OPTIMAL ESTIMATION FOR GENERAL GAUSSIAN PROCESSES

Tetsuya Takabatake
Graduate School of Engineering Science, University of Osaka

Jun Yu
Faculty of Business Administration, University of Macau

Chen Zhang
Faculty of Business Administration, University of Macau

This paper proposes a novel exact maximum likelihood (ML) estimation method for general Gaussian processes, where all parameters are estimated jointly. The exact ML estimator (MLE) is consistent and asymptotically normally distributed. We prove the local asymptotic normality (LAN) property of the sequence of statistical experiments for general Gaussian processes in the sense of Le Cam, thereby enabling optimal estimation and statistical inference. The results rely solely on the asymptotic behavior of the spectral density near zero, allowing them to be widely applied. The established optimality not only addresses the gap left by Adenstedt (1974) that proposed an infeasible efficient estimator for the long-run mean μ, but also enables us to evaluate the finite-sample performance of the commonly used plug-in MLE, in which the sample mean is substituted into the likelihood. Our simulation results show that the plug-in MLE performs nearly as well as the exact MLE, alleviating concerns that inefficient estimation of μ would compromise the efficiency of the remaining parameter estimates.

Keywords: General Gaussian processes, Maximum likelihood estimation, Local asymptotic normality.

We are grateful to Carsten Chong, Morten Nielsen, Peter Phillips for comments on the first version of the paper. Tetsuya Takabatake, Email t.takabatake.es@osaka-u.ac.jp. Jun Yu, Email: junyu@um.edu.mo. Chen Zhang, Email: chenzhang@um.edu.mo.

1. INTRODUCTION

Gaussian processes have been widely applied across a broad range of scientific and applied disciplines, including economics, finance, physics, hydrology, and telecommunications. One of their most extensively studied features is the long-memory property, which captures long-range dependence. The discrete-time autoregressive fractionally integrated moving average (ARFIMA) process was introduced by Granger (1980) and Hosking (1981) to model this feature. In economics and finance, long memory has been examined in a wide array of time series, including the real economy (Diebold and Rudebusch, 1989, 1991), stock returns (Lo, 1991, Liu and Jing, 2018), exchange rates (Diebold et al., 1991, Cheung, 1993), and volatility (Ding et al., 1993, Andersen and Bollerslev, 1997, Andersen et al., 2003). Various mechanisms have been proposed to explain the emergence of long memory, including cross-sectional aggregation (Granger, 1980, Abadir and Talmain, 2002), regime switching (Potter, 1976, Diebold and Inoue, 2001), marginalization (Chevillon et al., 2018), and network effects (Schennach, 2018).

More recently, a rapidly growing strand of literature has focused on continuous-time Gaussian processes, which can characterize local behavior and reproduce the rough sample paths observed in volatility and trading volume (Gatheral et al., 2018, Fukasawa et al., 2022, Wang et al., 2023, Bolko et al., 2023, Shi et al., 2024b, Chong and Todorov, 2025). Two prominent models in this class are the fractional Brownian motion (fBm)(Mandelbrot, 1965, Mandelbrot and Van Ness, 1968, Gatheral et al., 2018) and the fractional Ornstein–Uhlenbeck (fOU) process (Cheridito et al., 2003, Wang et al., 2023). When applied to volatility and trading volume, fractional Gaussian noise (fGn, the first-order difference of fBm) and fOU exhibit anti-persistence (or roughness) (Gatheral et al., 2018, Fukasawa and Takabatake, 2019, Shi et al., 2024a,b, Wang et al., 2024). Several studies have begun to investigate the micro-level origins of this roughness. For example, El Euch et al. (2018) showed that in highly endogenous markets, rough volatility may arise from a large number of split orders, while Jusselin and Rosenbaum (2020) demonstrated that rough volatility emerges naturally under the no-arbitrage condition.

The ML estimation method of non-centered stationary discrete- and continuous-time Gaussian models with long memory or anti-persistency, referred to as general Gaussian processes, is the focus of this paper.[1] The consideration of non-centered continuous-time Gaussian processes is motivated by their extensive application in finance. Since alternative memory properties are relevant across various applications, our goal is to develop an estimation method that does not impose prior restrictions on the memory type of the process. The choice of ML estimation is motivated by the practical need for accurate estimation of all model parameters, particularly when computing impulse response functions or performing forecasts. In such cases, suboptimal estimators—such as the semi-parametric methods of Geweke and Porter-Hudak (1983), Robinson (1995), Phillips and Shimotsu (2004), Shimotsu (2010), the method of moments in Wang et al. (2023), and the composite likelihood approach in Bennedsen et al. (2024)—are not recommended, even though they may offer certain advantages in other contexts. For example, Corsi (2009) criticized semi-parametric methods for producing significantly biased and inefficient estimates in forecasting applications with ARFIMA models. Moreover, although the ML and Whittle ML (Whittle, 1951, Fox and Taqqu, 1986) estimators are asymptotically equivalent, the ML estimation method generally demonstrates superior finite-sample performance (Sowell, 1992, Cheung and Diebold, 1994).[2]

Considerable progress has been made in developing ML estimation methods, extending from specific parametric models to general Gaussian processes based on discrete-time observations.[3] For example, Yajima (1985) established the consistency and asymptotic normality of the MLE for the ARFIMA$(0,d,0)$ model with $d \in (0,0.5)$, representing the long-memory case. These results were subsequently extended

[1] Hualde and Robinson (2011), Nielsen (2015), Hualde and Nielsen (2020) employ a conditional sum of squares (CSS) method for fractional time series. It has the same asymptotic variance as the ML method. However, the CSS method leverages the ARFIMA-specific structure, making implementations simpler and faster. Unfortunately, it cannot be directly applied to continuous-time processes such as fOU.

[2] Rao and Yang (2021) propose new frequency-domain quasi-likelihoods that improve the finite-sample behavior of the Whittle ML estimator for short-memory Gaussian processes.

[3] A parallel literature studies parameter estimation for continuous-time fractional models under continuous-record observations; see Kleptsyna and Le Breton (2002).

to stationary Gaussian processes with long memory by Dahlhaus (1989, 2006), and further generalized to general Gaussian processes by Lieberman et al. (2012). However, these existing methods adopt a two-stage procedure in which μ is first estimated by the sample mean and then substituted into the likelihood, yielding a so-called plug-in MLE for the remaining parameters. Some implementations apply MLE to demeaned data (Tsai and Chan, 2005, Shi and Yu, 2023), which is essentially equivalent to the plug-in MLE procedure. Regarding the optimality of this procedure, the sample mean is clearly not the efficient estimator for μ. While Dahlhaus (1989, 2006) argued for the efficiency of the plug-in MLE by showing that its asymptotic covariance matrix equals the inverse of the Fisher information matrix, they did not explicitly establish the existence of a Crameŕ–Rao lower bound. Cohen et al. (2013) derived the LAN property for centered stationary Gaussian processes with long memory or anti-persistence, providing a minimax lower bound for general estimators. However, their results do not imply the asymptotic efficiency of the plug-in MLE. Another concern is that the inefficiency in estimating μ may impair the finite-sample performance of the plug-in MLE. Cheung and Diebold (1994) showed that when μ is unknown, the finite-sample performance of the MLE for the other parameters deteriorates, even though their asymptotic variances remain the same as in the known-mean case. To date, the problem of obtaining theoretically optimal estimators for all parameters for general Gaussian processes within the ML framework remains unresolved. The only exception is Wang et al. (2024) that considered MLE for all parameters jointly, including μ, in the fOU process. However, their framework is model-specific and not applicable to other fractional models. Moreover, the minimax efficiency of the MLE was not addressed.

We introduce a novel exact ML method, a term that we adopt to distinguish from the plug-in ML method, for general Gaussian processes, where all parameters are estimated jointly. We establish the consistency and asymptotic normality of the exact MLE. These results extend those in Wang et al. (2024) from fOU to general Gaussian processes. Moreover, we establish the LAN property of the sequence of statistical experiments in the Le Cam sense. This result extends that in Cohen et al. (2013) from centered stationary Gaussian processes to “non-centered” ones within the long span asymptotics, which is different from several extensions under the high-frequency asymptotics recently made by Brouste and Fukasawa (2018), Fukasawa and Takabatake (2019), Szymanski (2024), Szymanski and Takabatake (2023), Chong and Mies (2025). The LAN property implies the efficiency of the exact MLE, addressing the gap left by Adenstedt (1974), which proposed an efficient but infeasible estimator for the long-run mean μ. The results rely solely on the asymptotic behavior of the spectral density near zero for a discrete record of observations, which allows for broad applicability. [4] Notably, for continuous-time processes, the convergence rate of the exact MLE of μ depends on the spectral density's aliasing effect as well as the memory parameter, contrasting with the ARFIMA processes, which rely only on the memory parameter.

To demonstrate the practical applicability of the proposed method, we conduct three Monte Carlo simulation studies in which our exact ML estimator is applied to three widely studied non-centered processes: the ARFIMA(0,d,0) model, fGn, and fOU. Overall, our exact estimator for μ outperforms the sample mean. Regarding the performance of the plug-in MLE, our simulation results show that the plug-in MLE performs nearly as well as the exact MLE, alleviating concerns that inefficient estimation of μ would compromise the efficiency of the remaining parameter estimates. In particular, for the ARFIMA(0,d,0) model, the gain in efficiency for estimating μ aligns with the theoretical result of Adenstedt (1974). We also conduct a forecasting horse race for realized volatility using the fOU process with three alternative estimators: the exact MLE, the plug-in MLE, and the change-of-frequency (CoF) estimator by Wang et al. (2023). As expected, the exact MLE delivers the best forecasting performance, followed by the plug-in MLE, which performs satisfactorily, and then the CoF estimator.

[4] To ensure our theoretical results are broadly applicable and consistent with discrete observations, we work with the spectral density corresponding to discrete-time data, regardless of whether the underlying process is continuous- or discrete-time. In a subsequent paper, we demonstrate how to verify sufficient conditions on the discrete-time spectral density provided in this paper using conditions on the continuous-time spectral density for a wide range of continuous-time processes, such as the continuous-time ARFIMA model (Tsai and Chan, 2005, Tsai, 2009), the fractionally integrated continuous-time ARMA process (Brockwell and Marquardt, 2005) and Matérn Gaussian process (Matérn, 1986).

To sum up, we contribute to the literature in the following aspects. First, we propose a novel exact ML estimation method for all parameters in a general stationary Gaussian process, establishing its consistency and asymptotic normality, in response to the rapidly growing literature on rough volatility and fractional continuous-time models. Second, we prove the LAN property of the sequence of statistical experiments for general stationary Gaussian processes in the Le Cam sense, providing a theoretical foundation of optimal estimation and reinforcing the discussion on efficiency. The LAN property we have established is also essential for building asymptotic optimality of statistical tests and selecting the order of models based on the likelihood function, see Remark 5 for further references. Third, our method serves as a benchmark for evaluating the finite-sample performance of the existing plug-in MLE. Although the performance gap between the plug-in MLE and the MLE with known μ can be substantial in finite samples (Cheung and Diebold, 1994), our analysis shows that this difference is not driven by inefficiencies in estimating μ.

The remainder of this paper is organized as follows. Section 2 presents the exact ML method and develops asymptotic properties. In Section 3, we provide several examples to which our results can be applied. Section 4 presents a Monte Carlo study to assess the performance of the estimation method. Section 5 concludes the paper. The proof of the main results is found in the Appendix and the proof of technical lemmas is found in the Online Appendix.

2. EXACT MLE AND ASYMPTOTIC PROPERTIES

2.1. *Notation and Exact MLE*

Let Θ_ξ be a convex domain of $\mathbb{R}^{p-1}$ with compact closure and set $\Theta := \Theta_\xi \times (0,\infty)$. Denote by $\mathring{\Theta}$ the set of all interior points of Θ. Write $\theta = (\xi,\sigma)^\top \in \Theta$ and $\vartheta = (\theta,\mu)^\top = (\xi,\sigma,\mu)^\top \in \Theta \times \mathbb{R}$. Denote by $\partial_z = \partial/\partial z$, $\partial_\omega = \partial/\partial\omega$ and $\partial_j := \partial/\partial\theta_j$ for $j \in \{1,\dots,p+1\}$. For notational simplicity, ∂_0, ∂_z^0 and ∂_ω^0 denote the identify operator. The derivative operators $\partial^k_{j_1,\dots,j_k}$ are recursively defined by $\partial^k_{j_1,\dots,j_k} := \partial_{j_1} \circ \partial^{k-1}_{j_2,\dots,j_k}$ for $j_1,\dots,j_k \in \{0,1,\dots,p+1\}$ and $k \in \mathbb{N}$. Moreover, $\mathbf{1}_n$ denotes a n-dimensional vector whose all elements are equal to 1 and, for an integrable function f on $[-\pi,\pi]$,

$\Sigma_n(f)$ denotes the symmetric Toeplitz matrix whose (i,j)th elements are equal to the $(i-j)$th Fourier coefficients of f.

Let us consider a stationary Gaussian time series $\{X_j^{\vartheta}\}_{j\in\mathbb{Z}}$ with mean μ and spectral density function $s^X(\omega,\theta)$. We may write $s^X_\theta(\omega) := s^X(\omega,\theta)$. Let us denote by $\vartheta_0 = (\xi_0, \sigma_0, \mu_0)^\top$ an interior point of $\Theta\times\mathbb{R}$, which may call a true value of the parameter ϑ, and we assume that we observe a realization of $X_1^{\vartheta_0},\ldots,X_n^{\vartheta_0}$. Let $s^X_\xi(\omega) = s^X_\theta(\omega)/\sigma^2$. Then, for each $\vartheta = (\theta,\mu)^\top \in \Theta\times\mathbb{R}$ and $n\in\mathbb{N}$, we denote by $\mathbb{P}^n_\vartheta$ a distribution on the Borel space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ under which a random vector $\mathbf{X}_n = (X_1,\ldots,X_n)^\top$ follows a n-dimensional Gaussian vector with mean vector $\mu\mathbf{1}_n$ and variance-covariance matrix $\Sigma_n(s^X_\theta)$. We also denote by $\gamma^X_\theta(\cdot)$ the auto-covariance function of X^ϑ.

Denote by $\ell_n(\vartheta)\equiv\ell_n(\xi,\sigma,\mu)$ the Gaussian log-likelihood function of the observations $\mathbf{X}_n$ under the distribution $\mathbb{P}^n_\vartheta$, which is given by

$$\ell_n(\vartheta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 - \frac{1}{2}\log\det\left[\Sigma_n(s^X_\xi)\right] - \frac{1}{2\sigma^2}(\mathbf{X}_n-\mu\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n), \tag{1}$$

and then the maximum likelihood estimator (MLE)[5] is defined by

$$\widehat{\vartheta}_n^{\mathrm{MLE}} := (\widehat{\xi}_n^{\mathrm{MLE}}, \widehat{\sigma}_n^{\mathrm{MLE}}, \widehat{\mu}_n^{\mathrm{MLE}})^\top \in \mathop{\arg\max}_{(\xi,\sigma,\mu)^\top\in\Theta_\xi\times(0,\infty)\times\mathbb{R}} \ell_n(\vartheta).$$

Notice that, from the definition of MLE, MLE satisfies the estimation equations:

$$\partial_\mu\ell_n(\vartheta) = 0 \ \text{ and } \ \partial_\sigma\ell_n(\vartheta) = 0,$$

which imply that the equations

$$\mu = \frac{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{X}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n} =: \mu_n(\xi) \ \text{ and } \ \sigma^2 = \frac{1}{n}(\mathbf{X}_n-\mu\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n) =: \sigma_n^2(\xi,\mu)$$

hold for any $(\xi,\sigma,\mu)^\top\in\Theta_\xi\times(0,\infty)\times\mathbb{R}$. Then MLE $\widehat{\xi}_n^{\mathrm{MLE}}$ is a maximizer of the function $\bar{\ell}_n(\xi) := \ell_n(\xi,\bar{\sigma}_n(\xi),\mu_n(\xi))$, where $\bar{\sigma}_n^2(\xi) := \sigma_n^2(\xi,\mu_n(\xi))$ and $\bar{\sigma}_n(\xi) := \sqrt{\bar{\sigma}_n^2(\xi)}$,

[5]We have derived an alternative expression of the likelihood function using the conditional likelihood based on the Bayes formula, which improves the computational efficiency by avoiding direct computations of the inverse and determinant of a large-scale covariance matrix. See B.9 for details.

over the parameter $\xi \in \Theta_\xi$ and MLEs $\widehat{\mu}_n^{\mathrm{MLE}}$ and $\widehat{\sigma}_n^{\mathrm{MLE}}$ satisfy the equations

$$\widehat{\mu}_n^{\mathrm{MLE}} = \mu_n(\widehat{\xi}_n^{\mathrm{MLE}}) \ \text{ and } \ \widehat{\sigma}_n^{\mathrm{MLE}} = \bar{\sigma}_n(\widehat{\xi}_n^{\mathrm{MLE}}) = \sqrt{\sigma_n^2(\widehat{\xi}_n^{\mathrm{MLE}}, \widehat{\mu}_n^{\mathrm{MLE}})}.$$

Therefore, we define our proposed estimator $\widehat{\vartheta}_n := (\widehat{\xi}_n, \widehat{\sigma}_n, \widehat{\mu}_n)^\top$ by

$$\widehat{\xi}_n \in \arg\max_{\xi \in \Theta_\xi} \bar{\ell}_n(\xi), \ \ \widehat{\sigma}_n := \bar{\sigma}_n(\widehat{\xi}_n), \ \text{ and } \ \widehat{\mu}_n := \mu_n(\widehat{\xi}_n), \tag{2}$$

and we call $\widehat{\vartheta}_n = (\widehat{\xi}_n, \widehat{\sigma}_n, \widehat{\mu}_n)^\top$ the *exact MLE* throughout this paper. In subsequent sections, we investigate the asymptotic properties of the exact MLE.

2.2. *Consistency and Asymptotic Normality of Exact MLE*

We first introduce several conditions on the spectral density function $s_\theta^X(\omega)$ summarized in the following assumption that is used to obtain asymptotic properties of the exact MLE and the likelihood ratio process; see Sections 2.2 and 2.3 for details.

Assumption 1: (1) *For each $\theta \in \Theta$, $s_\theta^X(\omega)$ is a non-negative integrable even function in ω on $[-\pi, \pi]$ with 2π-periodicity. Moreover, it satisfies*

- *for each $\omega \in [-\pi, \pi] \backslash \{0\}$, $s_\theta^X(\omega)$ is three times continuously differentiable in θ on the interior of Θ,*
- *for each $\theta \in \Theta$ and $j \in \{1, \ldots, p\}$, $s_\theta^X(\omega)$ and $\partial_j s_\theta^X(\omega)$ are continuously differentiable in ω on $[-\pi, \pi] \setminus \{0\}$.*

(2) *If θ_1 and θ_2 are distinct elements of Θ, the set $\{\omega \in [-\pi, \pi] : s_{\theta_1}^X(\omega) \neq s_{\theta_2}^X(\omega)\}$ has a positive Lebesgue measure.*

(3) *There exists a continuous function $\alpha_X : \Theta_\xi \to (-\infty, 1)$ such that for any $\iota > 0$ and some constants $c_{1,\iota}, c_{2,\iota}, c_{3,\iota} > 0$, which only depends on ι, the following conditions hold for every $(\theta, \omega) \in \Theta \times [-\pi, \pi] \backslash \{0\}$:*

(*a*) $c_{1,\iota} |\omega|^{-\alpha_X(\xi) + \iota} \leq s_\theta^X(\omega) \leq c_{2,\iota} |\omega|^{-\alpha_X(\xi) - \iota}$.

(*b*) *For any $j_1, j_2, j_3 \in \{0, 1, \ldots, p\}$,*

$$\left| \partial^3_{j_1, j_2, j_3} s_\theta^X(\omega) \right| \leq c_{3,\iota} |\omega|^{-\alpha_X(\xi) - \iota} \ \text{ and } \ \left| \partial_\omega \partial_{j_1} s_\theta^X(\omega) \right| \leq c_{3,\iota} |\omega|^{-\alpha_X(\xi) - 1 - \iota}.$$

Assumption 1 is the usual conditions on the "discrete-time" spectral density function for stationary Gaussian time series with long/short/anti-persistent memory used in the literature; see the assumptions in Fox and Taqqu (1986), Dahlhaus (1989, 2006), Lieberman et al. (2012), Cohen et al. (2013) and Fukasawa and Takabatake (2019) as references. The time series $\{X_j^{\vartheta}\}_{j\in\mathbb{Z}}$ is said to have long memory (or long-range dependence) if $0 < \alpha_X(\xi) < 1$, short memory if $\alpha_X(\xi) = 0$, and anti-persistence if $\alpha_X(\xi) < 0$. The range $\alpha_X(\xi) \le -1$ corresponds to noninvertibility, and our results cover this case as well. Since these memory properties are relevant across various applications, we do not impose prior restrictions on the memory type of the process.

Before stating our main results, we introduce additional notations. We write $\widehat{\theta}_n := (\widehat{\xi}_n, \widehat{\sigma}_n)^\top$ and $\widehat{\vartheta}_n = (\widehat{\theta}_n, \widehat{\mu}_n)^\top$. Define $p\times p$ dimensional matrix $\mathcal{F}_p(\theta)$ by

$$\mathcal{F}_p(\theta) := \left(\frac{1}{4\pi}\int_{-\pi}^{\pi} \partial_i \log s_\theta^X(\omega)\partial_j \log s_\theta^X(\omega)\,\mathrm{d}\omega\right)_{i,j=1,\dots,p} = \begin{pmatrix} \mathcal{F}_{p-1}(\xi) & a_{p-1}(\theta) \\ a_{p-1}(\theta)^\top & 2\sigma^{-2} \end{pmatrix} \tag{3}$$

where

$$a_{p-1}(\theta) := \frac{1}{2\pi\sigma}\int_{-\pi}^{\pi} \partial_\xi \log s_\xi^X(\omega)\,\mathrm{d}\omega,\ \ \mathcal{F}_{p-1}(\xi) := \left(\frac{1}{4\pi}\int_{-\pi}^{\pi} \partial_i \log s_\xi^X(\omega)\partial_j \log s_\xi^X(\omega)\,\mathrm{d}\omega\right)_{i,j=1,\dots,p-1}.$$

Then we assume the following condition on $\mathcal{F}_p(\theta)$.

ASSUMPTION 2: *The matrix $\mathcal{F}_p(\theta)$ is invertible for each $\theta \in \mathring{\Theta}$.*

Our first main result is a weak consistency and an asymptotic normality of the sequence of the exact MLEs $\{\widehat{\theta}_n\}_{n\in\mathbb{N}}$ defined in (2), that is a generalization of, for example, Theorems 3.1 and 3.2 in Dahlhaus (1989) and Theorem 1 in Lieberman et al. (2012) to the case of general Gaussian processes using the multi-step estimation procedure based on the exact MLE defined in (2).

THEOREM 1: *Under Assumptions 1 and 2, the sequence of the exact MLEs $\{\widehat{\theta}_n\}_{n\in\mathbb{N}}$ is consistent and asymptotically normal. That is, for each $\vartheta = (\theta,\mu)^\top \in \mathring{\Theta}\times\mathbb{R}$,*

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to \mathcal{N}_p(\mathbf{0}_p, \mathcal{F}_p(\theta)^{-1}) \ \text{ as } n\to\infty$$

in law under the distribution $\mathbb{P}_\vartheta^n$, where $\mathcal{F}_p(\theta)$ is the non-singular matrix defined in (3).

To prove the asymptotic normality of MLEs for joint estimation $\vartheta = (\theta, \mu)^\top$ as well as MLE for μ, we need to further assume the precise asymptotic behavior of the spectral density function $s_\theta^X(\omega)$ around the frequency $\omega = 0$ given below.

Assumption 3: *In addition to Assumption 1, we further assume that there exists a continuous function $c_X : \Theta_\xi \to (0, \infty)$ such that for each $(\xi, \sigma)^\top \in \mathring{\Theta}$,*

$$s_\theta^X(\omega) \sim \sigma^2 c_X(\xi) |\omega|^{-\alpha_X(\xi)} \ \ \textit{as } |\omega| \to 0.$$

Based on the matrix $\mathcal{F}_p(\theta)$ defined in (3), we further define

$$\Phi_n(\vartheta) := \begin{pmatrix} n^{-\frac{1}{2}} I_p & \mathbf{0}_p \\ \mathbf{0}_p^\top & n^{-\frac{1}{2}(1-\alpha_X(\xi))} \end{pmatrix} \text{ and } \mathcal{I}(\vartheta) := \begin{pmatrix} \mathcal{F}_p(\theta) & \mathbf{0}_p \\ \mathbf{0}_p^\top & \frac{2\pi\Gamma(1-\alpha_X(\xi))\sigma^2 c_X(\xi)}{\mathrm{B}(1-\alpha_X(\xi)/2, 1-\alpha_X(\xi)/2)} \end{pmatrix}, \tag{4}$$

where $B(\alpha, \beta)$ is the beta-function. Note that, under Assumption 2, the matrix $\mathcal{I}(\vartheta)$ is also invertible for each $\vartheta \in \Theta \times \mathbb{R}$. Moreover, we introduce a normalized score function $\zeta_n(\vartheta)$ and an observed Fisher information matrix $\mathcal{I}_n(\vartheta)$ defined by

$$\zeta_n(\vartheta) := \Phi_n(\vartheta)^\top \partial_\vartheta \ell_n(\vartheta) \ \text{ and } \ \mathcal{I}_n(\vartheta) := -\Phi_n(\vartheta)^\top \partial_\vartheta^2 \ell_n(\vartheta) \Phi_n(\vartheta). \tag{5}$$

Now we can state our second main result of the asymptotic normality of the MLE for the joint parameter $\vartheta = (\theta, \mu)^\top$, summarized in the following theorem.

Theorem 2: *Under Assumptions 2 and 3, the sequence of the exact MLEs $\{\widehat{\vartheta}_n\}_{n \in \mathbb{N}}$ satisfies the following asymptotic normality: for each $\vartheta \in \mathring{\Theta} \times \mathbb{R}$, we have*

$$\Phi_n(\vartheta)^{-1}(\widehat{\vartheta}_n - \vartheta) = \mathcal{I}_n(\vartheta)^{-1} \mathbb{1}_{\{\det[\mathcal{I}_n(\vartheta)] > 0\}} \zeta_n(\vartheta) + o_{\mathbb{P}_\vartheta^n}(1) \to \mathcal{N}_{p+1}\left(\mathbf{0}_{p+1}, \mathcal{I}(\vartheta)^{-1}\right) \tag{6}$$

in law under the distribution $\mathbb{P}_\vartheta^n$ as $n \to \infty$, where $\zeta_n(\vartheta)$ and $\mathcal{I}_n(\vartheta)$ are defined in (5).

Remark 1: *In the literature, the treatment of the long-run mean μ in general Gaussian processes remains unsatisfactory. A common approach is to assume that μ is known and set to zero (Sowell, 1992, Hualde and Robinson, 2011, Nielsen, 2015). Another approach is to use the sample mean as a consistent estimator and then optimize the plug-in objective function (Fox and Taqqu, 1986, Dahlhaus, 1989, 2006, Lieberman et al., 2012). Only a few exceptions exist; for example, Hualde and Nielsen (2020) considered the CSS estimator of μ*

for discrete-time models, while Wang et al. (2024) developed the exact MLE for all parameters in fOU. However, their methods are model-specific. Bennedsen et al. (2024) addressed unknown μ in continuous-time Gaussian processes within the maximum composite likelihood framework. In contrast, our theorem applies to both discrete- and continuous-time Gaussian processes, and provides rigorous discussions on efficiency.

