edu.mo.

1 Introduction

Estimating continuous-time diffusion models driven by the standard Brownian motion (Bm) with discretely sampled data has garnered considerable attention in the literature. The Markov property, inherent to Bm-driven diffusions, ensures that the log-likelihood function can be obtained as the sum of log transition probability densities. When the transition probability density has a closed-form expression, the likelihood function can be easily computed, enabling exact maximum likelihood (ML) estimation. If the transition probability density is not available in closed form, Aït-Sahalia (1999, 2002) provides a highly accurate method to approximate the transition probability density, facilitating the ML estimation. See, for example, Phillips and Yu (2009) for a literature review.

In the case where the drift function is affine and the diffusion function is constant, the diffusion model becomes the classical Ornstein-Uhlenbeck (OU) process. In this case, both the transition probability density and the ML estimator have closed-form expressions. The fractional OU (fOU) process studied in this paper is an extension of the OU process by replacing the standard Bm with a fractional Brownian motion (fBm). The standard Bm has independent increments. On the other hand, the increments of fBm can have a vibrant correlation structure. The fOU process has found wide applications in practice, including modeling and forecasting volatility and trading volume of financial assets (Gatheral et al., 2018; Fukasawa et al., 2022; Wang et al., 2023; Bolko et al., 2023; Bennedsen et al., 2024; Shi et al., 2024; Chong and Todorov, 2024), options pricing (Livieri et al., 2018; Bayer et al., 2016; Garnier and Sølna, 2018), variance swaps (Bayer et al., 2016), portfolio choice (Fouque and Hu, 2019), trading strategies (Glasserman and He, 2020), and hedging (Euch and Rosenbaum, 2018).

Due to the complex dependence structure of the increments of fBm (known as the fractional Gaussian noise or fGn), the fOU process, sampled at discrete points in time, is not Markovian. This non-Markovian property poses challenges in constructing ML estimation from discretely-sampled data. Recently, several alternative estimation methods have been proposed to estimate parameters in the fOU process based on discrete samples, including the method-of-moments (MM) method by Wang et al. (2023, WXY hereafter), the maximum composite likelihood (MCL) method by Bennedsen et al. (2023, 2024), and the approximate Whittle ML (AWML) method proposed by Shi et al. (2024). The MM and MCL methods are expected to be inefficient, as they only utilize limited information. The AWML method can yield asymptotic efficient estimators for the parameters in fOU, except the location parameter μ, which is usually estimated separately by the

sample mean or another consistent estimator *a priori*. However, the AWML estimators may not perform well in finite samples when the Whittle method approximates the exact likelihood function poorly. The concerns above motivates this paper to construct feasible exact ML estimation.

In this paper, we first provide two analytic formulae for the autocovariance function of the fOU process, which facilitate (i) exact ML estimation, (ii) checking the accuracy of an approximate likelihood method of Shi et al. (2024), and (iii) obtaining the optimal forecasts with fOU based on discretely sampled data. We show that the approximate likelihood method may give a poor finite-sample approximation, especially when the sample size is small and H is large. Regarding out-of-sample forecasting, the common practice in the literature is using the Euler scheme to discretize the continuous-record-based forecasting formula given by Fink et al. (2013) and then generating forecasts. This method is not optimal when only a discrete sample is available. Hence, the optimal forecasting formula for discrete samples proposed in this paper is fundamental for empirical studies related to forecasting.

The paper then develops a large-sample theory for the ML estimators of all the parameters in fOU under the long-span asymptotic scheme, where the sample size goes to infinity with a fixed sampling interval. Consistency and asymptotic normality are established. Compared to the asymptotic theory of the MCL estimators developed by Bennedsen et al. (2024) and that of the MM estimator studied by Wang et al. (2023), which is discontinuous at $H = 3/4$ for both the convergence rate and the asymptotic distribution, our asymptotic theory for the exact ML estimators uniformly applies to all the values of $H \in (0, 1)$, with the asymptotic covariance matrix being a continuous function of H. This feature greatly facilitates statistical inference, especially when the confidence interval of H includes $3/4$.

In addition, the newly developed large sample theory for the location parameter μ shows that the exact ML estimator is more efficient than the sample mean, especially when $H > 1/2$. The asymptotic theory for ML estimators of the long memory stationary process has been well studied in the literature, including the seminal works of Fox and Taqqu (1986), Dahlhaus (1989), and Lieberman (2012). However, these works either assume the location parameter of the process is known or is estimated prior by the sample mean or other consistent estimators with a specific convergence rate. Our asymptotic theory extends results in the literature when all parameters are estimated simultaneously by the ML approach.

With realistic parameter settings relevant to financial markets, we have conducted

comprehensive simulation studies to compare the finite sample performance of the ML estimation method with that of the existing estimation methods. The simulation results demonstrate improved forecasting accuracy using the ML estimators and the proposed optimal forecasting formula for discrete samples.

For empirical applications, we fit the fOU process to daily realized volatility (RV) and daily trading volume for ten exchange traded funds. Strong evidence of roughness for the logarithmic RV and trading volume series is found. Moreover, we compare the out-of-sample forecasting performance of RV using the fOU process with three different estimation methods (i.e., ML, MCL, and MM) and two different forecasting methods. It has been found that the forecasts generated by using the ML estimation method together with the proposed optimal forecast formula for discrete observations have the most minor mean squared error for all assets considered. The Diebold-Mariano test shows that the improvement relative to the forecasts from other combinations of estimation approach and forecasting formula are statistically significant. In addition, the results of the model confidence set test proposed by Hansen et al. (2011) suggest that the ML method together with the optimal forecast formula is always in the set of best predictive methods for all the assets.

The rest of the paper is structured as follows. In Section 2, we first introduce the fOU process and present two analytic formulae for the autocovariance of the fOU process. We then check the performance of the expressions against numerical methods, taking account of both accuracy and computational cost. Section 3 considers two applications of the analytic formulae for the autocovariance of fOU: investigating the distance of the Whittle approximation from the exact likelihood function and constructing the optimal forecasting formula with discrete samples, respectively. Section 4 studies the feasible exact ML estimation and develops the long-span asymptotic theory of the ML estimator. In Section 5, the finite sample performance of the ML method is compared with that of existing methods using simulated data. Moreover, the finite sample accuracy of the forecasts generated using the ML method and the proposed optimal forecasting formula is compared with that generated from combinations of alternative estimation approaches and forecasting formulae. Section 6 presents some empirical applications of the fOU process. Section 7 concludes. The proofs are given in the Appendix. The online supplement contains additional proof details, simulations, and empirical results.

Throughout the paper, we use $\xrightarrow{p}$, $\xrightarrow{d}$, $\overset{d}{=}$ to denote convergence in probability, convergence distribution, and distributional equivalence, respectively. For a matrix A, $|A|$ represents its determinant, and $\|A\| = \left(\mathrm{tr}\left(A^\top A\right)\right)^{1/2}$ denotes the Euclidean norm, where

the upper index $^\top$ denotes vector/matrix transpose. For a matrix series $\{A_j\}$, when $\|A_j - A\| \to 0$, it is claimed that A_j converges to A.

2 fOU Process

2.1 Some preliminaries of fOU

The standard OU process is driven by the standard Bm (W_t), defined by

$$dZ_t = \kappa(\mu - Z_t)dt + \sigma dW_t,\ Z_0 = O_p(1), \tag{2.1}$$

which has a unique path-wise solution given by

$$Z_t = e^{-\kappa t}Z_0 + \left(1 - e^{-\kappa t}\right)\mu + \sigma \int_0^t e^{-\kappa(t-s)}dW_s.$$

Under the conditions of $\kappa > 0$ and $Z_0 \sim \mathcal{N}\left(\mu, \sigma^2/\left(2\kappa\right)\right)$, Z_t is stationary and ergodic.

The fOU process is an extension of the standard OU process above by replacing W_t in (2.1) with an fBm, B_t^H for $H \in (0,1)$. An fBm is a zero mean Gaussian process with the autocovariance function of

$$\mathrm{Cov}\left(B_t^H, B_s^H\right) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right),\ \ \forall t, s \in [0,\infty)\ , \tag{2.2}$$

where $H \in (0,1)$ is called the Hurst parameter. When $H = 0.5$, B_t^H becomes a standard Bm, that is, $B_t^{0.5} \overset{d}{=} W_t$, which has independent increments. In contrast, whenever $H \neq 0.5$, B_t^H has stationary increments with a rich serial dependence structure. The increment sequence has long memory property (i.e., the summation of autocovariances diverges to infinity) when $H \in (0.5, 1)$, whereas it becomes antipersistent (i.e., the summation of autocovariances equals zero) when $H \in (0, 0.5)$. In addition, B_t^H is self-similar in the sense that $\forall a \in \mathbb{R}$, $B_{at}^H \overset{d}{=} |a|^H B_t^H$.

Strictly speaking, the fOU process is defined by the following differential equation:

$$dX_t = \kappa(\mu - X_t)dt + \sigma dB_t^H,\ \ X_0 = O_p(1)\,, \tag{2.3}$$

which has a unique path-wise solution as

$$X_t = e^{-\kappa t}X_0 + \left(1 - e^{-\kappa t}\right)\mu + \sigma \int_0^t e^{-\kappa(t-s)}dB_s^H. \tag{2.4}$$

Let $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1}e^{-y}dy$ denote the Gamma function. When $\kappa > 0$ and $X_0 \sim \mathcal{N}\left(\mu, \sigma^2\Gamma(2H+1)/\left(2\kappa^{2H}\right)\right)$, X_t is stationary and ergodic. Moreover, X_t is (locally)

Hölder continuous of order $H-\epsilon$ for any $\epsilon>0$. Hence, the fOU process with $H<0.5$ has sample paths rougher than those of the standard OU process with $H=0.5$. As a result, the fOU process with $H<0.5$ is referred to as a rough fOU process; see Gatheral et al. (2018).

Define $C(H):=\Gamma(2H+1)\sin(\pi H)$. Hult (2003) obtains the spectral density of the continuous-time fOU process as

$$f_X(\lambda;\beta)=\frac{\sigma^2}{2\pi}C(H)\,|\lambda|^{1-2H}(\kappa^2+\lambda^2)^{-1} \text{ for } \lambda\in(-\infty,\infty). \tag{2.5}$$

with $\beta\equiv(\sigma^2,\kappa,H)^\top$. In the literature, various methods have been proposed to estimate parameters of the fOU process based on a continuous record. A partial list of important contributions include Kleptsyna et al. (2000), Kleptsyna and Le Breton (2002), Hu and Nualart (2010), Hu et al. (2019), Xiao and Yu (2019a, b), Lohvinenko and Ralchenko (2017, 2019), Tanaka et al. (2020), Tanaka (2013). However, although fOU is specified in continuous time, in practice, observations are almost always available in a discrete sample. As a result, the above-mentioned estimation methods have been rarely used in practice.

In the present paper, we assume $\kappa>0$ and study the ML estimation approach and the optimal forecast of fOU based on discrete-time observations, say $\mathbf{X}=(X_0,X_\Delta,\cdots,X_{n\Delta})^\top$, where $n+1$ is the sample size and Δ is the sampling interval. The discrete time series $\{X_{j\Delta}\}_{j=0,1,2,\ldots,n}$ is a Gaussian stationary process with the following spectral density (Hult, 2003)

$$f_X^\Delta(\lambda;\beta)=\frac{\sigma^2}{2\pi}C(H)\,\Delta^{2H}\sum_{k=-\infty}^{\infty}\frac{|\lambda+2\pi k|^{1-2H}}{(\kappa\Delta)^2+(\lambda+2\pi k)^2} \text{ for } \lambda\in[-\pi,\pi]. \tag{2.6}$$

When $0.5<H<1$, it can be shown that

$$f_X^\Delta(\lambda;\beta)=|\lambda|^{1-2H}L_\kappa(\lambda)\to\infty \quad \text{as } \lambda\to 0$$

where $L_\kappa(\lambda)$ is a slowly varying function as $\lambda\to 0$ with the following expression and limit

$$\begin{aligned} L_\kappa(\lambda)&=\frac{\sigma^2}{2\pi}C(H)\,\Delta^{2H}\left\{\frac{1}{(\kappa\Delta)^2+\lambda^2}+|\lambda|^{2H-1}\sum_{k\neq 0}\frac{|\lambda+2\pi k|^{1-2H}}{(\kappa\Delta)^2+(\lambda+2\pi k)^2}\right\} \\ &\to\frac{\sigma^2}{2\pi}C(H)\,\Delta^{2H}\frac{1}{(\kappa\Delta)^2}. \end{aligned} \tag{2.7}$$

Hence, $\{X_{j\Delta}\}$ is a long-memory process. Whereas, in the case of $0 < H \leq 0.5$, it has

$$\lim_{\lambda \to 0} f_X^{\Delta}(\lambda;\beta) = f_X^{\Delta}(0;\beta) = \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \sum_{k=-\infty}^{\infty} \frac{|2\pi k|^{1-2H}}{(\kappa\Delta)^2 + (2\pi k)^2} \in (0,\infty),$$

which means $\{X_{j\Delta}\}$ becomes a short-memory weakly stationary process.

From the Gaussianity of fOU, the likelihood function is

$$L_n(\theta) = (2\pi)^{-(n+1)/2} |\sigma^2\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{X} - \mu\mathbf{1})^{\top} \Sigma^{-1} (\mathbf{X} - \mu\mathbf{1})\right), \tag{2.8}$$

where $\theta = (\mu, \sigma^2, \kappa, H)^{\top}$ contains all unknown parameters in fOU, $\mathbf{1} = (1, 1, \ldots, 1)^{\top}$, Σ is the covariance matrix of $\sigma^{-2}\mathbf{X}$, which is a symmetric Toeplitz matrix and defined by

$$\Sigma = \sigma^{-2} [\mathrm{Cov}(X_{i\Delta}, X_{s\Delta})]_{i,s=0,1,\cdots,n} := \sigma^{-2} \begin{pmatrix} \gamma_0 & \gamma_\Delta & \cdots & \gamma_{n\Delta} \\ \vdots & \vdots & \vdots & \vdots \\ \gamma_{n\Delta} & \gamma_{(n-1)\Delta} & \cdots & \gamma_0 \end{pmatrix}, \tag{2.9}$$

where $\gamma_{j\Delta}$ denotes the jth autocovariance of discretely sampled fOU.

Choosing θ to maximize $\ln L_n(\theta)$ yields the ML estimate. Garnier and Sølna (2018) (Eq. (6)) provide an expression of the autocovariance:

$$\gamma_{j\Delta} = \frac{\sigma^2}{2\kappa^{2H}} \left(\frac{1}{2}\int_{-\infty}^{\infty} e^{-|y|} |\kappa j\Delta + y|^{2H} dy - |\kappa j\Delta|^{2H}\right) \quad \text{for} \quad j = 0, 1, \ldots, n. \tag{2.10}$$

When $j = 0$, $\gamma_{j\Delta}$ becomes the variance of $X_{i\Delta}$ and can be written as

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{2\kappa^{2H}} \int_0^{\infty} e^{-y} y^{2H} dy = \frac{\sigma^2 \Gamma(2H+1)}{2\kappa^{2H}}. \tag{2.11}$$

2.2 Two analytic formulae for the autocovariance

Although Equation (2.10) gives an expression for each element in the covariance matrix Σ, it involves an integral over the interval of $(-\infty, +\infty)$. Generally this integral must be evaluated numerically. Numerical integrations for all potential parameter values make the ML estimation extremely time-consuming, especially when n is large. Moreover, numerical integrations potentially lead to large approximation errors, making the resulting MLE distant from the actual parameter values. This subsection provides two alternative analytic formulae for $\gamma_{j\Delta}$ to facilitate the calculation of the likelihood function.

Lemma 2.1 *Consider fOU defined in (2.3) with $\kappa > 0$ and the stationary initial condition of $X_0 \sim \mathcal{N}\left(\mu, \sigma^2\Gamma(2H+1)/\left(2\kappa^{2H}\right)\right)$. Let $\{X_0, X_\Delta, \cdots, X_{n\Delta}\}$ be a discrete sample.*

For any $j \geq 0$, it has
(a):

$$\gamma_{j\Delta} = \frac{\sigma^2 e^{-\kappa j\Delta}}{4\kappa^{2H}} \left\{ \Gamma(2H+1) - (\kappa j\Delta)^{2H} {}_1F_1(2H; 1+2H; \kappa j\Delta) + 2He^{2\kappa j\Delta}\Gamma(2H, \kappa j\Delta) \right\}, \tag{2.12}$$

where ${}_1F_1(\cdot;\cdot;\cdot)$ and $\Gamma(\cdot,\cdot)$ are, respectively, the confluent hypergeometric function of the first kind and the upper incomplete Gamma function defined by

$${}_1F_1(2H; 1+2H; \kappa j\Delta) = \sum_{n=0}^{\infty} \frac{2H}{2H+n} \frac{(\kappa j\Delta)^n}{n!}, \tag{2.13}$$

and

$$\Gamma(2H, \kappa j\Delta) = \int_{\kappa j\Delta}^{\infty} s^{2H-1}e^{-s}ds = \Gamma(2H) - \int_0^{\kappa j\Delta} s^{2H-1}e^{-s}ds\,;$$

(b):

$$\gamma_{j\Delta} = \frac{\sigma^2}{2\kappa^{2H}} \left\{ \cosh(\kappa j\Delta)\Gamma(2H+1) - (\kappa j\Delta)^{2H} {}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right) \right\} \tag{2.14}$$

where $\cosh(x) = [\exp(x) + \exp(-x)]/2$ is the hyperbolic cosine function and ${}_1F_2$ denotes the generalized hypergeometric function as

$${}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right) = \sum_{n=0}^{\infty} \frac{\Gamma(H+1/2)\Gamma(H+1)}{\Gamma(H+1/2+n)\Gamma(H+1+n)} \left(\frac{\kappa j\Delta}{2}\right)^{2n} \tag{2.15}$$

Remark 2.1 *As simulation results provided later suggest, the two analytic formulae have the same accuracy in computing $\gamma_{j\Delta}$ and yield more accurate results than well-known numerical integration methods. Note that $\gamma_{j\Delta}$ is a typical element in Σ, whose determinant and inverse must be calculated to evaluate the log-likelihood function in (2.8). Errors in approximating $\gamma_{j\Delta}$ by numerical integration methods may translate to the determinant and the inverse of Σ. As a result, they may potentially distort the ML estimate.*

Remark 2.2 *Between the two analytic formulae, (2.14) is faster to compute than (2.12) for two reasons. First, ${}_1F_2$ is faster to compute than ${}_1F_1$. That is because the term*

$$\frac{\Gamma(H+1/2)\Gamma(H+1)}{\Gamma(H+1/2+n)\Gamma(H+1+n)} = O\left(\frac{1}{n!n!}\right)$$

converges faster than $\frac{2H}{2H+n}\frac{1}{n!}$. Second, the formulae in (2.12) needs to calculate an extra term, $\Gamma(2H, \kappa j\Delta)$.

Remark 2.3 *Hult (2003) gives an alternative formula of the autocovariance $\gamma_{j\Delta}$ as*

$$\gamma_{j\Delta} = \sigma^2\Gamma(2H+1)\sin(\pi H)\left\{\frac{1}{2}\kappa^{-2H}\sec(\pi(1-2H)/2)\cosh(\kappa j\Delta) + \frac{(j\Delta)^{2H}\Gamma(-H)}{\sqrt{\pi}2^{2H+1}\Gamma(H+1/2)}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\right\}. \quad (2.16)$$

In the Online Supplement, we show how this formula is related to (2.14).

Remark 2.4 *Unlike the expression given in (2.10) where each element in Σ must be obtained by numerical integrations, the expressions given in (2.12), (2.14) and (2.16) suggest that we can calculate all the elements in Σ without relying on any numerical integration methods. All the special functions involved in the formulae presented in Lemma 2.1 and (2.16) have been well studied in the mathematics literature and can be accurately calculated by using built-in functions in standard softwares, such as MATLAB and R.*

2.3 Performance of alternative expressions

We now evaluate the accuracy and the computational cost of evaluating (2.10) numerically and calculating (2.12), (2.14) and (2.16). To do so, we set $\sigma = 1, H = 0.2, \Delta = 1/252$ and $\kappa = 1$ but allow j to vary from 0 to 2500. When evaluating (2.10) numerically, we use MATLAB commands quadgk and integral. The MATLAB command quadgk evaluates the integral numerically based on high-order global adaptive quadrature and default error tolerances. The MATLAB command integral evaluates the integral numerically based on global adaptive quadrature and default error tolerances. Moreover, we calculate the expressions given in (2.12), (2.14) and in (2.16) by using the MATLAB commands to evaluate ${}_1F_1$ and ${}_1F_2$, and the MATLAB commands gamma, igamma to evaluate $\Gamma(\cdot)$ and $\Gamma(\cdot,\cdot)$.

Table 1 reports autocovariances of fOU using different methods when j=0, 1, 100, 200, 300, 400, 500, 1000, 1500, 2000, 2500 while Table 2 reports the CPU time (in seconds) in calculating the autocovariance values for $j = 0, 1, ..., 2500$ when running MATLAB2021b in a 2.8 GHz Intel Core i7-10710U CPU with 16 GB of RAM and Windows 10. According to Table 1, the three analytic formulae given in (2.12), (2.14) and (2.16) always yield the identical values. However, both quadgk and integral give different values from those calculated from the analytic formulae, suggesting that numerical integrations lead to approximation errors. Between the two numerical methods, integral has smaller approximation errors than quadgk. However, according to Table 2, integral

is computationally more costly than quadgk. Among the three analytic formulae, (2.14) is the fastest to compute and (2.12) is the slowest, as expected. Computing (2.10) by quadgk is faster than by using the three analytical expressions. However, as shown in Table 1, quadgk leads to much larger approximation errors.

Table 1: The values of the autocovariances using different formulae when $\sigma = 1, H = 0.2, \Delta = 1/252$, $\kappa = 1$, and $j = 0, 1, 100, 200, 300, 400, 500, 1000, 1500, 2000, 2500$. The boldface shows the difference.

j	(2.10) by quadgk	(2.10) by integral	(2.12), (2.14), (2.16)
0	.443631908751538	.443631908751538	.443631908751538
1	.388882**037284944**	.388881**922491498**	.388881909561334
100	.11719**0968855458**	.11719**0530899574**	.117190522788601
200	.04586**9415544228**	.04586**9299134194**	.045869320276164
300	.01235**5758479432**	.01235**472615600**	.012355439277410
400	-.00411**2113948332**	-.00411**2210815218**	-.004112011199993
500	-.01189**0279222754**	-.01189**0441908578**	-.011890470208617
1000	-.0130**84013072070**	-.0130**91808741521**	-.013091823515404
1500	-.00762**0166025256**	-.00762**0537228084**	-.007620528689543
2000	-.00470**3526957885**	-.00470**5950339660**	-.004705938475695
2500	-.00320**818355352**	-.00320**893056313**	-.003208393362911

Table 2: CPU time (in seconds) of computing the autocovariances using different formulae when $\sigma = 1, H = 0.2, \Delta = 1/252$, $\kappa = 1$, and $j = 0, ..., 2500$.

j	(2.10) by quadgk	(2.10) by integral	(2.12)	(2.14)	(2.16)
from 0 to 2500	3.437500	41.250000	6.687500	4.218750	4.578125

To further understand the implications of the approximation errors, we apply quadgk, integral and (2.14) to calculate $\ln|\Sigma|$ and $\ln L_n(\theta)$ when $n = 2500$ and $H = 0.2, 0.4, 0.6, 0.8$. Table 3 reports the values of $\ln|\Sigma|$ and $\ln L_n(\theta)$ when $\sigma = 1, \Delta = 1/252$, $\kappa = 1$, calculated by applying three alternative methods: (2.10) by quadgk, (2.10) by integral, and (2.14). It is clear that the errors incurred by quadgk and integral in approximating the autocovariances lead to substantial errors in approximating $\ln|\Sigma|$ and $\ln L_n(\theta)$, especially for quadgk and when H is large. To strike the balance between the computational speed and numerical precision, in the rest of the paper, we will apply the analytical expression (2.14) to calculate autocovariance.