REMARK 2: *For ARFIMA processes, $\alpha_X(\xi)$ is a function of d (i.e. $\alpha_X(\xi) = 2d$). For fGn, we have $\alpha_X(\xi) = 2H - 1$ where Hurst parameter $H \in (0,1)$. However, for continuous-time models such as fOU, we have a) $\alpha_X(\xi) = 0$ when $H \in (0,1/2]$ and b) $\alpha_X(\xi) = 2H - 1$ when $H \in (1/2,1)$. This difference arises from the aliasing formula of the spectral density, which renders the spectral density bounded away from zero when $H \in (0,1/2]$. See Section 3 and Shi et al. (2024b) for more details. Therefore, for fOU, the exact MLE of μ converges at the $\sqrt{n}$ rate when $H \in (0,1/2]$, which differs from ARFIMA and fGn.*

REMARK 3: *The asymptotic normality properties in Theorems 1 and 2 show that the sequences of the plug-in MLEs of θ with nuisance parameter μ and the exact MLEs of $\vartheta = (\theta,\mu)^\top$ are respectively asymptotically efficient in the Fisher sense, in that their limiting covariance matrices equal the inverse of the Fisher information matrices, given by the limits of the sequences of the matrices $-n^{-1}\partial^2_\theta \ell_n((\theta,\mu_0)^\top)$ and $\mathcal{I}_n(\vartheta)$ defined in (5). These Fisher efficiencies have also been discussed in Dahlhaus (1989), Lieberman et al. (2012) for the plug-in MLE. However, these works do not establish the minimax optimality proved later in Corollary 5.*

One might wonder whether the minimax optimality can be deduced by passing to the limit from Cramér–Rao inequality formulated for possibly biased estimators. Unfortunately, this is not the case. Such finite-sample inequalities control pointwise variances but do not rule out the existence of superefficient points. A classical example is Hodges's estimator in the i.i.d. Gaussian location model, which is $\sqrt{n}$-consistent and has the same asymptotic variance as the MLE except at a single point where it is superefficient. This example illustrates that Cramér–Rao–type arguments alone are insufficient to rule out superefficient points, indicating that the existence of general lower bounds for estimators cannot be derived solely from such inequalities.

Le Cam (1953) showed that the set of superefficient points has Lebesgue measure zero, with subsequent extensions by Bahadur (1964) and Pfanzagl (1970). These results, how-

ever, rely on parametric i.i.d. models or other regularity assumptions. To the best of our knowledge, extensions of these results to statistical experiments induced by general Gaussian processes have not been established. Therefore, one cannot rely on Fisher efficiency or Cramér–Rao–type inequalities alone to establish minimax optimality in local neighborhoods of the true parameter. To overcome this limitation, one needs the LAN property together with Hájek–Le Cam's local asymptotic minimax theorem (Hájek, 1972, Le Cam, 1972), which ensures that no estimator can asymptotically achieve a smaller risk than the bound determined by the Fisher information in shrinking neighborhoods of the true parameter. This motivates the next subsection, where we establish the LAN property for statistical experiments induced by general Gaussian processes and then derive the minimax efficiency of the exact MLE as well as the plug-in MLE.

2.3. *Local Asymptotic Normality and Asymptotic Efficiency of the Exact MLE*

The concept of LAN is a cornerstone of modern asymptotic statistics. It was formally introduced by Le Cam (1960), based on previous contributions by Wald (1943) and the asymptotic theory of estimation developed by Le Cam (1953). The LAN property plays a central role not only in proving the asymptotic optimality of estimators but also in providing a general framework for statistical inference. In this subsection, we establish the LAN property for the sequence of statistical experiments $\{(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathbb{P}^n_\vartheta\}_{\vartheta\in\Theta\times\mathbb{R}})\}_{n\in\mathbb{N}}$ induced by general Gaussian processes. Building on this result, we then derive the local asymptotic minimax efficiency of the exact MLE. Further applications of the LAN property will also be discussed.

THEOREM 3: *Consider the sequence of rate matrices $\{\Phi_n(\vartheta)\}_{n\in\mathbb{N}}$ defined in (3). Under Assumptions 2 and 3, the family of distributions $\{\mathbb{P}^n_\vartheta\}_{\vartheta\in\Theta\times\mathbb{R}}$ satisfies the following LAN property at each $\vartheta \in \mathring{\Theta} \times \mathbb{R}$:*

$$\left|\log \frac{\mathrm{d}\mathbb{P}^n_{\vartheta+\Phi_n(\vartheta)u}}{\mathrm{d}\mathbb{P}^n_\vartheta} - \left(u^\top \zeta_n(\vartheta) - \frac{1}{2} u^\top \mathcal{I}(\vartheta) u\right)\right| = o_{\mathbb{P}^n_\theta}(1) \ \ \textit{as } n \to \infty,$$

where the invertible matrix $\mathcal{I}(\vartheta)$ is defined in (4) and the normalized score function $\zeta_n(\vartheta) = \Phi_n(\vartheta)^\top \ell_n(\vartheta)$ satisfies the convergence

$$\zeta_n(\vartheta) \to \mathcal{N}(0, \mathcal{I}(\vartheta)) \ \ \textit{as } n \to \infty$$

in law under the distribution $\mathbb{P}^n_\vartheta$.

REMARK 4: *Theorem 2.4 in Cohen et al. (2013) established the LAN property for centered stationary Gaussian time series with long-, short-, or anti-persistent memory under Assumptions 1 and 2. Theorem 3 extends this result to general Gaussian processes under the long-span asymptotics, in contrast to several recent works on high-frequency asymptotics (Brouste and Fukasawa, 2018, Fukasawa and Takabatake, 2019, Szymanski and Takabatake, 2023, Szymanski, 2024, Chong and Mies, 2025). For the centered case, the LAN property shown in Cohen et al. (2013) allows one to deduce a local asymptotic minimax lower bound through the Hájek–Le Cam local asymptotic minimax theorem (Hájek, 1972, Le Cam, 1972). However, their result applies only to centered Gaussian processes, and the extension to models with an unknown mean is not straightforward. As a consequence, it cannot be used to establish the minimax efficiency of either the exact MLE or the plug-in MLE in the non-centered setting.*

The LAN property provides local asymptotic minimax lower bounds for the risk of estimators of $\vartheta = (\theta, \mu)^\top$. In particular, the Hájek–Le Cam local asymptotic minimax theorem (see Hájek (1972), Ibragimov and Has'minskiĭ (1981), Le Cam (1972)) formalizes this bound, which we recall below.

THEOREM 4—Theorem II.12.1 in Ibragimov and Has'minskiĭ (1981): *Let* $\Theta \subset \mathbb{R}^d$ *be a parameter set, and let* θ_0 *be an interior point of* Θ. *Suppose that the family of distributions* $\{\mathbb{P}^n_\theta\}_{\theta\in\Theta}$ *satisfies the LAN property at* θ_0 *with a sequence of invertible* $d\times d$*-matrices* $\{\Phi_n(\theta_0)\}_{n\in\mathbb{N}}$ *and a* $d\times d$*-positive definite matrix* $\mathcal{I}(\theta_0)$. *Then, for any sequence of estimators* $\{\widehat{\theta}_n\}_{n\in\mathbb{N}}$ *and any symmetric nonnegative quasi-convex function* L *on* $\mathbb{R}^d$ *such that* $e^{-\varepsilon\|z\|^2_{\mathbb{R}^d}}L(z) \to 0$ *as* $\|z\|_{\mathbb{R}^d} \to \infty$ *for any* $\varepsilon > 0$, *we have*

$$\varliminf_{c\to\infty}\varliminf_{n\to\infty} \sup_{\theta\in\Theta:\|\Phi_n(\theta_0)^{-1}(\theta-\theta_0)\|_{\mathbb{R}^d}\le c} \mathbb{E}^n_\theta\left[L\left(\Phi_n(\theta_0)^{-1}(\widehat{\theta}_n-\theta)\right)\right] \ge (2\pi)^{-\frac{d}{2}}\int_{\mathbb{R}^d} L\left(\mathcal{I}(\theta_0)^{-\frac{1}{2}}z\right)\exp\left(-\frac{|z|^2}{2}\right)\mathrm{d}z.$$

Notice that we have already proved that the sequence of the exact MLEs $\{\widehat{\vartheta}_n\}_{n\in\mathbb{N}}$ defined in (2) satisfies the coupling property (6) in Theorem 2 so that, using the result in Section 7.12.(b) of Höpfner (2014) in addition to Theorem 3, we can conclude that the sequence of exact MLEs is asymptotically efficient in the local asymptotic minimax sense as well as in the Fisher sense, and then it attains the local asymptotic minimax bound of estimation given in Theorem 4. We summarize the aforementioned result in the following corollary.

Corollary 5—Asymptotic Minimax Optimality: *Consider the sequence of rate matrices $\{\Phi_n(\vartheta)\}_{n\in\mathbb{N}}$ defined in (3) and the matrix $\mathcal{I}(\vartheta)$ defined in (4). Under Assumptions 2 and 3, the sequence of the exact MLEs $\{\widehat{\vartheta}_n\}_{n\in\mathbb{N}}$ defined in (2) attains the local asymptotic minimax bound given in Theorem 4 at each $\vartheta_0 \in \mathring{\Theta}\times\mathbb{R}$. Namely, for any symmetric non-negative quasi-convex function L on $\mathbb{R}^{p+1}$ such that $e^{-\varepsilon\|z\|^2_{\mathbb{R}^{p+1}}}L(z)\to 0$ as $\|z\|_{\mathbb{R}^{p+1}}\to\infty$ for any $\varepsilon>0$, we obtain*

$$\lim_{n\to\infty}\sup_{\vartheta\in\Theta\times\mathbb{R}:\|\Phi_n(\vartheta_0)^{-1}(\vartheta-\vartheta_0)\|_{\mathbb{R}^{p+1}}\le c}\mathbb{E}^n_\vartheta\left[L\left(\Phi_n(\vartheta_0)^{-1}(\widehat{\vartheta}_n-\vartheta)\right)\right]=(2\pi)^{-\frac{p+1}{2}}\int_{\mathbb{R}^{p+1}}L\left(\mathcal{I}(\vartheta_0)^{-\frac{1}{2}}z\right)\exp\left(-\frac{|z|^2}{2}\right)\mathrm{d}z$$

for any $\vartheta_0\in\mathring{\Theta}\times\mathbb{R}$ and any constant $c\in(0,\infty)$.

Remark 5: *Corollary 5 highlights the role of the LAN property in establishing the minimax efficiency of the exact MLE. The relevance of LAN property, however, goes well beyond estimation. It provides a general framework for asymptotic inference. For example, it underlies the asymptotically uniformly most powerful unbiased (AUMPU) property of likelihood ratio tests (Choi et al., 1996) and supports consistency results for model selection criteria such as the (quasi) Bayesian information criterion (BIC) (Eguchi and Masuda, 2018). These potential applications illustrate the broader scope of the LAN property, whose detailed exploration is deferred to future research.*

2.4. *Comparison between Exact MLE and Plug-in MLE*

As an alternative estimator of θ, the plug-in MLE (PMLE) is defined by

$$\widehat{\theta}_n^{\mathrm{PMLE}}\in\arg\max_{\theta\in\Theta_*}\ell_n((\theta,\widetilde{\mu}_n)^\top)\tag{7}$$

using some compact set $\Theta_*\subset\Theta$ and a plug-in estimator $\widetilde{\mu}_n$. Under similar assumptions to Assumption 1, Dahlhaus (1989, 2006) and Lieberman et al. (2012) show that the plug-in MLE is consistent and asymptotically normal with the asymptotic variance-covariance matrix $\mathcal{F}_p(\theta_0)^{-1}$ under the distribution $\mathbb{P}^n_{\vartheta_0}$, when the plug-in estimator $\widetilde{\mu}_n$ satisfies

$$\widetilde{\mu}_n=\mu_0+o_{\mathbb{P}^n_{\vartheta_0}}(n^{-\frac{1}{2}(1-\alpha_X(\xi_0))})\ \text{ as } n\to\infty.\tag{8}$$

The condition given in (8), verified under Assumption 3, corresponds to the assumption on the estimator of μ in Theorem 3.2 of Dahlhaus (1989) for the long memory case $\alpha_X(\xi_0) \in (0,1)$, and to Assumption 5 of Lieberman et al. (2012) for the long/short/anti-persistent memory case $\alpha_X(\xi_0) \in (-\infty,1)$. The plug-in MLE shares the same convergence rate and asymptotic variance as our exact MLE of θ, implying that it is also asymptotically efficient.

Moreover, Theorem 3 combined with Theorem 4 yields the asymptotic minimax lower bound

$$\varliminf_{n\to\infty} \sup_{\|\Phi_n(\vartheta_0)^{-1}(\vartheta-\vartheta_0)\|_{\mathbb{R}^{p+1}}\le c} \mathbb{E}_\vartheta^n\left[n^{1-\alpha_X(\xi)}(\widetilde{\mu}_n-\mu)^2\right] \ge \frac{2\pi\sigma_0^2 c_X(\xi_0)\Gamma(1-\alpha_X(\xi_0))}{B(1-\alpha_X(\xi_0)/2, 1-\alpha_X(\xi_0)/2)} \tag{9}$$

for any $c>0$ and any sequence of estimators $\{\widetilde{\mu}_n\}_{n\in\mathbb{N}}$. This shows that the convergence rate in (8) coincides with the minimax optimal rate given in (9).

It is known that the best linear unbiased estimator (BLUE) of μ, denoted $\widehat{\mu}_n^{\mathrm{BLUE}}$, satisfies (8) for all $\alpha_X(\xi_0) \in (-\infty,1)$ (Adenstedt, 1974). However, $\widehat{\mu}_n^{\mathrm{BLUE}}$ is infeasible since it depends on the unknown true value ξ_0. Samarov and Taqqu (1988) proved that the widely used sample mean satisfies (8) when $\alpha_X(\xi_0) \in (-1,1)$, but it fails to attain the minimax optimal rate when $\alpha_X(\xi_0) \in (-\infty,-1]$, and even within $(-1,1)$ its asymptotic variance does not achieve the minimax lower bound.

The asymptotic inefficiency of the sample mean in fractional time series has been explicitly quantified. Adenstedt (1974) established the result for ARFIMA$(0,d,0)$, while Samarov and Taqqu (1988) extended it to general stationary Gaussian time series, including ARFIMA(p,d,q), fGn, and fOU. Their results are based on the asymptotic relative efficiency between $\widehat{\mu}_n^{\mathrm{BLUE}}$ and the sample mean $\overline{X}_n$, defined as

$$e(n,\vartheta) := \frac{\mathrm{Var}_\vartheta[\widehat{\mu}_n^{\mathrm{BLUE}}]}{\mathrm{Var}_\vartheta[\overline{X}_n]}.$$

Specifically, they showed that if $\alpha_X(\xi) \in (-1,1)$ and $d := \alpha_X(\xi)/2$, then under mild technical conditions verified by Assumption 3,

$$\lim_{n\to\infty} e(n,\vartheta) = \frac{-\pi d(1+2d)}{B(1-d,1-d)\sin(-\pi d)} = \frac{(2d+1)\Gamma(d+1)\Gamma(2-2d)}{\Gamma(1-d)}. \tag{10}$$

This limiting efficiency is always strictly less than 1, except in the trivial case $d=0$.

3. EXAMPLES

Our assumptions on the spectral density are very general and the results can be applied to many well-known processes, including but not limited to the ARFIMA(p,d,q) process with $|d|<1/2$, fGn with an unknown mean and fOU.

3.1. *Non-Centered Gaussian ARFIMA(p,d,q) Process*

The *non-centered* ARFIMA(p,d,q) process was introduce by Granger (1980) and Hosking (1981) independently. For notational simplicity, we start with the ARFIMA($0,d,0$) model. The non-centered ARFIMA($0,d,0$) model is specified as

$$X_j-\mu=\sigma(1-L)^{-d}\epsilon_j \text{ with } |d|<1/2, \tag{11}$$

where L is the lag operator, $(1-L)^{-d}$ is the fractional difference operator with the memory parameter d and $\epsilon_j \overset{\text{iid}}{\sim} N(0,1)$. It reduces to a Gaussian white noise when $d=0$. When $d\in(-1/2,1/2)$, ARFIMA($0,d,0$) is stationary and invertible (Bloomfield, 1985). Let $u_j:=(1-L)^{-d}\epsilon_j$ be the fractionally integrated process and $\gamma_u(k):=\mathrm{Cov}[u_j,u_{j+k}]$ be its kth-order auto-covariance. According to Hosking (1981), the auto-covariance function of $u=\{u_j\}_{j\in\mathbb{Z}}$ is expressed by

$$\gamma_u(k)=\frac{(-1)^k\Gamma(1-2d)}{\Gamma(k-d+1)\Gamma(1-k-d)},\quad k\in\mathbb{Z}. \tag{12}$$

The long-run variance covariance $\sum_{k=-\infty}^{\infty}\gamma_u(k)=\infty$ when $d\in(0,1/2)$ and $\sum_{k=-\infty}^{\infty}\gamma_u(k)=0$ when $d\in(-1/2,0)$. Therefore, $\{u_j\}_{j\in\mathbb{Z}}$ has a long memory if $d\in(0,1/2)$ and is anti-persistent if $d\in(-1/2,0)$. The spectral density of the model is given by

$$s_\theta^X(\omega)=\frac{\sigma^2}{2\pi}|1-e^{-i\omega}|^{-2d}\sim\frac{\sigma^2}{2\pi}|\omega|^{-2d}\quad\text{as } |\omega|\to 0.$$

The *non-centered* ARFIMA(p,d,q) process is defined by

$$\phi_\xi(L)(X_j-\mu)=\sigma\psi_\xi(L)u_j,\quad j\in\mathbb{Z}, \tag{13}$$

where $p,q\in\mathbb{N}\cup\{0\}$, $\xi:=(\phi_1,\ldots,\phi_p,\psi_1,\ldots,\psi_q)\in\mathbb{R}^{p+q}$, $\phi_\xi(z):=1-\phi_1z-\cdots-\phi_pz^p$ and $\psi_\xi(z):=1+\psi_1z+\cdots+\psi_qz^q$. Assume that for each ξ, the functions $\phi_\xi(z)$ and $\psi_\xi(z)$ have no common roots in $\mathbb{C}$, and that all their roots lie outside the unit circle. This implies that $\phi_\xi(z) \neq 0$ and $\psi_\xi(z) \neq 0$ for $|z| \leq 1$. Then, for $|d| < 1/2$, the difference equation in (13) admits a unique stationary solution $X = \{X_j\}_{j\in\mathbb{Z}}$ of the form

$$X_j = \mu + \sigma\phi_\xi(L)^{-1}\psi_\xi(L)(1-L)^{-d}\epsilon_j, \quad j \in \mathbb{Z}.$$

The spectral density function of ARFIMA(p,d,q) is given by

$$s_\theta^X(\omega) = \frac{\sigma^2}{2\pi}|1-e^{-i\omega}|^{-2d}\frac{|\psi_\xi(e^{-i\omega})|^2}{|\phi_\xi(e^{-i\omega})|^2}, \quad \omega \in (-\pi,\pi].$$

Since the assumption $\phi_\xi(z) \neq 0$ for $|z| \leq 1$ ensures that the spectral density function of ARMA(p,q),

$$f_{\mathrm{ARMA}}(\omega) := \frac{\sigma^2}{2\pi}\frac{|\psi_\xi(e^{-i\omega})|^2}{|\phi_\xi(e^{-i\omega})|^2},$$

is bounded away from zero on $[-\pi,\pi]$, the singularity of the spectral density of X in the vicinity of zero frequency is governed by the ARFIMA$(0,d,0)$ factor $|1-e^{-i\omega}|^{-2d}$. As $\omega \to 0$, it exhibits the asymptotic behavior

$$s_\theta^X(\omega) \sim \sigma^2 c_X(\xi)|\omega|^{-\alpha_X(\xi)}.$$

For the *non-centered* ARFIMA(p,d,q) process, $\alpha_X(\xi) = 2d, c_X(\xi) = (2\pi)^{-1}\left|\frac{\psi_\xi(1)}{\phi_\xi(1)}\right|^2$ in Assumptions 1 and 3. Hence, our results are applicable to the non-centered Gaussian ARFIMA(p,d,q) process. According to Theorem 2, when $d < 0$, the convergence rate for the exact MLE of μ is $\frac{1}{2}(1-\alpha_X(\xi)) > 1/2$, indicating superconsistency.

3.2. *Non-Centered Fractional Gaussian Noise*

The fBm with Hurst index $H \in (0,1)$, denoted by $B^H = \{B_t^H\}_{t\in\mathbb{R}}$, is a unique centered Gaussian process that is almost surely equal to zero at $t = 0$ and possesses both stationary increments and H-self-similarity properties. Specifically, these properties are expressed as

$$B_t^H - B_s^H \overset{d}{=} B_{t-s}^H - B_0^H \text{ and } B_{ct}^H \overset{d}{=} c^H B_t^H$$

for any $s,t \in \mathbb{R}$ and $c > 0$, where $\overset{d}{=}$ denotes equality in distribution. Mandelbrot and Van Ness (1968) demonstrated that fBm can be represented as a causal moving average process involving the past differential increments of a (two-sided) standard Brownian motion $B = \{B_t\}_{t\in\mathbb{R}}$. This representation is given by

$$B_t^H = \frac{1}{\Gamma(H+0.5)}\left\{\int_{-\infty}^{0}\left[(t-s)^{H-0.5}-(-s)^{H-0.5}\right]\mathrm{d}B_s + \int_0^t (t-s)^{H-0.5}\,\mathrm{d}B_s\right\},$$

where $\Gamma(x)$ denotes the gamma function,[6] which implies that fBm reduces to the standard Brownian motion when $H = 0.5$.

The sequence of increments of fBm is the (standard) fGn, denoted by $\{\epsilon_j\}_{j\in\mathbb{Z}}$. The fGn with mean μ and variance σ^2 is defined through

$$X_j := \mu + \sigma\epsilon_j = \mu + \sigma(B_j^H - B_{j-1}^H),\ \ j\in\mathbb{Z}.$$

From the definition of fBm, its covariance function is given by

$$\mathrm{Cov}[B_t^H, B_s^H] = \frac{1}{2}\left[|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right],\ \ \forall t,s\in\mathbb{R},$$

which yields the expression of the auto-covariance function of $X = \{X_j\}_{j\in\mathbb{Z}}$ by

$$\gamma_\theta^X(k) := \mathrm{Cov}[X_j, X_{j+k}] = \frac{\sigma^2}{2}\left[|k+1|^{2H} - 2|k|^{2H} + |k-1|^{2H}\right],\ \ k\in\mathbb{Z},\ \ \theta = (H,\sigma)^\top.$$

Notice that the Taylor expansion yields the following asymptotic expression:

$$\gamma_\theta^X(k) \sim \sigma^2 H(2H-1)|k|^{2H-2}\ \text{ as } |k|\to\infty. \tag{14}$$

The spectral density function of fGn is given by Sinai (1976):

$$s_\theta^X(\omega) = \sigma^2 C_H\{2(1-\cos(\omega))\}\sum_{k=-\infty}^{\infty}|2\pi k+\omega|^{-1-2H}\ \text{ for } \omega\in[-\pi,\pi], \tag{15}$$

where $C_H := (2\pi)^{-1}\Gamma(2H+1)\sin(\pi H)$. It can be shown that

$$s_\theta^X(\omega) \sim \sigma^2 C_H|\omega|^{1-2H},\ \text{when } \omega\to 0.$$

For the non-centered fGn, $\alpha_X(\xi)$ and $c_X(\xi)$ in Assumptions 1 and 3 are $\alpha_X(\xi) = 2H-1$ and $c_X(\xi) = C_H$. Hence, our results are applicable to the non-centered fractional

[6]This is also referred to as the Type I fBm.

Gaussian noise. According to Theorem 2, when $H < 0.5$, the convergence rate for the exact MLE of μ is $\frac{1}{2}(1-\alpha_X(\xi)) > 1/2$, implying superconsistency. This result echoes that for ARFIMA.

3.3. *Fractional Ornstein-Uhlenbeck Process*

The fOU is an extension of the classical Ornstein-Uhlenbeck (OU) process, where the driving noise is replaced by fBm with Hurst index $H \in (0,1)$. This process is particularly useful for modeling systems that exhibit long-range dependence and local self-similarity, which cannot be captured by the classical OU process. The stationary fOU process with a long-run mean μ has applications in various fields, including mathematical finance, physics, and time series modeling.

The fOU process $Y = \{Y_t\}_{t\in\mathbb{R}}$ is defined by a unique solution of the following linear SDE (Stochastic Differential Equation):

$$\mathrm{d}Y_t = -\kappa(Y_t - \mu)\,\mathrm{d}t + \sigma\,\mathrm{d}B_t^H, \;\; t \geq 0, \tag{16}$$

with initial condition Y_0, where $B^H = \{B_t^H\}_{t\in\mathbb{R}}$ is an fBm with Hurst index H. The explicit solution of this SDE is given by

$$Y_t = Y_0 e^{-\kappa t} + \mu(1 - e^{-\kappa t}) + \sigma \int_0^t e^{-\kappa(t-s)}\,\mathrm{d}B_s^H, \;\; t \geq 0, \tag{17}$$

where the above stochastic integral can be interpreted as the pathwise Riemann-Stieltjes integral or the Wiener integral associated with fBm for any $H \in (0,1)$.

The fOU process reduces to the classical OU process when $H = 0.5$ and to fBm when $\kappa = 0$. When $\kappa > 0$, the stationary solution of the SDE (16), denoted by $\bar{Y} = \{\bar{Y}_t\}_{t\in\mathbb{R}}$, is given by

$$\bar{Y}_t := \mu + \sigma \int_{-\infty}^t e^{-\kappa(t-s)}\,\mathrm{d}B_s^H, \;\; t \in \mathbb{R}. \tag{18}$$

For $t \geq 0$, the unique solution of the SDE (16) with the initial condition

$$Y_0 = \mu + \sigma \int_{-\infty}^0 e^s\,\mathrm{d}B_s^H$$

is exactly equal to the stationary solution $\bar{Y}$ in (18), and then the error between $\bar{Y}_t$ and Y_t with arbitrary initial condition Y_0 is expressed by

$$|Y_t - \bar{Y}_t| = |Y_0 - \bar{Y}_0| e^{-\kappa t}, \quad t \geq 0,$$

which implies that the error between the solutions (17) and (18) converges to zero exponentially as $t \to \infty$ for arbitrary initial condition Y_0. In the rest of this section, we consider the case where a data-generating process is the discretely and equidistantly observed time series from the stationary solution given in (18).

Consider a stationary time series $X = \{X_j\}_{j\in\mathbb{Z}}$ of the form $X_j := Y_{j\Delta}$ for $j \in \mathbb{Z}$ with the sampling frequency Δ. Write $\xi = (H,\kappa)^\top$ and $\theta = (\xi,\sigma)^\top$. Notice that the time series X is stationary and the following expression of its auto-covariance is available from Garnier and Sølna (2018) when $\kappa > 0$:

$$\gamma_\theta^X(k) = \frac{\sigma^2}{2\kappa^{2H}}\left(\frac{1}{2}\int_{-\infty}^{\infty} e^{-|s|}|\kappa k\Delta + s|^{2H}\mathrm{d}s - |\kappa k\Delta|^{2H}\right), \quad k \in \mathbb{Z}. \tag{19}$$

Cheridito et al. (2003) and Hult (2003) provide the spectral density function of the stationary solution $\bar{Y}$ given by

$$s_\theta^{\bar{Y}}(z) = \sigma^2 C_H |z|^{1-2H}(\kappa^2 + z^2)^{-1} \text{ for } z \in (-\infty,\infty). \tag{20}$$

Due to the aliasing formula of the spectral density function (*e.g.* see Priestley (1988)), the spectral density function of the discrete-time process $X = \{X_j\}_{j\in\mathbb{Z}}$ is given by

$$s_\theta^X(\omega) = \frac{1}{\Delta}\sum_{k\in\mathbb{Z}} s_\theta^{\bar{Y}}\left(\frac{\omega + 2\pi k}{\Delta}\right) = \sigma^2 C_H \Delta^{2H} \sum_{k=-\infty}^{\infty} \frac{|\omega + 2\pi k|^{1-2H}}{(\kappa\Delta)^2 + (\omega + 2\pi k)^2} \tag{21}$$

for $\omega \in [-\pi,\pi]$, see also Hult (2003). As $|\omega| \downarrow 0$, Shi et al. (2024b) has shown that

$$s_\theta^X(\omega) \sim \begin{cases} \sigma^2 C_H \Delta^{2H} \sum_{k=-\infty}^{\infty} \frac{|2\pi k|^{1-2H}}{(\kappa\Delta)^2 + (2\pi k)^2}, & \text{when } 0 < H \leq \frac{1}{2}, \\ \sigma^2 C_H \Delta^{2H-2}\kappa^{-2}|\omega|^{1-2H}, & \text{when } \frac{1}{2} < H < 1. \end{cases} \tag{22}$$

Hence, in Assumptions 1 and 3 for the fOU process, we have a) $\alpha_X(\xi) = 0$ and $c_X(\xi) = s_\theta^X(0)/\sigma^2$ when $H \in (0,1/2]$ and b) $\alpha_X(\xi) = 2H - 1$ and $c_X(\xi) = C_H \Delta^{2H-2}\kappa^{-2}$ when $H \in (1/2,1)$. Hence, our results are applicable to the fOU process. However, the function $\alpha_X(\xi)$ exhibits a sharp contrast compared to that of the ARFIMA and fGn. According to Theorem 2, the convergence rate for the exact MLE of μ in fOU is $\sqrt{n}$ when $H \leq 1/2$, as recently reported in (Wang et al., 2024).

4. MONTE CARLO STUDY

We consider three data-generating processes (DGPs): ARFIMA$(0,d,0)$, fGn and fOU. For simplicity, the long-run mean μ is set to 0 and the scale parameter σ is set to 1. The results for $\mu \pm 1$ are provided in Appendix B.11. The parameter d in the ARFIMA$(0,d,0)$ model takes 9 values: $\{-0.4,-0.3,\ldots,0,\ldots,0.4\}$, while the parameter H in fGn and fOU takes 9 values: $H = 0.1,0.2,\ldots,0.9$. The fOU has an additional parameter $\kappa = 10$. For each DGP, we compare our exact MLE with the two MLEs considered in Cheung and Diebold (1994). MLE1 refers to the case where μ is known, MLE2 refers to our exact MLE, and MLE3 refers to the plug-in MLE, where the sample mean is used as an estimator for μ. The number of replications is set to 1000. The sample size is set to 250 or 1000. Reported are the bias, standard error (Std), and root mean squared error (RMSE) across all replications for each method.

Table I reports the results for ARFIMA and Table II for fGn. From Tables I-II, we have the following findings. First, in terms of convergence rates, the performance of MLE2 aligns well with our asymptotic theory. For μ, the convergence rate is $n^{-(1-\alpha_X(\xi))/2}$, which becomes slower as d increases toward $1/2$ from $-1/2$ (or as H increases toward 1 from 0). A similar pattern can be observed for the sample mean, as it shares the same convergence rate as given in (8) and (9). For the remaining parameters, the convergence rate remains at $n^{1/2}$. Second, MLE2 of μ always performs better than MLE3 of μ, except when $d = 0$ in ARFIMA or $H = 1/2$ in fGn, where the two methods perform nearly identically. Third, interestingly, this superior performance in estimating μ by MLE2 does not translate into better performance in estimating other parameters. Using the true value of μ, MLE1 does not lead to better performance in estimating other parameters. The three ML methods lead to a similar finite sample performance for parameters other than μ.

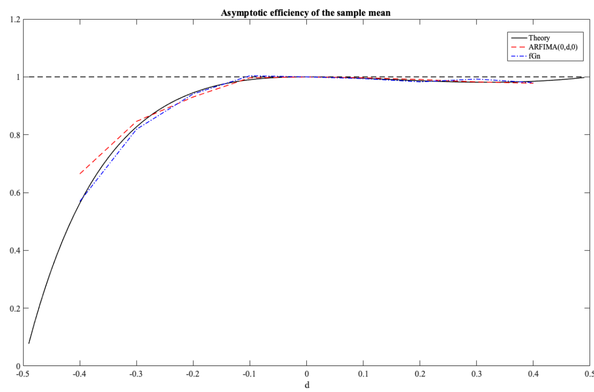

FIGURE 1.—Relative inefficiency of the sample mean over the exact MLE as a function of d for ARFIMA$(0,d,0)$ and fGn when $n = 1000$. For fGn, $d = H - 1/2$.

To see how the relative inefficiency of the sample mean over MLE2 of μ, the two dashed lines in Figure 1 plot the ratio of the sample variance of MLE2 for μ to that of MLE3 as a function of d for the two models when $n = 1000$ in ARFIMA and fGn. Clearly, the relative inefficiency goes up rapidly as d is closer to -0.5 in ARFIMA and fGn. For comparison, also plotted by the solid line is the theoretical asymptotic inefficiency given in (10). The red dashed line is closely aligned with the theory.

Tables III-IV report the results for fOU, from which we observe the following findings. First, in terms of convergence rates, the performance of MLE2 aligns well with our asymptotic theory. For μ, the convergence rate is $n^{1/2}$ when $H \leq 1/2$, and transitions to n^{1-H} when $H > 1/2$. The standard deviation of the estimator for μ decreases substantially as the sample size increases from 250 to 1000 when $H \leq 1/2$; however, as H approaches 1, the percentage reduction becomes markedly smaller. A similar pattern can be observed for the sample mean, as it shares the same convergence rate as given in (8) and (9). For the remaining parameters, the convergence rate remains at $n^{1/2}$. Second, we see a clear dominance of MLE2 over MLE3 in terms of finite sample performance of estimates of μ, when both n is large and H is near either zero or one. Third, this superior performance in estimating μ by MLE2 does not translate into a better performance in estimating other parameters. Using the true value of μ, MLE1 does not lead to better performance in estimating the other 3 parameters. The three ML methods lead to a similar finite sample performance for parameters other than μ.[7]

To see how the relative inefficiency of the sample mean over MLE2 of μ, Figure 2 plots the ratio of the sample variance of MLE2 for μ to that of MLE3 as a function of H for fOU when $n = 1000$.[8] When H is close to 0.5, the asymptotic efficiency of the sample mean is near 1. However, as H approaches 0 or 1, the relative inefficiency of the sample mean increases rapidly.

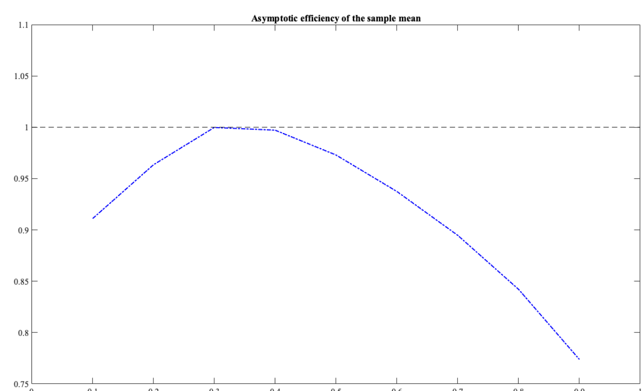

FIGURE 2.—Relative inefficiency of the sample mean over the exact MLE as a function of H for fOU when $n = 1000$.

[7] In the online supplement (Section B.10), we conduct a forecasting horse race for realized volatility using the fOU process with three alternative estimators: MLE2, MLE3, and the CoF estimator. As expected, MLE2 delivers the best forecasting performance, followed by MLE3 and then the CoF estimator.

[8] When the sample size goes up, our unreported simulation results show that the relative inefficiency goes down as predicted by our theory.

TABLE I

Bias and Std of alternative MLEs for ARFIMA$(0,d,0)$: $\mu = 0$ and $\sigma = 1$. MLE1 is MLE with known μ; MLE2 is our exact MLE; MLE3 is PMLE.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
$n = 250$										
		d = -0.40			d = -0.30			d = -0.20		
μ	Bias	-	-0.0004	-0.0003	-	-0.0005	-0.0010	-	0.0010	0.0012
	Std	-	0.0096	0.0113	-	0.0151	0.0162	-	0.0244	0.0248
d	Bias	-0.0029	-0.0140	-0.0074	-0.0029	-0.0157	-0.0121	-0.0056	-0.0194	-0.0179
	Std	0.0507	0.0510	0.0500	0.0518	0.0553	0.0542	0.0509	0.0536	0.0529
σ	Bias	-0.0001	-0.0022	-0.0012	-0.0019	-0.0042	-0.0038	-0.0025	-0.0049	-0.0047
	Std	0.0445	0.0445	0.0446	0.0447	0.0449	0.0449	0.0447	0.0446	0.0446
		d = -0.10			d = 0.00			d = 0.10		
μ	Bias	-	-0.0003	-0.0004	-	-0.0018	-0.0017	-	0.0028	0.0027
	Std	-	0.0394	0.0396	-	0.0637	0.0636	-	0.1024	0.1024
d	Bias	-0.0048	-0.0189	-0.0183	-0.0043	-0.0185	-0.0184	-0.0051	-0.0189	-0.0189
	Std	0.0492	0.0522	0.0519	0.0500	0.0528	0.0527	0.0492	0.0527	0.0527
σ	Bias	-0.0040	-0.0064	-0.0063	-0.0027	-0.0049	-0.0049	-0.0034	-0.0055	-0.0055
	Std	0.0454	0.0455	0.0455	0.0450	0.0451	0.0451	0.0443	0.0442	0.0442
		d = 0.20			d = 0.30			d = 0.40		
μ	Bias	-	-0.0013	-0.0022	-	0.0069	0.0093	-	0.0301	0.0286
	Std	-	0.1811	0.1824	-	0.3464	0.3509	-	0.6823	0.6887
d	Bias	-0.0046	-0.0191	-0.0191	-0.0049	-0.0220	-0.0219	-0.0129	-0.0301	-0.0299
	Std	0.0493	0.0529	0.0530	0.0467	0.0527	0.0528	0.0412	0.0462	0.0462
σ	Bias	-0.0047	-0.0067	-0.0067	-0.0027	-0.0049	-0.0049	-0.0034	-0.0051	-0.0050
	Std	0.0441	0.0442	0.0442	0.0430	0.0429	0.0429	0.0443	0.0443	0.0443
$n = 1000$										
		d = -0.40			d = -0.30			d = -0.20		
μ	Bias	-	-0.0001	-0.0001	-	0.0001	0.0001	-	-0.0000	0.0000
	Std	-	0.0028	0.0034	-	0.0052	0.0056	-	0.0088	0.0091
d	Bias	-0.0008	-0.0045	-0.0023	-0.0006	-0.0046	-0.0037	-0.0012	-0.0047	-0.0043
	Std	0.0253	0.0260	0.0258	0.0244	0.0248	0.0247	0.0252	0.0255	0.0254
σ	Bias	-0.0016	-0.0021	-0.0019	-0.0006	-0.0011	-0.0010	0.0002	-0.0003	-0.0002
	Std	0.0217	0.0218	0.0217	0.0226	0.0226	0.0226	0.0217	0.0217	0.0217
		d = -0.10			d = 0.00			d = 0.10		
μ	Bias	-	0.0012	0.0013	-	-0.0001	-0.0002	-	-0.0016	-0.0017
	Std	-	0.0165	0.0165	-	0.0315	0.0315	-	0.0597	0.0598
d	Bias	-0.0017	-0.0057	-0.0056	-0.0015	-0.0055	-0.0055	-0.0014	-0.0054	-0.0054
	Std	0.0242	0.0249	0.0249	0.0246	0.0250	0.0250	0.0243	0.0247	0.0247
σ	Bias	-0.0010	-0.0015	-0.0015	-0.0018	-0.0024	-0.0024	-0.0000	-0.0005	-0.0005
	Std	0.0221	0.0221	0.0221	0.0219	0.0218	0.0218	0.0228	0.0228	0.0228
		d = 0.20			d = 0.30			d = 0.40		
μ	Bias	-	-0.0033	-0.0031	-	-0.0061	-0.0077	-	0.0397	0.0366
	Std	-	0.1177	0.1182	-	0.2532	0.2555	-	0.5473	0.5534
d	Bias	-0.0036	-0.0078	-0.0078	-0.0025	-0.0070	-0.0070	-0.0045	-0.0089	-0.0088
	Std	0.0251	0.0259	0.0259	0.0240	0.0248	0.0248	0.0243	0.0255	0.0255
σ	Bias	-0.0013	-0.0018	-0.0018	-0.0019	-0.0023	-0.0023	-0.0012	-0.0016	-0.0015
	Std	0.0228	0.0228	0.0228	0.0224	0.0225	0.0225	0.0221	0.0221	0.0221

TABLE II

BIAS, STD AND RMSE OF ALTERNATIVE MLES FOR FGN: $\mu = 0$ AND $\sigma = 1$. MLE1 IS MLE WITH KNOWN μ; MLE2 IS OUR EXACT MLE; MLE3 IS PMLE.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
					$n = 250$					
		$H = 0.10$			$H = 0.20$			$H = 0.30$		
μ	Bias	-	0.0001	0.0001	-	0.0002	0.0001	-	-0.0000	-0.0000
	Std	-	0.0031	0.0041	-	0.0035	0.0039	-	0.0039	0.0040
H	Bias	0.0006	-0.0039	0.0002	-0.0024	-0.0079	-0.0060	-0.0027	-0.0098	-0.0091
	Std	0.0232	0.0236	0.0230	0.0302	0.0306	0.0302	0.0352	0.0362	0.0360
σ	Bias	0.0065	-0.0154	0.0042	-0.0033	-0.0313	-0.0221	-0.0017	-0.0382	-0.0345
	Std	0.1157	0.1149	0.1153	0.1519	0.1501	0.1498	0.1833	0.1824	0.1820
		$H = 0.40$			$H = 0.50$			$H = 0.60$		
μ	Bias	-	-0.0001	-0.0001	-	-0.0000	-0.0000	-	-0.0003	-0.0003
	Std	-	0.0038	0.0038	-	0.0040	0.0040	-	0.0040	0.0040
H	Bias	-0.0021	-0.0097	-0.0094	-0.0015	-0.0108	-0.0107	-0.0037	-0.0139	-0.0139
	Std	0.0389	0.0399	0.0398	0.0406	0.0425	0.0424	0.0414	0.0430	0.0430
σ	Bias	0.0083	-0.0319	-0.0308	0.0147	-0.0357	-0.0357	0.0080	-0.0497	-0.0499
	Std	0.2155	0.2125	0.2122	0.2348	0.2330	0.2327	0.2497	0.2449	0.2449
		$H = 0.70$			$H = 0.80$			$H = 0.90$		
μ	Bias	-	-0.0000	0.0000	-	-0.0002	-0.0002	-	0.0001	0.0000
	Std	-	0.0040	0.0040	-	0.0041	0.0042	-	0.0038	0.0038
H	Bias	-0.0018	-0.0129	-0.0129	-0.0005	-0.0139	-0.0139	-0.0067	-0.0212	-0.0210
	Std	0.0416	0.0437	0.0438	0.0402	0.0433	0.0433	0.0379	0.0405	0.0405
σ	Bias	0.0284	-0.0398	-0.0397	0.0483	-0.0437	-0.0430	0.0377	-0.0942	-0.0926
	Std	0.2771	0.2706	0.2709	0.3256	0.3173	0.3179	0.4020	0.3565	0.3577
					$n = 1000$					
		$H = 0.10$			$H = 0.20$			$H = 0.30$		
μ	Bias	-	-0.0000	-0.0000	-	-0.0000	-0.0001	-	-0.0001	-0.0001
	Std	-	0.0008	0.0011	-	0.0012	0.0014	-	0.0015	0.0015
H	Bias	0.0000	-0.0011	-0.0001	0.0006	-0.0011	-0.0007	-0.0006	-0.0027	-0.0025
	Std	0.0115	0.0116	0.0116	0.0152	0.0153	0.0153	0.0171	0.0175	0.0174
σ	Bias	0.0013	-0.0043	0.0005	0.0065	-0.0026	-0.0003	0.0012	-0.0099	-0.0089
	Std	0.0575	0.0575	0.0580	0.0787	0.0789	0.0788	0.0887	0.0895	0.0894
		$H = 0.40$			$H = 0.50$			$H = 0.60$		
μ	Bias	-	-0.0001	-0.0001	-	-0.0001	-0.0001	-	-0.0002	-0.0002
	Std	-	0.0018	0.0018	-	0.0020	0.0020	-	0.0023	0.0023
H	Bias	-0.0002	-0.0028	-0.0027	-0.0010	-0.0037	-0.0037	-0.0003	-0.0033	-0.0033
	Std	0.0192	0.0194	0.0194	0.0189	0.0192	0.0192	0.0206	0.0210	0.0210
σ	Bias	0.0022	-0.0116	-0.0113	-0.0017	-0.0169	-0.0169	0.0060	-0.0116	-0.0116
	Std	0.1037	0.1035	0.1034	0.1075	0.1072	0.1072	0.1230	0.1231	0.1231
		$H = 0.70$			$H = 0.80$			$H = 0.90$		
μ	Bias	-	0.0000	-0.0000	-	0.0000	0.0001	-	-0.0001	-0.0001
	Std	-	0.0027	0.0027	-	0.0031	0.0031	-	0.0034	0.0034
H	Bias	-0.0001	-0.0035	-0.0035	0.0007	-0.0030	-0.0030	-0.0018	-0.0065	-0.0064
	Std	0.0203	0.0208	0.0208	0.0203	0.0208	0.0208	0.0211	0.0216	0.0216
σ	Bias	0.0079	-0.0135	-0.0134	0.0185	-0.0089	-0.0086	0.0155	-0.0289	-0.0283
	Std	0.1302	0.1296	0.1296	0.1500	0.1490	0.1490	0.2205	0.2136	0.2138

TABLE III

Bias, Std and RMSE of alternative MLEs for fOU: $\kappa = 10$, $\sigma = 1$ and $n = 250$. MLE1 is MLE with known μ; MLE2 is our exact MLE; MLE3 is PMLE.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
		$H = 0.10$			$H = 0.20$			$H = 0.30$		
μ	Bias	-	-0.0021	-0.0022	-	0.0004	-0.0004	-	0.0014	0.0003
	Std	-	0.1166	0.1005	-	0.1047	0.0978	-	0.0984	0.0956
	RMSE	-	0.1166	0.1005	-	0.1047	0.0978	-	0.0984	0.0956
H	Bias	0.0036	0.0080	0.0081	0.0036	0.0081	0.0084	0.0038	0.0078	0.0082
	Std	0.0297	0.0307	0.0306	0.0402	0.0403	0.0403	0.0462	0.0455	0.0455
	RMSE	0.0300	0.0317	0.0317	0.0404	0.0411	0.0411	0.0463	0.0462	0.0462
κ	Bias	-0.2783	3.3820	3.3933	1.0258	4.3478	4.3914	1.7384	4.7370	4.7953
	Std	8.0909	10.0533	9.9980	7.3880	8.5961	8.5568	7.1869	8.0556	8.0456
	RMSE	8.0957	10.6069	10.5582	7.4588	9.6331	9.6179	7.3941	9.3451	9.3662
σ	Bias	0.0265	0.0521	0.0528	0.0376	0.0648	0.0663	0.0479	0.0721	0.0742
	Std	0.1596	0.1743	0.1743	0.2296	0.2409	0.2409	0.2774	0.2796	0.2794
	RMSE	0.1618	0.1819	0.1821	0.2327	0.2495	0.2499	0.2815	0.2887	0.2891
		$H = 0.40$			$H = 0.50$			$H = 0.60$		
μ	Bias	-	0.0020	0.0006	-	0.0018	0.0003	-	0.0022	0.0005
	Std	-	0.0939	0.0938	-	0.0899	0.0919	-	0.0862	0.0902
	RMSE	-	0.0940	0.0938	-	0.0899	0.0919	-	0.0863	0.0902
H	Bias	0.0041	0.0075	0.0079	0.0046	0.0071	0.0075	0.0048	0.0059	0.0065
	Std	0.0500	0.0488	0.0487	0.0527	0.0509	0.0509	0.0538	0.0516	0.0515
	RMSE	0.0501	0.0493	0.0493	0.0530	0.0514	0.0514	0.0541	0.0519	0.0519
κ	Bias	2.1900	4.8790	4.9283	2.4532	4.8017	4.8287	2.4756	4.4301	4.4500
	Std	7.1367	7.7810	7.7719	7.1795	7.5057	7.5117	7.0800	7.2298	7.2463
	RMSE	7.4651	9.1841	9.2027	7.5870	8.9102	8.9298	7.5003	8.4791	8.5035
σ	Bias	0.0577	0.0776	0.0801	0.0699	0.0831	0.0860	0.0801	0.0834	0.0875
	Std	0.3104	0.3088	0.3086	0.3485	0.3373	0.3378	0.3763	0.3568	0.3581
	RMSE	0.3157	0.3184	0.3188	0.3554	0.3474	0.3486	0.3847	0.3664	0.3686
		$H = 0.70$			$H = 0.80$			$H = 0.90$		
μ	Bias	-	0.0024	0.0006	-	0.0024	0.0005	-	0.0027	0.0004
	Std	-	0.0831	0.0892	-	0.0794	0.0882	-	0.0776	0.0873
	RMSE	-	0.0831	0.0892	-	0.0794	0.0882	-	0.0777	0.0873
H	Bias	0.0040	0.0039	0.0045	0.0038	0.0013	0.0023	0.0033	-0.0016	-0.0009
	Std	0.0531	0.0510	0.0509	0.0516	0.0494	0.0494	0.0415	0.0425	0.0422
	RMSE	0.0532	0.0512	0.0511	0.0518	0.0494	0.0494	0.0416	0.0425	0.0422
κ	Bias	2.0427	3.6373	3.6150	1.2981	2.2510	2.2808	-0.4930	-0.0475	-0.1009
	Std	6.7179	6.8232	6.8247	6.2694	6.1526	6.1850	3.9814	4.3394	4.3388
	RMSE	7.0215	7.7321	7.7230	6.4023	6.5514	6.5921	4.0118	4.3397	4.3399
σ	Bias	0.0862	0.0809	0.0850	0.1143	0.0839	0.0922	0.1704	0.1174	0.1247
	Std	0.3987	0.3789	0.3800	0.4584	0.4182	0.4205	0.5160	0.5082	0.5106
	RMSE	0.4079	0.3874	0.3894	0.4724	0.4266	0.4305	0.5434	0.5216	0.5256

TABLE IV

BIAS, STD AND RMSE OF ALTERNATIVE MLES FOR FOU: $\kappa = 10$, $\sigma = 1$ AND $n = 1000$. MLE1 IS MLE WITH KNOWN μ; MLE2 IS OUR EXACT MLE; MLE3 IS PMLE.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
			$H = 0.10$			$H = 0.20$			$H = 0.30$	
μ	Bias	-	0.0010	0.0008	-	0.0012	0.0011	-	0.0014	0.0012
	Std	-	0.0289	0.0303	-	0.0319	0.0325	-	0.0368	0.0368
	RMSE	-	0.0289	0.0303	-	0.0320	0.0326	-	0.0368	0.0368
H	Bias	0.0010	0.0016	0.0015	0.0008	0.0014	0.0014	0.0009	0.0014	0.0014
	Std	0.0145	0.0145	0.0144	0.0191	0.0190	0.0190	0.0222	0.0220	0.0220
	RMSE	0.0145	0.0145	0.0145	0.0192	0.0190	0.0190	0.0222	0.0220	0.0220
κ	Bias	-0.1356	0.4662	0.4094	0.3459	0.9577	0.9334	0.5840	1.1698	1.1679
	Std	3.7385	3.7798	3.7731	3.3303	3.3923	3.3894	3.3122	3.3685	3.3658
	RMSE	3.7410	3.8085	3.7953	3.3482	3.5249	3.5155	3.3633	3.5658	3.5626
σ	Bias	0.0065	0.0100	0.0094	0.0059	0.0095	0.0094	0.0067	0.0097	0.0099
	Std	0.0747	0.0748	0.0747	0.1021	0.1017	0.1018	0.1221	0.1215	0.1215
	RMSE	0.0750	0.0755	0.0753	0.1023	0.1022	0.1022	0.1223	0.1219	0.1219
			$H = 0.40$			$H = 0.50$			$H = 0.60$	
μ	Bias	-	0.0015	0.0014	-	0.0016	0.0015	-	0.0017	0.0016
	Std	-	0.0420	0.0420	-	0.0475	0.0481	-	0.0534	0.0551
	RMSE	-	0.0420	0.0421	-	0.0475	0.0481	-	0.0534	0.0551
H	Bias	0.0010	0.0014	0.0014	0.0012	0.0013	0.0014	0.0015	0.0012	0.0013
	Std	0.0243	0.0241	0.0241	0.0259	0.0255	0.0255	0.0269	0.0265	0.0265
	RMSE	0.0243	0.0241	0.0241	0.0259	0.0256	0.0256	0.0270	0.0266	0.0265
κ	Bias	0.7287	1.2778	1.2806	0.8155	1.3301	1.3322	0.8647	1.3207	1.3197
	Std	3.3593	3.3948	3.3937	3.4057	3.4304	3.4298	3.4398	3.4390	3.4412
	RMSE	3.4374	3.6273	3.6273	3.5020	3.6793	3.6795	3.5468	3.6839	3.6856
σ	Bias	0.0078	0.0099	0.0101	0.0093	0.0102	0.0103	0.0120	0.0105	0.0107
	Std	0.1388	0.1376	0.1375	0.1539	0.1521	0.1520	0.1700	0.1670	0.1669
	RMSE	0.1390	0.1380	0.1379	0.1542	0.1524	0.1524	0.1704	0.1673	0.1673
			$H = 0.70$			$H = 0.80$			$H = 0.90$	
μ	Bias	-	0.0017	0.0017	-	0.0017	0.0017	-	0.0018	0.0016
	Std	-	0.0597	0.0631	-	0.0663	0.0723	-	0.0728	0.0827
	RMSE	-	0.0597	0.0631	-	0.0663	0.0723	-	0.0728	0.0827
H	Bias	0.0015	0.0009	0.0010	0.0017	0.0004	0.0004	0.0029	0.0004	0.0007
	Std	0.0274	0.0271	0.0271	0.0271	0.0267	0.0267	0.0225	0.0225	0.0225
	RMSE	0.0274	0.0271	0.0271	0.0272	0.0267	0.0267	0.0227	0.0225	0.0225
κ	Bias	0.7907	1.2154	1.2087	0.6244	0.9236	0.9076	0.0312	0.1706	0.1774
	Std	3.4432	3.4458	3.4471	3.2790	3.2548	3.2523	2.0054	1.9812	1.9411
	RMSE	3.5329	3.6538	3.6529	3.3379	3.3833	3.3766	2.0056	1.9885	1.9492
σ	Bias	0.0140	0.0106	0.0109	0.0230	0.0123	0.0125	0.0530	0.0268	0.0296
	Std	0.1878	0.1848	0.1845	0.2200	0.2122	0.2124	0.2549	0.2464	0.2479
	RMSE	0.1883	0.1851	0.1849	0.2212	0.2126	0.2127	0.2603	0.2478	0.2497

5. CONCLUSION

Gaussian processes have gained significant attention due to their broad applicability across various scientific and applied disciplines. To obtain the MLE, two common approaches are typically employed. The first approach maximizes the likelihood assuming μ is known and set to 0, which results in an unrealistic MLE. The second approach uses the sample mean as an estimator for μ, leading to a plug-in MLE. However, both methods fail to address the inefficiency of the estimator for μ, and concerns have been raised about the finite sample performance of the plug-in MLE. Adenstedt (1974) proposed an efficient but infeasible estimator for μ.