Table 3: Values of $\ln|\Sigma|$ and $\ln L_n(\theta)$ when $\sigma = 1, \Delta = 1/252$, $\kappa = 1$, $H = 0.2, 0.4, 0.6, 0.8$, and $n = 2500$ obtained by alternative methods of computing the autocovariances. The boldface shows the difference.

H	Expression	(2.10) by quadgk	(2.10) by integral	(2.14)
0.2	$\ln\|\Sigma\|$	-8507.**250816421579657**	-8507.05**2292128733825**	-8507.053629693704352
	$\ln L_n(\theta)$	-5563.**078623025841807**	-5563.64**3675034223634**	-5563.647212954847419
0.4	$\ln\|\Sigma\|$	-1183**3.823834806182276**	-11832.647**803225803727**	-11832.642928066117747
	$\ln L_n(\theta)$	-3510.**705013957488518**	-3510.0**58645632319531**	-3510.005125368167683
0.6	$\ln\|\Sigma\|$	-159**98.984553859338121**	-15985.4**67690883881005**	-15985.430818915077907
	$\ln L_n(\theta)$	-286**9.805004141110658**	-2862.**548594308252177**	-2862.475203322524976
0.8	$\ln\|\Sigma\|$	-2**1381.235952217579324**	-20788.1**55230374970415**	-20788.131369394981448
	$\ln L_n(\theta)$	-2**484.097131805580375**	-2565.1**66811946502094**	-2565.154210472930117

3 Applications to Likelihood Approximation and Optimal Forecast

We consider two applications of the analytic formulae: (i) evaluating the accuracy of the approximate Whittle likelihood proposed by Shi et al. (2024), which motivates the use of exact ML estimation, and (ii) deriving the optimal forecast based on a discrete sample.

3.1 Likelihood approximation by Approximate Whittle Likelihood

Whittle (1951, 1954) proposed a way to approximate the log-likelihood function of a stationary model based on the spectral density. The log Whittle likelihood function takes the form of

$$l_W(\beta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\ln f_X^{\Delta}(\lambda;\beta) + \frac{P_n(\lambda)}{f_X^{\Delta}(\lambda;\beta)}\right)d\lambda, \tag{3.1}$$

where $\beta = (\sigma^2, \kappa, H)^\top$, $f_X^{\Delta}(\lambda;\beta)$ is the spectral density given in (2.6), and $P_n(\lambda)$ is the periodogram. When the location parameter μ is known and assumed to be zero, let $\mathrm{i} := \sqrt{-1}$ and then $P_n(\lambda)$ is

$$P_n(\lambda) = \frac{1}{2\pi(n+1)}\left|\sum_{j=0}^{n} X_{j\Delta}\exp(-\mathrm{i}j\lambda)\right|^2. \tag{3.2}$$

When μ is unknown that is often the case in practice, Shi et al. (2024) proposed to obtain the periodogram $P_n(\lambda)$ by using the sample mean $\bar{X}$ as[1]

$$P_n(\lambda) = \frac{1}{2\pi(n+1)} \left| \sum_{j=0}^{n} (X_{j\Delta} - \bar{X}) \exp(-\mathrm{i}j\lambda) \right|^2 . \tag{3.3}$$

The log Whittle likelihood function is derived from the following two well-known approximations:

$$\Sigma^{-1} \approx [a_{jk}]_{j,k=1}^{n+1} \text{ and } \ln|\Sigma| \approx (n+1)(2\pi)^{-1} \int_{-\pi}^{\pi} \ln f_X^{\Delta}(\lambda;\beta)\, d\lambda,$$

where $a_{jk} = (2\pi)^{-2} \int_{-\pi}^{\pi} f_X^{\Delta}(\lambda;\beta)^{-1} e^{\mathrm{i}(j-k)\lambda} d\lambda$.

However, the spectral density of the fOU process $f_X^{\Delta}(\lambda;\beta)$ given in (2.6) is not available in closed form in the sense that it involves an infinite summation, which converges at a slow rate when H is close to zero. Shi et al. (2024) proposed the modified Paxson approximation to calculate $f_X^{\Delta}(\lambda;\beta)$ and showed that it yields very small approximation errors.

The AWML estimate of β can be obtained by minimizing $l_W(\beta)$ with respect to β, under the constraints of $\sigma^2 > 0, \kappa > 0, H \in (0,1)$, which is denoted by $\widehat{\beta}_{AWML}$. The asymptotic theory for $\hat{\beta}_{AWML}$ is

$$\sqrt{n}\left(\hat{\beta}_{AWML} - \beta\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\frac{1}{4\pi}\int_{-\pi}^{\pi} \left\{\frac{\partial \ln f_X^{\Delta}(\lambda;\beta)}{\partial\beta}\right\}\left\{\frac{\partial \ln f_X^{\Delta}(\lambda;\beta)}{\partial\beta'}\right\} d\lambda\right)^{-1}\right). \tag{3.4}$$

For the weakly stationary first-order autoregressive model, Rao and Yang (2021) derive an analytic expression for the difference between the log likelihood function and the log Whittle likelihood function. Unfortunately, to the best of our knowledge, no analytic expression is available for the fOU model. However, the derived analytic expressions for covariances facilitate the calculation of the exact log likelihood, and hence, checking the accuracy of the approximate Whittle likelihood proposed by Shi et al. (2024) numerically. To investigate the difference between $\ln L_n(0,\beta)$ and $l_W(\beta)$, Table 4 reports $\ln L_n(0,\beta)/(n+1)$, $l_W(\beta)/(n+1)$, and $(\ln L_n(0,\beta) - l_W(\beta))/L_n(0,\beta)$ when $\mu = 0, \kappa = 1, \sigma = 1$, $\Delta = 1/252$, $n = 504, 2520$, $H = 0.1, 0.3, 0.5, 0.7$, with the data $\{X_{j\Delta}\}$ simulated from fOU. It clearly shows that $l_W(\beta)$ provides worse approximations

[1]While Shi et al. (2024) did not claim to use $\bar{X}$ to estimate μ nor develop the asymptotic theory for $\bar{X}$, since $I_n(\lambda)$ is calculated from $\bar{X}$, we will use it to estimate μ for AWML in our simulation and empirical studies.

to $\ln L_n(0,\beta)$ when H is large or when n is small. These differences are expected to have implications for the finite sample performance of the AWML method relative to the exact ML method, which will be carefully investigated in Section 5.

Table 4: The difference between $\ln L_n(0,\beta)/(n+1)$ and $l_W(\beta)/(n+1)$ when $\kappa = 1, \sigma = 1$, $\Delta = 1/252$, $n = 2520, 504$, $H = 0.1, 0.3, 0.5, 0.7$ and with data $\{X_{j\Delta}\}$ simulated from fOU.

n	H	$\frac{\ln L_n(0,\beta)}{n+1}$	$\frac{l_W(\beta)}{n+1}$	$\frac{((\ln L_n(0,\beta) - l_W(\beta)))}{\ln L_n(0,\beta)} \times 100$
2520	0.1	-0.485656	-0.484626	.0212
	0.3	-2.608224	-2.60467	.0136
	0.5	-4.873668	-4.83139	.0867
	0.7	-7.327926	-6.983444	4.701
504	0.1	-0.465799	-0.461355	.0954
	0.3	-2.410964	-2.333839	3.199
	0.5	-4.524476	-3.871145	14.44
	0.7	-6.851600	0.760280	111.10

3.2 Optimal forecast

When a continuous-time record of X_t over the period of $(0, T]$ is available, Fink et al. (2013) develop the formula of the conditional expectation and conditional variance of X_{T+h} with $h > 0$ to generate the optimal forecast.

When a discrete sample is available, WXY apply the Euler scheme to the formulae derived in Fink et al. (2013) to generate forecast.[2] With discrete-time observations, however, the formula of conditional mean derived by Fink et al. (2013) is not optimal for forecasting purposes anymore, for it does not minimize the root mean squared error (RMSE).

Let $\boldsymbol{\gamma}_{h\Delta}^{(0:n\Delta)} = \left(\text{Cov}\left(X_{(n+h)\Delta}, X_0\right), ..., \text{Cov}\left(X_{(n+h)\Delta}, X_{n\Delta}\right)\right)^\top$ be the vector of covariances between $X_{(n+h)\Delta}$ and $\mathbf{X}$. The expectation of $X_{(n+h)\Delta}$ conditional on the historical discrete-time observations $\mathbf{X}$ is

$$\mathbb{E}(X_{(n+h)\Delta}|\mathbf{X}) = \mu + \left(\gamma_{h\Delta}^{(0:n\Delta)}\right)^\top \Sigma^{-1}(\mathbf{X} - \mu\mathbf{1}), \tag{3.5}$$

which gives the optimal forecast of $X_{(n+h)\Delta}$ when $\mathbf{X}$ is available because it minimizes

[2]Recently, Gao et al. (2023) propose an alternative numerical method to approximate the conditional variance formula of Fink et al. (2013) (see **Algorithm** 1 in Gao et al. (2023)), which will be used in the rest of the paper.

the mean squared error (MSE) of the forecast errors, which is given by

$$\mathbb{E}\left[\left\{\mathbb{E}(X_{(n+h)\Delta}|\mathbf{X}) - X_{(n+h)\Delta}\right\}^2\right] = \gamma_{0\Delta} - \left(\gamma_{h\Delta}^{(0:n\Delta)}\right)^\top \Sigma^{-1}\left(\gamma_{h\Delta}^{(0:n\Delta)}\right). \tag{3.6}$$

Since elements in Σ and $\gamma_{h\Delta}^{(0:n\Delta)}$ are readily obtained from (2.14), our analytic formula facilitate the calculation of the optimal forecast when a discrete sample is available.

When the quantity of interest is $\exp\left(X_{(n+h)\Delta}\right)$ instead of $X_{(n+h)\Delta}$ (i.e., RV instead of log RV), as shown in WXY, one must also compute the conditional variance based on a discrete sample, which takes the same form as the MSE above, for $X_{(n+h)\Delta}$ and $\mathbf{X}$ are jointly normally distributed. Again, our analytic formula given in (2.14) facilitates the forecasting procedure for $\exp\left(X_{(n+h)\Delta}\right)$.

4 Exact ML Estimation and Asymptotic Theory

In Section 3 we have shown that the approximate Whittle method may provide poor approximations to the true log likelihood when H is large and n is small. This observation suggests that it is important to calculate the exact likelihood function and then to produce the exact ML estimate. The derived analytic formula of the autocovariance function $\gamma_{j\Delta}$ given in Lemma 2.1 significantly facilitates the construction of the exact ML estimator based on a discrete sample. This section, therefore, constructs the ML estimator in details and develops its long-span asymptotic theory. The long-span asymptotic scheme assumes $n \to \infty$ with a fixed Δ, which is the same scheme adopted in Shi et al. (2024) and Bennedsen et al. (2024), but different from the double asymptotic scheme considered in WXY that requires $\Delta \to 0$ simultaneously.

Consider the case where the location parameter μ is unknown. From (2.8), the log-likelihood function of the fOU process takes the form of

$$l_n(\theta) = -\frac{n+1}{2}\ln(2\pi) - \frac{1}{2}\ln\left|\sigma^2\Sigma\right| - \frac{1}{2\sigma^2}\left(\mathbf{X}-\mu\mathbf{1}\right)^\top \Sigma^{-1}\left(\mathbf{X}-\mu\mathbf{1}\right). \tag{4.1}$$

Note that the elements in Σ depend on κ and H only. Hence, we can profile the log-likelihood by

$$\mu(\kappa, H) = \frac{\mathbf{1}^\top\Sigma^{-1}\mathbf{X}}{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}, \tag{4.2}$$

$$\sigma^2(\kappa, H) = \frac{\left(\mathbf{X}-\mu(\kappa,H)\mathbf{1}\right)^\top\Sigma^{-1}\left(\mathbf{X}-\mu(\kappa,H)\mathbf{1}\right)}{n+1} = \frac{\mathbf{X}^\top\Sigma^{-1}\mathbf{X} - \frac{\left(\mathbf{1}^\top\Sigma^{-1}\mathbf{X}\right)^2}{\mathbf{1}^\top\Sigma^{-1}\mathbf{1}}}{n+1}. \tag{4.3}$$

Substituting (4.2) and (4.3) into (4.1) yields the following profile log-likelihood function

$$l_n(\kappa, H) = -\frac{n+1}{2}\ln(2\pi) - \frac{n+1}{2}\ln\sigma^2(\kappa,H) - \frac{1}{2}\ln|\Sigma| - \frac{n+1}{2}$$

$$\propto -\frac{n+1}{2}\ln\sigma^2(\kappa,H)-\frac{1}{2}\ln|\Sigma| \tag{4.4}$$

Maximizing the profile log-likelihood function above yields the ML estimators of κ and H, which can be identically written as

$$\left(\widehat{\kappa}_{ML},\widehat{H}_{ML}\right)=\arg\min_{\kappa,H}\left[(n+1)\ln\sigma^2(\kappa,H)+\ln|\Sigma|\right]. \tag{4.5}$$

Consequently, injecting $\left(\widehat{\kappa}_{ML},\widehat{H}_{ML}\right)$ into formulae (4.2) and (4.3) gives the ML estimators of μ and σ^2, respectively:

$$\widehat{\mu}_{ML}=\mu(\widehat{\kappa}_{ML},\widehat{H}_{ML}) \text{ and } \widehat{\sigma^2}_{ML}=\sigma^2(\widehat{\kappa}_{ML},\widehat{H}_{ML}). \tag{4.6}$$

In the rest of the paper, we use $\widehat{\theta}_{ML}\equiv\left(\widehat{\mu}_{ML},\widehat{\sigma^2}_{ML},\widehat{\kappa}_{ML},\widehat{H}_{ML}\right)^\top=\left(\widehat{\mu}_{ML},\widehat{\beta}_{ML}^\top\right)^\top$ to denote the MLE of θ with $\widehat{\beta}_{ML}\equiv\left(\widehat{\sigma^2}_{ML},\widehat{\kappa}_{ML},\widehat{H}_{ML}\right)^\top$.

The long-span asymptotic properties of $\widehat{\theta}_{ML}$ are provided in Theorem 4.1.

Theorem 4.1 *For $H\in(0,1)$, let $f_X^{\Delta}(\lambda;\beta)$ denote the spectral density given in (2.6), $g_n(H)=\begin{cases} n^{1-H} & \text{if } H>1/2,\\ n^{1/2} & \text{if } H\le 1/2\end{cases}$ and $A_n(\theta)=\mathrm{diag}(g_n(H),\sqrt{n},\sqrt{n},\sqrt{n})$. When $n\to\infty$ with a fixed Δ, it has*

$$A_n(\theta)\left(\widehat{\theta}_{ML}-\theta\right)\xrightarrow{d}N\left(0,\mathcal{I}^{-1}(\theta)\right), \tag{4.7}$$

where

$$\mathcal{I}^{-1}(\theta)=\begin{pmatrix} Avar_\mu & \mathbf{0}\\ \mathbf{0} & \left[\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\nabla\ln f_X^{\Delta}(\lambda;\beta)\right)\left(\nabla\ln f_X^{\Delta}(\lambda;\beta)\right)^\top d\lambda\right]^{-1}\end{pmatrix},$$

with

$$Avar_\mu=\begin{cases} 2\pi f_X^{\Delta}(0;\beta) & \text{if } H<1/2,\\ \frac{\Gamma(2-2H)}{B(3/2-H,3/2-H)}\frac{\sigma^2C(H)\Delta^{2H-2}}{\kappa^2} & \text{if } H\ge 1/2\end{cases},$$

$B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ being a Beta function.

Remark 4.1 *Theorem 4.1 shows that a unified asymptotic theory of $\widehat{\beta}_{ML}=\left(\widehat{\sigma^2}_{ML},\widehat{\kappa}_{ML},\widehat{H}_{ML}\right)^\top$ applies to all values of $H\in(0,1)$. The asymptotic covariance continuously changes in H. This feature facilitates statistical inference. In sharp contrast, for the long-span asymptotic theory of the MCL estimator in Bennedsen et al. (2024) and the double asymptotic theory of the MM estimator in Wang et al. (2023), both the convergence rate and the asymptotic distribution are discontinuous at $H=3/4$. Moreover, the asymptotic Rosenblatt distribution has to be applied for the MCL estimator and the MM estimator when $H>3/4$, making statistical inference difficult.*

Remark 4.2 *Also shown by Theorem 4.1 is that the exact ML estimator* $\widehat{\beta}_{ML}$ *has the same asymptotic theory as the AWML estimator* $\widehat{\beta}_{AWML}$ *in Shi et al. (2024). However, since AWML is based on the approximate likelihood function, which can be far away from the exact likelihood function, especially when* H *is large and* n *is small, as shown by the simulation results reported in Table 4. We expect* $\widehat{\beta}_{ML}$ *is more efficient than AWML of Shi et al. (2024) in finite sample, especially when* $H > 0.5$*. Moreover, the AWML method does not estimate* μ*.*

Remark 4.3 *Theorem 4.1 shows that, when* $H < 0.5$*, the convergence rate of* $\widehat{\mu}_{ML}$ *is* $\sqrt{n}$*, which does not depend on* H*. In contrast, the asymptotic theory developed in Adenstedt (1974) suggests that the convergence rate of the ML estimator of the location parameter is* n^{1-H} *for fGn with* $H \in (0, 0.5)$ *and for the ARFIMA*(p, d, q) *model with the memory parameter* $d = H - 0.5 \in (-0.5, 0)$*. While this difference appears to be surprising, it is caused by the fact that the fGn and the ARFIMA*(p, d, q) *process are antipersistent with a zero long-run variance when* $H \in (0, 0.5)$*. Whereas, when* $H < 0.5$*, the fOU process is a short memory stationary process with a well-defined strictly positive long-run variance as*

$$0 < 2\pi f_X^{\Delta}(0; \beta) = \sigma^2 C(H) \Delta^{2H} \sum_{k=-\infty}^{\infty} \frac{|2\pi k|^{1-2H}}{(\kappa\Delta)^2 + (2\pi k)^2} < \infty.$$

5 Monte Carlo Studies

This section is devoted to evaluating the finite-sample performance of the ML estimators when a discrete sample is simulated from Model (2.3). The simulation procedures for the fOU process are the same as in WXY. To examine the relative performance of the ML method, we apply three existing estimation methods to the simulated data: the MM method of WXY, the MCL method of Bennedsen et al. (2024), and the AWML method of Shi et al. (2024). The AWML method has been introduced in Section 3. We now briefly review the MM and MCL methods to improve the readability of comparison results presented below.

5.1 Alternative estimators

WXY use four moment conditions to construct the MM estimators for parameters in fOU, that is,

$$\begin{aligned}
\widehat{H}_{MM} =& \frac{1}{2}\log_2\left(\frac{\sum_{i=0}^{n-4}\left(X_{(i+4)\Delta}-2X_{(i+2)\Delta}+X_{i\Delta}\right)^2}{\sum_{i=0}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}\right),\\
\widehat{\sigma}_{MM} =& \sqrt{\frac{\sum_{i=0}^{n-2}\left(X_{(i+2)\Delta}-2X_{(i+1)\Delta}+X_{i\Delta}\right)^2}{(n+1)\left(4-2^{2\widehat{H}_{MM}}\right)\Delta^{2\widehat{H}_{MM}}}},\\
\widehat{\mu}_{MM} =& \frac{1}{n+1}\sum_{i=0}^{n}X_{i\Delta},\\
\widehat{\kappa}_{MM} =& \left(\frac{(n+1)\sum_{i=0}^{n}X_{i\Delta}^2-\left(\sum_{i=0}^{n}X_{i\Delta}\right)^2}{(n+1)^2\widehat{\sigma}_{MM}^2\widehat{H}_{MM}\Gamma\left(2\widehat{H}_{MM}\right)}\right)^{-0.5/\widehat{H}_{MM}},
\end{aligned}$$

where $\log_2(\cdot)$ is the base-2 logarithm.

Under some mild regularity conditions, WXY derive the asymptotic distributions of MM estimators with $T=n\Delta$ denoting the time span of the data:

$$\sqrt{n}\left(\widehat{H}_{MM}-H\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{(2\ln 2)^2}\right)\text{ as }\Delta\to 0;$$

$$\frac{\sqrt{n}}{\ln(1/\Delta)}\left(\widehat{\sigma}_{MM}-\sigma\right)\xrightarrow{d}\mathcal{N}\left(0,\frac{\Sigma_{11}+\Sigma_{22}-2\Sigma_{12}}{(2\ln 2)^2}\sigma^2\right)\text{ as }\Delta\to 0;$$

$$T^{1-H}\left(\widehat{\mu}_{MM}-\mu\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma^2/\kappa^2\right)\text{ as }\Delta\to 0, T\to\infty, T^{1-H}\Delta^H\to 0;$$

$$\begin{aligned}
\sqrt{T}\left(\widehat{\kappa}_{MM}-\kappa\right)&\xrightarrow{d}\mathcal{N}\left(0,\kappa\phi_H\right)\text{ as }\Delta\to 0, T\to\infty, \sqrt{T}\Delta^H\to 0\text{ for }H\in(0,3/4);\\
\frac{\sqrt{T}}{\ln(T)}\left(\widehat{\kappa}_{MM}-\kappa\right)&\xrightarrow{d}\mathcal{N}\left(0,\frac{16\kappa}{9\pi}\right)\text{ as }\Delta\to 0, T\to\infty, \sqrt{T}\Delta^H/\ln(T)\to 0\text{ for }H=3/4;\\
T^{2-2H}\left(\widehat{\kappa}_{MM}-\kappa\right)&\xrightarrow{d}\frac{-\kappa^{2H-1}}{H\Gamma(2H+1)}R\text{ as }\Delta\to 0, T\to\infty, T^{2-2H}\Delta^H\to 0\text{ for }H\in(3/4,1);
\end{aligned}$$

where R is a Rosenblatt random variable, and the expressions for Σ_{11}, Σ_{22}, Σ_{12}, ϕ_H are presented in WXY.