In this paper, we introduce a novel exact ML method for all parameters in general Gaussian processes with long-memory, short-memory, or anti-persistence properties. We prove that the exact MLE exhibits consistency and asymptotic normality. We also establish the LAN property of the sequence of statistical experiments for general Gaussian processes in the sense of Le Cam, which directly yields efficiency. Our method offers a comprehensive understanding of MLE for fractional Gaussian models. First, we show that the estimators for all parameters are optimal, effectively complementing the infeasible estimator for μ proposed by Adenstedt (1974). Second, we evaluate and compare the performance of the plug-in MLE, exact MLE and MLE with known μ. The plug-in MLE performs as well as the exact MLE for all parameters except for μ. The discrepancy between plug-in MLE and the MLE with known μ is not due to an inefficient estimator for μ.

The Whittle MLE is asymptotically equivalent to the exact MLE under certain regularity conditions. Although its finite-sample performance is generally inferior to that of the exact MLE, the performance gap narrows as the sample size increases. At the same time, the computational burden of the exact MLE increases substantially due to the need to invert the covariance matrix at each evaluation of the likelihood function—a step that the Whittle method avoids. The Whittle ML method remains an attractive alternative. However, existing theoretical results for the Whittle MLE primarily pertain to ARFIMA models and do not extend to continuous-time models. In future work, we aim to investigate the optimality of the Whittle MLE for general Gaussian processes.

REFERENCES

ABADIR, KARIM AND GABRIEL TALMAIN (2002): "Aggregation, persistence and volatility in a macro model," *The Review of Economic Studies*, 69 (4), 749–779. [2]

ADAMS, ROBERT A AND JOHN JF FOURNIER (2003): *Sobolev Spaces*, vol. 140, Elsevier. [2, 3]

ADENSTEDT, ROLF K (1974): "On large-sample estimation for the mean of a stationary random sequence," *The Annals of Statistics*, 1095–1107. [1, 5, 15, 28, 44, 13, 16]

ANDERSEN, TORBEN G. AND TIM BOLLERSLEV (1997): "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns," *The Journal of Finance*, 52 (3), 975–1005. [2]

ANDERSEN, TORBEN G., TIM BOLLERSLEV, FRANCIS X. DIEBOLD, AND PAUL LABYS (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71 (2), 579–625. [2, 17]

BAHADUR, R RAJ (1964): "On Fisher's bound for asymptotic variances," *The Annals of Mathematical Statistics*, 35 (4), 1545–1552. [11]

BENNEDSEN, MIKKEL, KIM CHRISTENSEN, AND PETER CHRISTENSEN (2024): "Composite likelihood estimation of stationary Gaussian processes with a view toward stochastic volatility," *Working paper*. [3, 11, 17]

BENNEDSEN, MIKKEL, ASGER LUNDE, AND MIKKO S. PAKKANEN (2022): "Decoupling the short-and long-term behavior of stochastic volatility," *J. Financ. Econom.*, 20 (5), 961—-1006. [17]

BLOOMFIELD, PETER (1985): "On series representations for linear predictors," *The Annals of Probability*, 226–233. [16]

BOLKO, ANINE E, KIM CHRISTENSEN, MIKKO S PAKKANEN, AND BEZIRGEN VELIYEV (2023): "A GMM approach to estimate the roughness of stochastic volatility," *Journal of Econometrics*, 235 (2), 745–778. [2]

BROCKWELL, PETER J. AND RICHARD A. DAVIS (1987): *Time Series: Theory and Methods*, Springer, New York. [17]

BROCKWELL, PETER J AND TINA MARQUARDT (2005): "Lévy-driven and fractionally integrated ARMA processes with continuous time parameter," *Statistica Sinica*, 477–494. [5]

BROUSTE, ALEXANDRE AND MASAAKI FUKASAWA (2018): "Local asymptotic normality property for fractional Gaussian noise under high-frequency observations," *Ann. Statist.*, 46 (5), 2045–2061. [5, 13]

CHERIDITO, PATRICK, HIDEYUKI KAWAGUCHI, AND MAKOTO MAEJIMA (2003): "Fractional Ornstein-Uhlenbeck processes," *Electronic Journal of Probability*, 8 (3), 1–14. [2, 20]

CHEUNG, YIN-WONG (1993): "Long Memory in Foreign-Exchange Rates," *Journal of Business & Economic Statistics*, 11 (1), 93–101. [2]

CHEUNG, YIN-WONG AND FRANCIS X DIEBOLD (1994): "On maximum likelihood estimation of the differencing parameter of fractionally-integrated noise with unknown mean," *Journal of Econometrics*, 62 (2), 301–316. [3, 4, 6, 21]

CHEVILLON, GUILLAUME, ALAIN HECQ, AND SÉBASTIEN LAURENT (2018): "Generating univariate fractional integration within a large VAR (1)," *Journal of Econometrics*, 204 (1), 54–65. [2]

Choi, Sungsub, W Jackson Hall, and Anton Schick (1996): "Asymptotically uniformly most powerful tests in parametric and semiparametric models," *The Annals of Statistics*, 24 (2), 841–861. [14]

Chong, Carsten H. and Fabian Mies (2025): "Likelihood Asymptotics of Stationary Gaussian Arrays," *arXiv preprint, arXiv:2502.09229*. [5, 13, 36, 4, 6, 15]

Chong, Carsten H and Viktor Todorov (2025): "A nonparametric test for rough volatility," *Journal of the American Statistical Association*, 1–23. [2]

Cohen, Serge, Fabrice Gamboa, Céline Lacaux, and Jean-Michel Loubes (2013): "LAN property for some fractional type Brownian motion," *ALEA: Latin American Journal Probability Mathematical Statistics*, 10 (1), 91–106. [4, 9, 13, 15, 16]

Corsi, Fulvio (2009): "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 7 (2), 174–196. [3]

Dahlhaus, Rainer (1989): "Efficient parameter estimation for self-similar processes," *The Annals of Statistics*, 17 (4), 1749–1766. [4, 9, 10, 11, 14, 15]

——— (2006): "Correction: Efficient parameter estimation for self-similar processes," *The Annals of Statistics*, 34 (2), 1045–1047. [4, 9, 10, 14]

Diebold, Francis X., Steven Husted, and Mark Rush (1991): "Real Exchange Rates under the Gold Standard," *Journal of Political Economy*, 99 (6), 1252–1271. [2]

Diebold, Francis X and Atsushi Inoue (2001): "Long memory and regime switching," *Journal of Econometrics*, 105 (1), 131–159. [2]

Diebold, Francis X. and Glenn D. Rudebusch (1989): "Long memory and persistence in aggregate output," *Journal of Monetary Economics*, 24 (2), 189–209. [2]

——— (1991): "Is Consumption Too Smooth? Long Memory and the Deaton Paradox," *The Review of Economics and Statistics*, 73 (1), 1–9. [2]

Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle (1993): "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, 1 (1), 83–106. [2]

Eguchi, Shoichi and Hiroki Masuda (2018): "Schwarz type model comparison for LAQ models," *Bernoulli*, 24 (3), 2278–2327. [14]

El Euch, Omar, Masaaki Fukasawa, and Mathieu Rosenbaum (2018): "The microstructual foundations of leverage effect and rough volatility," *Finance and Stochastics*, 22 (2), 241–280. [2]

Fox, Robert and Murad S. Taqqu (1986): "Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series," *The Annals of Statistics*, 14 (2), 517–532. [3, 9, 10]

Fukasawa, Masaaki and Tetsuya Takabatake (2019): "Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations," *Bernoulli*, 25 (3), 1870–1900. [2, 5, 9, 13, 39]

Fukasawa, Masaaki, Tetsuya Takabatake, and Rebecca Westphal (2022): "Consistent estimation for fractional stochastic volatility model under high-frequency asymptotics," *Math. Finance*, 32 (4), 1086–1132. [2]

Garnier, Josselin and Knut Sølna (2018): "Option pricing under fast-varying and rough stochastic volatility," *Annals of Finance*, 14 (4), 489–516. [20]

Gatheral, Jim, Thibault Jaisson, and Mathieu Rosenbaum (2018): "Volatility is rough," *Quantitative Finance*, 18 (6), 1–17. [2, 17]

Geweke, John and Susan Porter-Hudak (1983): "The estimation and application of long memory time series models," *Journal of Time Series Analysis*, 4 (4), 221–238. [3]

Granger, Clive WJ (1980): "Long memory relationships and the aggregation of dynamic models," *Journal of Econometrics*, 14 (2), 227–238. [2, 16]

Hájek, Jaroslav (1972): "Local asymptotic minimax and admissibility in estimation," in *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability*, University of California Press, vol. 1, 175–194. [12, 13]

Harville, David A (1998): "Matrix algebra from a statistician's perspective," . [40]

Höpfner, Reinhard (2014): *Asymptotic statistics: With a view to stochastic processes*, Walter de Gruyter. [13]

Hosking, J. R. M. (1981): "Fractional differencing," *Biometrika*, 68 (1), 165–176. [2, 16]

Hualde, Javier and M. Ø. Nielsen (2020): "Truncated sum of squares estimation of fractional time series models with deterministic trends," *Econometric Theory*, 36 (4), 751–772. [3, 10]

Hualde, Javier and Peter M. Robinson (2011): "Gaussian pseudo-maximum likelihood estimation of fractional time series models," *The Annals of Statistics*, 39 (6), 3152–3181. [3, 10]

Hult, Henrik (2003): "Topics on fractional Brownian motion and regular variation for stochastic processes," Ph.D. thesis, Matematik. [20]

Ibragimov, Il'dar A and Rafail Z Has'minskiĭ (1981): *Statistical Estimation: Asymptotic Theory*, vol. 16, Springer Science & Business Media. [13]

Jusselin, Paul and Mathieu Rosenbaum (2020): "No-arbitrage implies power-law market impact and rough volatility," *Mathematical Finance*, 30 (4), 1309–1336. [2]

Kawai, Reiichiro (2013): "Fisher information for fractional Brownian motion under high-frequency discrete sampling," *Communications in Statistics-Theory and Methods*, 42 (9), 1628–1636. [15]

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

Le Cam, Lucien (1953): "On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates," *University of California Publications in Statistics*, 1 (11), 277–330. [11, 12]

——— (1960): "Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses," *University of California Publications in Statistics*, 3, 37–98. [12]

——— (1972): "Limits of experiments," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press Berkeley-Los Angeles, vol. 1, 245–261. [12, 13]

Lieberman, Offer and Peter C. B. Phillips (2004): "Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra," *Journal of Time Series Analysis*, 25 (5), 733–753. [4]

Lieberman, Offer, Roy Rosemarin, and Judith Rousseau (2009): "Asymptotic theory for maximum likelihood estimation in stationary fractional Gaussian processes, under short-, long- and intermediate memory." *CiteSeerX*. [4]

——— (2012): "Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes," *Econometric Theory*, 28 (2), 457–470, (Its full version is available at http://www.ceremade.dauphine.fr/~rousseau/LRRlongmemo.html.). [4, 9, 10, 11, 14, 15, 35]

Liu, Guangying and BingYi Jing (2018): "On Estimation of Hurst Parameter Under Noisy Observations," *Journal of Business & Economic Statistics*, 36 (3), 483–492. [2]

Lo, Andrew W. (1991): "Long-Term Memory in Stock Market Prices," *Econometrica*, 59 (5), 1279–1313. [2]

Mandelbrot, Benoit (1965): "Self-similar error clusters in communication systems and the concept of conditional stationarity," *IEEE Transactions on Communication Technology*, 13 (1), 71–90. [2]

Mandelbrot, Benoit B and John W Van Ness (1968): "Fractional Brownian motions, fractional noises and applications," *SIAM Review*, 10 (4), 422–437. [2, 17]

Matérn, Bertil (1986): *Spatial Variation*, vol. 36 of *Lecture Notes in Statistics*, Berlin: Springer-Verlag, 2nd ed. [5]

Nielsen, Morten Ørregaard (2015): "Asymptotics for the Conditional-Sum-of-Squares Estimator in Multivariate Fractional Time-Series Models," *Journal of Time Series Analysis*, 36 (2), 154–188. [3, 10]

Pfanzagl, Johann (1970): "On the asymptotic efficiency of median unbiased estimates," *The Annals of Mathematical Statistics*, 1500–1509. [11]

Phillips, Peter C. B. and Katsumi Shimotsu (2004): "Local Whittle estimation in nonstationary and unit root cases," *The Annals of Statistics*, 32 (2), 656 – 692. [3]

Potter, Kenneth W (1976): "Evidence for nonstationarity as a physical explanation of the Hurst phenomenon," *Water Resources Research*, 12 (5), 1047–1052. [2]

Priestley, Maurice Bertram (1988): *The spectral analysis of time series*, Oxford University Press. [20]

Rao, Suhasini Subba and Junho Yang (2021): "Reconciling the Gaussian and Whittle likelihood with an application to estimation in the frequency domain," *The Annals of Statistics*, 49 (5), 2774–2802. [3]

Robinson, Peter M (1995): "Gaussian semiparametric estimation of long range dependence," *The Annals of Statistics*, 1630–1661. [3, 35]

Samarov, Alexander and Murad S Taqqu (1988): "On the efficiency of the sample mean in long-memory noise," *Journal of Time Series Analysis*, 9 (2), 191–200. [15]

Schennach, Susanne M (2018): "Long memory via networking," *Econometrica*, 86 (6), 2221–2248. [2]

Shi, Shuping and Jun Yu (2023): "Volatility Puzzle: Long Memory or Anti-persistency," *Management Science*, 69 (7), 3861–3883. [4]

Shi, Shuping, Jun Yu, and Chen Zhang (2024a): "Fractional Gaussian noise: Spectral density and estimation methods," *Journal of Time Series Analysis, forthcoming*. [2, 18]

——— (2024b): "On the spectral density of fractional Ornstein-Uhlenbeck processes," *Journal of Econometrics*, 245 (1), 105872. [2, 11, 20, 17]

Shimotsu, Katsumi (2010): "Exact local Whittle estimation of fractional integration with unknown mean and time trend," *Econometric Theory*, 501–540. [3]

Sinai, Yu G (1976): "Self-similar probability distributions," *Theory of Probability & Its Applications*, 21 (1), 64–80. [18]

Sowell, Fallaw (1992): "Maximum likelihood estimation of stationary univariate fractionally integrated time series models," *Journal of Econometrics*, 53 (1-3), 165–188. [3, 10]

Szymanski, Grégoire (2024): "Optimal estimation of the rough Hurst parameter in additive noise," *Stochastic Processes and their Applications*, 170, 104302. [5, 13]

Szymanski, Grégoire and Tetsuya Takabatake (2023): "Asymptotically Efficient Estimation for Fractional Brownian Motion with Additive Noise," *arXiv preprint, arXiv:2311.18669*. [5, 13]

Takabatake, Tetsuya (2024): "Corrigendum: Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra," *Journal of Time Series Analysis*, 45 (1), 158–160. [4]

Tsai, Henghsiu (2009): "On continuous-time autoregressive fractionally integrated moving average processes," *Bernoulli*, 15 (1), 178–194. [5]

Tsai, Henghsiu and K. S. Chan (2005): "Maximum Likelihood Estimation of Linear Continuous Time Long Memory Processes with Discrete Time Data," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 67 (5), 703–716. [4, 5]

Velasco, Carlos and Peter M Robinson (2000): "Whittle pseudo-maximum likelihood estimation for nonstationary time series," *Journal of the American Statistical Association*, 95 (452), 1229–1243. [35]

Wald, Abraham (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large," *Transactions of the American Mathematical Society*, 54 (3), 426–482. [12]

Wang, Xiaohu, Weilin Xiao, and Jun Yu (2023): "Modeling and forecasting realized volatility with the fractional Ornstein–Uhlenbeck process," *Journal of Econometrics*, 232 (2), 389–415. [2, 3, 5, 17, 18]

Wang, Xiaohu, Weilin Xiao, Jun Yu, and Chen Zhang (2024): "Maximum Likelihood Estimation of Fractional Ornstein-Uhlenbeck Process with Discretely Sampled Data," *Working Paper, University of Macau*. [2, 4, 11, 21, 17, 18]

Whittle, Peter (1951): *Hypothesis testing in time series analysis*, vol. 4, Almqvist & Wiksells boktr. [3]

Yajima, Yoshihiro (1985): "On estimation of long-memory time series models," *Australian Journal of Statistics*, 27 (3), 303–320. [3]

APPENDIX A: Appendix: Proof of Theorems

A.1. *Proof of Consistency in Theorem 1*

First, we introduce and summarize notations used in the proof. For a $\mathbb{R}$-valued function f on some set A, we write $f_{\pm}(a) := \max\{\pm f(a), 0\}$ for $a \in A$. Then we have $f = f_+ - f_-$. Let $\mathbf{Z}_n := \mathbf{X}_n - \mu_0 \mathbf{1}_n$. Recall that $\bar{\ell}_n(\xi) = \ell_n(\xi, \sigma_n(\xi), \mu_n(\xi))$ that can be written as

$$\bar{\ell}_n(\xi) = -\frac{n}{2}(1+\log(2\pi)) - \frac{n}{2}\log\bar{\sigma}_n^2(\xi) - \frac{1}{2}\log\det\Big[\Sigma_n(s_\xi^X)\Big]. \tag{23}$$

In addition, we introduce

$$\bar{\sigma}_n^2(\xi) := \sigma_n^2(\xi,\mu_0) = \frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top \Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n),\ \sigma^2(\xi) := \frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\frac{s_{\xi_0}^X(\omega)}{s_\xi^X(\omega)}\,\mathrm{d}\omega. \tag{24}$$

Write $\chi(\xi_1,\xi_2) := (\alpha_X(\xi_1) - \alpha_X(\xi_2))_+/(1-\alpha_X(\xi_2)_+)$ and $\chi_0(\xi_1) := \chi(\xi_0,\xi_1)$ for $\xi_1,\xi_2 \in \Theta_\xi$. Then we introduce a restricted parameter space of Θ_ξ by

$$\Theta_\xi(\iota) := \{\xi \in \Theta_\xi : \chi_0(\xi) \le 1-\iota\},\ \ \iota \in (0,1).$$

Set $L_n(\xi) := -\frac{2}{n}\bar{\ell}_n(\xi)$ and

$$L(\xi) := (1+\log(2\pi)) + \log\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{s_{\xi_0}^X(\omega)}{s_\xi^X(\omega)}\,\mathrm{d}\omega\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log s_\xi^X(\omega)\,\mathrm{d}\omega,$$

which is actually a limit function of $L_n(\xi)$ on $\Theta_\xi(\iota)$; see (27) for details. Note that $L(\xi)$ is finite in $\Theta_\xi(\iota)$ for any $\iota \in (0,1)$ and we can write

$$L(\xi) - L(\xi_0) = \log\left(\int_{-\pi}^{\pi}\frac{s_{\xi_0}^X(\omega)}{s_\xi^X(\omega)}\frac{\mathrm{d}\omega}{2\pi}\right) - \int_{-\pi}^{\pi}\log\frac{s_{\xi_0}^X(\omega)}{s_\xi^X(\omega)}\frac{\mathrm{d}\omega}{2\pi}, \tag{25}$$

so that Jensen's inequality of the strictly concave function and the identification condition on $\{s_\theta^X\}_{\theta\in\Theta}$ in Assumption 1 give the following.

$$\inf_{\xi\in\Theta_\xi(\iota),\|\xi-\xi_0\|_{\mathbb{R}^{p-1}}\ge\varepsilon} L(\xi) > L(\xi_0),\ \ \forall\varepsilon>0,\ \forall\iota\in(0,1). \tag{26}$$

To prove the consistency of $\{\widehat{\xi}_n\}_{n\in\mathbb{N}}$, we first prove the uniform convergence,

$$\sup_{\xi\in\Theta_\xi(\iota)}|L_n(\xi) - L(\xi)| = o_{\mathbb{P}_{\vartheta_0}^n}(1)\ \ \text{as } n\to\infty. \tag{27}$$

Note that we can write

$$L_n(\xi) - L(\xi) = \Big(\log\bar{\sigma}_n^2(\xi) - \log\sigma^2(\xi)\Big) + \left(\frac{1}{n}\log\det\Big[\Sigma_n(s_\xi^X)\Big] - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log s_\xi^X(\omega)\,\mathrm{d}\omega\right). \tag{28}$$

By the uniform convergence version of Szegö's theorem (Lieberman et al., 2012):

$$\sup_{\xi\in\Theta_\xi}\left|\frac{1}{n}\log\det\left[\Sigma_n(s_\xi^X)\right]-\frac{1}{2\pi}\int_{-\pi}^{\pi}\log s_\xi^X(\omega)\,\mathrm{d}\omega\right|=o(1)\ \text{ as } n\to\infty. \tag{29}$$

Moreover, we can also show the uniform convergence,

$$\sup_{\xi\in\Theta_\xi(\iota)}\left|\bar{\sigma}_n^2(\xi)-\sigma^2(\xi)\right|=o_{\mathbb{P}^n_{\vartheta_0}}(1)\ \text{ as } n\to\infty, \tag{30}$$

whose proof is left to Section B.3 in Online Appendix. Then we also obtain

$$\sup_{\xi\in\Theta_\xi(\iota)}\left|\log\bar{\sigma}_n^2(\xi)-\log\sigma^2(\xi)\right|=o_{\mathbb{P}^n_{\vartheta_0}}(1)\ \text{ as } n\to\infty, \tag{31}$$

which can be proved using (30) immediately. However, we also give a detailed proof in Section B.4 in Online Appendix for completeness. Then we conclude (27) using (25), (28), (29) and (31).

Now we give a proof of consistency of $\{\widehat{\xi}_n\}_{n\in\mathbb{N}}$ using (26) and (27). For each $\iota\in(0,1)$, we define

$$\widehat{\xi}_n(\iota)=\underset{\xi\in\Theta_\xi(\iota)}{\arg\max}\,\bar{\ell}_n(\xi)=\underset{\xi\in\Theta_\xi(\iota)}{\arg\min}\,L_n(\xi).$$

Similar to Robinson (1995), Velasco and Robinson (2000) and Lieberman et al. (2012), we divide the proof of consistency into the following two steps.

Step 1: We prove that for each $\iota\in(0,1)$, $\widehat{\xi}_n(\iota)$ is a consistent estimator of ξ, *i.e.* $\widehat{\xi}_n(\iota)\to\xi_0$ in $\mathbb{P}^n_{\vartheta_0}$-probability.

Proof of Step 1: The conclusion follows immediately from (26), (27) and the definition of $\widehat{\xi}_n(\iota)$. *Q.E.D.*

Step 2: We prove that there exists $\iota\in(0,1)$ such that $\widehat{\xi}_n-\widehat{\xi}_n(\iota)\to 0$ in $\mathbb{P}^n_{\vartheta_0}$-probability as $n\to\infty$.

Proof of Step 2: If $\Theta_\xi=\Theta_\xi(\iota)$ holds for some $\iota\in(0,1)$, the equality $\widehat{\xi}_n=\widehat{\xi}_n(\iota)$ holds so that we immediately conclude the assertion of Step 2. In the rest of the proof, we assume that $\Theta_\xi\setminus\Theta_\xi(\iota)$ is a nonempty set for any $\iota\in(0,1)$. Then, for any

$\iota \in (0,1)$ and $\epsilon, \epsilon_1 > 0$, we can show

$$\mathbb{P}^n_{\vartheta_0}\Big[\|\widehat{\xi}_n - \widehat{\xi}_n(\iota)\|_{\mathbb{R}^2} > \epsilon\Big] \le \mathbb{P}^n_{\vartheta_0}\Big[\inf_{\xi\in\Theta_\xi\setminus\Theta_\xi(\iota)} L_n(\xi) \le \inf_{\xi\in\Theta_\xi(\iota)} L_n(\xi)\Big]$$
$$\le \mathbb{P}^n_{\vartheta_0}\Big[\inf_{\xi\in\Theta_\xi\setminus\Theta_\xi(\iota)} L_n(\xi) \le L(\xi_0)+\epsilon_1\Big] + \mathbb{P}^n_{\vartheta_0}\Big[|L_n(\widehat{\xi}_n(\iota)) - L(\xi_0)| \ge \epsilon_1\Big]. \tag{32}$$

Note that for any $\epsilon_1, \epsilon_2 > 0$, the second term of (32) is dominated by

$$\mathbb{P}^n_{\vartheta_0}\Big[|L_n(\widehat{\xi}_n(\iota)) - L(\xi_0)| \ge \epsilon_1\Big] \le \mathbb{P}^n_{\vartheta_0}\Big[|\widehat{\xi}_n(\iota) - \xi_0| \ge \epsilon_2\Big] + \mathbb{P}^n_{\vartheta_0}\Big[\sup_{\xi\in\Theta_\xi(\iota)} |L_n(\xi) - L(\xi)| \ge \epsilon_1\Big]$$
$$+ \mathbb{P}^n_{\vartheta_0}\Big[|\widehat{\xi}_n(\iota) - \xi_0| < \epsilon_2, |L(\widehat{\xi}_n(\iota)) - L(\xi_0)| \ge \epsilon_1\Big]. \tag{33}$$

Then the continuity of $L(\xi)$ on $\Theta_\xi(\iota)$ shows that for any $\epsilon_1 > 0$, there exists $\epsilon_2 > 0$ such that the third term of (33) is equal to zero. Moreover, we can also note that the first and second terms of (33) are negligible as $n \to \infty$ using the result in Step 1 and the uniform convergence (27), respectively.