Remark 5.1 *The asymptotic theory of the MM estimators is more complicated than that of the ML estimators in three aspects. First, for the MM estimators, the asymptotic theory of H and σ is based on the in-fill asymptotic scheme, whereas the asymptotic theory of κ and μ is developed under the double scheme. In contrast, the asymptotic theory of all ML estimators is derived uniformly under the long-span asymptotic scheme. Second, the asymptotic theory of the MM estimator of κ is discontinuous in terms of both the convergence rate and the limiting distribution at $H = 3/4$, which makes the asymptotic theory challenging to apply when the confidence interval of H include $3/4$. On the other hand, the asymptotic theory of the ML estimator of κ is unique, with the asymptotic variance being a continuous function of the value of H. Third, the limit distribution for $\widehat{\kappa}_{MM}$ becomes non-standard when $H \in (3/4, 1)$ but remains Gaussian for $\widehat{\kappa}_{KL}$.*

Remark 5.2 *Since the asymptotic schemes used by MM and ML are different, we cannot compare their asymptotic efficiency based on the asymptotic theory. We will examine their finite-sample performance using simulated data. Since MM only uses limited information, it is expected that ML would be more efficient than MM.*

Remark 5.3 *The MM estimators have closed-form expressions, a feature making the MM estimation extremely easy to implement. Whereas, the ML estimates must be calculated via numerical optimizations, which are computationally more costly.*

Being concerned about the high computational cost of the exact ML method, Bennedsen et al. (2024) propose to use the MCL method of Lindsay (1988) to estimate parameters in fOU where the composite log-likelihood function is a weighted product of densities of marginal or conditional events. When μ is known, let $\widehat{\beta}_{MCL}$ be the MCL estimators of β and f be the probability density of $\mathbf{X}$ that is defined on a probability space (Ω, F, P). Suppose $(\mathcal{A}_m)_{m=1}^M$ is a collection of events with $\mathcal{A}_m \in F$ and the likelihood $L_m(\beta) \propto f(\mathbf{X} \in \mathcal{A}_m; \beta)$. Then the composite likelihood is defined as

$$CL(\beta) = \prod_{m=1}^{M} L_m(\beta)^{w_m}, \tag{5.1}$$

where $w_1, \ldots, w_M$ are nonnegative weights with $\sum_m w_m = 1$.[3] Consequently, the MCL estimator of β is

$$\widehat{\beta}_{MCL} = \arg\max_{\beta} \ln CL(\beta). \tag{5.2}$$

[3]As suggested by Bennedsen et al. (2024) we can set $w_k = 1/M$.

Under some mild regularity conditions, Bennedsen et al. (2024) derive the long-span asymptotic distributions of MCL. That is, as $n \to \infty$ with fixed Δ,

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}_{MCL}-\beta\right) &\xrightarrow{d} \mathcal{N}\left(0, G^{-1}(\beta)\right) \text{ for } H \in (0,3/4); \\
\frac{\sqrt{n}}{\sqrt{L_\gamma(n)}}\left(\widehat{\beta}_{MCL}-\beta\right) &\xrightarrow{d} \mathcal{N}\left(0, G^{-1}(\beta)\right) \text{ for } H = 3/4; \\
n^{2-2H} L_2^{-1/2}(n)\left(\widehat{\beta}_{MCL}-\beta\right) &\xrightarrow{d} U^{-1}(\beta)\,\Psi R \text{ for } H \in (3/4,1),
\end{aligned}$$

where R is a Rosenblatt random variable, and $U(\beta)$, $G(\beta)$, $L_\gamma(n)$, L_2 and Ψ are defined in Bennedsen et al. (2024).

Remark 5.4 *As in the case of MM, both the limit distribution and the rate of convergence of $\widehat{\beta}_{MCL}$ crucially depend on the true value of H. When $H \in (0,3/4)$, the rate of convergence is $n^{-1/2}$, and the limit distribution is Gaussian. When $H = 3/4$, the rate of convergence is $\left(\frac{\sqrt{n}}{\sqrt{L_\gamma(n)}}\right)^{-1}$. When $H \in (3/4,1)$, the convergence rate becomes n^{2H-2}, and the limit distribution becomes non-standard. This feature makes it challenging to use in practice when the confidence interval of H includes $3/4$. Moreover, since the rate of convergence of $\widehat{\beta}_{ML}$ is $n^{-1/2}$, ML is more efficient than MCL when $H \in [3/4,1)$.*

Remark 5.5 *In practice, μ is unknown. In the empirical study, Bennedsen et al. (2024) estimate μ by the sample mean. As in Shi et al. (2024), the asymptotic distribution of the sample mean is not considered. In principle, one could estimate μ by treating $CL(\beta)$ as a function of μ as well as β as in Bennedsen et al. (2023). If so, as remarked in Bennedsen et al. (2024), it is difficult to derive the asymptotic distribution for $\widehat{\beta}_{MCL}$.*

Remark 5.6 *Implementing the MCL approach requires a choice of M and $\{\mathcal{A}_m\}$. However, little is known about how to guide these choices in practice.*

5.2 Simulation results

This subsection will first conduct some simulation studies to investigate the finite sample performance of the proposed ML method and its relative performance against three existing estimation approaches: MM, MCL, and AWML. Then, we present other simulation results to demonstrate the finite sample properties of alternative forecasting approaches.

To do ML estimation, we need to maximize the log-likelihood function given in (4.1) numerically. The covariance matrix Σ is a real symmetric positive-definite Toeplitz matrix. Gohberg and Semencul (1972) provide formulae that express the inverse of a

Toeplitz matrix as a difference between products of triangular Toeplitz matrices (see, for example, Page 174 in Ben-Artzi and Shalom, 1986). Let $\mathbf{u} = (u_0, u_1, \ldots, u_n)^\top$ be the first column of Σ^{-1} and $\mathbf{v} = (0, u_n, \ldots, u_1)^\top$. The inverse matrix can be represented as

$$\Sigma^{-1} = \frac{1}{u_1}\left(\mathbf{L}(\mathbf{u})^\top \mathbf{L}(\mathbf{u}) - \mathbf{L}(\mathbf{v})^\top \mathbf{L}(\mathbf{v})\right), \tag{5.3}$$

where $\mathbf{L}(\mathbf{a})$ is a lower triangular Toeplitz matrix with the first column equal to $\mathbf{a}$. To obtain the value of $\mathbf{u}$, we solve the equation $\Sigma \cdot \mathbf{u} = (1, 0, \ldots, 0)^\top$ by applying the Levinson algorithm proposed by Zhang and Duhamel (1992).

To calculate the determinant of Σ, we adopt the algorithm of Dietrich and Osborne (1996) as

$$|\Sigma_{k+1}| = |\Sigma_k|\left(\gamma_0 - \boldsymbol{\gamma}_k^\top \Sigma_k^{-1} \boldsymbol{\gamma}_k\right), \quad k = 1, \ldots, n\,. \tag{5.4}$$

where Σ_k denote the upper-left $k \times k$ block of the covariance matrix Σ and $\boldsymbol{\gamma}_k$ denote the vector $(\gamma_1, \ldots, \gamma_k)^\top$, which is the first column of Σ_k.

In the first experiment, we simulate 1,000 sample paths from Model (2.3) with $\kappa = 4.446145$, $\mu = -2.465673$, $\sigma = 1.172012$, $\Delta = 1/250$, and H taking different values from 0.1 to 0.8.[4] When implementing MCL, we follow the suggestion of Bennedsen et al. (2023) by using the pairwise likelihood of all pairs of observations with a maximum of K periods distance between them for some fixed integer $K > 0$ such as $(X_\Delta, X_{2\Delta})$, $(X_{2\Delta}, X_{3\Delta})$,..., $\left(X_{(n-1)\Delta}, X_{n\Delta}\right)$, $(X_\Delta, X_{3\Delta})$,..., $\left(X_{(n-K)\Delta}, X_{n\Delta}\right)$ with $K = 5$. We also choose the equal weight.[5] Simulation results, including the mean and the standard deviation (SD), are reported in Table 5 when $n = 2500$ and Table 6 when $n = 500$.

According to Table 5, ML always performs better than AWML, followed by MCL, and then by MM for H, σ, κ in terms of the standard deviation. The improvement of ML over AWML becomes more apparent as H increases, consistent with the findings in Table 4. Moreover, although the ML estimate of μ is similar to the sample mean, it provides greater accuracy for estimating μ when $H \geq 1/2$. Similar conclusions can be obtained from Table 6 when the sample size is $n = 500$. Compared with Table 5, we can see that the improvement by ML over AWMLis more significant when the sample size is smaller. This is not surprising since the Whittle method gives a poorer approximation to the log-likelihood function when n becomes smaller. We also investigate the simulation results when H is fixed to be 0.260573 and the other parameters (σ, μ, and κ) take

[4]Note that $H = 0.260573, \kappa = 4.446145$, $\mu = -2.465673$, $\sigma = 1.172012$ are the estimated values by the ML method when Model (2.3) is fitted to the logarithmic daily RV of SPY index.

[5]Unlike Bennedsen et al. (2024) where μ is estimated by the sample mean, we estimate μ together with other parameters.

various values. The findings are qualitatively unchanged, and the simulation results are reported in the Online Supplement.

Next, we investigate the finite-sample properties of alternative forecasting formulae when a discrete sample is simulated from fOU. We set $H = 0.260573, \kappa = 4.446145, \mu = -2.465673, \sigma = 1.172012, n = 2500, \Delta = 1/250$ and assume that all parameters are known for generating forecasts. Thus, no estimation is needed. Then, we simulate 2500 observations for each replication. Alternative forecasting formulae are used to generate h-step-ahead-forecasts with $h = 1, 2, \ldots 5$, where the first $2500 - h$ observations are used to generate forecasts. Finally, we set the number of replications to 10000 and calculate RMSE. The theoretical RMSE of the optimal forecast can be obtained as the square root of the formula given in (3.6):

$$\text{Theoretical RMSE of the optimal forecast} = \sqrt{\gamma_{0\Delta} - \left(\gamma_{h\Delta}^{(0:n\Delta)}\right)^{\top} \Sigma^{-1} \left(\gamma_{h\Delta}^{(0:n\Delta)}\right)}.$$

All results are reported Table 7.

It is evident from Table 7 that the RMSE of the optimal forecast obtained from (3.5) is significantly lower than that of WXY and is close to the theoretical RMSE of the optimal forecast. This finding indicates that when a discrete sample is available, it is much better to use the conditional mean than the discretized formula of Fink et al. (2013) to perform out-of-sample forecasts.

Results in Table 7 are obtained from the fOU model with known parameters. In practice, parameters are always unknown. To compare the magnitude of the forecast error generated by alternative estimation methods and that generated by alternative forecasting formulae, we perform h-step-ahead-forecast for $h = 1, 2, \ldots 5$) using the following combinations of methods: (i) ML-estimated fOU together with the optimal forecast formula; (ii) MM-estimated fOU together with forecast formula from WXY; (iii) MM-estimated fOU together with the optimal forecast formula. The RMSE, obtained from 1000 replications, is reported in Table 8.[6] The ML estimate together with the optimal forecast formula consistently produces the lowest RMSE while the MM estimate with the forecast formula of WXY produces the highest RMSE.

[6]We also set $H = 0.260573, \kappa = 4.446145, \mu = -2.465673, \sigma = 1.172012, n = 2500, \Delta = 1/250$ for simulating paths and then simulate 2500 observations for each replication.

Table 5: Finite-sample properties of alternative estimation methods for (H, σ, μ, κ) with various values of $H \in [0.1, 0.8]$, $n = 2500$ and $\Delta = 1/250$.

		κ	H	μ	σ	::::	κ	H	μ	σ
Method	True value	4.446145	0.100000	-2.465673	1.172012	::::	4.446145	0.200000	-2.465673	1.172012
MM	Mean	8.156956	0.118790	-2.465673	1.172012	::::	6.712958	0.218491	-2.466795	1.186114
	SD	4.684829	0.026262	0.035257	0.178282	::::	2.417307	0.021499	0.039498	0.139869
MCL	Mean	4.377534	0.098558	-2.465673	1.172012	::::	4.719830	0.198960	-2.466795	1.059644
	SD	1.759199	0.010438	0.035257	0.060048	::::	1.488801	0.012786	0.039498	0.069837
AWML	Mean	4.386454	0.100685	-2.465673	1.122004	::::	4.728565	0.201129	-2.466795	1.071825
	SD	1.385191	0.008765	0.035257	0.050256	::::	1.134422	0.011305	0.039498	0.062491
ML	Mean	4.414050	0.099564	-2.465366	1.114195	::::	4.745714	0.199569	-2.467258	1.061235
	SD	1.302335	0.008337	0.035032	0.047689	::::	1.039778	0.010893	0.038043	0.059646
Method	True value	4.446145	0.300000	-2.465673	1.172012	::::	4.446145	0.400000	-2.465673	1.172012
MM	Mean	6.450341	0.320539	-2.469253	1.144823	::::	5.957557	0.417467	-2.468329	1.075316
	SD	2.046180	0.021207	0.046898	0.134499	::::	1.689577	0.024037	0.056664	0.141014
MCL	Mean	4.777123	0.297896	-2.469253	1.004397	::::	4.812129	0.397886	-2.468329	0.956567
	SD	1.398119	0.014596	0.046898	0.077195	::::	1.307882	0.015592	0.056664	0.081674
AWML	Mean	4.867030	0.301485	-2.469253	1.024141	::::	4.938722	0.401530	-2.468329	0.978457
	SD	1.147865	0.013025	0.046898	0.071010	::::	1.154986	0.014138	0.056664	0.076213
ML	Mean	4.943190	0.300428	-2.469217	1.015806	::::	4.950238	0.401050	-2.467967	0.970743
	SD	0.989681	0.012164	0.045083	0.069604	::::	0.989768	0.013187	0.055837	0.075642
Method	True value	4.446145	0.500000	-2.465673	1.172012	::::	4.446145	0.600000	-2.465673	1.172012
MM	Mean	5.726407	0.514194	-2.470754	1.009054	::::	5.696879	0.611020	-2.471359	0.948085
	SD	1.601530	0.024194	0.063526	0.142505	::::	1.473072	0.026258	0.079252	0.144096
MCL	Mean	5.298567	0.497599	-2.470754	0.910476	::::	5.175769	0.599687	-2.471359	0.877655
	SD	1.275910	0.016580	0.063526	0.086135	::::	1.247941	0.015706	0.079252	0.082508
AWML	Mean	4.760709	0.500665	-2.470754	0.932938	::::	4.540656	0.593493	-2.471359	0.867837
	SD	1.101151	0.013744	0.063526	0.072981	::::	1.137011	0.015748	0.079252	0.078219
ML	Mean	5.069599	0.501704	-2.471067	0.928959	::::	5.265544	0.604879	-2.471462	0.903227
	SD	0.867731	0.014314	0.061942	0.076531	::::	0.776515	0.015503	0.075429	0.072082
Method	True value	4.446145	0.700000	-2.465673	1.172012	::::	4.446145	0.800000	-2.465673	1.172012
MM	Mean	5.892033	0.708439	-2.468226	0.893418	::::	6.566816	0.803186	-2.471829	0.825635
	SD	1.621508	0.026426	0.092693	0.155205	::::	1.729948	0.023800	0.112849	0.147738
MCL	Mean	5.608306	0.702819	-2.468226	0.855271	::::	7.224908	0.813702	-2.471829	0.898949
	SD	1.395668	0.019954	0.092693	0.111477	::::	2.353120	0.026157	0.112849	0.190412
AWML	Mean	4.216008	0.674462	-2.468226	0.771517	::::	3.341247	0.718672	-2.471829	0.581523
	SD	1.106270	0.023748	0.092693	0.082672	::::	1.086899	0.051058	0.112849	0.112479
ML	Mean	5.529380	0.707212	-2.468218	0.875505	::::	5.813537	0.807403	-2.468201	0.842374
	SD	0.667337	0.016117	0.088656	0.080267	::::	0.470220	0.015519	0.106023	0.094978

Table 6: Finite-sample properties of alternative estimation methods for (H, σ, μ, κ) with various values of $H \in [0.1, 0.8]$, $n = 500$ and $\Delta = 1/250$.

		κ	H	μ	σ	::::	κ	H	μ	σ
Method	True value	4.446145	0.100000	-2.465673	1.172012	::::	4.446145	0.200000	-2.465673	1.172012
MM	Mean	4.890346	0.098944	-2.476788	1.255459	::::	5.600519	0.198973	-2.476219	1.245837
	SD	4.648676	0.071484	0.142294	0.498106	::::	5.033509	0.068318	0.149269	0.474832
MCL	Mean	5.321656	0.085616	-2.476788	1.120797	::::	5.672349	0.197606	-2.476219	1.174534
	SD	4.682424	0.049333	0.142294	0.242316	::::	4.087221	0.036959	0.149269	0.194905
AWML	Mean	5.376004	0.104625	-2.476788	1.205031	::::	6.760823	0.206847	-2.476219	1.225128
	SD	4.051980	0.019671	0.142294	0.119223	::::	3.661555	0.025436	0.149269	0.161697
ML	Mean	4.370875	0.098775	-2.476871	1.169990	::::	4.840050	0.195802	-2.476658	1.151613
	SD	1.899732	0.016819	0.128469	0.097643	::::	1.582566	0.022568	0.129050	0.131427
Method	True value	4.446145	0.300000	-2.465673	1.172012	::::	4.446145	0.400000	-2.465673	1.172012
MM	Mean	6.108450	0.298812	-2.479240	1.236516	::::	6.418008	0.398458	-2.480264	1.227081
	SD	4.626418	0.065207	0.156573	0.454829	::::	4.248231	0.062580	0.164935	0.436086
MCL	Mean	6.211983	0.300797	-2.479240	1.187773	::::	6.686706	0.402660	-2.480264	1.202163
	SD	3.860853	0.031967	0.156573	0.197988	::::	3.837283	0.034619	0.164935	0.226809
AWML	Mean	7.356816	0.308567	-2.479240	1.245800	::::	7.680359	0.409084	-2.480264	1.267572
	SD	3.599389	0.028792	0.156573	0.193882	::::	3.664320	0.030713	0.164935	0.220704
ML	Mean	4.989880	0.293346	-2.480837	1.135453	::::	5.015213	0.391997	-2.483494	1.125673
	SD	1.413288	0.025424	0.153477	0.150430	::::	1.323559	0.027568	0.163325	0.167783
Method	True value	4.446145	0.500000	-2.465673	1.172012	::::	4.446145	0.600000	-2.465673	1.172012
MM	Mean	6.752937	0.497856	-2.480863	1.218922	::::	7.203506	0.597288	-2.478460	1.211124
	SD	3.969060	0.060020	0.175457	0.430581	::::	3.937838	0.056607	0.186765	0.426452
MCL	Mean	7.345850	0.505780	-2.480863	1.227848	::::	8.270592	0.611950	-2.478460	1.286097
	SD	4.058504	0.037319	0.175457	0.259587	::::	4.575060	0.041023	0.186765	0.325562
AWML	Mean	6.689375	0.499791	-2.480863	1.232196	::::	5.554368	0.574889	-2.478460	1.142277
	SD	3.433822	0.025843	0.175457	0.190099	::::	3.109976	0.032238	0.186765	0.200919
ML	Mean	5.041528	0.490945	-2.486004	1.117037	::::	5.050404	0.590785	-2.485469	1.113263
	SD	1.240879	0.029039	0.169596	0.182116	::::	1.155138	0.029616	0.174071	0.194905
Method	True value	4.446145	0.700000	-2.465673	1.172012	::::	4.446145	0.800000	-2.465673	1.172012
MM	Mean	7.665527	0.696968	-2.479808	1.211898	::::	8.467536	0.797285	-2.481418	1.239616
	SD	3.834607	0.053985	0.196360	0.441561	::::	4.009864	0.051876	0.204945	0.531494
MCL	Mean	10.216937	0.725907	-2.479808	1.466824	::::	14.017574	0.846455	-2.481418	1.861672
	SD	6.252785	0.049906	0.196360	0.552863	::::	7.253496	0.049643	0.204945	0.698690
AWML	Mean	4.833263	0.625256	-2.479808	0.973818	::::	3.748846	0.632161	-2.481418	0.702181
	SD	2.782346	0.046913	0.196360	0.272416	::::	2.244837	0.041092	0.204945	0.147202
ML	Mean	5.030314	0.691695	-2.487919	1.118168	::::	4.916638	0.794567	-2.490307	1.145867
	SD	1.070439	0.030520	0.178050	0.217682	::::	1.144566	0.032204	0.179683	0.139612

Table 7: RMSE by the alternative forecasting methods of fOU for h-day-ahead-forecast.

h	1	2	3	4	5
WXY	0.28198	0.30920	0.39320	0.46225	0.55091
Optimal forecasts	0.20021	0.22527	0.24292	0.26828	0.28951
Theoretical RMSE of the optimal forecast	0.22048	0.26020	0.28081	0.30540	0.32085

Table 8: RMSE by the alternative estimation approaches and forecasting methods of fOU for h-day-ahead-forecast.

h	1	2	3	4	5
MM+WXY	0.30657	0.32874	0.36024	0.46515	0.51902
MM+optimal forecast	0.25943	0.27560	0.31933	0.33432	0.38616
ML+optimal forecast	0.21918	0.25905	0.27602	0.29334	0.31546

6 Empirical Studies

6.1 Estimation results

We fit the fOU model given in (2.3) to the logarithmic daily RV series for the Standard and Poor's (S&P) 500 index exchange-traded fund (ETF) and the nine industry ETFs, with the tick symbols SPY, XLY, XLP, XLE, XLF, XLV, XLI, XLB, XLK, and XLU. The sample period is from January 5, 2016, to December 31, 2020. We download the data from the Risk Lab of Dacheng Xiu.[7] The left panel of Table 9 provides the estimation results for the fOU Model using four different estimation methods: ML, AWML, MM, and MCL. The estimated H is less than 0.5 in all cases, suggesting all RV series are rough.[8] Interestingly, in most cases, the ML estimate of each parameter takes a value closer to the AMWL estimates and further away from their MCL and MM counterparts. The ML estimates of H range from 0.203121 for XLP to 0.260573 for SPY.

We also fit Model (2.3) to the logarithmic daily trading volume series for the same asset over the same period.[9] The results are reported in the right panel of Table 9. The empirical results are generally similar to what was found for the log RV series. For example, all estimates of H are much less than 0.5; the point estimates from ML are close to those from AWML. Moreover, the point estimates of H for volume series are

[7] `https://dachxiu.chicagobooth.edu/\#risklab`.

[8] However, it is important to note that this finding does not necessarily imply that the integrated volatility (IV) series is rough, for there are estimation errors in RV. Since one of our interests in modeling RV is to forecast future RV not IV, it is important for us to find a good model for RV for this particular purpose, as it is typically done in the literature; see, for example, Andersen et al. (2003).

[9] The trading volume data are downloaded from Yahoo Finance at the daily frequency.

smaller than those for their RV counterparts — the ML estimates of H range between 0.105369 for XLU and 0.186572 for XLI.

6.2 Forecasting performance of fOU

We now compare the performance of fOU in forecasting RV after it is fitted to the log RV series by alternative estimation methods: ML, MCL, and MM. When forecasting future RV, we replace the underlying parameters in fOU with these alternative estimates and use both the optimal forecasting formula and the forecasting formula of WXY. Then, we evaluate the forecasting performance of the competing approaches based on RMSE. We further check the statistical significance between competing methods using the Diebold-Mariano (DM) test of Diebold and Mariano (1995) and the model confidence set (MCS) of Hansen et al. (2011), respectively.

We split the sample period into two subperiods: from January 4, 2016, to December 31, 2020, and from January 4, 2021, to December 30, 2022. For each day in the second period, the fOU model is fitted to observations over the most recent 5 years using one of three estimation methods (ML, MCL, and MM), then, h-day-ahead (with $h = 1, 5$) forecasts of daily RVs are obtained using one of the two forecasting formulae. In total, we have six alternative combinations: (i) the ML estimate together with the optimal forecasting formula, (ii) the MCL estimate together with the optimal forecasting formula, (iii) the MM estimate together with the optimal forecasting formula, (iv) the ML estimate together with the forecasting formula of WXY, (v) the MCL estimate together with the forecasting formula of WXY, and (vi) the MM estimate together with the forecasting formula of WXY.

Table 10 reports RMSE from the six combinations above. The best result is highlighted in boldface.[10] It is evident that the ML estimates together with the optimal forecasting formula always performs the best.

To investigate if forecasts from the ML estimate together with the optimal forecasting formula are statistically significantly different from those of other methods, Table 11 reports the DM statistic based on the squared forecast errors and the p-value (in parenthesis) with the benchmark being the ML estimate together with the optimal forecasting formula (boldface means statistically significant at the 10% level). According to the DM test, regardless of forecasting horizons, forecasts by the ML estimate together with the optimal forecasting formula are statistically different from those by four out

[10] By Gaussianity of fOU, the h-step-ahead predictor of $RV_t = \exp(X_t)$ is $\widehat{RV}_{t+h} = \exp\left(\mathbb{E}[X_{t+h} \mid \mathcal{F}_t] + \frac{1}{2}\operatorname{Var}[X_{t+h} \mid \mathcal{F}_t]\right)$ where $\mathcal{F}$ denotes the sigma algebra of $X_0, X_1, \ldots, X_t$.

Table 9: Empirical results for $\ln(RV)$ and $\ln(volume)$ of SPY, XLY, XLP, XLE, XLF, XLV, XLI, XLB, XLK and XLU.

Name	Method	$\ln(RV)$				$\ln(volume)$			
		κ	H	μ	σ	κ	H	μ	σ
SPY	ML	4.446145	0.260573	-2.465673	1.172012	4.215973	0.157989	0.003779	0.797597
	AWML	4.606861	0.253570	-2.462327	1.491415	5.389794	0.166988	0.007365	1.491415
	MM	5.823982	0.298035	-2.462327	1.453865	4.175881	0.165010	0.007365	0.824279
	MCL	2.577467	0.246340	-2.462327	1.090190	4.357776	0.165575	0.007365	0.829422
XLB	ML	0.966455	0.204695	-2.091105	0.722440	12.728688	0.179482	0.001678	1.114306
	AWML	1.038395	0.211580	-2.134613	0.979263	11.446955	0.175527	0.006391	1.092911
	MM	1.963080	0.225311	-2.134613	0.810170	16.305692	0.192629	0.006391	1.196384
	MCL	1.152012	0.202513	-2.134613	0.715899	16.265946	0.192508	0.006391	1.192435
XLE	ML	2.137154	0.249324	-1.750971	0.751446	0.075829	0.130348	0.098003	0.696664
	AWML	1.550811	0.250046	-1.780489	0.869432	4.648608	0.145728	0.019849	0.753164
	MM	9.152197	0.383350	-1.780489	1.624795	8.281561	0.255476	0.019849	1.385716
	MCL	1.609472	0.215789	-1.780489	0.637802	5.392826	0.151157	0.019849	0.778478
XLF	ML	2.018703	0.222482	-1.952330	0.791754	8.033162	0.167030	0.003403	0.899647
	AWML	2.969242	0.233993	-2.044372	1.207683	9.092342	0.170730	0.002399	0.920720
	MM	8.361772	0.305953	-2.044372	1.266752	14.813948	0.205067	0.002399	1.106793
	MCL	2.385198	0.225746	-2.044372	0.809569	9.153149	0.172800	0.002399	0.928295
XLI	ML	1.459491	0.216333	-2.189787	0.813539	13.149749	0.186572	0.000596	1.064284
	AWML	1.709495	0.220056	-2.214138	1.046251	14.454120	0.192757	0.004057	1.101224
	MM	1.219409	0.210275	-2.214138	0.785424	17.077446	0.197713	0.004057	1.121563
	MCL	1.420062	0.216198	-2.214138	0.813290	22.161701	0.215806	0.004057	1.246909
XLK	ML	1.668544	0.220104	-2.143940	0.940939	9.847038	0.169502	0.000818	0.985028
	AWML	2.219921	0.225226	-2.162469	1.218899	8.432843	0.164798	0.000616	0.963668
	MM	5.903950	0.276237	-2.162469	1.286163	12.125780	0.187591	0.000616	1.092152
	MCL	2.731487	0.229932	-2.162469	0.994593	8.096917	0.164745	0.000616	0.959579
XLP	ML	0.524153	0.203121	-2.331152	0.744684	8.222255	0.141522	0.002616	0.883951
	AWML	0.619573	0.212002	2.384502	1.110369	9.101073	0.146887	0.004432	0.908799
	MM	0.701563	0.133317	-2.384502	0.502381	17.487814	0.134950	0.004432	0.844310
	MCL	0.636387	0.208451	2.384502	0.762438	15.317032	0.171262	0.004432	1.031199
XLU	ML	0.867652	0.218259	-2.021525	0.664237	0.865405	0.105369	-0.033946	0.645621
	AWML	0.975011	0.219525	-2.051622	0.939547	7.077768	0.123771	-0.002357	0.710282
	MM	0.905570	0.197875	-2.051622	0.595602	21.778640	0.182866	-0.002357	0.975192
	MCL	1.204449	0.208877	-2.051622	0.632280	11.383995	0.127858	-0.002357	0.772601
XLV	ML	2.842219	0.225034	-2.245561	0.816055	9.203585	0.165271	-0.000349	0.944645
	AWML	2.478489	0.226379	-2.247078	0.976172	9.037310	0.165609	0.004742	0.948439
	MM	3.012977	0.239137	-2.247078	0.885550	24.995825	0.238078	0.004742	1.417276
	MCL	1.764479	0.213311	-2.247078	0.767777	9.605133	0.166686	0.004742	0.953507
XLY	ML	2.607289	0.223217	-2.269736	0.908920	1.106239	0.121481	0.010246	0.768686
	AWML	2.134027	0.223342	-2.279545	1.182996	4.892674	0.133971	0.004820	0.820746
	MM	3.850248	0.257546	-2.279545	1.103246	7.493364	0.084882	0.004820	0.623389
	MCL	1.930542	0.221440	-2.279545	0.902707	5.534828	0.143275	0.004820	0.857084

Table 10: RMSE for h-day-ahead-forecast of RV of fOU using three different estimation methods and two different forecasting methods.

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	0.3315	0.3241	0.3114	0.3171	0.3001	0.3084	0.3144	0.3131	0.3184	0.3019
MCL+WXY	0.3272	0.3216	0.3054	0.3092	0.2994	0.2947	0.3090	0.3034	0.3107	0.2969
ML+WXY	0.3091	0.2904	0.3005	0.3059	0.2813	0.2867	0.2938	0.2911	0.2919	0.2906
MM+optimal	0.3155	0.3136	0.2931	0.2833	0.2824	0.2809	0.2892	0.2968	0.2917	0.2845
MCL+optimal	0.2973	0.2803	0.2751	0.2695	0.2743	0.2780	0.2657	0.2701	0.2793	0.2779
ML+optimal	**0.2905**	**0.2753**	**0.2673**	**0.2634**	**0.2699**	**0.2622**	**0.2516**	**0.2585**	**0.2527**	**0.2598**
Panel B: $h = 5$										
MM+WXY	0.3718	0.3693	0.3618	0.3677	0.3675	0.3563	0.3543	0.3596	0.3660	0.3574
MCL+WXY	0.3536	0.3468	0.3490	0.3485	0.3416	0.3358	0.3430	0.3347	0.3309	0.3472
ML+WXY	0.3624	0.3546	0.3521	0.3516	0.3482	0.3391	0.3488	0.3448	0.3521	0.3488
MM+optimal	0.3501	0.3460	0.3483	0.3434	0.3389	0.3317	0.3405	0.3308	0.3343	0.3433
MCL+optimal	0.3456	0.3356	0.3294	0.3327	0.3324	0.3220	0.3342	0.3251	0.3236	0.3397
ML+optimal	**0.3342**	**0.3306**	**0.3229**	**0.3246**	**0.3369**	**0.3239**	**0.3245**	**0.3234**	**0.3146**	**0.3287**

of five methods for all assets. They are statistically different from the MCL estimate together with the optimal forecasting formula for two assets.

To determine whether the predictive model belongs to the set of 'best' predictive models or not, Table 12 reports the p-value of MSC obtained from $2,000$ bootstrap iterations with a block length of 12. Values in boldface indicate that the model belongs to the confidence set of the best models. Moreover, the method with a p-value smaller than 10% should be removed from the best models's set. From Table 12, at the 1-day horizon, we can see that MM-estimate-with-the-forecasting-formula-of-WXY, the MCL-estimate-with-the-forecasting-formula-of-WXY, the ML-estimate-with-the-forecasting-formula-of-WXY, and the MM-estimate-with-the-optimal-forecasting-formula are always rejected. The MCL-estimate-with-the-optimal-forecasting-formula is rejected in one case. Whereas, the ML-estimate-together-with-the-optimal-forecasting-formula is not rejected in any case. Similar conclusions can be drawn at the 5-day horizon. We also investigate the performance of alternative methods in forecasting $\ln(RV)$ of these selected ten ETFs. These results are qualitatively unchanged and are reported in the Online Supplement.

7 Conclusion

How to estimate fOU has received a great deal of attention in the statistics literature, where a common assumption is that a continuous record of observations is available. In recent years, fOU has been found successful in modeling volatility and trading volume.

Table 11: DM statistic for h-day-ahead-forecast of RV of fOU using three different estimation methods and two different forecasting methods (the benchmark model is ML with optimal).

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	-3.5949	-3.2622	-3.6028	-3.7112	-3.2217	-3.1174	-3.2967	-3.3188	-3.4242	-3.5079
	(**0.0002**)	(**0.0006**)	(**0.0002**)	(**0.0001**)	(**0.0006**)	(**0.0009**)	(**0.0005**)	(**0.0005**)	(**0.0003**)	(**0.0002**)
MCL+WXY	-2.8147	-2.9058	-2.7270	-2.9134	-2.6324	-2.6975	-2.2785	-2.5469	-2.9575	-2.9649
	(**0.0024**)	(**0.0018**)	(**0.0032**)	(**0.0018**)	(**0.0042**)	(**0.0035**)	(**0.0113**)	(**0.0054**)	(**0.0016**)	(**0.0015**)
ML+WXY	-1.6576	-1.5286	-2.0572	-1.9854	-2.0430	-1.6419	-1.9218	-1.8138	-2.0244	-1.6369
	(**0.0487**)	(**0.0632**)	(**0.0198**)	(**0.0236**)	(**0.0205**)	(**0.0503**)	(**0.0273**)	(**0.0349**)	(**0.0215**)	(**0.0508**)
MM+optimal	-2.0246	-2.4706	-2.1793	-2.2472	-2.3003	-2.1062	-2.0945	-2.4157	-2.2922	-2.4595
	(**0.0215**)	(**0.0067**)	(**0.0147**)	(**0.0123**)	(**0.0107**)	(**0.0176**)	(**0.0181**)	(**0.0079**)	(**0.0109**)	(**0.0070**)
MCL+optimal	-1.2194	-1.0159	-1.1385	-1.0231	-1.0486	-1.4117	-1.1908	-1.1585	-1.4751	-1.0172
	(0.1114)	(0.1548)	(0.1275)	(0.1531)	(0.1472)	(**0.0790**)	(0.1169)	(0.1233)	(**0.0701**)	(0.1545)
Panel A: $h = 5$										
MM+WXY	-3.4121	-2.0637	-2.5538	-2.0923	-2.1943	-3.6469	-3.3897	-2.6342	-3.9004	-2.0689
	(**0.0003**)	(**0.0195**)	(**0.0053**)	(**0.0182**)	(**0.0141**)	(**0.0001**)	(**0.0003**)	(**0.0042**)	(**0.0000**)	(**0.0193**)
MCL+WXY	-2.4387	-1.9650	-2.1655	-1.9952	-2.1869	-2.4898	-2.4456	-2.6063	-2.7094	-1.8547
	(**0.0074**)	(**0.0247**)	(**0.0152**)	(**0.0230**)	(**0.0144**)	(**0.0064**)	(**0.0072**)	(**0.0046**)	(**0.0034**)	(**0.0318**)
ML+WXY	-1.8407	-1.5243	-1.8143	-1.2435	-1.9293	-1.3500	-1.9166	-1.5211	-1.6160	-1.4733
	(**0.0328**)	(**0.0637**)	(**0.0348**)	(**0.1068**)	(**0.0268**)	(**0.0885**)	(**0.0276**)	(**0.0641**)	(**0.0530**)	(**0.0703**)
MM+optimal	-2.2513	-1.7551	-2.0060	-1.8352	-1.9869	-2.4593	-2.0472	-1.6386	-1.6493	-1.7575
	(**0.0122**)	(**0.0396**)	(**0.0224**)	(**0.0332**)	(**0.0235**)	(**0.0070**)	(**0.0203**)	(**0.0506**)	(**0.0495**)	(**0.0394**)
MCL+optimal	-1.1510	-1.3187	-1.0636	-1.0923	-1.0509	-1.0476	-1.3043	-1.2029	-1.1925	-1.0507
	(0.1249)	(**0.0936**)	(0.1437)	(0.1373)	(0.1466)	(0.1474)	(**0.0961**)	(0.1145)	(0.1165)	(0.1467)

However, like most economic and financial variables, volatility and trading volume are measured at discrete time points. As a result, estimating fOU with a discrete sample becomes a new research focus in the fOU literature.

In the present paper, we first derive two analytical formulae for the autocovariance function of discretely sampled observations from fOU. These formulae facilitate calculating the likelihood function, the ML estimates, and the optimal forecast formula proposed in the paper in terms of accuracy and computational cost. Applying the derived analytical formula, we investigate how well the Whittle likelihood can approximate the true likelihood. Under empirically realistic settings, it is shown that the Whittle approximation can be far away from the exact likelihood function, especially when the sample size is small, say 500, or when H is large. Therefore, the AWML estimator for the parameters in fOU studied by Shi et al. (2024) may have a poor finite sample performance, although is asymptotic efficient. The MM and MCL estimators proposed in the literature are not efficient, for only using limit information of the likelihood function.

We, therefore, propose using the exact ML method to estimate the parameters in fOU and develop a long-span asymptotic theory for the ML estimator. Unlike the existing estimation approaches where the location parameter μ is estimated separately from other

Table 12: p-values of MSC for h-day-ahead-forecast of RV of fOU using three different estimation methods and two different forecasting methods (the benchmark model is ML with optimal).

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MCL+WXY	0.0178	0.0139	0.0124	0.0140	0.0110	0.0113	0.0194	0.0196	0.0158	0.0106
ML+WXY	0.0429	0.0976	**0.1321**	0.0715	0.0843	0.0669	**0.1149**	0.0711	0.0448	0.0951
MM+optimal	0.0735	0.0853	0.0687	0.0636	0.0647	0.0506	0.0709	0.0232	0.0218	0.0695
MCL+optimal	**0.2392**	**0.1893**	**0.2404**	**0.2567**	**0.1833**	**0.1676**	**0.1945**	**0.1978**	0.0936	**0.1262**
ML+optimal	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Panel B: $h = 5$										
MM+WXY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MCL+WXY	0.0143	0.0231	0.0501	0.0115	0.0364	0.0465	0.0344	0.0389	0.0219	0.0309
ML+WXY	0.0628	0.0768	**0.1121**	0.0831	0.0843	0.0618	**0.1185**	0.0999	0.0850	0.0830
MM+optimal	0.0330	0.0341	0.0821	0.0336	0.0561	0.0548	0.0709	0.0709	0.0661	0.0375
MCL+optimal	**0.2489**	**0.2278**	**0.2959**	**0.2473**	**0.2700**	**0.2181**	**0.2149**	**0.2738**	**0.1625**	**0.1700**
ML+optimal	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**

parameters, the ML approach simultaneously estimates μ and other parameters in fOU. It has been proved that the ML estimator is consistent and has asymptotic normality for all values of $H \in (0, 1)$, with the convergence rate and the asymptotic variance being continuous functions of H. This feature facilitates statistical inference and is in sharp contrast with the asymptotics of the MM and MCL estimators, whose asymptotic theories have a jump at the point of $H = 3/4$.

Simulations are performed to demonstrate the feasibility and effectiveness of the ML estimation approach and the optimal forecast formula proposed when a discrete sample is available. Simulation results show that the ML method outperforms the MM, MCL, and AWML methods in finite samples, and the proposed optimal forecast formula performs better than the other forecasting methods applied in the literature.

When fitting the fOU process to the RV and trading volume series of ten ETFs, strong evidence of $H < 0.5$ is found. Moreover, empirical studies also show that forecasts generated using the optimal forecasting formula with parameter estimates from the ML approach are significantly more accurate than those generated using combinations of other forecasting formulas and estimation methods.

In practice, we suggest empirical researchers to first use MCL or AWML or even MM methods to obtain initial estimates of parameters in fOU. Then our exact ML method is used to the final estimates of all four parameters in fOU. When implementing the exact ML method, the MCL or AWML or MM estimates can serve as the initial value during

the numerical optimization.

There is much room for future research. First, our results focus on the stationary fOU process. The exact ML estimator for the explosive fOU process has not been investigated. We plan to pursue this line of research in future work. Second, this paper considers the univariate Gaussian fOU process. It is worthwhile to consider relaxing the Gaussianity assumption and extending the results to the multivariate case. Both generalizations seem rather complicated. Third, investigating the finite sample bias of the ML estimators would be meaningful work.

8 Appendix

Proof of Lemma 2.1: (a) Let us first prove (2.12). Let $X_0 = \mu + \sigma \int_{-\infty}^{0} e^{\kappa s} dB_s^H$, which has the distribution of $\mathcal{N}\left(\mu, \sigma^2 H\Gamma\left(2H\right)/\kappa^{2H}\right)$ when $\kappa < 0$. Then the fOU process defined in (2.3) has the following discretization

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

where the second equation is obtained by using integration by parts.

Let $\widetilde{X}_t = X_t - \mu, S_t = \sigma B_t^H - \kappa\sigma e^{-\kappa t} \int_0^t e^{\kappa s} B_s^H ds$. It is straightforward to get that

$$\widetilde{X}_t = e^{-\kappa t}\widetilde{X}_0 + S_t\,, \tag{8.1}$$

and

$$\begin{aligned} &\mathrm{Cov}\left(X_t, X_s\right) = \mathrm{Cov}\left(\widetilde{X}_t, \widetilde{X}_s\right) \\ =& \mathrm{Cov}\left(S_t, S_s\right) + e^{-\kappa t}\,\mathrm{Cov}\left(\widetilde{X}_0, S_s\right) + e^{-\kappa s}\,\mathrm{Cov}\left(S_t, \widetilde{X}_0\right) + e^{-\kappa(t+s)}\,\mathrm{Var}\left(\widetilde{X}_0\right). \end{aligned} \tag{8.2}$$

As $\{X_t\}$ is a covariance stationary process when $\kappa < 0$, without losing generality, we only derive the expression of $\mathrm{Cov}\left(X_t, X_s\right)$ for $s < t$. From (2.2), we have

$$\begin{aligned} \mathrm{Cov}(S_t, S_s) &= \mathbb{E}\left[\left(-\kappa\sigma e^{-\kappa t}\int_0^t e^{\kappa u} B_u^H du + \sigma B_t^H\right)\left(-\kappa\sigma e^{-\kappa s}\int_0^s e^{\kappa v} B_v^H dv + \sigma B_s^H\right)\right] \\ &= \frac{\sigma^2}{2}\sum_{n=1}^{10} I_n, \end{aligned}$$

where

$$I_1 = -\kappa e^{-\kappa t}\int_0^t e^{\kappa u} s^{2H} du,\ I_2 = -\kappa e^{-\kappa t}\int_0^t e^{\kappa u} u^{2H} du,\ I_3 = \kappa e^{-\kappa t}\int_0^t e^{\kappa u}|u-s|^{2H} du,$$

$$I_4 = -\kappa e^{-\kappa s}\int_0^s e^{\kappa v}t^{2H}dv,\ I_5 = -\kappa e^{-\kappa s}\int_0^s e^{\kappa v}v^{2H}dv,\ I_6 = \kappa e^{-\kappa s}\int_0^s e^{\kappa v}(t-v)^{2H}dv,$$
$$I_7 = t^{2H}+s^{2H}-(t-s)^{2H},\quad I_8 = \kappa^2 e^{-\kappa(t+s)}\int_0^t e^{\kappa v}dv\int_0^s e^{\kappa u}u^{2H}du,$$
$$I_9 = \kappa^2 e^{-\kappa(t+s)}\int_0^s e^{\kappa u}du\int_0^t e^{\kappa v}v^{2H}dv,$$
$$I_{10} = -\kappa^2 e^{-\kappa(t+s)}\int_0^t\int_0^s e^{\kappa(u+v)}|u-v|^{2H}\,du\,dv.$$

It is easy to get

$$I_1 = s^{2H}\left(e^{-\kappa t}-1\right)\ \text{ and }\ I_2 = -t^{2H}+2He^{-\kappa t}\int_0^t e^{\kappa u}u^{2H-1}du.$$

Using the change-of-variable technique by letting $z=|s-u|$, we can have

$$\begin{aligned}
I_3 =&\kappa e^{-\kappa t}\left(\int_0^s e^{\kappa u}(s-u)^{2H}du+\int_s^t e^{\kappa u}(u-s)^{2H}du\right)\\
=&\kappa e^{-\kappa t+\kappa s}\left(\int_0^s e^{-\kappa z}z^{2H}dz+\int_0^{t-s} e^{\kappa z}z^{2H}dz\right)\\
=&-e^{-\kappa t+\kappa s}\left(e^{-\kappa s}s^{2H}-2H\int_0^s e^{-\kappa z}z^{2H-1}dz-e^{\kappa(t-s)}(t-s)^{2H}\right.\\
&\left.+2H\int_0^{t-s}e^{\kappa z}z^{2H-1}dz\right)\\
=&-e^{-\kappa t}s^{2H}+(t-s)^{2H}+2He^{-\kappa t+\kappa s}\left(\int_0^s e^{-\kappa z}z^{2H-1}dz-\int_0^{t-s}e^{\kappa z}z^{2H-1}dz\right).
\end{aligned}$$