Finally, we evaluate the first term of (32). Recall that $\chi_0(\xi) = (\alpha_X(\xi_0) - \alpha_X(\xi))_+/(1 - \alpha_X(\xi)_+)$ for $\xi \in \Theta_\xi$. Take sufficiently small $\iota \in (0,1)$ satisfying $1 - \iota > \alpha_X(\xi_0)_+$. Notice that $g(x) := (a-x)_+/(1-x_+)$ for $a := \alpha_X(\xi_0) \in (-\infty, 1)$ is a continuous and monotonically decreasing function on $(-\infty, a]$ and $g(0) = \alpha_X(\xi_0)_+ < 1-\iota$. So we get a) $\alpha_X(\xi) < 0$ for any $\xi \in \Theta_\xi \setminus \Theta_\xi(\iota)$, b) $\alpha_X(\xi_1) \ge \alpha_X(\xi_2)$ for any $\xi_1 \in \Theta_\xi(\iota)$ and $\xi_2 \in \Theta_\xi \setminus \Theta_\xi(\iota)$, and c) there exists some point $\xi_1(\iota) \in \Theta_\xi(\iota)$ such that $\alpha(\xi_1(\iota)) = -1 + \alpha_X(\xi_0) - \iota$ holds since $\Theta_\xi \setminus \Theta_\xi(\iota)$ is not empty. Since $\alpha_X(\xi_1) \ge \alpha_X(\xi_2)$ for any $\xi_1 \in \Theta_\xi(\iota)$ and $\xi_2 \in \Theta_\xi \setminus \Theta_\xi(\iota)$, using Lemma 3 of Chong and Mies (2025), we can show that there exists a constant $C_1 > 0$ such that for any $\xi_1 \in \Theta_\xi(\iota)$ and $\xi_2 \in \Theta_\xi \setminus \Theta_\xi(\iota)$,

$$\frac{\bar{\sigma}^2_n(\xi_1)}{\bar{\sigma}^2_n(\xi_2)} \le \sup_{\mathbf{x}\in\mathbb{R}^n} \frac{\|\Sigma_n(s^X_{\xi_1})^{-\frac{1}{2}}\mathbf{x}\|_{\mathbb{R}^n}}{\|\Sigma_n(s^X_{\xi_2})^{-\frac{1}{2}}\mathbf{x}\|_{\mathbb{R}^n}} = \|\Sigma_n(s^X_{\xi_1})^{-\frac{1}{2}}\Sigma_n(s^X_{\xi_2})^{\frac{1}{2}}\|_{\mathrm{op}} \le \frac{1}{C_1}. \tag{34}$$

Set $\xi_1(\iota) := -1 + \alpha_X(\xi_0) - \iota$ and

$$r_{n,1}(\xi) := \log\bar{\sigma}^2_n(\xi) - \log\sigma^2(\xi), \;\; r_{n,2}(\xi) := \frac{1}{n}\log\det\Big[\Sigma_n(s^X_\xi)\Big] - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log s^X_\xi(\omega)\,\mathrm{d}\omega.$$

Since $\alpha_X(\xi_1(\iota)) \ge \alpha_X(\xi_2)$ for any $\xi_2 \in \Theta_\xi \setminus \Theta_\xi(\iota)$, the inequality (34) yields

$$L_n(\xi_2) \ge (1+\log(2\pi)) + \log C_1 + \log\bar{\sigma}^2_n(\xi_1(\iota)) + \frac{1}{n}\log\det\Big[\Sigma_n(s^X_{\xi_2})\Big]$$

$$\geq (1+\log(2\pi)) + \log C_1 + \log \sigma^2(\xi_1(\iota)) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log s_{\xi_2}^X(\omega)\,\mathrm{d}\omega + r_{n,1}(\xi_1(\iota)) + r_{n,2}(\xi_2),$$

and

$$\log \sigma^2(\xi_1(\iota)) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log s_{\xi_2}^X(\omega)\,\mathrm{d}\omega \geq \log\left(\frac{c_+}{2\pi c_-}\int_{-\pi}^{\pi} |\omega|^{-1+\iota}\,\mathrm{d}\omega\right) + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left(c_-|\omega|^{-\alpha_X(\xi_2)}\right)\mathrm{d}\omega$$

$$\geq \log\left(\frac{c_+}{\pi c_-}\right) + \log\left(\frac{\pi^\iota}{\iota}\right) + \log c_- - \alpha_X(\xi_1(\iota))(\log\pi - 1).$$

Therefore, on the set $A_1(\delta_1) \cap A_2(\delta_2)$ with $A_1(\delta) := \{\sup_{\xi\in\Theta_\xi(\iota)} |r_{n,1}(\xi)| < \delta\}$ and $A_2(\delta) := \{\sup_{\xi\in\Theta_\xi} |r_{n,2}(\xi)| < \delta\}$, we obtain $\inf_{\xi\in\Theta_\xi\setminus\Theta_\xi(\iota)} L_n(\xi) \geq L(\iota,\delta_1,\delta_2)$, where

$$L(\iota,\delta_1,\delta_2) := (1+\log(2\pi)) + \log C_1 + \log\left(\frac{c_+}{\pi c_-}\right) + \log\left(\frac{\pi^\iota}{\iota}\right) + \log c_- - \alpha_X(\xi_1(\iota))(\log\pi - 1) - (\delta_1+\delta_2).$$

Since $L(\iota,\delta_1,\delta_2)$ diverges to infinity as $\iota \to 0$, for any $\epsilon_1,\delta_1,\delta_2 > 0$, there exists $\iota \equiv \iota(\epsilon_1,\delta_1,\delta_2) \in (0,1)$ such that $L(\iota,\delta_1,\delta_2) > L(\xi_0) + \epsilon_1$. Then, as $n \to \infty$, we obtain

$$\mathbb{P}^n_{\vartheta_0}\Big[\inf_{\xi\in\Theta_\xi\setminus\Theta_\xi(\iota)} L_n(\xi) \leq L(\xi_0) + \epsilon_1\Big]$$

$$\leq \sum_{j=1}^{2} \mathbb{P}^n_{\vartheta_0}\Big[A_j(\delta_j)^c\Big] + \mathbb{P}^n_{\vartheta_0}\Big[A_1(\delta_1) \cap A_2(\delta_2) \cap \Big\{L(\iota,\delta_1,\delta_2) \leq L(\xi_0) + \epsilon_1\Big\}\Big] \to 0$$

using (29) and (31). This completes the proof of Step 2 and consistency. *Q.E.D.*

A.2. *Proof of Asymptotic Normality in Theorem 1*

Before proving the asymptotic normality of the sequence of the exact MLEs, we summarize notations used in the proof and prepare several limit theorems repeatedly used in the proof. Recall that

$$\bar{\ell}_n(\xi) = -\frac{n}{2}(1+\log(2\pi)) - \frac{n}{2}\log\bar{\sigma}_n^2(\xi) - \frac{1}{2}\log\det\Big[\Sigma_n(s_\xi^X)\Big],$$

Then we can show

$$-\frac{2}{n}\partial_i\bar{\ell}_n(\xi) = \frac{1}{\bar{\sigma}_n^2(\xi)}\partial_i\bar{\sigma}_n^2(\xi) + \frac{1}{n}\mathrm{Tr}\Big[\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Big], \tag{35}$$

$$-\frac{2}{n}\partial_{i,j}^2\bar{\ell}_n(\xi) = -\left(-\frac{\partial_i\bar{\sigma}_n^2(\xi)}{\bar{\sigma}_n^2(\xi)}\right)\left(-\frac{\partial_j\bar{\sigma}_n^2(\xi)}{\bar{\sigma}_n^2(\xi)}\right) + \frac{1}{\bar{\sigma}_n^2(\xi)}\partial_{i,j}^2\bar{\sigma}_n^2(\xi)$$
$$+\frac{1}{n}\mathrm{Tr}\Big[\Sigma_n(\partial_{i,j}^2 s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Big] - \frac{1}{n}\mathrm{Tr}\Big[\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Big]. \tag{36}$$

Here notice that $-\frac{2}{n}\partial_i\bar{\ell}_n(\xi)$ in the expression (35) can be decomposed by

$$-\frac{2}{n}\partial_i\bar{\ell}_n(\xi)=\left(\frac{1}{\bar{\sigma}_n^2(\xi)}-\frac{1}{\sigma_0^2}\right)\partial_i\bar{\sigma}_n^2(\xi)+\frac{1}{\sigma_0^2}\partial_i\bar{\sigma}_n^2(\xi)+\frac{1}{n}\mathrm{Tr}\left[\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\right]$$

so that we obtain the expression

$$\begin{aligned}n^{-\frac{1}{2}}\partial_\xi\bar{\ell}_n(\xi_0)&=-\frac{\sqrt{n}}{2\sigma_0^2}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2)\left(-\frac{\partial_\xi\bar{\sigma}_n^2(\xi_0)}{\bar{\sigma}_n^2(\xi_0)}\right)-\frac{\sqrt{n}}{2\sigma_0^2}\left(\partial_\xi\bar{\sigma}_n^2(\xi_0)-\mathbb{E}_{\vartheta_0}^n\left[\partial_\xi\tilde{\sigma}_n^2(\xi_0)\right]\right)\\&=-\frac{\sqrt{n}}{2\sigma_0^2}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2)a_{p-1}(\xi_0)-\frac{\sqrt{n}}{2\sigma_0^2}\left(\partial_\xi\bar{\sigma}_n^2(\xi_0)-\mathbb{E}_{\vartheta_0}^n\left[\partial_\xi\tilde{\sigma}_n^2(\xi_0)\right]\right)+o_{\mathbb{P}_{\vartheta_0}^n}(1),\end{aligned}\tag{37}$$

where we used (30) and Lemma 5 in the last equality. Moreover, combining (37) with Lemma 6, we can show that

$$n^{-\frac{1}{2}}\partial_\xi\bar{\ell}_n(\xi_0)\to\mathcal{N}(0,V_{p-1}(\xi_0))\tag{38}$$

in law under the distribution $\mathbb{P}_{\vartheta_0}^n$ as $n\to\infty$, where

$$\begin{aligned}V_{p-1}(\xi_0)&:=\lim_{n\to\infty}\mathrm{Var}_{\vartheta_0}^n\left[-\frac{\sqrt{n}}{2\sigma_0^2}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2)a_{p-1}(\xi_0)-\frac{\sqrt{n}}{2\sigma_0^2}\left(\partial_\xi\bar{\sigma}_n^2(\xi_0)-\mathbb{E}_{\vartheta_0}^n\left[\partial_\xi\tilde{\sigma}_n^2(\xi_0)\right]\right)\right]\\&=\frac{1}{2}a_{p-1}(\xi_0)a_{p-1}(\xi_0)^\top+\mathcal{F}_{p-1}(\xi_0)+2a_{p-1}(\xi_0)\left(-\frac{1}{2}a_{p-1}(\xi_0)\right)^\top=\mathcal{G}_{p-1}(\xi_0).\end{aligned}$$

Moreover, using consistency of $\{\widehat{\xi}_n\}_{n\in\mathbb{N}}$, we can show that $n^{-\frac{1}{2}}\partial_\xi\bar{\ell}_n(\widehat{\xi}_n)=o_{\mathbb{P}_{\vartheta_0}^n}(1)$ as $n\to\infty$ so that, using the Taylor theorem and (38), we obtain, as $n\to\infty$,

$$\int_0^1\mathcal{G}_{p-1,n}(\xi_0+u(\widehat{\xi}_n-\xi_0))\,\mathrm{d}u\,\sqrt{n}(\widehat{\xi}_n-\xi_0)=\frac{1}{\sqrt{n}}\partial_\xi\bar{\ell}_n(\xi_0)-\frac{1}{\sqrt{n}}\partial_\xi\bar{\ell}_n(\widehat{\xi}_n)\to\mathcal{N}(0,V_{p-1}(\xi_0))\tag{39}$$

in law under the distribution $\mathbb{P}_{\vartheta_0}^n$, where $\mathcal{G}_{p-1,n}(\xi):=-n^{-1}\partial_\xi^2\bar{\ell}_n(\xi)$. Then, combining the uniform convergence in (54) with the continuity of the function $\xi\mapsto\mathcal{G}^{i,j}(\xi,\theta_0)$ at $\xi=\xi_0$ and using (39) and Slutsky's lemma, we obtain the stochastic expansion

$$\mathcal{G}_{p-1}(\xi_0)\,\sqrt{n}(\widehat{\xi}_n-\xi_0)=\frac{1}{\sqrt{n}}\partial_\xi\bar{\ell}_n(\xi_0)+o_{\mathbb{P}_{\vartheta_0}^n}(1)\ \text{ as } n\to\infty.\tag{40}$$

Moreover, using Taylor's theorem and Lemma 5, we can also show that

$$\sqrt{n}(\bar{\sigma}_n(\widehat{\xi}_n)-\sigma_0)=\frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\widehat{\xi}_n)-\sigma_0^2)+o_{\mathbb{P}_{\vartheta_0}^n}(1)\ \text{ as } n\to\infty\tag{41}$$

and

$$
\begin{aligned}
\frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\widehat{\xi}_n)-\sigma_0^2) =& \frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\widehat{\xi}_n)-\bar{\sigma}_n^2(\xi_0)) + \frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2) \\
=& \frac{\sigma_0^2}{2}\left\langle \sqrt{n}(\widehat{\xi}_n-\xi_0), \frac{1}{\sigma_0^3}\int_0^1 \partial_\xi \bar{\sigma}_n^2(\xi_0+v(\widehat{\xi}_n-\xi_0))\,\mathrm{d}v \right\rangle_{\mathbb{R}^{p-1}} + \frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2) + o_{\mathbb{P}^n_{\vartheta_0}}(1) \\
=& \frac{\sigma_0^2}{2}\left\langle \sqrt{n}(\widehat{\xi}_n-\xi_0), -a_{p-1}(\theta_0) \right\rangle_{\mathbb{R}^{p-1}} + \frac{\sqrt{n}}{2\sigma_0}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2) + o_{\mathbb{P}^n_{\vartheta_0}}(1) \qquad (42)
\end{aligned}
$$

as $n \to \infty$. Combining (37) with (40), as $n \to \infty$, we get

$$
\begin{aligned}
\sqrt{n}(\widehat{\xi}_n-\xi_0) =& -\frac{\sigma_0^2}{2}\mathcal{G}_{p-1}(\xi_0)^{-1}a_{p-1}(\theta_0)\frac{\sqrt{n}}{\sigma_0^3}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2) \\
& - \mathcal{G}_{p-1}(\xi_0)^{-1}\frac{\sqrt{n}}{2\sigma_0^2}\left(\partial_\xi \bar{\sigma}_n^2(\xi_0) - \mathbb{E}^n_{\vartheta_0}\left[\partial_\xi \tilde{\sigma}_n^2(\xi_0)\right]\right) + o_{\mathbb{P}^n_{\vartheta_0}}(1). \qquad (43)
\end{aligned}
$$

Using (41) and (42), as $n \to \infty$, we obtain

$$
\begin{aligned}
\sqrt{n}(\bar{\sigma}_n(\widehat{\xi}_n)-\sigma_0) =& \left\{\frac{\sigma_0^4}{4}a_{p-1}(\theta_0)^\top \mathcal{G}_{p-1}(\xi_0)^{-1}a_{p-1}(\theta_0) + \frac{\sigma_0^2}{2}\right\}\frac{\sqrt{n}}{\sigma_0^3}(\bar{\sigma}_n^2(\xi_0)-\sigma_0^2) + o_{\mathbb{P}^n_{\vartheta_0}}(1) \\
& + \frac{\sigma_0^2}{2}\left\langle -\frac{\sqrt{n}}{2\sigma_0^2}\left(\partial_\xi \bar{\sigma}_n^2(\xi_0) - \mathbb{E}^n_{\vartheta_0}\left[\partial_\xi \tilde{\sigma}_n^2(\xi_0)\right]\right), -\mathcal{G}_{p-1}(\xi_0)^{-1}a_{p-1}(\theta_0)\right\rangle_{\mathbb{R}^{p-1}}. \qquad (44)
\end{aligned}
$$

Therefore, using (43) and (44), the estimation error $\sqrt{n}(\widehat{\theta}_n-\theta_0) = (\sqrt{n}(\widehat{\xi}_n - \xi_0), \sqrt{n}(\widehat{\sigma}_n-\sigma_0))^\top$ is expressed by

$$
\begin{aligned}
\sqrt{n}(\widehat{\theta}_n-\theta_0) =& \begin{pmatrix} \mathcal{G}_{p-1}(\xi_0)^{-1} & -\frac{\sigma_0^2}{2}a_{p-1}(\theta_0)^\top\mathcal{G}_{p-1}(\xi_0)^{-1} \\ -\frac{\sigma_0^2}{2}a_{p-1}(\theta_0)^\top\mathcal{G}_{p-1}(\xi_0)^{-1} & \frac{\sigma_0^4}{4}a_{p-1}(\theta_0)^\top\mathcal{G}_{p-1}(\xi_0)^{-1}a_{p-1}(\theta_0)+\frac{\sigma_0^2}{2} \end{pmatrix}\bar{\zeta}_n + o_{\mathbb{P}^n_{\vartheta_0}}(1) \\
=& \mathcal{F}_p(\theta_0)^{-1}\bar{\zeta}_n + o_{\mathbb{P}^n_{\vartheta_0}}(1), \qquad (45)
\end{aligned}
$$

where the last equality can be proved in a similar way to the proof of Lemma 4 in Fukasawa and Takabatake (2019). Combining the expression (45) with Lemma 6, we complete the proof of Theorem 1.

A.3. *Proof of Theorem 2*

Before proving Theorem 2, we prove the asymptotic normality of MLE of μ. The following proposition is needed with its proof found in Section B.5 in Online Supplement.

Proposition 1: *Assume that Assumption 3 holds and a sequence of estimators* $\{\widehat{\xi}_n\}_{n\in\mathbb{N}}$ *satisfying that the sequence of rescaled estimation errors* $\{\sqrt{n}(\widehat{\xi}_n-\xi_0)\}_{n\in\mathbb{N}}$ *is stochastically bounded under the sequence of distributions* $\{\mathbb{P}^n_{\vartheta_0}\}_{n\in\mathbb{N}}$. *Then we can show that*

$$n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\widehat{\xi}_n)-\mu_0)\to\mathcal{N}\left(0,\frac{2\pi\sigma_0^2 c_X(\xi_0)\Gamma(1-\alpha_X(\xi_0))}{B(1-\alpha_X(\xi_0)/2,1-\alpha_X(\xi_0)/2)}\right)$$

in law under the distribution $\mathbb{P}^n_{\vartheta_0}$ *as* $n\to\infty$ *for any interior point* ϑ_0 *of* Θ.

Recall that $\Phi_n(\vartheta)=\mathrm{diag}(n^{-\frac{1}{2}}I_p,n^{-\frac{1}{2}(1-\alpha_X(\xi_0))})$ and

$$\sigma_n^2(\xi,\mu)=\frac{1}{n}(\mathbf{X}_n-\mu\mathbf{1}_n)^\top\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n),\ \ \tilde{\sigma}_n^2(\xi)=\sigma_n^2(\xi,\mu_0),\ \ \mu_n(\xi)=\frac{\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}\mathbf{X}_n}{\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n},$$

$$\ell_n(\vartheta)=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\sigma^2-\frac{1}{2}\log\det\left[\Sigma_n(s_\xi^X)\right]-\frac{n}{2\sigma^2}\sigma_n^2(\xi,\mu).$$

Then, using the formulas of derivatives of the log-determinant and the inverse matrix (*e.g.* see Harville (1998)), the first-order derivatives of the Gaussian log-likelihood function $\ell_n(\vartheta)$ with respect to parameters can be written as

$$\begin{cases}\partial_\xi\ell_n(\vartheta)=-\frac{1}{2}\mathrm{Tr}\left[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\right]-\frac{n}{2\sigma^2}\partial_\xi\sigma_n^2(\xi,\mu),\\ \partial_\sigma\ell_n(\vartheta)=-\frac{n}{\sigma}+\frac{n}{\sigma^3}\sigma_n^2(\xi,\mu)\ \ \partial_\mu\ell_n(\vartheta)=\frac{1}{\sigma^2}\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n)\end{cases}. \tag{46}$$

Then we can write

$$\partial_\xi\ell_n(\vartheta_0)=-\frac{n}{2\sigma_0^2}\left(\partial_\xi\tilde{\sigma}_n^2(\xi_0)-\mathbb{E}^n_{\vartheta_0}[\partial_\xi\tilde{\sigma}_n^2(\xi_0)]\right),$$

$$\partial_\sigma\ell_n(\vartheta_0)=\frac{n}{\sigma_0^3}\left(\tilde{\sigma}_n^2(\xi_0)-\sigma_0^2\right),\ \ \partial_\mu\ell_n(\vartheta_0)=\frac{1}{\sigma_0^2}(\mathbf{1}_n^\top\Sigma_n(s_{\xi_0}^X)^{-1}\mathbf{1}_n)(\mu_n(\xi_0)-\mu_0),$$

so that the normalized score function $\zeta_n(\vartheta_0)=\Phi_n(\vartheta_0)^\top\partial_\vartheta\ell_n(\vartheta_0)$ is expressed by

$$\zeta_n(\vartheta_0)=\mathrm{diag}\left(-\frac{\sqrt{n}}{2\sigma_0^2}I_{p-1},\frac{\sqrt{n}}{\sigma_0^3},\frac{(\mathbf{1}_n^\top\Sigma_n(s_{\xi_0}^X)^{-1}\mathbf{1}_n)}{\sigma_0^2 n^{\frac{1}{2}(1-\alpha_X(\xi_0))}}\right)\begin{pmatrix}\partial_\xi\tilde{\sigma}_n^2(\xi_0)-\mathbb{E}^n_{\vartheta_0}[\partial_\xi\tilde{\sigma}_n^2(\xi_0)]\\ \tilde{\sigma}_n^2(\xi_0)-\sigma_0^2\\ \mu_n(\xi_0)-\mu_0\end{pmatrix}. \tag{47}$$

Then we can prove the following central limit theorem

$$\mathcal{L}(\zeta_n(\vartheta)|\mathbb{P}^n_\vartheta) \to \mathcal{N}(0, \mathcal{I}(\vartheta)), \text{ as } n \to \infty, \tag{48}$$

whose proof is left to Section A.5 in Online Supplement. Moreover, we can show

$$\sqrt{n}\begin{pmatrix} \partial_\xi \bar{\sigma}^2_n(\xi_0) - \mathbb{E}^n_{\vartheta_0}[\partial_\xi \bar{\sigma}^2_n(\xi_0)] \\ \bar{\sigma}^2_n(\xi_0) - \sigma^2_0 \end{pmatrix} = \sqrt{n}\begin{pmatrix} \partial_\xi \tilde{\sigma}^2_n(\xi_0) - \mathbb{E}^n_{\vartheta_0}[\partial_\xi \tilde{\sigma}^2_n(\xi_0)] \\ \tilde{\sigma}^2_n(\xi_0) - \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi_0)] \end{pmatrix} + o_{\mathbb{P}^n_{\vartheta_0}}(1) \text{ as } n \to \infty. \tag{49}$$

whose proof is left to Section A.6 in Online Supplement. Combining the expression of the estimation error $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ in (45) with that of $\zeta_n(\vartheta_0)$ in (47) and using the equalities in (65) and (49), we conclude

$$\Phi_n(\vartheta_0)^{-1}(\widehat{\vartheta}_n - \vartheta_0) = \zeta_n(\vartheta_0) + o_{\mathbb{P}^n_{\vartheta_0}}(1) \text{ as } n \to \infty.$$

We complete the proof of Theorem 2.

A.4. *Proof of Theorem 3*

Recall that $\Phi_n(\vartheta) = \operatorname{diag}(n^{-\frac{1}{2}} I_p, n^{-\frac{1}{2}(1-\alpha_X(\xi))})$. We first give an outline of the proof of Theorem 3. Using the Taylor theorem, the log-likelihood ratio is written as

$$\begin{aligned}
\log \frac{\mathrm{d}\mathbb{P}^n_{\vartheta+\Phi_n(\vartheta)u}}{\mathrm{d}\mathbb{P}^n_\vartheta}(\mathbf{X}_n) &= u^\top \zeta_n(\vartheta) - \frac{1}{2}\int_0^1 (1-z)\partial^2_\vartheta \ell_n(\vartheta + z\Phi_n(\vartheta)u)\left[(\Phi_n(\vartheta)u)^{\otimes 2}\right] \mathrm{d}z \\
&= u^\top \zeta_n(\vartheta) - \frac{1}{2}\int_0^1 (1-z)\mathcal{I}_n(\vartheta + z\Phi_n(\vartheta)u)\left[(R_n(\vartheta, z\Phi_n(\vartheta)u)u)^{\otimes 2}\right] \mathrm{d}z,
\end{aligned} \tag{50}$$

where $R_n(\vartheta, v) := \Phi_n(\vartheta+v)^{-1}\Phi_n(\vartheta)$ for $v \in \mathbb{R}^{p+1}$. Since we have the CLT for the stochastic sequence $\{\zeta_n(\vartheta)\}_{n=1}^\infty$ in (48) and the inequality

$$\begin{aligned}
&\left|\int_0^1 (1-z)\mathcal{I}_n(\vartheta + z\Phi_n(\vartheta)u)\left[(R_n(\vartheta, z\Phi_n(\vartheta)u)u)^{\otimes 2}\right] \mathrm{d}z - u^\top \mathcal{I}(\vartheta)u\right| \\
&\le \|R_n(\vartheta, z\Phi_n(\vartheta)u)^\top \mathcal{I}_n(\vartheta + z\Phi_n(\vartheta)u)R_n(\vartheta, z\Phi_n(\vartheta)u) - \mathcal{I}(\vartheta)\|_1 \|u\|^2_{\mathbb{R}^{p+1}},
\end{aligned}$$

where $\|A\|_1 := \sum_{i,j=1}^p |a_{ij}|$ for a $p\times p$-matrix $A = (a_{ij})_{i,j=1,\dots,p}$, we can conclude Theorem 3 using the triangle inequality, multiplicativity of the matrix norm $\|\cdot\|_1$, and uniform continuity of $\mathcal{I}(\vartheta)$ on compact subsets of Θ, if we can prove

$$\sup_{v\in\mathbb{R}^{p+1}:\|v\|_{\mathbb{R}^{p+1}}\le c\|\Phi_n(\vartheta)\|_{\mathrm{op}}} \|R_n(\vartheta, v) - I_{p+1}\|_1 = o_{\mathbb{P}^n_\vartheta}(1) \text{ as } n \to \infty, \tag{51}$$

$$\|\mathcal{I}_n(\vartheta) - \mathcal{I}(\vartheta)\|_1 = o_{\mathbb{P}^n_\vartheta}(1) \ \text{ as } n \to \infty, \tag{52}$$

$$\sup_{u \in \mathbb{U}_{n,c}(\vartheta)} \|\mathcal{I}_n(\vartheta + \Phi_n(\vartheta)u) - \mathcal{I}_n(\vartheta)\|_1 = o_{\mathbb{P}^n_\vartheta}(1) \ \text{ as } n \to \infty, \tag{53}$$

for any $c > 0$, where $\mathbb{U}_n(\vartheta) := \Phi_n^{-1}(\Theta)(\Theta - \vartheta) = \{u \in \mathbb{R}^{p+1} : \vartheta + \Phi_n(\vartheta)u \in \Theta\}$ and $\mathbb{U}_{n,c}(\vartheta) := \{u \in \mathbb{U}_n(\vartheta) : \|u\|_{\mathbb{R}^{p+1}} \le c\}$ for $c > 0$. In the rest of the Appendix, we try to prove the above three results.