Similarly, simple calculations with the usage of the change-of-variable technique and integration by parts yield the results of

$$\begin{aligned}
I_4 &= t^{2H}\left(e^{-\kappa s}-1\right),\\
I_5 &= -s^{2H}+2He^{-\kappa s}\int_0^s e^{\kappa v}v^{2H-1}dv,\\
I_6 &= -e^{-\kappa s}t^{2H}+(t-s)^{2H}+2He^{-\kappa s+\kappa t}\int_{t-s}^t e^{-\kappa z}z^{2H-1}dz,\\
I_8 &= (1-e^{-\kappa t})s^{2H}-2He^{-\kappa s}(1-e^{-\kappa t})\int_0^s e^{\kappa u}u^{2H-1}du,\\
I_9 &= (1-e^{-\kappa s})t^{2H}-2He^{-\kappa t}(1-e^{-\kappa s})\int_0^t e^{\kappa v}v^{2H-1}dv.
\end{aligned}$$

To calculate the term I_{10} that involves a double integral, we first get

$$I_{10} = -\kappa^2 e^{-\kappa(t+s)}\left(\int_0^s\int_0^s e^{\kappa(u+v)}\,|u-v|^{2H}\,du\,dv+\int_s^t\int_0^s e^{\kappa(u+v)}(v-u)^{2H}\,du\,dv\right)$$

$$= -\kappa^2 e^{-\kappa(t+s)} \left(2\int_0^s \int_0^v e^{\kappa(u+v)}(v-u)^{2H}\, du\, dv + \int_s^t \int_0^s e^{\kappa(u+v)}(v-u)^{2H}\, du\, dv \right)$$
$$= I_{10}^{(1)} + I_{10}^{(2)}.$$

where $I_{10}^{(1)} \equiv -2\kappa^2 e^{-\kappa(t+s)} \int_0^s \int_0^v e^{\kappa(u+v)}(v-u)^{2H}\, du\, dv$ and $I_{10}^{(2)} \equiv -\kappa^2 e^{-\kappa(t+s)} \int_s^t \int_0^s e^{\kappa(u+v)}(v-u)^{2H}\, du\, dv$. Letting $z = v - u$, it has

$$\begin{aligned}
I_{10}^{(1)} &= -2\kappa^2 e^{-\kappa(t+s)} \int_0^s \int_0^v e^{2\kappa v} e^{-\kappa z} z^{2H} dz dv \\
&= -2\kappa^2 e^{-\kappa(t+s)} \int_0^s \int_z^s e^{2\kappa v} e^{-\kappa z} z^{2H} dv dz \\
&= -2\kappa^2 e^{-\kappa(t+s)} \int_0^s e^{-\kappa z} z^{2H} \frac{e^{2\kappa s} - e^{2\kappa z}}{2\kappa} dz \\
&= -\kappa e^{-\kappa(t+s)} \left(e^{2\kappa s} \int_0^s e^{-\kappa z} z^{2H} dz - \int_0^s e^{\kappa z} z^{2H} dz \right) \\
&= 2e^{-\kappa t} s^{2H} - 2He^{-\kappa(t-s)} \int_0^s e^{-\kappa z} z^{2H-1} dz - 2He^{-\kappa(t+s)} \int_0^s e^{\kappa z} z^{2H-1} dz,
\end{aligned}$$

where the second equation is obtained by changing the order of the integration. To simplify $I_{10}^{(2)}$, we derive the result under the condition of $t > 2s$. The same result can be obtained when $s < t < 2s$ by taking the same procedure with some tedious calculations, which we omit here for simplicity. Again, by letting $z = v - u$, it has

$$\begin{aligned}
I_{10}^{(2)} &= -\kappa^2 e^{-\kappa(t+s)} \int_s^t \int_{v-s}^v e^{-\kappa z + 2\kappa v} z^{2H} dz dv \\
&= -\kappa^2 e^{-\kappa(t+s)} \left\{ \left(\int_0^s \int_s^{z+s} + \int_s^{t-s} \int_z^{z+s} + \int_{t-s}^t \int_z^t \right) e^{-\kappa z + 2\kappa v} z^{2H} dv dz \right\} \\
&= -\kappa^2 e^{-\kappa t - \kappa s} \left(\int_0^s e^{-\kappa z} z^{2H} \frac{e^{2\kappa(z+s)} - e^{2\kappa s}}{2\kappa}\, dz \right. \\
&\left. + \int_s^{t-s} e^{-\kappa z} z^{2H} \frac{e^{2\kappa(z+s)} - e^{2\kappa z}}{2\kappa} dz + \int_{t-s}^t e^{-\kappa z} z^{2H} \frac{e^{2\kappa t} - e^{2\kappa z}}{2\kappa}\, dz \right) \\
&= -\frac{\kappa}{2} e^{-\kappa(t-s)} \int_0^{t-s} e^{\kappa z} z^{2H} dz + \frac{\kappa}{2} e^{-\kappa(t+s)} \int_s^t e^{\kappa z} z^{2H} dz \\
&+ \frac{\kappa}{2} e^{-\kappa(t-s)} \int_0^s e^{-\kappa z} z^{2H} dz - \frac{\kappa}{2} e^{\kappa(t-s)} \int_{t-s}^t e^{-\kappa z} z^{2H} dz \\
&= e^{-\kappa s} t^{2H} - e^{-\kappa t} s^{2H} - He^{-\kappa s + \kappa t} \int_{t-s}^t e^{-\kappa z} z^{2H-1} dz + He^{-\kappa t + \kappa s} \int_0^s e^{-\kappa z} z^{2H-1} dz \\
&- He^{-\kappa t - \kappa s} \int_s^t e^{\kappa z} z^{2H-1} dz + He^{-\kappa t + \kappa s} \int_0^{t-s} e^{\kappa z} z^{2H-1} dz - (t-s)^{2H}.
\end{aligned}$$

where the second equation is from changing the order of the integration. Finally, we get

$$\begin{aligned}\mathrm{Cov}(S_t, S_s) &= \frac{\sigma^2}{2}\sum_{n=1}^{10} I_n \\ &= \frac{H\sigma^2}{2}\left(-e^{-\kappa(t-s)}\int_0^{t-s} e^{\kappa z} z^{2H-1}dz + e^{\kappa(t-s)}\int_{t-s}^{t} e^{-\kappa z} z^{2H-1}dz\right. \\ &\quad - e^{-\kappa(t+s)}\int_s^t e^{\kappa z} z^{2H-1}dz + e^{-\kappa(t-s)}\int_0^s e^{-\kappa z} z^{2H-1}dz \\ &\quad \left. +2e^{-\kappa(t+s)}\int_0^t e^{\kappa z} z^{2H-1}dz\right). \end{aligned} \tag{8.3}$$

By taking a similar procedure as above, we can derive

$$\begin{aligned}\mathrm{Cov}\left(S_t, \widetilde{X}_0\right) &= \frac{H\sigma^2}{2}\left(-\frac{\Gamma(2H)}{\kappa^{2H}}e^{-\kappa t} - e^{-\kappa t}\int_0^t e^{\kappa z} z^{2H-1}dz + e^{\kappa t}\int_t^\infty e^{-\kappa z} z^{2H-1}dz\right), \\ \mathrm{Cov}\left(S_s, \widetilde{X}_0\right) &= \frac{H\sigma^2}{2}\left(-\frac{\Gamma(2H)}{\kappa^{2H}}e^{-\kappa s} - e^{-\kappa s}\int_0^s e^{\kappa z} z^{2H-1}dz + e^{\kappa s}\int_s^\infty e^{-\kappa z} z^{2H-1}dz\right), \\ \mathrm{Var}\left(\widetilde{X}_0\right) &= \frac{\sigma^2 H\Gamma(2H)}{\kappa^{2H}}.\end{aligned}$$

Substituting the formulae of $\mathrm{Cov}(S_t, S_s)$, $\mathrm{Cov}\left(S_t, \widetilde{X}_0\right)$, $\mathrm{Cov}\left(S_s, \widetilde{X}_0\right)$, and $\mathrm{Var}\left(\widetilde{X}_0\right)$ derived above into Equation (8.2), together with the result of $\int_0^\infty e^{-\kappa z} z^{2H-1}dz = \frac{\Gamma(2H)}{\kappa^{2H}}$, we can have

$$\mathrm{Cov}\left(X_t, X_s\right) = A_1 + A_2 + A_3, \tag{8.4}$$

where

$$\begin{aligned}A_1 &= \frac{H\sigma^2\Gamma(2H)}{2\kappa^{2H}}e^{-\kappa(t-s)}, \\ A_2 &= -\frac{H\sigma^2}{2}e^{-\kappa(t-s)}\int_0^{t-s} e^{\kappa z} z^{2H-1}dz, \\ A_3 &= \frac{H\sigma^2}{2}e^{\kappa(t-s)}\int_{t-s}^{+\infty} e^{-\kappa z} z^{2H-1}dz.\end{aligned}$$

Letting $x = \kappa z$, we then have

$$\begin{aligned}A_2 &= -\frac{H\sigma^2 e^{-\kappa(t-s)}}{2\kappa^{2H}}\int_0^{\kappa(t-s)} e^x x^{2H-1}dx \\ &= -\frac{H\sigma^2 e^{-\kappa(t-s)}\left(\kappa\left(t-s\right)\right)^{2H}}{2\kappa^{2H}}\mathrm{B}\left(1, 2H\right){}_1F_1\left(2H; 1+2H; \kappa\left(t-s\right)\right) \\ &= -\frac{H\sigma^2 e^{-\kappa(t-s)}\left(\kappa\left(t-s\right)\right)^{2H}}{2\kappa^{2H}}\frac{1}{2H}{}_1F_1\left(2H; 1+2H; \kappa\left(t-s\right)\right),\end{aligned} \tag{8.5}$$

where the second equation comes from Gradshteyn and Ryzhik (2007) (see, Eq ET II 187(14) of 3.383 on page 347)

$$\int_0^u x^{\nu-1}(u-x)^{\mu-1}e^{\beta x}dx = \mathrm{B}(\mu,\nu)u^{\mu+\nu-1}{}_1F_1(\nu;\mu+\nu;\beta u)\,, \tag{8.6}$$

with $\mathrm{B}(\cdot,\cdot)$ and ${}_1F_1(\cdot;\cdot;\cdot)$ denoting the Beta function and the confluent hypergeometric function of the first kind, and the third equation is from

$$\mathrm{B}(1,2H) = \frac{\Gamma(1)\Gamma(2H)}{\Gamma(2H+1)} = \frac{1}{2H}\,.$$

Similarly, by letting $x=\kappa z$, we get

$$A_3 = \frac{H\sigma^2 e^{\kappa(t-s)}}{2\kappa^{2H}}\int_{\kappa(t-s)}^{+\infty} e^{-x}x^{2H-1}dx = \frac{H\sigma^2 e^{\kappa(t-s)}}{2\kappa^{2H}}\Gamma(2H,\kappa(t-s))\,. \tag{8.7}$$

where $\Gamma(\cdot,\cdot)$ is the upper incomplete Gamma function.

Finally, substituting (8.5) and (8.7) into (8.4), we can have

$$\begin{aligned}\mathrm{Cov}(X_t,X_s) =& A_1+A_2+A_3\\ =&\frac{H\sigma^2e^{-\kappa(t-s)}}{2\kappa^{2H}}\Big[\Gamma(2H)-\frac{(\kappa(t-s))^{2H}}{2H}{}_1F_1(2H;1+2H;\kappa(t-s))\\ &+e^{2\kappa(t-s)}\Gamma(2H,\kappa(t-s))\Big].\end{aligned} \tag{8.8}$$

Replacing t and s in Equation (8.8) with $(t+j)\,\Delta$ and $t\Delta$, respectively, the analytical expression in (2.12) is obtained.

(b): First, it is easy to verify that

$$\begin{aligned}\Gamma(2H;\kappa j\Delta) &= \Gamma(2H)-\int_0^{\kappa j\Delta}t^{2H-1}e^{-t}dt = \Gamma(2H)-(\kappa j\Delta)^{2H}\int_0^1 s^{2H-1}e^{-\kappa j\Delta s}ds\\ &= \Gamma(2H)-\frac{(\kappa j\Delta)^{2H}}{2H}{}_1F_1(2H;2H+1;-\kappa j\Delta)\,,\end{aligned} \tag{8.9}$$

where the second equation is obtained by letting $t=\kappa j\Delta s$ and the third equation comes from the definition of the ${}_1F_1$ function.

Using (8.9), we can have

$$\begin{aligned}\mathrm{Cov}\left(X_{t\Delta},X_{(t+j)\Delta}\right) =&\frac{H\sigma^2e^{-\kappa j\Delta}}{2\kappa^{2H}}\Big\{\Gamma(2H)-\frac{(\kappa j\Delta)^{2H}}{2H}{}_1F_1(2H;2H+1;\kappa j\Delta)\\ &e^{2\kappa j\Delta}\Gamma(2H)-e^{2\kappa j\Delta}\frac{(\kappa j\Delta)^{2H}}{2H}{}_1F_1(2H;2H+1;-\kappa j\Delta)\Big\}\,.\end{aligned} \tag{8.10}$$

Moreover, by letting $s = 1 - t$ and using the definition of the ${}_1F_1$ function, it is easy to show that

$$\begin{aligned} {}_1F_1(2H; 2H+1; \kappa j\Delta) &= 2H \int_0^1 t^{2H-1} e^{\kappa j\Delta t} dt = 2He^{\kappa j\Delta} \int_0^1 (1-s)^{2H-1} e^{-\kappa j\Delta s} ds \\ &= e^{\kappa j\Delta} {}_1F_1(1; 2H+1; -\kappa j\Delta)\,. \end{aligned} \tag{8.11}$$

Similarly, we can prove that

$${}_1F_1(2H; 2H+1; -\kappa j\Delta) = e^{-\kappa j\Delta} {}_1F_1(1; 2H+1; \kappa j\Delta)\,. \tag{8.12}$$

Substituting (8.11) and (8.12) into (8.10), we can rewrite $\mathrm{Cov}\left(X_{t\Delta}, X_{(t+j)\Delta}\right)$ as

$$\begin{aligned} \mathrm{Cov}\left(X_{t\Delta}, X_{(t+j)\Delta}\right) =& \frac{H\sigma^2}{\kappa^{2H}} \left\{ \frac{e^{-\kappa j\Delta} + e^{\kappa j\Delta}}{2} \Gamma(2H) \right. \\ & \left. - \frac{(\kappa j\Delta)^{2H}}{2H} \left[{}_1F_1(1; 2H+1; -\kappa j\Delta) + {}_1F_1(1; 2H+1; \kappa j\Delta) \right] \right\}. \end{aligned} \tag{8.13}$$

Moreover, using the well-known result of ${}_1F_1(a; b; z) = \sum_{n=0}^{\infty} \frac{(a)_n}{(b)_n} \frac{z^n}{n!}$, we can obtain the following results

$${}_1F_1(1; 2H+1; -\kappa j\Delta) = \sum_{n=0}^{\infty} \frac{(1)_n}{(2H+1)_n} \frac{(-\kappa j\Delta)^n}{n!} = \sum_{n=0}^{\infty} \frac{\Gamma(2H+1)}{\Gamma(2H+1+n)} (-\kappa j\Delta)^n\,, \tag{8.14}$$

and

$${}_1F_1(1; 2H+1; \kappa j\Delta) = \sum_{n=0}^{\infty} \frac{\Gamma(2H+1)}{\Gamma(2H+1+n)} (\kappa j\Delta)^n\,, \tag{8.15}$$

where $(2H+1)_n = (2H+1)(2H+2)\cdots(2H+n) = \frac{\Gamma(2H+1+n)}{\Gamma(2H+1)}$.

Therefore, using (8.14) and (8.15), we have

$$\begin{aligned} & {}_1F_1(1; 2H+1; -\kappa j\Delta) + {}_1F_1(1; 2H+1; \kappa j\Delta) \\ =& 2\sum_{n=0}^{\infty} \frac{\Gamma(2H+1)}{\Gamma(2H+1+2n)} (\kappa j\Delta)^{2n} \\ =& 2\sum_{n=0}^{\infty} \frac{\Gamma(H+1/2)\Gamma(H+1)}{\Gamma(H+n+1/2)\Gamma(H+n+1)} \left(\frac{\kappa^2 (j\Delta)^2}{4} \right)^n \\ =& 2{}_1F_2(1; H+1/2, H+1; \frac{1}{4}\kappa^2 (j\Delta)^2)\,, \end{aligned} \tag{8.16}$$

where the second equation is obtained by using $\Gamma(2x) = \frac{2^{2x-1}}{\sqrt{\pi}}\Gamma(x)\Gamma\left(x+\frac{1}{2}\right)$ (see, Gradshteyn and Ryzhik, 2007, Eq. (8.335.1)) and the third equation comes from the definition of the generalized hypergeometric function ${}_1F_2$ as

$$\begin{aligned}
&{}_1F_2(1; H+1/2, H+1; \frac{1}{4}\kappa^2 (j\Delta)^2) \\
=&\sum_{n=0}^{\infty} \frac{(1)_n}{(H+1/2)_n (H+1)_n} \frac{1}{n!} \left(\frac{\kappa^2 (j\Delta)^2}{4}\right)^n \\
=&\sum_{n=0}^{\infty} \frac{\left(\kappa^2 (j\Delta)^2/4\right)^n}{[(H+1/2)\cdots(H+1/2+n-1)]\,[(H+1)\cdots(H+1+n-1)]} \\
=&\sum_{n=0}^{\infty} \frac{\Gamma(H+1/2)\Gamma(H+1)}{\Gamma(H+n+1/2)\Gamma(H+n+1)} \left(\frac{\kappa^2 (j\Delta)^2}{4}\right)^n .
\end{aligned}$$

Finally, using (8.13) and (8.16), we get that the covariance function given in (2.14). This completes the proof of Lemma 2.1.

Proof of Theorem 4.1: The fOU process $\{X_{j\Delta}\}$ is a short-memory stationary process when $0 < H \leq 0.5$. The asymptotic properties of the ML estimate for short stationary processes have been well established in the literature; see, e.g., Hannan (1973). Hence, we focus on the proof for the long-memory case where $0.5 < H < 1$.

Define $\mathcal{F}_n(\theta) \equiv -l_n''(\theta)$, minus one multiplying the second-order derivative of the log-likelihood function $l_n(\theta)$ defined in (4.1) with respect to the parameter vector $\theta = (\mu, \sigma^2, \kappa, H)^\top$. Let $A_n(\theta) = \text{diag}(n^{1-H}, \sqrt{n}, \sqrt{n}, \sqrt{n})$. Sweeting (1980) proves that the ML estimate $\widehat{\theta}_{ML}$ has asymptotic normality as

$$A_n(\theta)\left(\widehat{\theta}_{ML} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \mathcal{I}^{-1}(\theta)\right), \tag{8.17}$$

if the following two conditions are satisfied:

(C1) When $n \to \infty$, it has

$$\mathcal{I}_n(\theta) \equiv \{A_n(\theta)\}^{-1} \mathcal{F}_n(\theta) \left[\{A_n(\theta)\}^{-1}\right]^\top \xrightarrow{p} \mathcal{I}(\theta), \tag{8.18}$$

where $\mathcal{I}(\theta)$ is a positive definite matrix with probability one.

(C2) For all $c > 0$, it has

$$\sup \left\| \{A_n(\theta)\}^{-1} A_n(\theta^*) - \mathbf{I}_4 \right\| \xrightarrow{p} 0, \tag{8.19}$$

where $\mathbf{I}_4$ is the 4×4 identity matrix, the sup is over the set $\left\|\{A_n(\theta)\}^\top (\theta^* - \theta)\right\| \leq c$ with $\|\cdot\|$ denoting the Euclidean norm of a matrix, and

$$\sup \left\| \{A_n(\theta)\}^{-1} \left[\mathcal{F}_n(\Theta) - \mathcal{F}_n(\theta)\right] \left[\{A_n(\theta)\}^{-1}\right]^\top \right\| \xrightarrow{p} 0, \tag{8.20}$$

where $\mathcal{F}_n(\Theta)$ is defined as $\mathcal{F}_n$ with row i evaluated at θ_i^*, for $i = 1, 2, 3, 4$, and the sup is over the set $\left\| \{A_n(\theta)\}^\top (\theta_i^* - \theta) \right\| \leq c$.

To prove the two conditions above being satisfied for the fOU process with $0.5 < H < 1$, we first list some asymptotic properties of the Toeplitz matrice $\sigma^2\Sigma$, which will be frequently applied later. For each $\delta > 0$ with a constant K independent of $\beta = (\sigma^2, \kappa, H)^\top$ and n, as $n \to \infty$, it has

$$n^{2H-2}\left(\mathbf{1}^\top\Sigma^{-1}\mathbf{1}\right) \to \frac{B(3/2-H, 3/2-H)}{\Gamma(2-2H)}\frac{\kappa^2}{C(H)\Delta^{2H-2}}, \tag{8.21}$$

$$\mathbf{1}^\top\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}\mathbf{1} \leq K \cdot n^{2-2H+\delta}, \tag{8.22}$$

$$\mathbf{1}^\top\Sigma^{-1}\frac{\partial\Sigma}{\partial H}\Sigma^{-1}\frac{\partial\Sigma}{\partial H}\Sigma^{-1}\mathbf{1} \leq K \cdot n^{2-2H+\delta}, \tag{8.23}$$

$$\mathbf{1}^\top\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\kappa\partial\kappa}\Sigma^{-1}\mathbf{1} \leq K \cdot n^{2-2H+\delta}, \tag{8.24}$$

where the first limit can be obtained from Theorem 5.2 in Adenstedt (1974), and the other three inequalities come from Lemma 5.4 (d) in Dahlhaus (1989).

We start from proving the condition (C1). Consider the elements in the first row of $\mathcal{I}_n(\theta)$, i.e., $-\left(\frac{\partial^2 l_n(\theta)}{\partial\mu\partial\mu}, \frac{\partial^2 l_n(\theta)}{\partial\sigma^2\partial\mu}, \frac{\partial^2 l_n(\theta)}{\partial\kappa\partial\mu}, \frac{\partial^2 l_n(\theta)}{\partial H\partial\mu}\right)$. When $n \to \infty$, using (8.21), we have

$$-n^{2H-2}\frac{\partial^2 l_n(\theta)}{\partial\mu\partial\mu} = n^{2H-2}\frac{1}{\sigma^2}\mathbf{1}^\top\Sigma^{-1}\mathbf{1} \to \frac{B(3/2-H, 3/2-H)}{\Gamma(2-2H)}\frac{\kappa^2}{\sigma^2 C(H)\Delta^{2H-2}}, \tag{8.25}$$

and

$$-n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial\sigma^2\partial\mu} = n^{H-3/2}\frac{1}{\sigma^4}\mathbf{1}^\top\Sigma^{-1}(\mathbf{X} - \mu\mathbf{1}) \xrightarrow{p} 0, \tag{8.26}$$

where the last limit comes from the fact that

$$\begin{cases} \mathbb{E}\left(n^{H-3/2}\dfrac{\partial^2 l_n(\theta)}{\partial\sigma^2\partial\mu}\right) = 0, \\ \mathrm{Var}\left(n^{H-3/2}\dfrac{\partial^2 l_n(\theta)}{\partial\sigma^2\partial\mu}\right) = n^{2H-3}\dfrac{1}{\sigma^8}\mathbf{1}^\top\Sigma^{-1}\mathbf{1} = n^{2H-3}\dfrac{1}{\sigma^8}n^{2-2H} \to 0, \end{cases}$$

as $n \to \infty$.