A.5. *Proof of* (48)

Set $\mathbf{Z}_n := \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}}(\mathbf{X}_n - \mu_0 \mathbf{1}_n) \sim \mathcal{N}(0, I_n)$. For $\mathbf{u}_p = (u_1, \ldots, u_p)^\top \in \mathbb{R}^p$ and $u_{p+1} \in \mathbb{R}$, we write

$$J_n(u_1, \ldots, u_{p+1}) := \begin{pmatrix} \mathbf{u}_p^\top & u_{p+1} \end{pmatrix} \times \mathrm{diag}\left(-\frac{\sqrt{n}}{2\sigma_0^2} I_{p-1}, \frac{\sqrt{n}}{\sigma_0^3}, \frac{(\mathbf{1}_n^\top \Sigma_n(s^X_{\theta_0})^{-1}\mathbf{1}_n)}{\sigma_0^2 n^{\frac{1}{2}(1-\alpha_X(\xi_0))}} \right) \times \begin{pmatrix} \partial_\xi \tilde{\sigma}^2_n(\xi_0) - \mathbb{E}^n_{\vartheta_0}[\partial_\xi \tilde{\sigma}^2_n(\xi_0)] \\ \tilde{\sigma}^2_n(\xi_0) - \sigma_0^2 \\ \mu_n(\xi_0) - \mu_0 \end{pmatrix}.$$

Then $J_n(u_1, \ldots, u_{p+1})$ is rewritten as

$$\begin{aligned} J_n(u_1, \ldots, u_{p+1}) &= \frac{1}{2\sqrt{n}} \sum_{j=1}^{p-1} u_j \mathbf{Z}_n^\top \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \Sigma_n(\partial_j s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \mathbf{Z}_n \\ &\quad + \frac{1}{\sigma_0 \sqrt{n}} u_p \mathbf{Z}_n^\top \mathbf{Z}_n + \frac{(\mathbf{1}_n^\top \Sigma_n(s^X_{\theta_0})^{-1}\mathbf{1}_n)}{n^{\frac{1}{2}(1-\alpha_X(\xi_0))}} u_{p+1} \frac{(\Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}}\mathbf{1}_n)^\top \mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s^X_{\theta_0})^{-1}\mathbf{1}_n} \\ &= \frac{1}{\sqrt{n}} \mathbf{Z}_n^\top \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \Sigma_n(g^{u_1,\ldots,u_p}_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \mathbf{Z}_n + n^{-\frac{1}{2}(1-\alpha_X(\xi_0))} u_{p+1} (\Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}}\mathbf{1}_n)^\top \mathbf{Z}_n, \end{aligned}$$

where

$$g^{u_1,\ldots,u_p}_{\theta_0}(\omega) := \frac{1}{2} \sum_{j=1}^{p-1} u_j \partial_j s^X_{\theta_0}(\omega) + \frac{1}{\sigma_0} u_p s^X_{\theta_0}(\omega).$$

Since the matrix $\Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \Sigma_n(g^{u_1,\ldots,u_p}_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}}$ is symmetric, there exists a nth square matrix V_n such that V_n is an orthogonal matrix and

$$V_n \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \Sigma_n(g^{u_1,\ldots,u_p}_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} V_n^\top = \mathrm{diag}(\lambda_{1,n}, \ldots, \lambda_{n,n}),$$

where $\{\lambda_{j,n}\}_{j=1,\ldots,n}$ are eigenvalues of the matrix $\Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \Sigma_n(g^{u_1,\ldots,u_p}_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}}$. Then we set a n-dimensional random vector $\mathbf{W}_n = (W_{1,n}, \ldots, W_{n,n})^\top$ and a n-dimensional (non-random) vector $\mathbf{A}_n = (A_{1,n}, \ldots, A_{n,n})^\top$ by

$$\mathbf{W}_n := V_n \mathbf{Z}_n \text{ and } \mathbf{A}_n := u_{p+1} V_n \Sigma_n(s^X_{\theta_0})^{-\frac{1}{2}} \mathbf{1}_n.$$

Notice that $\mathbf{W}_n$ is a n-dimensional standard Gaussian vector and we can write

$$J_n(u_1, \ldots, u_{p+1}) = \sum_{j=1}^n \left\{ n^{-\frac{1}{2}} \lambda_{j,n}(W^2_{j,n} - 1) + n^{-\frac{1}{2}(1-\alpha_X(\xi_0))} A_{j,n} W_{j,n} \right\}.$$

Set $U_{j,n} := n^{-\frac{1}{2}} \lambda_{j,n}(W^2_{j,n} - 1) + \sigma_0^{-2} n^{-\frac{1}{2}(1-\alpha_X(\xi_0))} A_{j,n} W_{j,n}$. Notice that $\{U_{j,n}\}_{j=1,\ldots,n}$ is an independent triangle array with mean zero and variance

$$\begin{aligned}
\mathrm{Var}[U_{j,n}] =& n^{-1} \lambda^2_{j,n} \mathbb{E}[(W^2_{j,n} - 1)^2] + n^{-(1-\alpha_X(\xi_0))} A^2_{j,n} \mathbb{E}[W^2_{j,n}] \\
&+ 2 n^{-\frac{1}{2}} n^{-\frac{1}{2}(1-\alpha_X(\xi_0))} \lambda_{j,n} A_{j,n} \mathbb{E}[(W^2_{j,n} - 1) W_{j,n}] \\
=& 2 n^{-1} \lambda^2_{j,n} + n^{-(1-\alpha_X(\xi_0))} A^2_{j,n},
\end{aligned}$$

where we used the facts that

$$\mathbb{E}[(W^2_{j,n} - 1)^2] = \mathbb{E}[W^4_{j,n}] - 2\mathbb{E}[W^2_{j,n}] + 1 = 2, \quad \mathbb{E}[(W^2_{j,n} - 1) W_{j,n}] = 0.$$

Then we can also show that

$$\begin{aligned}
\mathrm{Var}\left[\sum_{j=1}^n U_{j,n}\right] &= \sum_{j=1}^n \mathrm{Var}[U_{j,n}] = \frac{2}{n} \sum_{j=1}^n \lambda^2_{j,n} + n^{-(1-\alpha_X(\xi_0))} \sum_{j=1}^n A^2_{j,n} \\
&= \frac{2}{n} \mathrm{Tr}[\mathrm{diag}(\lambda_{1,n}, \ldots, \lambda_{n,n})^2] + n^{-(1-\alpha_X(\xi_0))} \|\mathbf{A}_n\|^2_{\mathbb{R}^n} \\
&= \frac{2}{n} \mathrm{Tr}[(\Sigma_n(g^{u_1,\ldots,u_p}_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1})^2] + u^2_{p+1} n^{-(1-\alpha_X(\xi_0))} (\mathbf{1}_n^\top \Sigma_n(s^X_{\theta_0})^{-1} \mathbf{1}_n) \\
&= \mathbf{u}_p^\top \begin{pmatrix} \widetilde{\mathcal{F}}_{p-1,n}(\theta_0) & \frac{1}{\sigma_0 n} \mathrm{Tr}[\Sigma_n(\partial_\xi s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1}] \\ \text{sym.} & 2\sigma_0^{-2} \end{pmatrix} \mathbf{u}_p + u^2_{p+1} n^{-(1-\alpha_X(\xi_0))} (\mathbf{1}_n^\top \Sigma_n(s^X_{\theta_0})^{-1} \mathbf{1}_n),
\end{aligned}$$

where

$$\widetilde{\mathcal{F}}_{p-1,n}(\theta_0) := \frac{1}{2n} \begin{pmatrix} \mathrm{Tr}[(\Sigma_n(\partial_1 s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1})^2] & \ldots & \mathrm{Tr}[\Sigma_n(\partial_1 s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1} \Sigma_n(\partial_{p-1} s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1}] \\ \vdots & \ddots & \vdots \\ \mathrm{Tr}[\Sigma_n(\partial_{p-1} s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1} \Sigma_n(\partial_1 s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1}] & \ldots & \mathrm{Tr}[(\Sigma_n(\partial_{p-1} s^X_{\theta_0}) \Sigma_n(s^X_{\theta_0})^{-1})^2] \end{pmatrix},$$

so that Lemma 4 and Theorems 4.1 and 5.2 in Adenstedt (1974) yield

$$\lim_{n\to\infty} \mathrm{Var}\left[\sum_{j=1}^{n} U_{j,n}\right] = \mathbf{u}_p^\top \mathcal{F}_p(\theta_0)\mathbf{u}_p + u_{p+1}^2 \frac{2\pi\sigma_0^2 c_X(\xi_0)\Gamma(1-\alpha_X(\xi_0))}{B(1-\alpha_X(\xi_0)/2, 1-\alpha_X(\xi_0)/2)} = \mathbf{u}_{p+1}^\top \mathcal{I}(\theta_0)\mathbf{u}_{p+1}.$$

Moreover, we can show that

$$\mathbb{E}[U_{j,n}^4] \le 4n^{-2}\mathbb{E}[|\lambda_{j,n}(W_{j,n}^2-1)|^4] \ + 4\sigma_0^{-4} n^{2(1-\alpha_X(\xi_0))}(\mathbf{1}_n^\top \Sigma_n(s_{\xi_0}^X)^{-1}\mathbf{1}_n)^{-4}\mathbb{E}[|A_{j,n}W_{j,n}|^4].$$

Here notice that, since $\{W_{j,n}\}_{j=1}^n$ is an independent centered sequence for each $n \in \mathbb{N}$, we can show

$$\mathbb{E}\Big[\Big|\sum_{j=1}^n A_{j,n}W_{j,n}\Big|^4\Big] = \sum_{j_1,j_2,j_3,j_4=1}^{n} \prod_{i=1}^4 A_{j_i,n}\mathbb{E}\Big[\prod_{k=1}^4 W_{j_k,n}\Big], = \sum_{j=1}^n A_{j,n}^4 \mathbb{E}[W_{j,n}^4] + 3\left(\sum_{j=1}^n A_{j,n}^2 \mathbb{E}[W_{j,n}^2]\right)^2,$$

which implies

$$\begin{aligned}\sum_{j=1}^n \mathbb{E}[|A_{j,n}W_{j,n}|^4] =& \mathbb{E}\Big[\Big|\sum_{j=1}^n A_{j,n}W_{j,n}\Big|^4\Big] - 3\left(\sum_{j=1}^n |A_{j,n}|^2\right)^2 \\ =& u_{p+1}^4\left(\mathbb{E}[|(\Sigma_n(s_{\theta_0}^X)^{-\frac{1}{2}}\mathbf{1}_n)^\top \mathbf{Z}_n|^4] - 3\|\mathbf{A}_n\|_{\mathbb{R}^n}^2\right) \\ =& u_{p+1}^4\left(3(\mathbf{1}_n^\top\Sigma_n(s_{\theta_0}^X)^{-1}\mathbf{1}_n)^2 - 3(\mathbf{1}_n^\top\Sigma_n(s_{\theta_0}^X)^{-1}\mathbf{1}_n)^2\right) = 0,\end{aligned}$$

where we used $(\Sigma_n(s_{\theta_0}^X)^{-\frac{1}{2}}\mathbf{1}_n)^\top \mathbf{Z}_n \sim \mathcal{N}(0, \mathbf{1}_n^\top\Sigma_n(s_{\theta_0}^X)^{-1}\mathbf{1}_n)$. Moreover, we can show

$$\mathbb{E}[(W_{j,n}^2-1)^4] = \mathbb{E}[W_{j,n}^8 - 4W_{j,n}^6 + 6W_{j,n}^4 - 4W_{j,n}^2 + 1] = 60,$$

so that we obtain

$$\sum_{j=1}^n \mathbb{E}[U_{j,n}^4] \le \frac{240}{n^2}\sum_{j=1}^n \lambda_{j,n}^4 = \frac{240}{n^2}\mathrm{Tr}[\mathrm{diag}(\lambda_{1,n},\dots,\lambda_{n,n})^4] = \frac{240}{n^2}\mathrm{Tr}[(\Sigma_n(g_{\theta_0}^{u_1,\dots,u_p})\Sigma_n(s_{\xi_0}^X)^{-1})^4] \to 0$$

using Lemma 4. Therefore, we have succeeded in verifying Lindeberg's condition. So we conclude the result.

A.6. *Proof of* (49)

Recall that

$$e_{n,1}(\xi) := \bar{\sigma}_n^2(\xi) - \tilde{\sigma}_n^2(\xi) = -n^{-1}(\mu_n(\xi)-\mu_0)^2\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n,$$

$$\partial_\xi e_{n,1}(\xi) = -2n^{-1}(\mu_n(\xi) - \mu_0)\partial_\xi \mu_n(\xi)\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n - n^{-1}(\mu_n(\xi) - \mu_0)^2 \mathbf{1}_n^\top \partial_\xi \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n,$$

see (61) and (62). Notice that, using Lemmas 1 and 2 and the Cauchy-Schwarz inequality, we can show that

$$\sqrt{n} e_{n,1}(\xi_0) = o_{\mathbb{P}^n_{\vartheta_0}}(1) \text{ and } \sqrt{n}\partial_\xi e_{n,1}(\xi_0) = o_{\mathbb{P}^n_{\vartheta_0}}(1) \text{ as } n \to \infty.$$

Moreover, $\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi_0)] = n^{-1}\mathrm{Tr}[\Sigma_n(s^X_{\vartheta_0})\Sigma_n(s^X_{\xi_0})^{-1}] = \sigma_0^2$ holds so that we complete the proof of (49).

APPENDIX B: Online Supplement to "Optimal Estimation for General Gaussian Processes" by Takabatake, Yu, and Zhang (Not for Publication)

B.1. *Some Useful Lemmas*

In this subsection, we summarize preliminary results used in the proof of theorems in Section 2, whose proofs are left to Section B.2. The following two lemmas are useful to prove Theorems 1 and 2 and are frequently used in their proofs. Recall that we write $\chi(\xi_1,\xi_2) = (\alpha_X(\xi_1) - \alpha_X(\xi_2))_+/(1-\alpha_X(\xi_2)_+) \in [0,1)$ and $\chi_0(\xi_1) = \chi(\xi_0,\xi_1)$ for $\xi_1,\xi_2 \in \Theta_\xi$.

Lemma 1: *Under Assumption 1, we can show that for any $i,j,k \in \{1,2,\dots,p-1\}$, $\varepsilon > 0$ and $\theta \in \Theta$,*

$$
\begin{aligned}
\left|\mathbf{1}_n^\top \partial_i \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n\right| &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\varepsilon},\\
\left|\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n\right| &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\varepsilon},\\
\left|\mathbf{1}_n^\top \partial_{i,j,k}^3 \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n\right| &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\varepsilon},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\Sigma_n(s_{\theta_0}^X)\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\chi_0(\xi)},\\
\mathbf{1}_n^\top \partial_i\Sigma_n(s_\theta^X)^{-1}\Sigma_n(s_{\theta_0}^X)\partial_i\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\chi_0(\xi)+\varepsilon},\\
\mathbf{1}_n^\top \partial_{i,j}^2\Sigma_n(s_\theta^X)^{-1}\Sigma_n(s_{\theta_0}^X)\partial_{i,j}^2\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\chi_0(\xi)+\varepsilon},\\
\mathbf{1}_n^\top \partial_{i,j,k}^3\Sigma_n(s_\theta^X)^{-1}\Sigma_n(s_{\theta_0}^X)\partial_{i,j,k}^3\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n &\le C(\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)n^{\chi_0(\xi)+\varepsilon}.
\end{aligned}
$$

Lemma 2: *For any $q \in \mathbb{N}$, there exists a positive constant C_q such that for any $i,j,k \in \{1,2,\dots,p-1\}$, $\varepsilon > 0$ and $\theta \in \Theta$,*

$$
\begin{aligned}
&E_{\vartheta_0}^n\left[\left|\partial_{i,j}^2\mu_n(\xi)\right|^q\right] \le C_q(\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n)^{-\frac{q}{2}}n^{\frac{q}{2}\chi_0(\xi)+\varepsilon},\\
&\mathbb{E}_{\vartheta_0}^n\left[\left|\partial_{i,j,k}^3\mu_n(\xi)\right|^q\right] \le C_q(\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n)^{-\frac{q}{2}}n^{\frac{q}{2}\chi_0(\xi)+\varepsilon}.
\end{aligned}
$$

To prove consistency and the asymptotic normality of the exact MLE in Theorem 1, we need to verify uniform convergence of $\sigma^2(\xi)$ and its derivatives. Then we repeatedly use the following Sobolev inequality, which can be proved using Theorem 4.12 of Adams and Fournier (2003) and the Fubini theorem.

LEMMA 3—The Sobolev Inequality: *Let $d \in \mathbb{N}$, $\mathring{\Theta}_*$ be a bounded open cube in $\mathbb{R}^d$, Θ_* be the closure of Θ_*, and $\{(\mathcal{X}_n, \mathcal{A}_n, \mathbb{P}_*^n)\}_{n\in\mathbb{N}}$ be a sequence of complete probability spaces. Assume that $\{u_n(\theta, x_n)\}_{(\theta,x_n)\in\Theta_*\times\mathcal{X}_n}$ is a sequence of pathwise continuously differentiable random fields, i.e. for each $n \in \mathbb{N}$, it holds that*

- *the function $\theta \mapsto u_n(\theta, \omega_n)$ is continuously differentiable on $\mathring{\Theta}_*$ and uniformly continuous on Θ_* for $\mathbb{P}_*^n$-a.s. $x_n \in \mathcal{X}_n$,*
- *the functions $u_n(\cdot,\cdot)$ and $\partial_\theta u_n(\cdot,\cdot)$ on $\Theta_* \times \mathcal{X}_n$ are $\mathcal{B}(\Theta_*) \otimes \mathcal{X}_n$-measurable, where $\mathcal{B}(\Theta_*)$ denotes the Borel σ-algebra on the set Θ_*.*

Then for any $q \in \mathbb{N}$ satisfying $q > d$, there exists a positive constant $C_1 = C_1(q,d)$, which is independent of n, x_n and $u_n(\cdot,\cdot)$, such that

$$\sup_{\theta'\in\Theta_*} \left|u_n(\theta', x_n)\right|^q \le C_1 \left[\int_{\Theta_*} \left|u_n(\theta', x_n)\right|^q \, \mathrm{d}\theta' + \int_{\Theta_*} \left\|\partial_\theta u_n(\theta', x_n)\right\|_{\mathbb{R}^d}^q \, \mathrm{d}\theta' \right]$$

hold for any $n \in \mathbb{N}$ and $\mathbb{P}_*^n$-a.s. $\omega_n \in \mathcal{X}_n$. In particular, we get

$$\mathbb{E}^{\mathbb{P}_*^n}\left[\sup_{\theta'\in\Theta_*} \left|u_n(\theta',\cdot)\right|^q\right] \le C_2 \sup_{\theta'\in\Theta_*} \left\{ \mathbb{E}^{\mathbb{P}_*^n}\left[\left|u_n(\theta',\cdot)\right|^q\right] + \mathbb{E}^{\mathbb{P}_*^n}\left[\left\|\partial_\theta u_n(\theta',\cdot)\right\|_{\mathbb{R}^d}^q\right] \right\}$$

for the positive constant $C_2 := C_1(q,d) m_d(\Theta_*)$, where $m_d(A)$ denotes the Lebesgue measure of a measurable set A of $\mathbb{R}^d$.

REMARK 6: We provide a clarification regarding the proof of Lemma 3. In the Sobolev inequality and the Sobolev embedding theorem, the geometric structure of the domain and its boundary plays an essential role. The version of the Sobolev embedding theorem given in Theorem 4.12 of Adams and Fournier (2003) is stated under the assumption that the domain Θ_* satisfies the *strong locally Lipschitz condition*; see Section 4.9 of Adams and Fournier (2003) for its definition. However, since

Θ_* is bounded, it suffices to assume that Θ_* has a *locally Lipschitz boundary*, that is, for each point $\theta_* \in \partial\Theta_*$, there exists a neighborhood $U(\theta_*)$ such that $U(\theta_*) \cap \partial\Theta_*$ is the graph of a Lipschitz continuous function. In particular, any bounded open cube in $\mathbb{R}^d$ has a locally Lipschitz boundary. Therefore, Theorem 4.12 in Adams and Fournier (2003) directly implies the result stated in Lemma 3.

Next, we provide a precise approximation error bound for the trace of the product of Toeplitz matrices and the inverses of (possibly different) Toeplitz matrices. The entries of these matrices are defined via the Fourier transforms of the spectral density function and their derivatives with respect to the model parameters. This result is particularly useful for evaluating the cumulants of quadratic forms of Gaussian vectors arising from stationary Gaussian time series, as well as the cumulants of their derivatives with respect to model parameters, where the matrix in the quadratic form is given by the inverse of a Toeplitz matrix associated with a spectral density function.

LEMMA 4: *Let $q \in \mathbb{N}$, $\{\alpha_j\}_{j=1}^{2q}$ be a family of continuous functions on Θ_ξ to $(-\infty, 1)$, $\{f_j\}_{j=1}^{2q}$ be a family of 2π-periodic even measurable functions on $\mathbb{R}$ satisfying the following conditions for each $j \in \{1, \dots, 2q\}$ and $r \in \{1, \dots, q\}$:*

(*a*) *f_j is continuously differentiable on $\mathbb{R} \setminus (2\pi\mathbb{Z})$,*

(*b*) *f_{2r} is non-negative almost everywhere and the set $\{x \in [-\pi, \pi] : f_{2r}(x) > 0\}$ has a positive Lebesgue measure,*

(*c*) *there exists constants $c_+ > c_- > 0$ such that, for any $\theta = (\xi, \sigma)^\top \in \Theta$, $x \in [-\pi, \pi]$, the following inequalities hold:*

$$|x|^{\alpha_{2r-1}(\xi)} |f_{2r-1}(x, \theta)| + |x|^{\alpha_j(\xi)+1} \left| \frac{\mathrm{d}f_j}{\mathrm{d}x}(x, \theta) \right| \le c_+ \ \text{ and } \ |x|^{\alpha_{2r}(\xi)} f_{2r}(x, \theta) \ge c_-.$$

For $\underline{\xi} = (\xi_1, \dots, \xi_{2q})^\top \in \Theta_\xi^{2q}$, we write

$$\chi_q(\underline{\xi}) := \sum_{r=1}^{q} \frac{(\alpha_{2r-1}(\xi_{2r-1}) - \alpha_{2r}(\xi_{2r}))_+}{1 - \alpha_{2r}(\xi_{2r})_+} \vee \frac{(\alpha_{2r+1}(\xi_{2r+1}) - \alpha_{2r}(\xi_{2r}))_+}{1 - \alpha_{2r}(\xi_{2r})_+},$$

where $\alpha_{2q+1}(\xi) := \alpha_1(\xi)$ and $\xi_{2q+1} := \xi_1$ for notational simplicity, and then we introduce a restricted parameter space $\Theta_\xi^*(\iota)$ defined by $\Theta_\xi^*(\iota) := \{\underline{\xi} = (\xi_1, \dots, \xi_{2q}) \in \Theta_\xi^{2q} : \chi_q(\underline{\xi}) \le 1 - \iota\}$

for $\iota \in (0,1)$. *Then we obtain*

$$n^{-\chi_q(\underline{\xi})-\epsilon}\left|\mathrm{Tr}\left[\prod_{r=1}^{q}\Sigma_n(f_{2r-1,\theta_{2r-1}})\Sigma_n(f_{2r,\theta_{2r}})^{-1}\right]-\frac{n}{2\pi}\int_{-\pi}^{\pi}\prod_{r=1}^{q}\frac{f_{2r-1,\theta_{2r-1}}(x)}{f^n_{2r,\theta_{2r}}(x)}\mathrm{d}x\right|=o(1)$$

as $n\to\infty$ *uniformly on compact subsets of* $\Theta^*_{\underline{\xi}}(\iota)$ *for any* $\epsilon>0$ *and* $\iota\in(0,1)$, *where we write* $f_{j,\theta}(x)\equiv f_j(x,\theta)$ *for* $(x,\theta)^\top\in[-\pi,\pi]\times\Theta$.

We provide a remark on the proof of Lemma 4. The proof relies critically on Theorem 1 in the supplement of Takabatake (2024) and Lemma 3 of Chong and Mies (2025). As a result, following the proof steps of Theorem 3 in Lieberman et al. (2009), the assertion of Lemma 4 can now be justified using Theorem 1 in Takabatake (2024), in place of Theorem 2 in Lieberman and Phillips (2004), and Lemma 3 of Chong and Mies (2025), in place of Lemma 2 in the Online Supplement of Lieberman et al. (2012).

Here we prepare the uniform convergence of $\partial_i\bar{\sigma}^2_n(\xi)$ and $\partial^2_{i,j}\bar{\sigma}^2_n(\xi)$ on $\Theta_\xi(\iota)$.

Lemma 5: *Under Assumption 1, we can show*

$$\sup_{\xi\in\Theta_\xi(\iota)}\left|\partial_i\bar{\sigma}^2_n(\xi)-\partial_i\sigma^2(\xi)\right|=o_{\mathbb{P}^n_{\vartheta_0}}(1),\ \sup_{\xi\in\Theta_\xi(\iota)}\left|\partial^2_{i,j}\bar{\sigma}^2_n(\xi)-\partial^2_{i,j}\sigma^2(\xi)\right|=o_{\mathbb{P}^n_{\vartheta_0}}(1)$$

for $i,j=1,\dots,p-1$ *and* $\iota\in(0,1)$, *where* $\sigma^2(\xi)$ *is defined in* (24). *In particular, we write*

$$\partial_i\sigma^2(\xi)=\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\left(-\frac{\partial_i s^X_\xi(\omega)}{s^X_\xi(\omega)}\right)\left(\frac{s^X_{\xi_0}(\omega)}{s^X_\xi(\omega)}\right)\mathrm{d}\omega=-\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\partial_i\log s^X_\xi(\omega)\left(\frac{s^X_{\xi_0}(\omega)}{s^X_\xi(\omega)}\right)\mathrm{d}\omega,$$

$$\partial^2_{i,j}\sigma^2(\xi)=\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}\left(-\frac{\partial^2_{i,j}s^X_\xi(\omega)}{s^X_\xi(\omega)}\right)\left(\frac{s^X_{\xi_0}(\omega)}{s^X_\xi(\omega)}\right)\mathrm{d}\omega+\frac{\sigma_0^2}{2\pi}\int_{-\pi}^{\pi}2\left(\partial_i\log s^X_\xi(\omega)\right)\left(\partial_j\log s^X_\xi(\omega)\right)\left(\frac{s^X_{\xi_0}(\omega)}{s^X_\xi(\omega)}\right)\mathrm{d}\omega.$$

Define $(p-1)$th matrix-valued continuous functions $\mathcal{G}_{p-1,n}(\xi):=(\mathcal{G}^{i,j}_n(\xi))_{i,j=1,\dots,p-1}$ and $\mathcal{G}_{p-1}(\xi,\theta_0):=(\mathcal{G}^{i,j}(\xi,\theta_0))_{i,j=1,\dots,p-1}$ by $\mathcal{G}^{i,j}_n(\xi):=-\frac{1}{n}\partial^2_{i,j}\bar{\ell}_n(\xi)$ and

$$\begin{aligned}\mathcal{G}^{i,j}(\xi,\theta_0):=&-\frac{1}{2}\left(-\frac{\partial_i\sigma^2(\xi)}{\sigma^2(\xi)}\right)\left(-\frac{\partial_j\sigma^2(\xi)}{\sigma^2(\xi)}\right)+\frac{1}{2\sigma^2(\xi)}\partial^2_{i,j}\sigma^2(\xi)\\&+\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial^2_{i,j}s^X_\xi(\omega)}{s^X_\xi(\omega)}\mathrm{d}\omega-\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\partial_i\log s^X_\xi(\omega)\right)\left(\partial_j\log s^X_\xi(\omega)\right)\mathrm{d}\omega\end{aligned}$$

for each $i,j=1,\dots,p-1$. Notice that, using the expressions of $\partial_i\sigma^2(\xi_0)$ and $\partial^2_{i,j}\sigma^2(\xi_0)$ in Lemma 5, we can write

$$\mathcal{G}_{p-1}(\xi_0) := \mathcal{G}_{p-1}(\xi_0,\theta_0) = -\frac{1}{2}a_{p-1}(\xi_0)a_{p-1}(\xi_0)^\top + \mathcal{F}_{p-1}(\xi_0).$$

Moreover, using the expression in (36), Lemma 4, the uniform convergence in (30) and Lemma 5, we obtain, for $i,j=1,\dots,p-1$,

$$\sup_{\xi\in\Theta_\xi(\iota)} |\mathcal{G}_n^{i,j}(\xi) - \mathcal{G}^{i,j}(\xi,\theta_0)| = o_{\mathbb{P}^n_{\vartheta_0}}(1) \ \text{ as } n\to\infty. \tag{54}$$

Moreover, we also prepare the following central limit theorem, whose proof is omitted since a stronger result than Lemma 6 is proved in (48) and it can be proved as a corollary of (48).

Lemma 6: *Under Assumption 1, we can show*

$$\bar{\zeta}_n := \mathrm{diag}\left(-\frac{\sqrt{n}}{2\sigma_0^2}I_{p-1}, \frac{\sqrt{n}}{\sigma_0^3}\right)\begin{pmatrix}\partial_\xi\bar{\sigma}_n^2(\xi_0) - \mathbb{E}^n_{\vartheta_0}[\partial_\xi\tilde{\sigma}_n^2(\xi_0)] \\ \bar{\sigma}_n^2(\xi_0) - \sigma_0^2\end{pmatrix} \to \mathcal{N}(0,\mathcal{F}_p(\theta_0)) \ \textit{ as } n\to\infty,$$

where the matrix $\mathcal{F}_p(\theta)$ is defined in (3).

B.2. *Proof of Lemmas in Section B.1*

Proof of Lemma 1: First, by the chain rule, we have

$$\begin{aligned}
\partial_i\Sigma_n(s_\xi^X)^{-1} =& -\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1},\\
\partial^2_{i,j}\Sigma_n(s_\xi^X)^{-1} =& \Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\\
&-\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial^2_{i,j}s_\xi^X)\Sigma_n(s_\xi^X)^{-1} + \Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1},\\
\partial^3_{i,j,k}\Sigma_n(s_\xi^X)^{-1} =& -\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_k s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\\
&+\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial^2_{j,k}s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\\
&-\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_k s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\\
&+\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial^2_{i,k}s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\\
&-\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_j s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_i s_\xi^X)\Sigma_n(s_\xi^X)^{-1}\Sigma_n(\partial_k s_\xi^X)\Sigma_n(s_\xi^X)^{-1}
\end{aligned}$$

$$
\begin{aligned}
&+\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}-\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^3_{i,j,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\\
&+\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\\
&-\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\\
&+\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\\
&-\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\\
&+\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{j,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\\
&-\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}.
\end{aligned}
$$

In the rest of the proof, we only prove the second assertion because, from the above expressions of the derivatives $\partial_i\Sigma_n(s^X_\xi)^{-1}$, $\partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}$ and $\partial^3_{i,j,k}\Sigma_n(s^X_\xi)^{-1}$, we can see that the other assertions can be proved similarly. Notice that we can show

$$
\begin{aligned}
\left|\mathbf{1}_n^\top\partial^2_{i,j}\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\right| \le& \mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\Sigma_n(|\partial^2_{i,j}s^X_\theta|)\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\\
&+2\sum_{m_1,m_2\in\{+,-\}}\left|\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\Sigma_n((\partial_j s^X_\theta)_{m_2})\Sigma_n(s^X_\theta)^{-1}\Sigma_n((\partial_i s^X_\theta)_{m_1})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\right|,
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{m_1,m_2\in\{+,-\}}\left|\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\Sigma_n((\partial_j s^X_\theta)_{m_2})\Sigma_n(s^X_\theta)^{-1}\Sigma_n((\partial_i s^X_\theta)_{m_1})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\right|\\
&\le\sum_{m_1,m_2\in\{+,-\}}\left\|\Sigma_n(s^X_\theta)^{-\frac12}\Sigma_n((\partial_j s^X_\theta)_{m_1})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\right\|_{\mathbb{R}^n}\left\|\Sigma_n(s^X_\theta)^{-\frac12}\Sigma_n((\partial_i s^X_\theta)_{m_1})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n\right\|_{\mathbb{R}^n}\\
&\le\sum_{m_1,m_2\in\{+,-\}}\left\|\Sigma_n(s^X_\theta)^{-\frac12}\Sigma_n((\partial_j s^X_\theta)_{m_1})^{\frac12}\right\|^2_{\mathrm{op}}\left\|\Sigma_n(s^X_\theta)^{-\frac12}\Sigma_n((\partial_i s^X_\theta)_{m_1})^{\frac12}\right\|^2_{\mathrm{op}}\left\|\Sigma_n(s^X_\theta)^{-\frac12}\mathbf{1}_n\right\|^2_{\mathbb{R}^n},
\end{aligned}
$$

so that we conclude the second assertions using Lemma 3 of Chong and Mies (2025). Therefore, we finish the proof. *Q.E.D.*

Proof of Lemma 2: Since we have

$$
\mu_n(\xi)-\mu_0=\frac{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n-\mu_0\mathbf{1}_n)}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\sim\mathcal{N}\left(0,\frac{\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\Sigma_n(s^X_{\theta_0})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n}{(\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n)^2}\right) \tag{55}
$$

under the distribution $\mathbb{P}^n_{\vartheta_0}$, we can show that

$$
\mathbb{E}^n_{\vartheta_0}\left[\left|\mu_n(\xi)-\mu_0\right|^q\right]=2^{\frac q2}\Gamma\left(\frac{q+1}{2}\right)\pi^{-\frac12}\left|\frac{(\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\Sigma_n(s^X_{\theta_0})\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n)^{\frac12}}{\mathbf{1}_n^\top\Sigma_n(s^X_\theta)^{-1}\mathbf{1}_n}\right|^q
$$

so that the first assertion follows from Lemma 1.

Set $\mathbf{Z}_n := \mathbf{X}_n - \mu_0 \mathbf{1}_n$. First, note that,

$$\partial_i \mu_n(\xi) = \frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{X}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \mu_n(\xi)\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} = \frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - (\mu_n(\xi)-\mu_0)\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n},$$

$$\begin{aligned}\partial_{i,j}^2 \mu_n(\xi) =& \frac{\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) - \partial_j\mu_n(\xi)\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} \\ &-(\mu_n(\xi)-\mu_0)\left[\frac{\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\right], \end{aligned} \tag{56}$$

$$\begin{aligned}\partial_{i,j,k}^3 \mu_n(\xi) =& \frac{\mathbf{1}_n^\top \partial_{i,j,k}^3 \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_k \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) \\ &-\left(\frac{\mathbf{1}_n^\top \partial_{i,k}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_{j,k}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) \\ &+2\left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_k \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) - \partial_{j,k}^2\mu_n(\xi)\left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) \\ &-\partial_j\mu_n(\xi)\left[\frac{\mathbf{1}_n^\top \partial_{i,k}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_k \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\right] \\ &-\partial_k\mu_n(\xi)\left[\frac{\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\right] \\ &-(\mu_n(\xi)-\mu_0)\left[\frac{\mathbf{1}_n^\top \partial_{i,j,k}^3 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n} - \left(\frac{\mathbf{1}_n^\top \partial_{i,j}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_k \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\right. \\ &\qquad -\left(\frac{\mathbf{1}_n^\top \partial_{i,k}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) - \left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_{j,k}^2 \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right) \\ &\qquad \left.+2\left(\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_j \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\left(\frac{\mathbf{1}_n^\top \partial_k \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n}\right)\right]. \end{aligned} \tag{57}$$

Then, we can show that

$$\begin{aligned}\mathbb{E}_{\vartheta_0}^n\left[\left|\partial_i\mu_n(\xi)\right|^q\right] \le& 2^q\left(\mathbb{E}_{\vartheta_0}^n\left[\left|\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\theta^X)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}\right|^q\right] + \mathbb{E}_{\vartheta_0}^n\left[\left|\mu_n(\xi)-\mu_0\right|^q\right]\left|\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}\right|^q\right) \\ =& 2^q\left(\left|\frac{(\mathbf{1}_n^\top \partial_i \Sigma_n(s_\theta^X)^{-1}\Sigma_n(s_{\vartheta_0}^X)\partial_i\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)^{\frac{1}{2}}}{\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}\right|^q + \mathbb{E}_{\vartheta_0}^n\left[\left|\mu_n(\xi)-\mu_0\right|^q\right]\left|\frac{\mathbf{1}_n^\top \partial_i \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}{\mathbf{1}_n^\top \Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n}\right|^q\right)\end{aligned}$$

so that we conclude the second assertion of Lemma 2 using Lemma 1 and the first assertion of Lemma 2. Moreover, using Lemma 1, we can also show that there exists a positive constant C_q such that for any $\theta \in \Theta$ and $\varepsilon > 0$,

$$\mathbb{E}^n_{\vartheta_0}\left[\left|\partial^2_{i,j}\mu_n(\xi)\right|^q\right] \le C_q\left(\mathbb{E}^n_{\vartheta_0}\left[\left|\frac{\mathbf{1}_n^\top \partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\right|^q\right] + \mathbb{E}^n_{\vartheta_0}\left[\left|\frac{\mathbf{1}_n^\top \partial_i\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\right|^q\right] n^\varepsilon\right)$$
$$+ C_q n^\varepsilon\left(\mathbb{E}^n_{\vartheta_0}\left[\left|\partial_j\mu_n(\xi)\right|^q\right] + \mathbb{E}^n_{\vartheta_0}\left[\left|\mu_n(\xi)-\mu_0\right|^q\right]\right),$$

$$\mathbb{E}^n_{\vartheta_0}\left[\left|\partial^3_{i,j,k}\mu_n(\xi)\right|^q\right]$$
$$\le C_q \max_{i,j,k\in\{1,\dots,p\}}\left(\mathbb{E}^n_{\vartheta_0}\left[\left|\frac{\mathbf{1}_n^\top \partial^3_{i,j,k}\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\right|^q\right] + \mathbb{E}^n_{\vartheta_0}\left[\left|\frac{\mathbf{1}_n^\top \partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\right|^q\right] n^\varepsilon + \mathbb{E}^n_{\vartheta_0}\left[\left|\frac{\mathbf{1}_n^\top \partial_i\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n}\right|^q\right] n^\varepsilon\right)$$
$$+ C_q n^\varepsilon \max_{i,j\in\{1,\dots,p\}}\left(\mathbb{E}^n_{\vartheta_0}\left[\left|\partial^2_{i,j}\mu_n(\xi)\right|^q\right] + \mathbb{E}^n_{\vartheta_0}\left[\left|\partial_i\mu_n(\xi)\right|^q\right] + \mathbb{E}^n_{\vartheta_0}\left[\left|\mu_n(\xi)-\mu_0\right|^q\right]\right).$$

Since we can compute the absolute moments of Gaussian random variables by

$$\mathbb{E}^n_{\vartheta_0}\left[\left|\mathbf{1}_n^\top \partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n\right|^q\right] = 2^{\frac{q}{2}}\Gamma\left(\frac{q+1}{2}\right)\pi^{-\frac{1}{2}}(\mathbf{1}_n^\top \partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}\Sigma_n(s^X_{\theta_0})\partial^2_{i,j}\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n)^{\frac{q}{2}},$$
$$\mathbb{E}^n_{\vartheta_0}\left[\left|\mathbf{1}_n^\top \partial^3_{i,j,k}\Sigma_n(s^X_\xi)^{-1}\mathbf{Z}_n\right|^q\right] = 2^{\frac{q}{2}}\Gamma\left(\frac{q+1}{2}\right)\pi^{-\frac{1}{2}}(\mathbf{1}_n^\top \partial^3_{i,j,k}\Sigma_n(s^X_\xi)^{-1}\Sigma_n(s^X_{\theta_0})\partial^3_{i,j,k}\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n)^{\frac{q}{2}},$$

we can also conclude the third and fourth assertions of Lemma 2 using Lemma 1 and the other assertion of Lemma 2 similarly. Therefore, the proof is complete. *Q.E.D.*

Proof of Lemma 5: Recall that

$$\bar{\sigma}^2_n(\xi) = \frac{1}{n}\left(\mathbf{X}_n - \mu_n(\xi)\mathbf{1}_n\right)^\top \Sigma_n(s^X_\xi)^{-1}\left(\mathbf{X}_n - \mu_n(\xi)\mathbf{1}_n\right), \quad \mu_n(\xi) = \frac{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{X}_n}{\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n},$$

$e_{n,1}(\xi) := \bar{\sigma}^2_n(\xi) - \tilde{\sigma}^2_n(\xi)$, $e_{n,2}(\xi) := \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] - \sigma^2(\xi)$, $e_{n,3}(\xi) := \tilde{\sigma}^2_n(\xi) - \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)]$, and

$$\partial_\xi e_{n,1}(\xi) = -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_\xi\mu_n(\xi)\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n - n^{-1}(\mu_n(\xi)-\mu_0)^2\mathbf{1}_n^\top\partial_\xi\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n,$$

see (62). In the rest of the proof, we will show the error terms $e_{n,i}(\xi)$, $i = 1,2,3$ are negligible uniformly in $\xi \in \Theta_\xi(\iota)$ as $n \to \infty$.

First, we evaluate the first term $e_{n,1}(\xi)$. Notice that we can show

$$
\begin{aligned}
\partial^2_{i,j} e_{n,1}(\xi) = & -2n^{-1}\partial_j\mu_n(\xi)\partial_i\mu_n(\xi)(\mathbf{1}_n^\top\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n) - 2n^{-1}(\mu_n(\xi)-\mu_0)\partial^2_{i,j}\mu_n(\xi)(\mathbf{1}_n^\top\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_i\mu_n(\xi)(\mathbf{1}_n^\top\partial_j\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_j\mu_n(\xi)(\mathbf{1}_n^\top\partial_i\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n) - n^{-1}(\mu_n(\xi)-\mu_0)^2\mathbf{1}_n^\top\partial^2_{i,j}\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n,\\
\partial^3_{i,j,k} e_{n,1}(\xi) = & -2n^{-1}\Big(\partial^2_{j,k}\mu_n(\xi)\partial_i\mu_n(\xi)+\partial_j\mu_n(\xi)\partial^2_{i,k}\mu_n(\xi)\Big)\mathbf{1}_n^\top\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n\\
& -2n^{-1}\partial_j\mu_n(\xi)\partial_i\mu_n(\xi)(\mathbf{1}_n^\top\partial_k\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}\Big\{\partial_k\mu_n(\xi)\partial^2_{i,j}\mu_n(\xi)+(\mu_n(\xi)-\mu_0)\partial^3_{i,j,k}\mu_n(\xi)\Big\}\mathbf{1}_n^\top\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial^2_{i,j}\mu_n(\xi)(\mathbf{1}_n^\top\partial_k\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}\Big\{\partial_k\mu_n(\xi)\partial_i\mu_n(\xi)+(\mu_n(\xi)-\mu_0)\partial^2_{i,k}\mu_n(\xi)\Big\}(\mathbf{1}_n^\top\partial_j\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_i\mu_n(\xi)(\mathbf{1}_n^\top\partial^2_{j,k}\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}\Big\{\partial_k\mu_n(\xi)\partial_j\mu_n(\xi)+(\mu_n(\xi)-\mu_0)\partial^2_{j,k}\mu_n(\xi)\Big\}(\mathbf{1}_n^\top\partial_i\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_j\mu_n(\xi)(\mathbf{1}_n^\top\partial^2_{i,k}\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -2n^{-1}(\mu_n(\xi)-\mu_0)\partial_k\mu_n(\xi)(\mathbf{1}_n^\top\partial^2_{i,j}\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n)\\
& -n^{-1}(\mu_n(\xi)-\mu_0)^2\mathbf{1}_n^\top\partial^3_{i,j,k}\Sigma_n(s_\theta^X)^{-1}\mathbf{1}_n.
\end{aligned}
$$

Similar calculations to (67) using the Sobolev inequality in Lemma 3 and the Fubini Theorem yield that for each $q>p-1$ and $\iota\in(0,1)$, there exists a positive constant C_q such that for any $i,j\in\{1,2,\dots,p-1\}$,

$$
\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\partial^2_{i,j}e_{n,1}(\xi)|^q\right]\le C_q\sup_{\xi\in\Theta_\xi(\iota)}\left(\mathbb{E}^n_{\vartheta_0}\left[|\partial^2_{i,j}e_{n,1}(\xi)|^q\right]+\sum_{k=1}^p\mathbb{E}^n_{\vartheta_0}\left[|\partial^3_{i,j,k}e_{n,1}(\xi)|^q\right]\right). \quad (58)
$$

Using the above expressions of $\partial^2_{i,j}e_{n,1}(\xi)$ and $\partial^3_{i,j,k}e_{n,1}(\xi)$, Lemmas 1 and 2 and the Cauchy-Schwarz inequality, the RHS of (58) goes to zero as $n\to\infty$ so that we conclude $e_{n,1}(\xi)$ is negligible uniformly in $\xi\in\Theta_\xi(\iota)$ as $n\to\infty$.

Next, we evaluate the second term $e_{n,2}(\xi)$. Since $\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)]=\sigma_0^2n^{-1}\mathrm{Tr}[\Sigma_n(s^X_{\xi_0})\Sigma_n(s^X_\xi)^{-1}]$, we can show

$$
\partial_i\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = -\frac{\sigma_0^2}{n}\mathrm{Tr}[\Sigma_n(s^X_{\xi_0})\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}],
$$

$$\partial^2_{i,j}\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = \frac{2\sigma_0^2}{n}\mathrm{Tr}[\Sigma_n(s^X_{\xi_0})\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}]$$
$$-\frac{\sigma_0^2}{n}\mathrm{Tr}[\Sigma_n(s^X_{\xi_0})\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}],$$

where we used the facts that $\mathrm{Tr}[A] = \mathrm{Tr}[A^\top]$ and $\mathrm{Tr}[AB] = \mathrm{Tr}[BA]$ hold for any square matrices A and B in the last inequality, so that, using the expressions of $\partial_i\sigma^2(\xi)$ and $\partial^2_{i,j}\sigma^2(\xi)$ in Lemma 5 and Lemma 4, we conclude $\partial_i e_{n,2}(\xi)$ and $\partial^2_{i,j}e_{n,2}(\xi)$ vanish uniformly on $\Theta_\xi(\iota)$ as $n \to \infty$.

Finally, we consider the third term $e_{n,3}(\xi)$. The Sobolev inequality in Lemma 3 yields that for any $\iota \in (0,1)$ and $2q > p-1$, there exists a positive constant C_{2q} such that for any $i,j,k = 1,\dots,p-1$,

$$\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\partial_i e_{n,3}(\xi)|^{2q}\right] \le C_{2q}\sup_{\xi\in\Theta_\xi(\iota)}\left(\mathbb{E}^n_{\vartheta_0}\left[|\partial_i e_{n,3}(\xi)|^{2q}\right] + \sum_{j=1}^{p-1}\mathbb{E}^n_{\vartheta_0}\left[|\partial^2_{i,j}e_{n,3}(\xi)|^{2q}\right]\right), \tag{59}$$

$$\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\partial^2_{i,j}e_{n,3}(\xi)|^{2q}\right] \le C_{2q}\sup_{\xi\in\Theta_\xi(\iota)}\left(\mathbb{E}^n_{\vartheta_0}\left[|\partial^2_{i,j}e_{n,3}(\xi)|^{2q}\right] + \sum_{j=1}^{p-1}\mathbb{E}^n_{\vartheta_0}\left[|\partial^3_{i,j,k}e_{n,3}(\xi)|^{2q}\right]\right). \tag{60}$$

Notice that we can show

$$\partial_i\tilde{\sigma}^2_n(\xi) = -\frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$

$$\partial^2_{i,j}\tilde{\sigma}^2_n(\xi) = \frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$-\frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n),$$

$$\partial^3_{i,j,k}\tilde{\sigma}^2_n(\xi) = -\frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$+\frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$-\frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$+\frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{j,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$-\frac{2}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_i s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_j s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$+\frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$-\frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^3_{i,j,k}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n)$$
$$+\frac{1}{n}(\mathbf{X}_n - \mu_0\mathbf{1}_n)^\top\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial^2_{i,j}s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_k s^X_\xi)\Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n),$$

which implies

$$\partial_i \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = \mathbb{E}^n_{\vartheta_0}[\partial_i \tilde{\sigma}^2_n(\xi)],\ \partial^2_{i,j}\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = \mathbb{E}^n_{\vartheta_0}[\partial^2_{i,j}\tilde{\sigma}^2_n(\xi)] \text{ and } \partial^3_{i,j,k}\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = \mathbb{E}^n_{\vartheta_0}[\partial^3_{i,j,k}\tilde{\sigma}^2_n(\xi)].$$

Then we can show that the quantities in the RHS of the inequalities (59) and (60) vanish as $n \to \infty$ using Lemma 4 similarly to the proof of (64) in the proof of (30). Therefore, we finish the proof of Lemma 5. *Q.E.D.*

B.3. *Proof of* (30)

Fix $\iota \in (0,1)$. Recall that

$$\tilde{\sigma}^2_n(\xi) := \sigma^2_n(\xi, \mu_0) = \frac{1}{n}(\mathbf{X}_n - \mu_0 \mathbf{1}_n)^\top \Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n).$$

We decompose the error $\bar{\sigma}^2_n(\xi) - \sigma^2(\xi)$ into the following three terms:

$$e_{n,1}(\xi) := \bar{\sigma}^2_n(\xi) - \tilde{\sigma}^2_n(\xi),\ \ e_{n,2}(\xi) := \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] - \sigma^2(\xi),\ \text{ and }\ e_{n,3}(\xi) := \tilde{\sigma}^2_n(\xi) - \mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)].$$

In the rest of the proof, we will show that the error terms $e_{n,i}(\xi)$, $i = 1,2,3$, vanish uniformly on $\Theta_\xi(\iota)$ as $n \to \infty$.

We first consider $e_{n,1}(\xi)$. Note that the error term $e_{n,1}(\xi)$ is written as

$$\begin{aligned} e_{n,1}(\xi) &= -2n^{-1}(\mu_n(\xi) - \mu_0)\mathbf{1}_n^\top \Sigma_n(s^X_\xi)^{-1}(\mathbf{X}_n - \mu_0\mathbf{1}_n) + n^{-1}(\mu_n(\xi) - \mu_0)^2 \mathbf{1}_n^\top \Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n \\ &= -n^{-1}(\mu_n(\xi) - \mu_0)^2 \mathbf{1}_n^\top \Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n, \end{aligned} \tag{61}$$

so that, by the chain rule, its first-order derivatives with respect to ξ is

$$\partial_\xi e_{n,1}(\xi) = -2n^{-1}(\mu_n(\xi) - \mu_0)\partial_\xi \mu_n(\xi)\mathbf{1}_n^\top \Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n - n^{-1}(\mu_n(\xi) - \mu_0)^2 \mathbf{1}_n^\top \partial_\xi \Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n. \tag{62}$$

Moreover, the Sobolev inequality in Lemma 3 yields that for any $\iota \in (0,1)$ and $2q > p-1$, there exists a positive constant C_{2q} such that

$$\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi \in \Theta_\xi(\iota)} |e_{n,1}(\xi)|^{2q}\right] \le C_{2q} \sup_{\xi \in \Theta_\xi(\iota)} \left(\mathbb{E}^n_{\vartheta_0}\left[|e_{n,1}(\xi)|^{2q}\right] + \sum_{j=1}^{p-1} \mathbb{E}^n_{\vartheta_0}\left[|\partial_j e_{n,1}(\xi)|^{2q}\right]\right). \tag{63}$$

Then, using the above expressions of $e_{n,1}(\xi)$ and $\partial_\xi e_{n,1}(\xi)$, Lemmas 1 and 2 and the Cauchy-Schwarz inequality, we can show that the quantity in the RHS of the inequality (63) converges to zero as $n \to \infty$. Hence, we conclude $e_{n,1}(\xi)$ vanishes uniformly on $\Theta_\xi(\iota)$ as $n \to \infty$.

Next, we consider $e_{n,2}(\xi)$. Notice that we have the equality $\mathbb{E}^n_{\vartheta_0}[\tilde{\sigma}^2_n(\xi)] = \sigma_0^2 n^{-1}\mathrm{Tr}[\Sigma_n(s^X_{\xi_0})\Sigma_n(s^X_\xi)^{-1}]$ so that, using Lemma 4, we conclude $e_{n,2}(\xi)$ vanishes uniformly on $\Theta_\xi(\iota)$ as $n \to \infty$.

Finally, we consider $e_{n,3}(\xi)$. The Sobolev inequality in Lemma 3 yields that for any $\iota \in (0,1)$ and $2q > p-1$, there exists a positive constant C_{2q} such that

$$\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)} |e_{n,3}(\xi)|^{2q}\right] \le C_{2q} \sup_{\xi\in\Theta_\xi(\iota)}\left(\mathbb{E}^n_{\vartheta_0}\left[|e_{n,3}(\xi)|^{2q}\right] + \sum_{j=1}^{p-1}\mathbb{E}^n_{\vartheta_0}\left[|\partial_j e_{n,3}(\xi)|^{2q}\right]\right). \tag{64}$$

Then we can show that the quantity in the RHS of the inequality (64) vanishes as $n \to \infty$ using Lemma 4 because we know that 1) the moments $\mathbb{E}^n_{\vartheta_0}\left[|e_{n,3}(\xi)|^{2q}\right]$ and $\mathbb{E}^n_{\vartheta_0}[|\partial_j e_{n,3}(\xi)|^{2q}]$ can be expressed by linear combinations of cumulants up to the order $2q$ using the Leonov-Shiryaev formula, 2) $e_{n,3}(\xi)$ and $\partial_j e_{n,3}(\xi)$ are centralized quadratic forms of Gaussian vector so that for each $r \ge 2$, its rth order cumulants can be expressed by $\mathrm{cum}_r[e_{n,3}(\xi)] = n^{-r}c_r\mathrm{Tr}[(\Sigma_n(s^X_\xi)^{-1}\Sigma_n(s^X_{\xi_0}))^r]$ and $\mathrm{cum}_r[\partial_j e_{n,3}(\xi)] = n^{-r}c_r\mathrm{Tr}[(\Sigma_n(s^X_\xi)^{-1}\Sigma_n(\partial_\xi s^X_\xi)\Sigma_n(s^X_\xi)^{-1}\Sigma_n(s^X_{\xi_0}))^r]$ for some positive constant c_r, and 3) the first order cumulants of $e_{n,3}(\xi)$ and $\partial_j e_{n,3}(\xi)$ are equal to zero. Therefore, we conclude $e_{n,3}(\xi)$ also vanishes uniformly on $\Theta_\xi(\iota)$ as $n \to \infty$. This completes the proof of (30).