Moreover, as $n \to \infty$, using (8.22), we also have

$$\begin{aligned} -n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial\kappa\partial\mu} &= -n^{H-3/2}\frac{1}{\sigma^2}\mathbf{1}^\top\frac{\partial\Sigma^{-1}}{\partial\kappa}(\mathbf{X} - \mu\mathbf{1}) \\ &= n^{H-3/2}\frac{1}{\sigma^2}\mathbf{1}^\top\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}(\mathbf{X} - \mu\mathbf{1}) \xrightarrow{p} 0, \end{aligned} \tag{8.27}$$

because

$$\mathbb{E}\left(n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial\kappa\partial\mu}\right) = 0, \ \mathrm{Var}\left(n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial\kappa\partial\mu}\right) = n^{2H-3}\frac{1}{\sigma^4}\mathbf{1}^\top\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}\mathbf{1} \to 0.$$

Similarly, as $n \to \infty$, it can be proved that

$$-n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial H \partial \mu} \xrightarrow{p} 0\,. \tag{8.28}$$

From A.2, we can see that $f_X^{\Delta}(\lambda;\beta)$ satisfies the Assumption 2.1 of Cohen et al. (2013). Hence, for the other elements in the matrix $\mathcal{I}_n(\theta)$, by using Lemma 2.6 in Cohen et al. (2013), as $n \to \infty$, it is easy to get

$$\begin{aligned}
&-\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{n}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{n}} \end{pmatrix}
\begin{pmatrix} \frac{\partial^2 \ell_n(\theta)}{\partial \sigma^2 \partial \sigma^2} & \frac{\partial^2 \ell_n(\theta)}{\partial \sigma^2 \partial \kappa} & \frac{\partial^2 \ell_n(\theta)}{\partial \sigma^2 \partial H} \\ \frac{\partial^2 \ell_n(\theta)}{\partial \sigma^2 \partial \kappa} & \frac{\partial^2 \ell_n(\theta)}{\partial \kappa \partial \kappa} & \frac{\partial^2 \ell_n(\theta)}{\partial H \partial \kappa} \\ \frac{\partial^2 \ell_n(\theta)}{\partial \sigma^2 \partial H} & \frac{\partial^2 \ell_n(\theta)}{\partial \kappa \partial H} & \frac{\partial^2 \ell_n(\theta)}{\partial H \partial H} \end{pmatrix}
\begin{pmatrix} \frac{1}{\sqrt{n}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{n}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{n}} \end{pmatrix} \\
&\xrightarrow{p} \frac{1}{4\pi}\int_{-\pi}^{\pi} \left(\nabla \ln f_X^{\Delta}(\lambda;\beta)\right)\left(\nabla \ln f_X^{\Delta}(\lambda;\beta)\right)^{\top} d\lambda\,.
\end{aligned} \tag{8.29}$$

Therefore, using (8.25)–(8.29), we can obtain that the condition (C1) is satisfied as

$$\mathcal{I}_n(\theta) \xrightarrow{p} \mathcal{I}(\theta) = \begin{pmatrix} \frac{B(3/2-H,3/2-H)}{\Gamma(2-2H)}\frac{\kappa^2}{\sigma^2 C(H)\Delta^{2H-2}} & 0 \\ 0 & \frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\nabla \ln f_X^{\Delta}(\lambda;\beta)\right]\left[\nabla \ln f_X^{\Delta}(\lambda;\beta)\right]^{\top} d\lambda \end{pmatrix}.$$

To prove the first part of the condition (C2), note that the set over which the sup is sought is

$$\begin{aligned}
&\left\|\{A_n(\theta)\}^{\top}(\theta^* - \theta)\right\| \\
&= \sqrt{n^{2-2H}(\mu^* - \mu)^2 + n(\sigma^{2*} - \sigma^2)^2 + n(\kappa^* - \kappa)^2 + n(H^* - H)^2} \le c\,.
\end{aligned} \tag{8.30}$$

Using (8.30), we have $n(H^* - H)^2 \to 0$, as $n \to \infty$. Therefore, it has

$$\log\left(n^{H-H^*}\right) = (H - H^*)\log(n) \to 0 \quad \text{and} \quad n^{H-H^*} \to 1\,.$$

As a result, for any θ^* in the set of $\left\|\{A_n(\theta)\}^{\top}(\theta^* - \theta)\right\| \le c$, using (8.30), as $n \to \infty$, we obtain

$$\left\|\{A_n(\theta)\}^{-1} A_n(\theta^*) - \mathbf{I}_4\right\| = \left\|\mathrm{diag}\left(n^{H-H^*} - 1, 0, 0, 0\right)\right\| \to 0\,.$$

Therefore, (8.19) follows and the first part of the condition (C2) is proved.

We now turn to prove the second part of the condition (C2), i.e., (8.20). Let us first consider the elements in the first row of the matrix $\{A_n(\theta)\}^{-1}\left[\mathcal{F}_n(\Theta) - \mathcal{F}_n(\theta)\right]\left[\{A_n(\theta)\}^{-1}\right]^{\top}$, i.e.,

$$\left(n^{2H-2}\frac{\partial^2 l_n(\theta)}{\partial \mu \partial \mu} \quad n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial \sigma^2 \partial \mu} \quad n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial \kappa \partial \mu} \quad n^{H-3/2}\frac{\partial^2 l_n(\theta)}{\partial H \partial \mu}\right)$$

$$- \left(n^{2H-2} \frac{\partial^2 l_n(\theta_1^*)}{\partial\mu\partial\mu} \quad n^{H-3/2} \frac{\partial^2 l_n(\theta_1^*)}{\partial\sigma^2\partial\mu} \quad n^{H-3/2} \frac{\partial^2 l_n(\theta_1^*)}{\partial\kappa\partial\mu} \quad n^{H-3/2} \frac{\partial^2 l_n(\theta_1^*)}{\partial H\partial\mu} \right), \tag{8.31}$$

where θ_1^* satisfies $\left\| \{A_n(\theta)\}^\top (\theta_1^* - \theta) \right\| \leq c$.

When $n \to \infty$, it has

$$n^{2H-2} \left[\frac{\partial^2 l_n(\theta)}{\partial\mu\partial\mu} - \frac{\partial^2 l_n(\theta_1^*)}{\partial\mu\partial\mu} \right] = n^{2H-2} \left(\frac{1}{\sigma^2} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} - \frac{1}{\sigma^2} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \Big|_{\theta=\theta_1^*} \right) \to \mathbf{0}, \tag{8.32}$$

where the last limit comes from the facts of the continuity of $\frac{1}{2\sigma^2} \mathbf{1}^\top \Sigma^{-1} \mathbf{1}\big|_{\theta=\theta_1^*}$ when θ_1^* goes to the true value of the parameter and $n^{2H-2} \frac{1}{\sigma^2} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} = O(1)$ as $n \to \infty$.

With (8.26) and the fact of $\mathbf{X} - \mu_1^* \mathbf{1} = (\mathbf{X} - \mu\mathbf{1}) + (\mu - \mu_1^*)\mathbf{1}$, as $n \to \infty$, we can have

$$\begin{aligned}
& - n^{H-3/2} \frac{\partial^2 l_n(\theta_1^*)}{\partial\sigma^2\partial\mu} \\
=& n^{H-3/2} \left(\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} (\mathbf{X} - \mu_1^* \mathbf{1}) \right) \\
=& n^{H-3/2} \left(\left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} \right] (\mathbf{X} - \mu\mathbf{1}) + \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \Big|_{\theta=\theta_1^*} \right] (\mu - \mu_1^*) \right) \\
=& n^{H-3/2} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} \right] (\mathbf{X} - \mu\mathbf{1}) + n^{H-3/2} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \Big|_{\theta=\theta_1^*} \right] (\mu - \mu_1^*) \\
\xrightarrow{p}& 0,
\end{aligned} \tag{8.33}$$

where the first limit can be obtained from the results of

$$\begin{aligned}
\mathrm{Var} \left(n^{H-3/2} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} \right] (\mathbf{X} - \mu\mathbf{1}) \right) &= n^{2H-3} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} \right] \cdot \sigma^2 \Sigma \cdot \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \Big|_{\theta=\theta_1^*} \right]^\top \\
&= n^{-1} \left[n^{2H-2} \frac{1}{\sigma^6} \mathbf{1}^\top \Sigma^{-1} \mathbf{1}^\top + o(1) \right] = O\left(n^{-1}\right),
\end{aligned}$$

and the second limit follows from

$$n^{H-3/2} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \Big|_{\theta=\theta_1^*} \right] (\mu - \mu_1^*) = n^{2H-2} \left[\frac{1}{\sigma^4} \mathbf{1}^\top \Sigma^{-1} \mathbf{1} \Big|_{\theta=\theta_1^*} \right] \left\{ n^{1-H} (\mu - \mu_1^*) \right\} n^{-1/2} = o\left(n^{-1/2}\right).$$

Therefore, using (8.26) and (8.33), as $n \to \infty$, we get

$$n^{H-3/2} \left(\frac{\partial^2 l_n(\theta)}{\partial\sigma^2\partial\mu} - \frac{\partial^2 l_n(\theta_1^*)}{\partial\sigma^2\partial\mu} \right) \xrightarrow{p} 0. \tag{8.34}$$

Similarly, as $n \to \infty$, it can be proved that

$$n^{H-3/2} \left(\frac{\partial^2 l_n(\theta)}{\partial\kappa\partial\mu} - \frac{\partial^2 l_n(\theta_1^*)}{\partial\kappa\partial\mu} \right) \xrightarrow{p} 0, n^{H-3/2} \left(\frac{\partial^2 l_n(\theta)}{\partial H\partial\mu} - \frac{\partial^2 l_n(\theta_1^*)}{\partial H\partial\mu} \right) \xrightarrow{p} 0. \tag{8.35}$$

Using (8.32), (8.34), and (8.35), we can see that the elements in the first row of the matrix, $\{A_n(\theta)\}^{-1}\left[\mathcal{F}_n(\Theta)-\mathcal{F}_n(\theta)\right]\left[\{A_n(\theta)\}^{-1}\right]^\top$, converge to zero in probability as $n\to\infty$. By applying the same procedure above, the elements in the first column of the matrix $\{A_n(\theta)\}^{-1}\left[\mathcal{F}_n(\Theta)-\mathcal{F}_n(\theta)\right]\left[\{A_n(\theta)\}^{-1}\right]^\top$, can be proved to converge to zero in probability as $n\to\infty$.

Now, consider the other elements of $\{A_n(\theta)\}^{-1}\left[\mathcal{F}_n(\Theta)-\mathcal{F}_n(\theta)\right]\left[\{A_n(\theta)\}^{-1}\right]^\top$, i.e.,

$$\frac{1}{n}\left\{\begin{pmatrix} \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial\sigma^2} & \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial H} \\ \frac{\partial^2\ell_n(\theta_3^*)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta_3^*)}{\partial\kappa\partial\kappa} & \frac{\partial^2\ell_n(\theta_3^*)}{\partial H\partial\kappa} \\ \frac{\partial^2\ell_n(\theta_4^*)}{\partial\sigma^2\partial H} & \frac{\partial^2\ell_n(\theta_4^*)}{\partial\kappa\partial H} & \frac{\partial^2\ell_n(\theta_4^*)}{\partial H\partial H} \end{pmatrix} - \begin{pmatrix} \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\sigma^2} & \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial H} \\ \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial\kappa\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial H\partial\kappa} \\ \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial H} & \frac{\partial^2\ell_n(\theta)}{\partial\kappa\partial H} & \frac{\partial^2\ell_n(\theta)}{\partial H\partial H} \end{pmatrix}\right\}, \tag{8.36}$$

where $\left\|\{A_n(\theta)\}^\top(\theta_i^*-\theta)\right\|\le c$, for $i=2,3,4$. Denote

$$\widetilde{\mathbf{X}}\equiv\mathbf{X}-\mu\mathbf{1},\ \theta_i^*\equiv(\mu_i^*,\beta_i^*)'\ \text{ with }\ \beta_i^*:=(\sigma_i^{2*},\kappa_i^*,H_i^*)',$$

where μ is the true value of the location parameter, making $\widetilde{\mathbf{X}}$ to have a zero mean.

Using the relationship between $\widetilde{\mathbf{X}}$ and $\mathbf{X}$, we can obtain

$$\begin{aligned}(\mathbf{X}-\mu_i^*\mathbf{1})^\top\Sigma^{-1}(\mathbf{X}-\mu_i^*\mathbf{1}) =& (\mathbf{X}-\mu\mathbf{1}+\mu\mathbf{1}-\mu_i^*\mathbf{1})^\top\Sigma^{-1}(\mathbf{X}-\mu\mathbf{1}+\mu\mathbf{1}-\mu_i^*\mathbf{1})\\ =&\widetilde{\mathbf{X}}^\top\Sigma^{-1}\widetilde{\mathbf{X}}+2(\mu-\mu_i^*)\mathbf{1}^\top\Sigma^{-1}\widetilde{\mathbf{X}}+(\mu-\mu_i^*)^2\mathbf{1}^\top\Sigma^{-1}\mathbf{1},\end{aligned}$$

and

$$\begin{aligned} l_n(\theta)|_{\theta=\theta_i^*} &= \left\{-\frac{1}{2}\ln\left|\sigma^2\Sigma\right|-\frac{1}{2\sigma^2}(\mathbf{X}-\mu\mathbf{1})^\top\Sigma^{-1}(\mathbf{X}-\mu\mathbf{1})\right\}\Big|_{\theta=\theta_i^*}\\ &= \left\{-\frac{1}{2}\ln\left|\sigma^2\Sigma\right|-\frac{1}{2\sigma^2}\widetilde{\mathbf{X}}^\top\Sigma^{-1}\widetilde{\mathbf{X}}\right\}\Big|_{\beta=\beta_i^*}\\ &\quad-\left\{\frac{1}{2\sigma^2}\left[2(\mu-\mu_i^*)\mathbf{1}^\top\Sigma^{-1}\widetilde{\mathbf{X}}+(\mu-\mu_i^*)^2\mathbf{1}^\top\Sigma^{-1}\mathbf{1}\right]\right\}\Big|_{\beta=\beta_i^*}\\ &= l_n(\beta)|_{\beta=\beta_i^*}+G_n(\beta)|_{\beta=\beta_i^*}, \end{aligned}\tag{8.37}$$

with

$$\begin{aligned} l_n(\beta)|_{\beta=\beta_i^*} &\equiv \left\{-\frac{1}{2}\ln\left|\sigma^2\Sigma\right|-\frac{1}{2\sigma^2}\widetilde{\mathbf{X}}^\top\Sigma^{-1}\widetilde{\mathbf{X}}\right\}\Big|_{\beta=\beta_i^*},\\ G_n(\beta)|_{\beta=\beta_i^*} &= -\left\{\frac{1}{2\sigma^2}\left[2(\mu-\mu_i^*)\mathbf{1}^\top\Sigma^{-1}\widetilde{\mathbf{X}}+(\mu-\mu_i^*)^2\mathbf{1}^\top\Sigma^{-1}\mathbf{1}\right]\right\}\Big|_{\beta=\beta_i^*}. \end{aligned}$$

From (8.37), we can see that the difference of the second-order derivatives of the log-likelihood function $\ell_n(\cdot)$ at different points can be rewritten as

$$\frac{\partial^2\ell_n(\theta_i^*)}{\partial\cdot\partial\cdot}-\frac{\partial^2\ell_n(\theta)}{\partial\cdot\partial\cdot}=\frac{\partial^2\ell_n(\beta_i^*)}{\partial\cdot\partial\cdot}-\frac{\partial^2\ell_n(\beta)}{\partial\cdot\partial\cdot}+\frac{\partial^2G_n(\beta_i^*)}{\partial\cdot\partial\cdot}-\frac{\partial^2G_n(\beta)}{\partial\cdot\partial\cdot}. \tag{8.38}$$

Therefore, using (8.36) and (8.38), can we simplize the matrix expression as

$$\begin{aligned}
&\frac{1}{n}\left\{\begin{pmatrix} \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial\sigma^2} & \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta_2^*)}{\partial\sigma^2\partial H} \\ \frac{\partial^2\ell_n(\theta_3^*)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta_3^*)}{\partial\kappa\partial\kappa} & \frac{\partial^2\ell_n(\theta_3^*)}{\partial H\partial\kappa} \\ \frac{\partial^2\ell_n(\theta_4^*)}{\partial\sigma^2\partial H} & \frac{\partial^2\ell_n(\theta_4^*)}{\partial\kappa\partial H} & \frac{\partial^2\ell_n(\theta_4^*)}{\partial H\partial H} \end{pmatrix} - \begin{pmatrix} \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\sigma^2} & \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial H} \\ \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial\kappa\partial\kappa} & \frac{\partial^2\ell_n(\theta)}{\partial H\partial\kappa} \\ \frac{\partial^2\ell_n(\theta)}{\partial\sigma^2\partial H} & \frac{\partial^2\ell_n(\theta)}{\partial\kappa\partial H} & \frac{\partial^2\ell_n(\theta)}{\partial H\partial H} \end{pmatrix}\right\} \\
=&\frac{1}{n}\left(\frac{\partial^2\ell_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2\ell_n(\beta)}{\partial\cdot\partial\cdot} + \frac{\partial^2 G_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2 G_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3} \\
=&\frac{1}{n}\left(\frac{\partial^2\ell_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2\ell_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3} + \frac{1}{n}\left(\frac{\partial^2 G_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2 G_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3}. \qquad (8.39)
\end{aligned}$$

Using Lemma 2.7 in Cohen et al. (2013) and the mean value theorem, it is can be easily proved that

$$\sup\frac{1}{n}\left\|\left(\frac{\partial^2\ell_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2\ell_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3}\right\| \xrightarrow{p} 0 \quad \text{as } n\to\infty, \qquad (8.40)$$

where the sup is sought over the set of β_i^* satisfying

$$\sqrt{n\left(\sigma^{2*}-\sigma^2\right)^2 + n\left(\kappa^*-\kappa\right)^2 + n\left(H^*-H\right)^2} \le c\,. \qquad (8.41)$$

Hence, in the following, we only need to prove that every element in $n^{-1}\left(\frac{\partial^2 G_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2 G_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3}$ converges to zero in probability as $n\to\infty$, which in turn leads to

$$\sup\frac{1}{n}\left\|\left(\frac{\partial^2 G_n(\beta_i^*)}{\partial\cdot\partial\cdot} - \frac{\partial^2 G_n(\beta)}{\partial\cdot\partial\cdot}\right)_{3\times3}\right\| \xrightarrow{p} 0\,. \qquad (8.42)$$

Next, we just present the proof of $n^{-1}\left(\frac{\partial^2 G_n(\beta_3^*)}{\partial\kappa\partial\kappa} - \frac{\partial^2 G_n(\beta)}{\partial\kappa\partial\kappa}\right) \xrightarrow{p} 0$ in details. The other elements going to zero can be proved in a similar way and hence omitted due to the space limit.

Note that

$$\frac{\partial^2 G_n(\beta)}{\partial\kappa\partial\kappa} = -\frac{1}{2\sigma^2}\left[2\left(\mu-\mu_i^*\right)\mathbf{1}^\top\frac{\partial^2\Sigma^{-1}}{\partial\kappa\partial\kappa}\widetilde{\mathbf{X}} + \left(\mu-\mu_i^*\right)^2\mathbf{1}^\top\frac{\partial^2\Sigma^{-1}}{\partial\kappa\partial\kappa}\mathbf{1}^\top\right] \qquad (8.43)$$

where

$$\frac{\partial^2\Sigma^{-1}}{\partial\kappa\partial\kappa} = 2\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1}\frac{\partial\Sigma}{\partial\kappa}\Sigma^{-1} - \Sigma^{-1}\frac{\partial^2\Sigma}{\partial\kappa\partial\kappa}\Sigma^{-1}\,.$$

From Lemma 5.4 (d) in Dahlhaus (1989), it has

$$\mathbf{1}^\top\Sigma^{-1}\frac{\partial^2\Sigma}{\partial\kappa\partial\kappa}\Sigma^{-1}\mathbf{1} \le K\cdot n^{2-2H+\delta}\,,$$

$$\mathbf{1}^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \kappa} \Sigma^{-1} \frac{\partial \Sigma}{\partial \kappa} \Sigma^{-1} \mathbf{1} \leq K \cdot n^{2-2H+\delta},$$

and hence

$$\mathbf{1}^\top \frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa} \mathbf{1} = O\left(n^{2-2H}\right). \tag{8.44}$$

Using the condition of $n^{1-H}(\mu - \mu_i^*) \to 0$, (8.43) and (8.44), we have

$$\frac{\partial^2 G_n(\beta)}{\partial \kappa \partial \kappa} = -\frac{(\mu - \mu_i^*)}{\sigma^2} \mathbf{1}^\top \frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa} \widetilde{\mathbf{X}} + o(1). \tag{8.45}$$

Consequently, we can write

$$\begin{aligned} &\frac{1}{n}\left(\frac{\partial^2 G_n(\beta_3^*)}{\partial \kappa \partial \kappa} - \frac{\partial^2 G_n(\beta)}{\partial \kappa \partial \kappa}\right) \\ =&\frac{(\mu - \mu_i^*)}{n} \mathbf{1}^\top \left[\left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right) - \left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}\right] \widetilde{\mathbf{X}} + o(1). \end{aligned} \tag{8.46}$$

From Hölder inequality, it is easy to get

$$\begin{aligned} &n^{-1}\mathbf{1}^\top \left[\left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right) - \left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}\right] \widetilde{\mathbf{X}} \\ \leq &\sqrt{n^{-1}\mathbf{1}^\top \left[\left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right) - \left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}\right] \left[\left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right) - \left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}\right]^\top \mathbf{1}} \\ &\times \sqrt{n^{-1}\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}}} \xrightarrow{p} 0, \end{aligned} \tag{8.47}$$

where the last limit comes from the facts of $n^{-1}\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} = O_p(1)$ due to the ergodicity of the fOU process, and the continuity of $\left(\frac{1}{2\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}$ when $\beta_3^* \to \beta$, which makes every term of the matrix $\left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right) - \left(\frac{1}{\sigma^2}\frac{\partial^2 \Sigma^{-1}}{\partial \kappa \partial \kappa}\right)\Big|_{\beta=\beta_3^*}$ shrinks to zero as $n \to \infty$.

Therefore, substituting (8.47) into (8.46), $n \to \infty$, we have

$$\frac{1}{n}\left(\frac{\partial^2 G_n(\beta_3^*)}{\partial \kappa \partial \kappa} - \frac{\partial^2 G_n(\beta)}{\partial \kappa \partial \kappa}\right) = (\mu - \mu_i^*)\, o_p(1) + o(1) \xrightarrow{p} 0,$$

and finally, the part 2 of (C2) as

$$\sup \left\| \{A_n(\theta)\}^{-1} \left[\mathcal{F}_n(\Theta) - \mathcal{F}_n(\theta)\right] \left[\{A_n(\theta)\}^{-1}\right]^\top \right\| \xrightarrow{p} 0,$$

with the sup being sought over the set $\left|\{A_n(\theta)\}^\top (\theta_i^* - \theta)\right| \leq c$ is proved.

References

Adenstedt, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. Annals of Statistics, 2(6), 1095–1107.

Aït-Sahalia, Y. (1999). Transition Densities for Interest Rate and Other Nonlinear Diffusions. Journal of Finance, 54(4), 1361–1395.

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusion: A closed-form approximation approach. Econometrica, 70(1), 223–262.

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

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

Ben-Artzi, A., and Shalom, T. (1986). On inversion of Toeplitz and close to Toeplitz matrices. Linear algebra and its Applications, 75(3), 173–192.

Bennedsen, M., Christensen, K., and Christensen, P. (2023). Likelihood-based estimation of rough stochastic volatility models. Aarhus University, manuscript.

Bennedsen, M., Christensen, K., and Christensen, P. (2024). Composite likelihood estimation of stationary Gaussian processes with a view toward stochastic volatility. Working paper, arXiv preprint arXiv:2403.12653.

Bolko, A. E., Christensen, K., Pakkanen, M. S., and Veliyev, B. (2023). A GMM approach to estimate the roughness of stochastic volatility. Journal of Econometrics. 235(2), 745–778.

Chong, C. H., and Todorov, V. (2024). A nonparametric test for rough volatility. arxiv preprint arxiv:2407.10659.

Cohen, S., Gamboa, F., Lacaux, C., and Loubes, J. M. (2013). LAN property for some fractional type Brownian motion. ALEA: Latin American Journal of Probability and Mathematical Statistics, 10(1), 91-106.

Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. Annals of Statistics, 17(4), 1749–1766.

Diebold, F. X., and Mariano, R. S. (2002). Comparing predictive accuracy. Journal of Business Economic and Statistics, 20(1), 134–144.

Dietrich, C. R., and Osborne, M. R. (1996). $O(n \ln^2 n)$ determinant computation of a Toeplitz matrix and fast variance estimation. Applied Mathematics Letters, 9(2), 29–31.

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

Fink, H., Klüppelberg, C., and Zähle, M. (2013). Conditional distributions of processes related to fractional Brownian motion. Journal of Applied Probability, 50(1), 166–183.

Fox, R., Taqqu, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Annals of Statistics, 14(2), 517–532.

Fouque, J. P., and Hu, R. (2019). Optimal portfolio under fractional stochastic environment. Mathematical Finance, 29(3), 697–734.

Fukasawa, M., Takabatake, T., and Westphal, R. (2022). Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics. Mathematical Finance, 32(4), 1086–1132.

Gao, F., Liu, S., Oosterlee, C. W., and Temme, N. M. (2023). Evaluation of integrals with fractional Brownian motion for different Hurst indices. International Journal of Computer Mathematics, 100(4), 847–866.

Garnier, J., and Sølna, K. (2018). Option pricing under fast-varying and rough stochastic volatility. Annals of Finance, 14(4), 489–516.

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

Glasserman, P., and He, P. (2020). Buy rough, sell smooth. Quantitative Finance, 20(3), 363–378.

Gohberg, I., Semencul, A. (1972). On the inversion of finite Toeplitz matrices and their continuous analogues. Mat. Issled., 2(1), 201–233 (Russian).

Gradshteyn, I. S., and Ryzhik, I. M. (2007). Table of Integrals, Series, and Products. Seventh Edition, New York, NY, USA: Academic.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. Journal of Applied Probability, 10(1), 130–145.

Hansen, P.R., Lunde, A., and Nason, J. M. (2011). The model confidence set. Econometrica, 79(2), 453–497.

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.

Hult, H. (2003). Topics on fractional Brownian motion and regular variation for stochastic processes (Doctoral dissertation, Matematik).

Kleptsyna, M. L., and Le Breton, A. (2002). Statistical analysis of the fractional Ornstein–Uhlenbeck type process. Statistical Inference for Stochastic Processes, 5(3), 229–248.

Kleptsyna, M. L., Le Breton, A., and Roubaud, M. C. (2000). Parameter estimation and optimal filtering for fractional type stochastic systems. Statistical Inference for Stochastic Processes, 3(1-2), 173–182.

Lieberman, O., Rosemarin, R., and Rousseau, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. Econometric Theory, 28(2), 457–470.

Lindsay, B. (1988). Composite likelihood methods. Contemporary Mathematics, 80(1), 220–239.

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

Lohvinenko, S., and Ralchenko, K. (2017). Maximum likelihood estimation in the fractional Vasicek model. Lithuanian Journal of Statistics, 56(1), 77–87.

Lohvinenko, S., Ralchenko, K., (2019). Maximum likelihood estimation in the non-ergodic fractional Vasicek model. Modern Stochastics: Theory and Applications, 6(3), 377–395.

Phillips, P.C.B. and Yu, J. (2009). Maximum Likelihood and Gaussian Estimation of Continuous Time Models. In Finance Handbook of Financial Time Series, 497–530.

Rao, S. S., and Yang, J. (2021). Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain. Annals of Statistics, 49(5), 2774–2802.

Shi, S., Yu, J., and Zhang, C. (2024). On the Spectral Density of Fractional Ornstein-Uhlenbeck Processes, Journal of Econometrics, 245, 105872.

Sweeting, T. J. (1980). Uniform asymptotic normality of the maximum likelihood estimator. Annals of Statistics, 8(6), 1375–1381.

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., Xiao, W., and Yu, J. (2020). Maximum likelihood estimation for the fractional Vasicek model. Econometrics, 8(3), 32.

Wang, X., Phillips, P.C.B., and Yu, J. (2011). Bias in Estimating Multivariate and Univariate Diffusions. Journal of Econometrics, 161, 228–245.

Wang, X., Xiao, W., and Yu, J. (2023). Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process. Journal of Econometrics. 232, 389–415.

Whittle, P. (1951). Hypothesis testing in time series analysis, Volume 4. Almqvist and Wiksells boktr.

Whittle, P. (1954). On stationary processes in the plane. Biometrika, 41(3-4), 434–449.

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.

Zhang, H. M., and Duhamel, P. (1992). On the methods for solving Yule-Walker equations. IEEE Transactions on signal processing, 40(12), 2987–3000.

Online Supplement to "Maximum Likelihood Estimation of Fractional Ornstein-Uhlenbeck Process with Discretely Sampled Data" by Wang, Xiao, Yu and Zhang (not for publication)

This online appendix contains some additional proof details omitted in the main paper due to space limit, a set of Monte Carlo results, and a set of empirical results. The Monte Carlo simulations aim to determine how a change in one parameter affects the estimates of the other parameters in fOU. The target of the empirical results is to show that all the empirical conclusions drawn in the main paper are qualitatively unchanged when $\ln(RV)$ is predicted.

A.1: Connecting Hult's formula (2.16) with our formula (2.14)

Using two well known equalities $\csc(\vartheta) = \sec\left(\frac{\pi}{2} - \vartheta\right)$ and $\sin(\vartheta) = 1/\csc(\vartheta)$, we can write the first term of (2.16) as

$$\sigma^2\Gamma(2H+1)\sin(\pi H)\frac{1}{2\kappa^{2H}}\sec(\frac{\pi(1-2H)}{2})\cosh(\kappa j\Delta) = \frac{\sigma^2\Gamma(2H+1)}{2\kappa^{2H}}\cosh(\kappa j\Delta)\,,$$

which reduces to the first term of (2.14).

For the second term of (2.16), using the Legendre duplication formula, $\Gamma(\vartheta)\Gamma\left(\vartheta+\frac{1}{2}\right) = 2^{1-2\vartheta}\sqrt{\pi}\Gamma(2\vartheta)$, and the Euler's feflection formula (See Eq FI II 430 of 8.334 on page 876 from Gradshteyn and Ryzhik (2007)), $\Gamma(1-\vartheta)\Gamma(\vartheta) = \frac{\pi}{\sin(\pi\vartheta)}$, we can write the second term of (2.16) as

$$\begin{aligned}
&\sigma^2\Gamma(2H+1)\sin(\pi H)\frac{(j\Delta)^{2H}\Gamma(-H)}{\sqrt{\pi}2^{2H+1}\Gamma(H+1/2)}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&\frac{\sigma^2\Gamma(2H+1)\sin(\pi H)}{\sqrt{\pi}2^{2H+1}}\frac{(j\Delta)^{2H}\Gamma(-H)\Gamma(H)}{\Gamma(H+1/2)\Gamma(H)}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&\frac{\sigma^2\Gamma(2H+1)\sin(\pi H)}{\sqrt{\pi}2^{2H+1}}\frac{(j\Delta)^{2H}\Gamma(-H)\Gamma(H)}{2^{1-2H}\sqrt{\pi}\Gamma(2H)}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&\frac{\sigma^2\Gamma(2H+1)\sin(\pi H)}{2^2\pi}\frac{(j\Delta)^{2H}\Gamma(-H)\Gamma(H)2H}{\Gamma(2H)2H}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&\frac{\sigma^2\Gamma(2H+1)\sin(\pi H)}{2\pi}\frac{(j\Delta)^{2H}\Gamma(-H)\Gamma(H)H}{\Gamma(2H+1)}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&-\frac{\sigma^2\sin(\pi H)(j\Delta)^{2H}\Gamma(1-H)\Gamma(H)}{2\pi}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right)\\
=&-\frac{\sigma^2(j\Delta)^{2H}}{2}{}_1F_2\left(1; H+\frac{1}{2}, H+1; \frac{(\kappa j\Delta)^2}{4}\right),
\end{aligned}$$

which reduces to the second term of (2.14).

A.2: $f_X^{\Delta}(\lambda;\beta)$ satisfies the Assumption 2.1 of Cohen et al. (2013)

As stated in Cohen et al. (2013), the Assumption 2.1 of Cohen et al. (2013) corresponds to *Assumptions 1, 2 and 4* in Lieberman et al. (2012), except that some smoothness property on the derivative of order three of $f_X^{\Delta}(\lambda;\beta)$ is imposed. Consequently, for the sake of convenience, we will prove that the *Assumptions 1, 2 and 4* defined in Lieberman et al. (2012) are all satisfied for our stationary fOU process. We focus on the proof of the case where $H \in (1/2, 1)$. The same procedure can be extended straightforwardly to the case of $H \in (0, 1/2)$, which we will briefly discuss at the end of this proof, but omit tedious details for simplicity. In the following, we repeat the *Assumptions 1, 2 and 4* defined in Lieberman et al. (2012) in italics whenever necessary to make our proof easy to read and self-contained.

Assumption 1: There exists $\alpha(\beta) \in (-\infty, 1)$ *such that* $f_X^{\Delta}(\lambda;\beta) \sim |\lambda|^{-\alpha(\beta)} g_\beta(\lambda)$ *as* $\lambda \to 0$*, where* $g_\beta(\lambda)$ *is a positive function that varies slowly at* $\lambda = 0$*. The functions* $f_X^{\Delta}(\lambda;\beta)$*,* $f_X^{\Delta}(\lambda;\beta)^{-1}$ *and* $\partial f_X^{\Delta}(\lambda;\beta)/\partial\lambda$ *are continuous at all* (λ, β)*,* $\lambda \neq 0$*. For each* $\delta > 0$*, it has* $f_X^{\Delta}(\lambda;\beta) = O\left(|\lambda|^{-\alpha(\beta)-\delta}\right)$*,* $f_X^{\Delta}(\lambda;\beta)^{-1} = O\left(|\lambda|^{\alpha(\beta)-\delta}\right)$*, and* $\partial f_X^{\Delta}(\lambda;\beta)/\partial\lambda = O\left(|\lambda|^{-\alpha(\beta)-1-\delta}\right)$.

To prove Assumption 1, we first get that when $H \in (1/2, 1)$, the spectral density function of discretely sampled fOU in (2.6) can be rewritten as

$$\begin{aligned} f_X^{\Delta}(\lambda;\beta) &= |\lambda|^{1-2H} \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \left\{ \frac{1}{(\kappa\Delta)^2 + \lambda^2} + |\lambda|^{2H-1} \sum_{j\neq 0} \phi_j(\lambda, \kappa, H) \right\} \\ &= |\lambda|^{1-2H} g_\beta(\lambda) , \end{aligned}$$

where $g_\beta(\lambda) = \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \left\{ \frac{1}{(\kappa\Delta)^2+\lambda^2} + |\lambda|^{2H-1} \sum_{j\neq 0} \phi_j(\lambda, \kappa, H) \right\}$.

Under the condition of $H \in (1/2, 1)$, it has $\alpha(\beta) := 2H - 1 \in (0, 1)$, and hence, $|\lambda|^{2H-1} \to 0$ and

$$g_\beta(\lambda) \to \frac{\sigma^2}{2\pi} C(H) \Delta^{2H} \frac{1}{(\kappa\Delta)^2} \quad \text{as} \quad \lambda \to 0 . \tag{8.48}$$

On the one hand, we can directly obtain that $\phi_0(\lambda, H, \kappa) = \frac{|\lambda|^{1-2H}}{(\Delta\kappa)^2+|\lambda|^2}$ is twice continuously differentiable and positive at all $(\lambda, H), \lambda \neq 0$. Moreover, a standard calculation shows that

$$\frac{\partial}{\partial\lambda}\phi_0(\lambda, H, \kappa) = \frac{1-2H}{(\Delta\kappa)^2 + |\lambda|^2} |\lambda|^{-2H} - \frac{2}{\left((\Delta\kappa)^2 + |\lambda|^2\right)^2} |\lambda|^{2-2H},$$

$$
\begin{aligned}
\frac{\partial}{\partial H}\phi_0(\lambda,H,\kappa) =& \frac{-2|\lambda|^{1-2H}\ln(|\lambda|)}{(\Delta\kappa)^2+|\lambda|^2},\\
\frac{\partial}{\partial \kappa}\phi_0(\lambda,H,\kappa) =& -\frac{2|\lambda|^{1-2H}\Delta^2\kappa}{\left((\Delta\kappa)^2+|\lambda|^2\right)^2},\\
\frac{\partial^2}{\partial H^2}\phi_0(\lambda,H,\kappa) =& \frac{4|\lambda|^{1-2H}\ln^2(|\lambda|)}{(\Delta\kappa)^2+|\lambda|^2},\\
\frac{\partial^2}{\partial \kappa^2}\phi_0(\lambda,H,\kappa) =& \frac{6\Delta^4\kappa^2-2\Delta^2\lambda^2}{\left((\Delta\kappa)^2+|\lambda|^2\right)^3}|\lambda|^{1-2H},\\
\frac{\partial^2}{\partial \kappa\partial H}\phi_0(\lambda,H,\kappa) =& \frac{4\Delta^2\kappa|\lambda|^{1-2H}\ln(|\lambda|)}{\left((\Delta\kappa)^2+|\lambda|^2\right)^2}.
\end{aligned}
$$

We can see that all the functions above are continuous at all $(\lambda,H,\kappa),\lambda\neq 0$ and the discontinuity in $\lambda\neq 0$ is removable. Then for each $\delta>0$, as $\lambda\to 0$, we have

$$
\begin{aligned}
\frac{\phi_0(\lambda,H,\kappa)}{|\lambda|^{-(2H-1)-\delta}} =& \frac{|\lambda|^\delta}{(\Delta\kappa)^2+|\lambda|^2}\to 0,\\
\frac{\phi_0^{-1}(\lambda,H,\kappa)}{|\lambda|^{2H-1-\delta}} =& |\lambda|^\delta\left((\Delta\kappa)^2+|\lambda|^2\right)\to 0,\\
\frac{\frac{\partial}{\partial\lambda}\phi_0(\lambda,H,\kappa)}{|\lambda|^{-(2H-1)-\delta}} =& \frac{1-2H}{(\Delta\kappa)^2+|\lambda|^2}|\lambda|^\delta-\frac{2}{\left((\Delta\kappa)^2+|\lambda|^2\right)^2}|\lambda|^{2+\delta}\to 0.
\end{aligned}
$$

Consequently, we have

$$
\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.49}
$$

$$
\phi_0^{-1}(\lambda,H,\kappa) = O\left(|\lambda|^{2H-1-\delta}\right), \tag{8.50}
$$

$$
\frac{\partial}{\partial\lambda}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-2H-\delta}\right), \tag{8.51}
$$

and

$$
\frac{\partial}{\partial H}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.52}
$$

$$
\frac{\partial}{\partial \kappa}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.53}
$$

$$
\frac{\partial^2}{\partial H^2}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.54}
$$

$$
\frac{\partial^2}{\partial \kappa^2}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.55}
$$

$$
\frac{\partial^2}{\partial \kappa\partial H}\phi_0(\lambda,H,\kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right). \tag{8.56}
$$

On the other hand, for $k \neq 0$, we can see that $\phi_k(\lambda, H, \kappa)$ is nonnegative, twice continuously differentiable with

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\phi_k(\lambda, H, \kappa) =& \frac{(1-2H)|\lambda+2k\pi|^{-2H}}{(\Delta\kappa)^2+|\lambda+2k\pi|^2} - \frac{2|\lambda+2k\pi|^{2-2H}}{((\Delta\kappa)^2+|\lambda+2k\pi|^2)^2},\\
\frac{\partial}{\partial H}\phi_k(\lambda, H, \kappa) =& \frac{-2|\lambda+2k\pi|^{1-2H}\ln(|\lambda+2k\pi|)}{(\Delta\kappa)^2+|\lambda+2k\pi|^2},\\
\frac{\partial}{\partial\kappa}\phi_k(\lambda, H, \kappa) =& -\frac{2|\lambda+2k\pi|^{1-2H}\Delta^2\kappa}{((\Delta\kappa)^2+|\lambda+2k\pi|^2)^2},\\
\frac{\partial^2}{\partial H^2}\phi_k(\lambda, H, \kappa) =& \frac{4|\lambda+2k\pi|^{1-2H}\ln^2(|\lambda+2k\pi|)}{(\Delta\kappa)^2+|\lambda+2k\pi|^2},\\
\frac{\partial^2}{\partial\kappa^2}\phi_k(\lambda, H, \kappa) =& \frac{6\Delta^4\kappa^2-2\Delta^2|\lambda+2k\pi|^2}{((\Delta\kappa)^2+|\lambda+2k\pi|^2)^3}|\lambda+2k\pi|^{1-2H},\\
\frac{\partial^2}{\partial\kappa\partial H}\phi_k(\lambda, H, \kappa) =& \frac{4\Delta^2\kappa|\lambda+2k\pi|^{1-2H}\ln(|\lambda+2k\pi|)}{((\Delta\kappa)^2+|\lambda+2k\pi|^2)^2}.
\end{aligned}$$

Using the results above and the definition of $\phi_k(\lambda, H, \kappa)$, for $\delta > 0$ and as $\lambda \to 0$, we can see that

$$\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.57}$$

$$\frac{\partial}{\partial\lambda}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-2H-\delta}\right), \tag{8.58}$$

$$\frac{\partial}{\partial H}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.59}$$

$$\frac{\partial}{\partial\kappa}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.60}$$

$$\frac{\partial^2}{\partial H^2}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.61}$$

$$\frac{\partial^2}{\partial\kappa^2}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.62}$$

$$\frac{\partial^2}{\partial\kappa\partial H}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right). \tag{8.63}$$

Notice that $\frac{\partial}{\partial\lambda}\phi_k(\lambda, H, \kappa) \leq 0$ with $H \in (\frac{1}{2}, 1)$. Then, we have

$$\begin{aligned}
\phi_k(\lambda, H, \kappa) \leq& \phi_k(-\pi, H, \kappa)\\
=& \frac{|(2k-1)\pi|^{1-2H}}{(\Delta\kappa)^2+|\lambda+2k\pi|^2}\\
\leq& |(2k-1)\pi|^{-1-2H}\\
\leq& 2\pi|k|^{-1-2H}
\end{aligned}$$

$$\leq 2\pi|k|^{-1-2H_m},$$

where $H_m = \inf_{\beta\in\Theta} H > 0$.

Consequently, we have $\sum_{k\neq 0} 2\pi|k|^{-1-2H_m} < \infty$ and by Weierstrass's M-test, the series $\sum_{k\neq 0}\phi_k(\lambda, H, \kappa)$ converges uniformly. As a result, $\sum_{k\neq 0}\phi_k(\lambda, H, \kappa)$ is continuous, and hence also bounded, at all (λ, H, κ). Similar arguments show that the derivatives of $\sum_{k\neq 0}\phi_k(\lambda, H, \kappa)$ are given as the infinite sum over $k \neq 0$ of the corresponding derivatives of the summands $\phi_k(\lambda, H, \kappa)$, and that they are also continuous and bounded at all (λ, H, κ).

From (8.49)–(8.63), we have that $\sum_{k\in\mathbb{Z}}\phi_k(\lambda, H, \kappa)$ and derivatives of $\sum_{k\in\mathbb{Z}}\phi_k(\lambda, H, \kappa)$ satisfy the following conditions

$$\sum_{k\in\mathbb{Z}}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.64}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial\lambda}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-2H-\delta}\right), \tag{8.65}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial H}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.66}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial\kappa}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.67}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial H^2}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.68}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial\kappa^2}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.69}$$

$$\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial H\partial\kappa}\phi_k(\lambda, H, \kappa) = O\left(|\lambda|^{-(2H-1)-\delta}\right). \tag{8.70}$$

Now, using (8.64) and (8.65), we can see that

$$f_X^{\Delta}(\lambda; \beta) = O\left(|\lambda|^{-(2H-1)-\delta}\right), \tag{8.71}$$

$$\frac{\partial}{\partial\lambda} f_X^{\Delta}(\lambda; \beta) = O\left(|\lambda|^{-2H-\delta}\right). \tag{8.72}$$

It is easy to check, as $\lambda \to 0$,

$$\frac{\phi_0(\lambda, H, \kappa)}{|\lambda|^{1-2H}} = \frac{1}{(\Delta\kappa)^2 + |\lambda|^2} \to \frac{1}{(\Delta\kappa)^2},$$
$$\frac{\phi_k(\lambda, H, \kappa)}{|\lambda|^{1-2H}} = \frac{1}{(\Delta\kappa)^2 + |\lambda + 2k\pi|^2}\frac{|\lambda|^{2H-1}}{|\lambda + 2k\pi|^{2H-1}} \to 0.$$

Let $C_H = C(H)/(2\pi) = \dfrac{\Gamma(2H+1)\sin(\pi H)}{2\pi}$. Then, using these results above, we have

$$\frac{f_X^{\Delta}(\lambda;\beta)}{|\lambda|^{1-2H}} = \frac{\sigma^2 C_H \Delta^{2H}(\phi_0(\lambda,H,\kappa) + \sum_{k\neq 0}\phi_k(\lambda,H,\kappa))}{|\lambda|^{1-2H}} \to \frac{\sigma^2 C_H \Delta^{2H}}{(\Delta\kappa)^2} < \infty.$$

Thus, we have

$$\frac{f_X^{\Delta}(\lambda;\beta)^{-1}}{|\lambda|^{2H-1-\delta}} = \frac{|\lambda|^{\delta}}{f_X^{\Delta}(\lambda;\beta)/|\lambda|^{1-2H}}. \tag{8.73}$$

Combining (8.48), (8.71)–(8.73) with the continuity of $f_X^{\Delta}(\lambda;\beta)$, we obtain that $f_X^{\Delta}(\lambda;\beta)$ satisfies **Assumption 1** of Lieberman et al. (2012).

Assumption 2: $\partial f_X^{\Delta}(\lambda;\beta)/\partial\beta_j$ and $\partial^2 f_X^{\Delta}(\lambda;\beta)/\partial\beta_j\partial\beta_k$ are continuous at all (λ,β), $\lambda\neq 0$, $\frac{\partial}{\partial\beta_j} f_X^{\Delta}(\lambda;\beta) = O\left(|\omega|^{-\alpha(\beta)-\delta}\right)$ with $1\le j\le p$, $\frac{\partial^2}{\partial\beta_j\partial\beta_k} f_X^{\Delta}(\lambda;\beta) = O\left(|\omega|^{-\alpha(\beta)-\delta}\right)$ with $1\le j,k\le p$ and $\frac{\partial^3}{\partial\beta_j\partial\beta_k\partial\beta_l} f_X^{\Delta}(\lambda;\beta) = O\left(|\lambda|^{-\alpha(\beta)-\delta}\right)$, with $1\le j,k,l\le p$.

To prove Assumption 2, by (8.66)-(8.70), we can easily deduce that

$$\begin{aligned}
\frac{\partial}{\partial H} f_X^{\Delta}(\lambda;\beta) &= \sigma^2\left[\frac{\partial(C_H\Delta^{2H})}{\partial H}\sum_{k\in\mathbb{Z}}\phi_k + C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial H}\phi_k\right] \\
&= O\left(|\lambda|^{-(2H-1)-\delta}\right), \qquad (8.74)\\
\frac{\partial}{\partial \kappa} f_X^{\Delta}(\lambda;\beta) &= \sigma^2 C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial\kappa}\phi_k = O\left(|\lambda|^{-(2H-1)-\delta}\right), \qquad (8.75)\\
\frac{\partial^2}{\partial H^2} f_X^{\Delta}(\lambda;\beta) &= \sigma^2\left[\frac{\partial^2(C_H\Delta^{2H})}{\partial H^2}\sum_{k\in\mathbb{Z}}\phi_k + 2\frac{\partial(C_H\Delta^{2H})}{\partial H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial H}\phi_k + C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial H^2}\phi_k\right] \\
&= O\left(|\lambda|^{-(2H-1)-\delta}\right), \qquad (8.76)\\
\frac{\partial^2}{\partial \kappa^2} f_X^{\Delta}(\lambda;\beta) &= \sigma^2\left[\frac{\partial(C_H\Delta^{2H})}{\partial H}\sum_{k\in\mathbb{Z}}\phi_k + C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial\kappa^2}\phi_k\right] \\
&= O\left(|\lambda|^{-(2H-1)-\delta}\right), \qquad (8.77)\\
\frac{\partial^2}{\partial\kappa\partial H} f_X^{\Delta}(\lambda;\beta) &= \sigma^2\left[\frac{\partial(C_H\Delta^{2H})}{\partial H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial\kappa}\phi_k + C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial^2}{\partial H\partial\kappa}\phi_k\right] \\
&= O\left(|\lambda|^{-(2H-1)-\delta}\right). \qquad (8.78)
\end{aligned}$$

Since $f_X^{\Delta}(\lambda;\beta)$ is linear with respect to σ^2, we have

$$\frac{\partial}{\partial\sigma^2} f_X^{\Delta}(\lambda;\beta) = C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\phi_k(\lambda,H,\kappa), \tag{8.79}$$

$$\frac{\partial^2}{\partial\sigma^2\partial\sigma^2} f_X^{\Delta}(\lambda;\beta) = 0, \tag{8.80}$$

$$\frac{\partial^2}{\partial\sigma^2\partial H} f_X^{\Delta}(\lambda;\beta) = \frac{\partial(C_H\Delta^{2H})}{\partial H}\sum_{k\in\mathbb{Z}}\phi_k(\lambda,H,\kappa) + C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial H}\phi_k(\lambda,H,\kappa), \tag{8.81}$$

$$\frac{\partial^2}{\partial\sigma^2\partial\kappa} f_X^{\Delta}(\lambda;\beta) = C_H\Delta^{2H}\sum_{k\in\mathbb{Z}}\frac{\partial}{\partial\kappa}\phi_k(\lambda,H,\kappa). \tag{8.82}$$

Using (8.74)-(8.82), we obtain that **Assumption 2** of Lieberman et al. (2012) follows.

Assumption 4: The function $\alpha(\beta)$ is continuous, and the constants appearing in the $O(\cdot)$ above can be chosen independently of β (not of δ).

Using the result of $\alpha(\beta) = 2H - 1$, we can easily obtain the results of **Assumption 4**.

Hence **Assumption 1**, **Assumption 2** and **Assumption 4** in Lieberman et al. (2012) are fulfilled for the spectral density of fOU process, $f_X^{\Delta}(\lambda;\beta)$, for $H \in (1/2, 1)$. For $H \in (0, 1/2)$, using similar arguments as the case of $H \in (1/2, 1)$ with $\alpha(\beta) = 0$, we can also show that **Assumption 1**, **Assumption 2** and **Assumption 4** in Lieberman et al. (2012) are fulfilled. Moreover, it is easy to obtain the smoothness property on the derivative of three order of $f_X^{\Delta}(\lambda;\beta)$ for all $H \in (0, 1)$. Consequently, Assumption 2.1 in Cohen et al. (2013) are fulfilled for all $H \in (0, 1)$.

A.3: Simulations for fixed H and various values of σ, μ, and κ

In this experiment, we fix H to 0.260573 and allow the other parameters (σ, μ, and κ) to take different values to determine how a change in one parameter affects the estimates of the other parameters. Table 13 reports the simulation results when $\kappa = 4.446145$, $\mu = -2.465673$, and σ varies from 0.75 to 1.5. Table 14 reports the simulation results when $\kappa = 4.446145$, $\sigma = 1.172021$, and μ varies from -2.5 to 0.5. Table 15 reports the simulation results when $\mu = -2.465673$, $\sigma = 1.172021$, and κ varies from 1 to 10. According to Tables 13-15, ML always performs better than the other three methods in estimating H, σ, κ in terms of the standard deviation. The alternative estimators of μ have a similar finite-sample performance.

Forecasting results for $\ln(RV)$

The main paper forecasts RVs for the Standard and Poor's (S&P) 500 index ETF and the nine industry ETFs. In this subsection, we compare the performance of the competing

Table 13: Finite-sample properties of alternative estimation methods for (H, σ, μ, κ) when $H = 0.260573$, $T = 10$, $\Delta = 1/250$ with various values of σ.

		κ	H	μ	σ	::::	κ	H	μ	σ
Method	True value	4.446145	0.260573	-2.465673	0.750000	::::	4.446145	0.260573	-2.465673	1.000000
MM	Mean	6.482180	0.280680	-2.464634	0.745912	::::	6.533847	0.281600	-2.468015	1.000003
	SD	2.346232	0.026090	0.028622	0.103559	::::	2.202818	0.025526	0.039426	0.136363
MCL	Mean	4.736685	0.260140	-2.464634	0.660453	::::	4.706750	0.259778	-2.468015	0.879312
	SD	1.447678	0.013427	0.028622	0.045268	::::	1.421765	0.013283	0.039426	0.061482
AWML	Mean	4.783731	0.261750	-2.464634	0.662134	::::	4.743560	0.260423	-2.468015	0.881415
	SD	1.124392	0.012255	0.028622	0.041165	::::	1.298635	0.012456	0.039426	0.057362
ML	Mean	4.777911	0.261367	-2.464362	0.663226	::::	4.792363	0.260869	-2.467470	0.882593
	SD	0.986943	0.011514	0.027339	0.037803	::::	1.028730	0.012159	0.038241	0.054805
Method	True value	4.446145	0.260573	-2.465673	1.250000	::::	4.446145	0.260573	-2.465673	1.500000
MM	Mean	6.348864	0.280466	-2.468772	1.242599	::::	6.475826	0.281198	-2.470099	1.497698
	SD	2.273947	0.026257	0.048727	0.176225	::::	2.304653	0.026077	0.057656	0.209800
MCL	Mean	4.590842	0.259216	-2.468772	1.095389	::::	4.697268	0.259947	-2.470099	1.320399
	SD	1.474072	0.013528	0.048727	0.076820	::::	1.487458	0.013205	0.057656	0.090503
AWML	Mean	4.680732	0.258145	-2.468772	1.099112	::::	4.559140	0.260980	-2.470099	1.338108
	SD	1.322748	0.013379	0.048727	0.071574	::::	1.269776	0.011651	0.057656	0.084067
ML	Mean	4.616638	0.260284	-2.468778	1.099433	::::	4.718207	0.261001	-2.469418	1.325072
	SD	1.061964	0.012504	0.046272	0.069913	::::	1.087483	0.011867	0.053866	0.080545

Table 14: Finite-sample properties of alternative estimation methods for (H, σ, μ, κ) when $H = 0.260573$, $T = 10$, $\Delta = 1/250$ with various values of μ.

		κ	H	μ	σ	::::	κ	H	μ	σ
Method	True value	4.446145	0.260573	-2.500000	1.172012	::::	4.446145	0.260573	-1.500000	1.172012
MM	Mean	6.387549	0.280293	-2.502425	1.164356	::::	6.557786	0.280944	-1.507268	1.168321
	SD	2.128346	0.025208	0.045411	0.155490	::::	2.346244	0.026223	0.042684	0.165097
MCL	Mean	4.633282	0.259075	-2.502425	1.027643	::::	4.760308	0.259858	-1.507268	1.031093
	SD	1.408072	0.013218	0.045411	0.070849	::::	1.440295	0.013230	0.042684	0.070822
AWML	Mean	7.720884	0.192844	-2.502425	1.018842	::::	4.733229	0.259874	-1.507268	1.032021
	SD	1.303314	0.012313	0.045411	0.067321	::::	1.190782	0.012424	0.042684	0.064719
ML	Mean	4.704257	0.260290	-2.502018	1.031934	::::	4.767662	0.260916	-1.507925	1.034495
	SD	1.122367	0.011553	0.042531	0.061377	::::	1.043506	0.011693	0.040872	0.061317
Method	True value	4.446145	0.260573	-0.500000	1.172012	::::	4.446145	0.260573	0.500000	1.172012
MM	Mean	6.449656	0.281795	-0.495466	1.500560	::::	6.380326	0.280981	0.496312	1.167651
	SD	2.215225	0.025240	0.058120	0.198412	::::	2.219293	0.025869	0.048198	0.161037
MCL	Mean	4.604618	0.259538	-0.495466	1.316249	::::	4.599735	0.259537	0.496312	1.028776
	SD	1.422735	0.012374	0.058120	0.083962	::::	1.414295	0.013592	0.048198	0.072817
AWML	Mean	4.614885	0.260175	-0.495466	1.246918	::::	4.628527	0.259832	0.496312	1.032918
	SD	1.195485	0.011151	0.058120	0.065218	::::	1.223542	0.011465	0.048198	0.068321
ML	Mean	4.632349	0.260666	-0.495156	1.321244	::::	4.645695	0.261065	0.496755	1.034763
	SD	1.068652	0.010982	0.056782	0.073103	::::	1.111736	0.011760	0.046513	0.061723

Table 15: Finite-sample properties of alternative estimation methods for (H, σ, μ, κ) when $H = 0.260573$, $T = 10$, $\Delta = 1/250$ with various values of κ.

		κ	H	μ	σ	::::	κ	H	μ	σ
Method	True value	1.000000	0.260573	-2.465673	1.172012	::::	4.000000	0.260573	-2.465673	1.172012
MM	Mean	1.771108	0.275012	-2.471662	1.129448	::::	5.655800	0.278912	-2.471042	1.154490
	SD	1.049706	0.028485	0.199653	0.175615	::::	2.265258	0.026882	0.055927	0.170342
MCL	Mean	1.241364	0.259833	-2.471662	1.027760	::::	4.103789	0.259484	-2.471042	1.026788
	SD	0.633910	0.012649	0.199653	0.066936	::::	1.359120	0.011771	0.055927	0.065335
AWML	Mean	1.335375	0.260573	-2.471662	1.038502	::::	4.300459	0.262798	-2.471042	1.034784
	SD	1.220856	0.012757	0.199653	0.060939	::::	1.130115	0.011541	0.055927	0.059933
ML	Mean	1.494851	0.263265	-2.469936	1.044580	::::	4.418231	0.262342	-2.470825	1.040536
	SD	0.687877	0.011052	0.183393	0.058929	::::	1.035405	0.010274	0.055291	0.057587
Method	True value	7.000000	0.260573	-2.465673	1.172012	::::	10.000000	0.260573	-2.465673	1.172012
MM	Mean	9.432022	0.280405	-2.469224	1.161178	::::	12.333309	0.276733	-2.467195	1.138530
	SD	2.781223	0.021658	0.030181	0.144839	::::	3.087544	0.022344	0.019818	0.149289
MCL	Mean	7.290524	0.260941	-2.469224	1.036934	::::	10.564819	0.260861	-2.467195	1.037270
	SD	1.895625	0.013418	0.030181	0.073588	::::	2.257180	0.014246	0.019818	0.078490
AWML	Mean	6.822431	0.260994	-2.469224	1.043618	::::	10.480387	0.273741	-2.467195	1.037254
	SD	1.348381	0.010399	0.030181	0.067411	::::	1.466019	0.011672	0.019818	0.067287
ML	Mean	7.195729	0.261431	-2.468857	1.037218	::::	10.080197	0.259238	-2.467167	1.025906
	SD	1.247612	0.010680	0.028115	0.058322	::::	1.300650	0.011368	0.018669	0.061477

models in forecasting $\ln(RV)$. We split the sample period into two periods. The first period is between January 4, 2016 and December 31, 2020 and the second period is between January 4, 2021 and December 30, 2022. Similar to the main paper, on each day in the second period, h-day-ahead (with $h = 1, 5$) forecasts of daily $\ln(RV)$ are obtained from the following methods, namely MM with WXY, MCL with WXY, ML with WXY, MM with optimal, MCL with optimal, and ML with optimal. The rolling window estimation framework is also adopted. Table 16 reports the root mean squared error (RMSE) of each candidate model for h-day-ahead-forecast of $\ln(RV)$s with the best result highlighted in boldface for each h. Interestingly, ML with the optimal formula always performs the best, followed by MCL with the optimal formula. This result is consistent with the forecasting results of RVs.

To investigate if forecasts from the ML estimate with the optimal forecasting formula are statistically significantly more accurate than those of other estimation methods and forecasting formulas, Table 17 reports the Diebold-Mariano (DM) statistic based on the squared forecast errors and the p-value (in parenthesis) with the benchmark being ML with the optimal forecast (boldface means statistically significant at the 10% level). According DM, the forecast from the ML estimate with the optimal forecasting formula is always statistically different from the MM estimate with WXY's formula, the MCL estimate with WXY's formula of Wang et al. (2023), the ML estimate with WXY's

Table 16: RMSE for h-day-ahead-forecast of $\ln(RV)$ of fOU using three different estimation methods and two different forecasting methods.

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	0.3072	0.2919	0.2822	0.2962	0.2783	0.2694	0.2686	0.2601	0.2610	0.2607
MCL+WXY	0.2995	0.2851	0.2601	0.2819	0.2724	0.2617	0.2544	0.2568	0.2592	0.2505
ML+WXY	0.2846	0.2768	0.2387	0.2618	0.2499	0.2439	0.2487	0.2462	0.2534	0.2410
MM+optimal	0.2937	0.2612	0.2596	0.2775	0.2654	0.2594	0.2598	0.2657	0.2641	0.2679
MCL+optimal	0.2774	0.2259	0.2296	0.2574	0.2373	0.2315	0.2298	0.2304	0.2391	0.2330
ML+optimal	**0.2703**	**0.2240**	**0.2211**	**0.2502**	**0.2361**	**0.2303**	**0.2207**	**0.2280**	**0.2244**	**0.2272**
Panel B: $h = 5$										
MM+WXY	0.3446	0.3460	0.3347	0.3450	0.3368	0.3337	0.3289	0.3275	0.3309	0.3325
MCL+WXY	0.3390	0.3451	0.3264	0.3373	0.3280	0.3211	0.3251	0.3243	0.3284	0.3240
ML+WXY	0.3276	0.3340	0.3244	0.3350	0.3186	0.3166	0.3163	0.3107	0.3164	0.3123
MM+optimal	0.3346	0.3406	0.3258	0.3352	0.3231	0.3180	0.3229	0.3138	0.3206	0.3184
MCL+optimal	0.3163	0.3255	0.3149	0.3251	0.3076	0.3162	0.3152	0.3078	0.3082	0.3076
ML+optimal	**0.3119**	**0.3224**	**0.3139**	**0.3197**	**0.3054**	**0.3012**	**0.3084**	**0.3005**	**0.3046**	**0.3050**

formula of Wang et al. (2023), and the MM estimates with the optimal forecasting formula regardless the forecasting horizon. It is almost always different from the MCL estimate with the optimal forecasting formula regardless the forecasting horizon.

To determine whether the predictive model belongs to the set of "best" predictive model or not, we employ the model confidence set (MCS). Table 18 reports the p-value of the semi-quadratic statistic obtained from $2,000$ bootstrap iterations with a block length of 12. Values in boldface denote that the model belongs to the confidence set of the best models. From Table 18, we can see that the MM estimate with WXY formula, the MCL estimate with WXY's formula and the MM estimate with optimal forecasting formula are always rejected regardless of the $\ln(RV)$ series and forecasting horizon. The ML estimate with WXY's formula is rejected in all but four cases. The MCL estimate with optimal forecasting formula is rejected in a few cases. Most importantly, in no case, ML-estimated with optimal forecasting formula can be rejected. Similar conclusions can be obtained at the 5-day horizon.

Table 17: DM statistic for h-day-ahead-forecast of $\ln(RV)$ of fOU using three different estimation methods and two different forecasting methods (the benchmark model is ML with optimal).

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	-4.5116	-3.8385	-3.5269	-3.2693	-3.3666	-3.2396	-3.1542	-3.8318	-3.5463	-4.6066
	(**0.0000**)	(**0.0000**)	(**0.0002**)	(**0.0005**)	(**0.0003**)	(**0.0005**)	(**0.0008**)	(**0.0000**)	(**0.0002**)	(**0.0000**)
MCL+WXY	-3.0971	-3.2719	-3.1773	-2.4528	-2.8837	-2.7089	-2.2910	-2.3784	-2.3762	-1.8678
	(**0.0000**)	(**0.0000**)	(**0.0000**)	(**0.0071**)	(**0.0020**)	(**0.0034**)	(**0.0110**)	(**0.0087**)	(**0.0087**)	(**0.0309**)
ML+WXY	-1.5788	-1.3418	-1.5416	-1.7029	-1.8347	-1.5122	-1.4158	-1.9957	-1.9257	-1.5220
	(**0.0572**)	(**0.0898**)	(**0.0616**)	(**0.0443**)	(**0.0333**)	(**0.0652**)	(**0.0784**)	(**0.0230**)	(**0.0271**)	(**0.0640**)
MM+optimal	-2.6540	-2.3950	-2.7799	-2.0675	-2.0138	-1.9010	-2.9484	-3.0341	-2.5463	-2.6597
	(**0.0040**)	(**0.0083**)	(**0.0027**)	(**0.0193**)	(**0.0220**)	(**0.0287**)	(**0.0016**)	(**0.0012**)	(**0.0054**)	(**0.0039**)
MCL+optimal	-1.1982	-1.0351	-1.4686	-1.5823	-1.4006	-1.2531	-1.1705	-1.0721	-1.1042	-1.0678
	(0.1154)	(0.1503)	(**0.0710**)	(**0.0568**)	(**0.0807**)	(0.1051)	(0.1209)	(0.1418)	(0.1348)	(0.1428)
Panel A: $h = 5$										
MM+WXY	-3.3756	-3.1275	-3.2530	-3.3495	-3.4455	-3.4796	-3.2736	-3.0693	-3.0746	-3.1288
	(**0.0004**)	(**0.0009**)	(**0.0006**)	(**0.0004**)	(**0.0003**)	(**0.0003**)	(**0.0005**)	(**0.0011**)	(**0.0011**)	(**0.0009**)
MCL+WXY	-2.8407	-2.2543	-2.8143	-2.2435	-2.9293	-2.3500	-2.1966	-2.2511	-2.6160	-2.4733
	(**0.0023**)	(**0.0121**)	(**0.0024**)	(**0.0124**)	(**0.0017**)	(**0.0094**)	(**0.0140**)	(**0.0122**)	(**0.0044**)	(**0.0067**)
ML+WXY	-1.3067	-1.3949	-1.2046	-1.3749	-2.0173	-2.0687	-1.2844	-1.5998	-1.4599	-1.2001
	(**0.0957**)	(**0.0815**)	(0.1142)	(**0.0846**)	(**0.0218**)	(**0.0193**)	(**0.0995**)	(**0.0548**)	(**0.0722**)	(0.1151)
MM+optimal	-1.5901	-2.1619	-1.3208	-1.4020	-1.2482	-2.2660	-1.6710	-1.6781	-1.4876	-1.2299
	(**0.0559**)	(**0.0153**)	(**0.0933**)	(**0.0805**)	(0.1060)	(**0.0117**)	(**0.0474**)	(**0.0467**)	(**0.0684**)	(0.1094)
MCL+optimal	-1.2735	-1.1482	-1.1784	-1.3128	-1.1434	-1.0918	-1.1842	-1.3945	-1.2179	-1.0406
	(0.1014)	(0.1255)	0.1193	(**0.0946**)	(0.1264)	(0.1375)	(0.1182)	(**0.0816**)	(0.1116)	(0.1490)

Table 18: p-values of MSC for h-day-ahead-forecast of $\ln(RV)$ of fOU using three different estimation methods and two different forecasting methods (the benchmark model is ML with optimal).

Time series	SPY	XLB	XLE	XLF	XLI	XLK	XLP	XLU	XLV	XLY
Panel A: $h = 1$										
MM+WXY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MCL+WXY	0.0025	0.0077	0.0053	0.0296	0.0384	0.0365	0.0694	0.0631	0.0897	0.0015
ML+WXY	0.0285	**0.1188**	0.0933	0.0885	**0.1077**	**0.1045**	**0.1349**	**0.1050**	**0.1715**	0.0165
MM+optimal	0.0248	0.0339	0.0328	0.0418	0.0536	0.0611	0.0899	0.0747	0.0904	0.0035
MCL+optimal	**0.3025**	**0.3622**	**0.2245**	**0.2545**	**0.3188**	**0.1365**	**0.2754**	**0.2241**	**0.2075**	0.0455
ML+optimal	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**
Panel B: $h = 5$										
MM+WXY	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
MCL+WXY	0.0021	0.0030	0.0047	0.0023	0.0084	0.0019	0.0023	0.0017	0.0023	0.0044
ML+WXY	0.0911	**0.1423**	0.0930	0.0785	**0.1405**	**0.1180**	0.0939	0.0911	**0.1258**	0.0709
ML+WXY	0.0870	0.0537	0.0517	0.0285	0.0491	0.0599	0.0643	0.0416	0.0394	0.0690
MCL+optimal	**0.2061**	**0.3488**	**0.3000**	**0.2960**	**0.1809**	**0.1920**	**0.1235**	**0.2719**	**0.2611**	0.0953
ML+optimal	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**	**1.0000**