B.4. *Proof of* (31)

Using the Taylor theorem, we can write

$$\log\bar{\sigma}^2_n(\xi) - \log\sigma^2(\xi) = (\bar{\sigma}^2_n(\xi) - \sigma^2(\xi))\int_0^1\left(\sigma^2(\xi) + u(\bar{\sigma}^2_n(\xi) - \sigma^2(\xi))\right)^{-1}\mathrm{d}u$$

so that we obtain

$$|\log\bar{\sigma}^2_n(\xi) - \log\sigma^2(\xi)| \le |\bar{\sigma}^2_n(\xi) - \sigma^2(\xi)|\left|\sigma^2(\xi) - |\bar{\sigma}^2_n(\xi) - \sigma^2(\xi)|\right|^{-1}.$$

Set $\sigma^2(\xi_*(\iota)) := \inf_{\xi\in\Theta_\xi(\iota)}\sigma^2(\xi) > 0$ so that we can take $\epsilon_1 \in (0, \sigma^2(\xi_*(\iota)))$. Then, for any $\epsilon > 0$, we can show the inequality

$$\begin{aligned}\mathbb{P}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\log\bar{\sigma}^2_n(\xi)-\log\sigma^2(\xi)|>\epsilon\right] \leq&\, \mathbb{P}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\bar{\sigma}^2_n(\xi)-\sigma^2(\xi)|>\epsilon_1\right]\\ &+\mathbb{P}^n_{\vartheta_0}\left[\sup_{\xi\in\Theta_\xi(\iota)}|\bar{\sigma}^2_n(\xi)-\sigma^2(\xi)|>(\sigma^2(\xi_*(\iota))-\epsilon_1)\epsilon\right],\end{aligned}$$

which concludes (31) using (30).

B.5. *Proof of Proposition 1*

Under Assumption 3, Theorems 4.1 and 5.2 in Adenstedt (1974) yield the convergence

$$n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\xi_0)-\mu_0)\to\mathcal{N}\left(0,\frac{2\pi\sigma_0^2 c_X(\xi_0)\Gamma(1-\alpha_X(\xi_0))}{B(1-\alpha_X(\xi_0)/2,1-\alpha_X(\xi_0)/2)}\right)$$

in law under the distribution $\mathbb{P}^n_{\vartheta_0}$ as $n\to\infty$ for any interior point ϑ_0 of Θ. Thus it suffices to prove that

$$n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\widehat{\xi}_n)-\mu_n(\xi_0)) = o_{\mathbb{P}^n_{\vartheta_0}}(1) \ \text{ as } n\to\infty. \tag{65}$$

Notice that the sequence $\{\sqrt{n}(\hat{\xi}_n-\xi_0)\}_{n\in\mathbb{N}}$ is stochastically bounded from the assumption. Set $\overline{\xi}_n := \sqrt{n}(\hat{\xi}_n-\xi_0)$ for $n\in\mathbb{N}$. Then, using the Cauchy-Schwarz inequality and the Chebyshev inequality, for any $\varepsilon>0$ and $M>0$, we obtain

$$\begin{aligned}&\mathbb{P}^n_{\vartheta_0}\left[\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\widehat{\xi}_n)-\mu_n(\xi_0))\right|\geq\varepsilon\right]\\ &\leq \mathbb{P}^n_{\vartheta_0}[\|\overline{\xi}_n\|_{\mathbb{R}^{p-1}}\geq M]+\mathbb{P}^n_{\vartheta_0}\left[\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\xi)-\mu_n(\xi_0))\right|\geq\varepsilon\right]\\ &\leq \mathbb{P}^n_{\vartheta_0}[\|\overline{\xi}_n\|_{\mathbb{R}^{p-1}}\geq M]+\varepsilon^{-q}\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\xi)-\mu_n(\xi_0))\right|^q\right]\end{aligned}$$

so that (65) follows once we have proved that for any $M>0$ and $q>p-1$,

$$\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\xi)-\mu_n(\xi_0))\right|^q\right]=o(1)\ \text{ as } n\to\infty. \tag{66}$$

Notice that Lemma 3 and the Fubini theorem yield that for any $q>p-1$, it holds

$$\begin{aligned}
&\mathbb{E}^n_{\vartheta_0}\left[\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}(\mu_n(\xi)-\mu_n(\xi_0))\right|^q\right]\\
&\le C_{q,1}(n^{-1/2}M)^{q-1}\mathbb{E}^n_{\vartheta_0}\left[\left(\sum_{j=1}^{p-1}\left\|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}\partial_j\mu_n(\cdot)\right\|_{L^q(B_{n^{-1/2}M}(\xi_0))}\right)^q\right]\\
&\le C_{q,2}(n^{-1/2}M)^q\sum_{j=1}^{p-1}\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\mathbb{E}^n_{\vartheta_0}\left[\left|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}\partial_j\mu_n(\xi)\right|^q\right],
\end{aligned} \tag{67}$$

and Lemma 2 gives the inequality

$$\begin{aligned}
&n^{-\frac{q}{2}}\sum_{j=1}^{p-1}\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}\mathbb{E}^n_{\vartheta_0}\left[|n^{\frac{1}{2}(1-\alpha_X(\xi_0))}\partial_j\mu_n(\xi)|^q\right]\\
&\le C_{q,3}n^{-\frac{q}{2}}\sum_{j=1}^{p-1}\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}n^{\frac{q}{2}(1-\alpha_X(\xi_0))}(\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n)^{-\frac{q}{2}}n^{\frac{q}{2}\chi_0(\xi)+\varepsilon}.
\end{aligned} \tag{68}$$

Since we can show the inequality

$$\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n=\left\|\Sigma_n(s^X_\xi)^{-\frac{1}{2}}\Sigma_n(s^X_{\xi_0})^{\frac{1}{2}}\Sigma_n(s^X_{\xi_0})^{-\frac{1}{2}}\mathbf{1}_n\right\|^2_{\mathbb{R}^n}\le\left\|\Sigma_n(s^X_\xi)^{-\frac{1}{2}}\Sigma_n(s^X_{\xi_0})^{\frac{1}{2}}\right\|^2_{\mathrm{op}}\left\|\Sigma_n(s^X_{\xi_0})^{-\frac{1}{2}}\mathbf{1}_n\right\|^2_{\mathbb{R}^n},$$

which follows from the definition of the operator norm, we can further evaluate the last quantity in the inequality (68) up to a constant multiplication by

$$\begin{aligned}
&n^{-\frac{q}{2}}\sum_{j=1}^{p-1}\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}n^{\frac{q}{2}(1-\alpha_X(\xi_0))}(\mathbf{1}_n^\top\Sigma_n(s^X_\xi)^{-1}\mathbf{1}_n)^{-\frac{q}{2}}n^{\frac{q}{2}\chi_0(\xi)+\varepsilon}\\
&\le n^{-\frac{q}{2}}\sum_{j=1}^{p-1}\sup_{\xi\in B_{n^{-1/2}M}(\xi_0)}n^{\frac{q}{2}(1-\alpha_X(\xi_0))}(\mathbf{1}_n^\top\Sigma_n(s^X_{\xi_0})^{-1}\mathbf{1}_n)^{-\frac{q}{2}}n^{\frac{3q}{2}\chi_0(\xi)+\varepsilon},
\end{aligned}$$

where we use Assumption 1 and Lemma 3 of Chong and Mies (2025) in the last inequality. Then we conclude (66) using Theorems 4.1 and 5.2 in Adenstedt (1974). Therefore, the proof is complete.

B.6. *Proof of* (51)

Notice that we have $R_n(\vartheta,\vartheta') = \mathrm{diag}(I_p, n^{-\frac{1}{2}\{\alpha_X(\xi+\xi')-\alpha_X(\xi)\}})$ for any $\vartheta' = (\xi', \sigma', \mu')^\top$, and

$$\|\Phi_n(\vartheta)\|_{\mathrm{op}} = \max\{n^{-\frac{1}{2}}, n^{-\frac{1}{2}(1-\alpha_X(\xi))}\} = n^{-\frac{1}{2}\min\{1,1-\alpha_X(\xi)\}},$$

so that we conclude (51) using the continuous differentiability of $\alpha_X(\xi)$.

B.7. *Proof of* (52)

Recall that $\sigma_n^2(\xi,\mu) = \frac{1}{n}(\mathbf{X}_n - \mu\mathbf{1}_n)^\top \Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n - \mu\mathbf{1}_n)$ and $\tilde\sigma_n^2(\xi) = \sigma_n^2(\xi,\mu_0)$. Using the expressions of the score function in (46), the second-order derivatives of the log-likelihood function $\ell_n(\vartheta)$ are given by

$$\begin{cases} \partial_\xi^2 \ell_n(\vartheta) = \frac{n}{2\sigma^2}\left(\partial_\xi^2\sigma_n^2(\xi,\mu) - \frac{\sigma^2}{n}\mathrm{Tr}\left[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi^2 s_\xi)\right]\right) + \frac{1}{2}\mathrm{Tr}\left[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\right], \\ \partial_\xi\partial_\sigma \ell_n(\vartheta) = \partial_\sigma\partial_\xi \ell_n(\vartheta) = \frac{n}{\sigma^3}\partial_\xi\sigma_n^2(\xi,\mu), \ \ \partial_\sigma^2\ell_n(\vartheta) = -\frac{n}{\sigma^4}\left(3\sigma_n^2(\xi,\mu) - \sigma^2\right), \\ \partial_\xi\partial_\mu \ell_n(\vartheta) = \partial_\mu\partial_\xi \ell_n(\vartheta) = -\frac{1}{\sigma^2}\mathbf{1}_n^\top \partial_\xi\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n - \mu\mathbf{1}_n), \ \ \partial_\mu^2\ell_n(\vartheta) = -\frac{1}{\sigma^2}(\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n), \\ \partial_\sigma\partial_\mu \ell_n(\vartheta) = \partial_\mu\partial_\sigma \ell_n(\vartheta) = -\frac{2}{\sigma^3}\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n - \mu\mathbf{1}_n). \end{cases} \tag{69}$$

Then we can show (52) similarly to the proof of Lemma 2.6 in Cohen et al. (2013) and Lemma 3.4 in Kawai (2013) using Lemma 4. So we omit the detailed proofs.

B.8. *Proof of* (53)

Using the Taylor theorem, it suffices to prove that

$$n^{-\frac{1}{2}(r_1+r_2)-\frac{1}{2}(1-\alpha_X(\xi))r_3} \sup_{u\in\mathbb{U}_{n,c}(\vartheta)}\left|\partial_{\xi_i}^{r_1}\partial_\sigma^{r_2}\partial_\mu^{r_3}\ell_n(\vartheta+\Phi_n(\vartheta)u)\right| = o_{\mathbb{P}_\vartheta^n}(1) \ \text{ as } n\to\infty \tag{70}$$

for any $c>0$, $i = 1,\ldots,p-1$ and $r_1,r_2,r_3 \in \{1,2,3\}$ with $r_1+r_2+r_3 = 3$. Using the expressions of the second order derivatives of $\ell_n(\vartheta)$ in (69), we get the expressions of the third-order derivatives of $\ell_n(\vartheta)$ by

$$\partial_\xi^3\ell_n(\vartheta) = \frac{n}{2\sigma^2}\partial_\xi^3\sigma_n^2(\xi,\mu) - \frac{1}{2}\mathrm{Tr}\left[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi^3 s_\xi)\right] + \frac{1}{2}\mathrm{Tr}\left[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi^2 s_\xi)\right]$$

$$+\frac{1}{2}\partial_\xi \mathrm{Tr}\Big[\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\Sigma_n(s_\xi)^{-1}\Sigma_n(\partial_\xi s_\xi)\Big],$$

and

$$\partial_\xi^2\partial_\sigma \ell_n(\vartheta) = \frac{n}{\sigma^3}\partial_\xi^2 \sigma_n^2(\xi,\mu),\;\; \partial_\xi^2\partial_\mu \ell_n(\vartheta) = -\frac{1}{\sigma^2}\mathbf{1}_n^\top \partial_\xi^2\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n),$$
$$\partial_\sigma^2\partial_\xi \ell_n(\vartheta) = -\frac{3n}{\sigma^4}\partial_\xi \sigma_n^2(\xi,\mu),\;\; \partial_\sigma^2\partial_\mu \ell_n(\vartheta) = \frac{2}{\sigma^3}(\mathbf{1}_n^\top\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n),$$
$$\partial_\mu^2\partial_\xi \ell_n(\vartheta) = -\frac{1}{\sigma^2}(\mathbf{1}_n^\top \partial_\xi\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n),\;\; \partial_\mu^2\partial_\sigma \ell_n(\vartheta) = \frac{2}{\sigma^3}(\mathbf{1}_n^\top \partial_\xi\Sigma_n(s_\xi^X)^{-1}\mathbf{1}_n),$$
$$\partial_\xi\partial_\sigma\partial_\mu \ell_n(\vartheta) = -\frac{2}{\sigma^3}\mathbf{1}_n^\top\partial_\xi\Sigma_n(s_\xi^X)^{-1}(\mathbf{X}_n-\mu\mathbf{1}_n),\;\; \partial_\sigma^3\ell_n(\vartheta) = \frac{2n}{\sigma^5}\left(6\sigma_n^2(\xi,\mu)-\sigma^3\right),\;\; \partial_\mu^3\ell_n(\vartheta) = 0,$$

so that (70) follows from similar arguments to the proofs of Lemma 2.7 in Cohen et al. (2013) using Lemmas 1 and 4 as well as Theorems 4.1 and 5.2 in Adenstedt (1974). This completes the proof of Theorem 3 as well as that of (53).

B.9. *Alternative expression of MLE*

In this section, we derive an easily tractable alternative expression of the log-likelihood function $\ell_n(\vartheta)$, which is useful to quickly compute the exact MLE. Denote by $p_\vartheta(x_1,\dots,x_n)$ the Gaussian likelihood function of the distribution $\mathbb{P}_\vartheta^n$ and by $p_\vartheta(x_j|x_1,\dots,x_{j-1})$ the conditional likelihood function of the distribution of X_j^ϑ conditional on the j-dimensional vector $(X_1^\vartheta,X_2^\vartheta,\dots,X_{j-1}^\vartheta)$. By expressing the likelihood function $p_\vartheta(x_1,\dots,x_n)^\top$ of the joint distribution as a product of the conditional likelihood functions $p_\vartheta(x_j|x_1,\dots,x_{j-1})$ and using the closed-form expression of their conditional Gaussian likelihood functions $p_\vartheta(x_j|x_1,\dots,x_{j-1})$, the log-likelihood function $\ell_n(\vartheta)$ can be expressed by

$$\ell_n(\vartheta) = \log p_\vartheta(X_1,...,X_n) = \log p_\vartheta(X_1)\prod_{j=2}^n \log p_\vartheta(X_j|X_1,...,X_{j-1}) = -\frac{1}{2}\sum_{j=1}^n \log v_j(\theta) - \frac{1}{2}\sum_{j=1}^n \frac{(X_j-\eta_j(\vartheta))^2}{v_j(\theta)},$$

where $\eta_1(\vartheta) := \mu$, $v_1(\theta) := \mathrm{Var}[X_1^\vartheta] = \sigma^2$, $\eta_j(\vartheta) := \mathbb{E}[X_j^\vartheta|\mathbf{X}_{j-1}]$ and $v_j(\theta) := \mathrm{Var}[X_j^\vartheta|\mathbf{X}_{j-1}]$ for $j\in\{2,3,\dots,n\}$, which can be written as

$$\eta_j(\vartheta) = \sum_{i=1}^{j-1}\phi_{j,i}(\xi)X_{j-i} + w_j(\xi)\mu \;\;\text{and}\;\; v_j(\theta) = \gamma_\theta^X(0)\Pi_{i=1}^{j-1}\left(1-\phi_{i,i}(\xi)^2\right),$$

where $w_1(\xi) := 1$, $w_j(\xi) := \left(1 - \sum_{i=1}^{j-1} \phi_{j,i}(\xi)\right)$, and

$$\phi_{1,1}(\xi) := \gamma_\theta^X(1)/\gamma_\theta^X(0), \phi_{j,j}(\xi) := \left[\gamma_\theta^X(j) - \sum_{i=1}^{j-1} \phi_{j-1,i}(\xi)\gamma_\theta^X(j-i)\right] v_{j-1}(\theta)^{-1}, \phi_{j,i}(\xi) := \phi_{j-1,i}(\xi) - \phi_{j,j}(\xi)\phi_{j-1,j-i}(\xi)$$

for $i \in \{1, \dots, j-1\}$ and $j \in \{2, \dots, n\}$. Here, $\phi_{j,i}(\xi)$ are partial linear regression coefficients. Notice that the above expression of $\ell_n(\vartheta)$ can be rewritten as

$$\ell_n(\vartheta) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\sum_{j=1}^{n}\log\bar{v}_j(\xi) - \frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{(Z_j(\xi) - w_j(\xi)\mu)^2}{\bar{v}_j(\xi)}$$

using the notation $\bar{v}_j(\xi) := \sigma^{-2}v_j(\theta)$ and $Z_1(\xi) := X_1$, $Z_j(\xi) := X_j - \sum_{i=1}^{j-1}\phi_{j,i}(\xi)X_{j-i}$ for $j \in \{2, 3, \dots, n\}$. Then we can see that the MLE also satisfies the estimating equations

$$\mu = \left(\sum_{j=1}^{n}\frac{w_j(\xi)^2}{\bar{v}_j(\xi)}\right)^{-1}\sum_{j=1}^{n}\frac{w_j(\xi)}{\bar{v}_j(\xi)}Z_j(\xi) \text{ and } \sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\frac{(Z_j(\xi) - w_j(\xi)\mu)^2}{\bar{v}_j(\xi)}$$

for any $(\xi, \sigma, \mu) \in \Theta_\xi \times (0, \infty) \times \mathbb{R}$. Therefore, from the uniqueness of the maximum values of σ and μ on $(0, \infty)$ and $\mathbb{R}$ respectively, we obtain the equalities

$$\mu_n(\xi) = \left(\sum_{j=1}^{n}\frac{w_j(\xi)^2}{\bar{v}_j(\xi)}\right)^{-1}\sum_{j=1}^{n}\frac{w_j(\xi)}{\bar{v}_j(\xi)}Z_j(\xi) \text{ and } \sigma_n^2(\xi, \mu) = \frac{1}{n}\sum_{j=1}^{n}\frac{(Z_j(\xi) - w_j(\xi)\mu)^2}{\bar{v}_j(\xi)}.$$

Notice that the coefficients $\phi_{j,i}(\xi)$ and the corresponding conditional variance $\bar{v}_j(\xi)$ can be calculated by the Durbin-Lenvinson recursive algorithm (Brockwell and Davis, 1987, Chapter 5).

B.10. *An Financial Application*

Fractional models have been employed to model realized volatility (RV) (Andersen et al., 2003, Bennedsen et al., 2022, 2024) and trading volume (Shi et al., 2024b, Wang et al., 2024). A recently introduced model is the fOU process (Gatheral et al., 2018, Wang et al., 2023). For instance, Wang et al. (2023) demonstrate that fOU outperforms traditional fractional models such as ARFIMA$(1, d, 0)$ and fBm. In this section, we demonstrate that our exact MLE further enhances the forecasting accuracy of fOU. Compared to the existing literature, we also incorporate the CoF method from Wang et al. (2023). However, since the CoF method is less efficient than the AWML approach proposed by Shi et al. (2024a) and Wang et al. (2024), we anticipate that forecasting accuracy will rank as follows: MLE2 (exact MLE), followed by MLE3 (plug-in MLE), and then CoF.

We apply our method to Dow Jones 30 (DJ30) stocks. The daily realized volatility of the DJ30 stocks, spanning September 15, 2012, to August 28, 2021, is obtained from Risk Lab of Dacheng Xiu.[9] We assume that the log RV follows an fOU process and set $\Delta = 1/250$ to reflect 250 trading days per year. A four-year rolling window is employed to fit the model, estimate the parameters, and generate h-day-ahead forecasts of RV. The results, presented in Tables V-VII, align with our expectations. Both MLE approaches outperform the CoF method, with improvements ranging from approximately 2% to 15%. The exact MLE slightly enhances the performance of the plug-in MLE. It is not surprising that we find the plug-in MLE exhibits good finite-sample performance compared to our exact MLE.

[9]see https://dachxiu.chicagobooth.edu/#risklab.

TABLE V

RMSE of the alternative estimation methods for h-day-ahead forecasts for realized volatility (RV) with a four-year rolling window between September 15, 2012 and August 28, 2021.

		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
AAPL	MLE2	0.0611	0.0715	0.0782	0.0836	0.0877
	MLE3	0.0612	0.0716	0.0783	0.0836	0.0878
	CoF	0.0639	0.0779	0.0874	0.0950	0.1007
ALD	MLE2	0.0543	0.0635	0.0712	0.0770	0.0816
	MLE3	0.0545	0.0638	0.0716	0.0774	0.0820
	CoF	0.0555	0.0661	0.0749	0.0813	0.0863
AMGN	MLE2	0.0691	0.0770	0.0819	0.0871	0.0909
	MLE3	0.0692	0.0771	0.0820	0.0872	0.0909
	CoF	0.0703	0.0796	0.0853	0.0911	0.0954
AXP	MLE2	0.0577	0.0693	0.0774	0.0842	0.0895
	MLE3	0.0580	0.0697	0.0778	0.0847	0.0900
	CoF	0.0605	0.0757	0.0861	0.0944	0.1006
BA	MLE2	0.0922	0.1081	0.1190	0.1287	0.1371
	MLE3	0.0925	0.1084	0.1194	0.1292	0.1376
	CoF	0.0965	0.1192	0.1352	0.1483	0.1587
BEL	MLE2	0.0485	0.0560	0.0616	0.0665	0.0706
	MLE3	0.0486	0.0561	0.0618	0.0667	0.0707
	CoF	0.0497	0.0587	0.0654	0.0708	0.0750
CAT	MLE2	0.0595	0.0675	0.0733	0.0783	0.0822
	MLE3	0.0596	0.0677	0.0735	0.0785	0.0824
	CoF	0.0594	0.0687	0.0754	0.0813	0.0858
CHV	MLE2	0.0525	0.0638	0.0725	0.0789	0.0847
	MLE3	0.0527	0.0642	0.0730	0.0795	0.0853
	CoF	0.0532	0.0650	0.0742	0.0810	0.0868
CRM	MLE2	0.0630	0.0739	0.0807	0.0861	0.0905
	MLE3	0.0631	0.0739	0.0808	0.0861	0.0905
	CoF	0.0663	0.0817	0.0927	0.1015	0.1086
CSCO	MLE2	0.0537	0.0635	0.0714	0.0774	0.0821
	MLE3	0.0538	0.0637	0.0715	0.0775	0.0822
	CoF	0.0554	0.0675	0.0771	0.0842	0.0896

TABLE VI

RMSE of the alternative estimation methods for h-day-ahead forecasts for realized volatility (RV) with four-year rolling window between September 15, 2012 and August 28, 2021.

		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
DIS	MLE2	0.0556	0.0657	0.0739	0.0801	0.0853
	MLE3	0.0558	0.0660	0.0743	0.0805	0.0857
	CoF	0.0579	0.0705	0.0804	0.0876	0.0934
GS	MLE2	0.0513	0.0630	0.0709	0.0772	0.0821
	MLE3	0.0513	0.0631	0.0710	0.0773	0.0822
	CoF	0.0526	0.0660	0.0751	0.0821	0.0873
HD	MLE2	0.0538	0.0628	0.0695	0.0754	0.0804
	MLE3	0.0540	0.0630	0.0697	0.0757	0.0807
	CoF	0.0549	0.0656	0.0736	0.0804	0.0859
IBM	MLE2	0.0499	0.0583	0.0642	0.0691	0.0732
	MLE3	0.0500	0.0585	0.0644	0.0694	0.0734
	CoF	0.0519	0.0632	0.0713	0.0777	0.0826
INTC	MLE2	0.0634	0.0742	0.0823	0.0884	0.0930
	MLE3	0.0635	0.0744	0.0824	0.0885	0.0931
	CoF	0.0648	0.0777	0.0872	0.0943	0.0994
JNJ	MLE2	0.0533	0.0605	0.0665	0.0714	0.0751
	MLE3	0.0534	0.0606	0.0667	0.0716	0.0753
	CoF	0.0546	0.0620	0.0680	0.0729	0.0765
JPM	MLE2	0.0519	0.0639	0.0730	0.0803	0.0858
	MLE3	0.0521	0.0641	0.0732	0.0806	0.0861
	CoF	0.0532	0.0668	0.0768	0.0846	0.0903
KO	MLE2	0.0444	0.0529	0.0597	0.0648	0.0689
	MLE3	0.0446	0.0531	0.0600	0.0650	0.0691
	CoF	0.0458	0.0558	0.0637	0.0693	0.0738
MCD	MLE2	0.0485	0.0583	0.0662	0.0725	0.0776
	MLE3	0.0488	0.0586	0.0666	0.0728	0.0779
	CoF	0.0509	0.0629	0.0722	0.0791	0.0845
MMM	MLE2	0.0492	0.0571	0.0628	0.0675	0.0709
	MLE3	0.0493	0.0572	0.0629	0.0677	0.0711
	CoF	0.0496	0.0592	0.0664	0.0723	0.0766

TABLE VII

RMSE of the alternative estimation methods for h-day-ahead forecasts for realized volatility (RV) with four-year rolling window between September 15, 2012 and August 28, 2021.

		$h=1$	$h=2$	$h=3$	$h=4$	$h=5$
MRK	MLE2	0.0565	0.0648	0.0714	0.0760	0.0797
	MLE3	0.0566	0.0650	0.0716	0.0762	0.0799
	CoF	0.0584	0.0694	0.0776	0.0830	0.0871
MSFT	MLE2	0.0527	0.0631	0.0711	0.0770	0.0819
	MLE3	0.0528	0.0633	0.0713	0.0771	0.0821
	CoF	0.0531	0.0646	0.0737	0.0802	0.0857
NIKE	MLE2	0.0565	0.0671	0.0746	0.0808	0.0854
	MLE3	0.0566	0.0673	0.0749	0.0811	0.0857
	CoF	0.0582	0.0717	0.0810	0.0882	0.0932
PG	MLE2	0.0547	0.0649	0.0737	0.0806	0.0861
	MLE3	0.0549	0.0651	0.0740	0.0809	0.0864
	CoF	0.0560	0.0673	0.0769	0.0841	0.0897
SPC	MLE2	0.0546	0.0643	0.0715	0.0770	0.0818
	MLE3	0.0548	0.0645	0.0718	0.0774	0.0822
	CoF	0.0559	0.0665	0.0743	0.0803	0.0853
UNH	MLE2	0.0570	0.0665	0.0728	0.0786	0.0834
	MLE3	0.0571	0.0667	0.0731	0.0790	0.0838
	CoF	0.0577	0.0682	0.0752	0.0816	0.0869
V	MLE2	0.0468	0.0570	0.0648	0.0709	0.0760
	MLE3	0.0470	0.0572	0.0651	0.0712	0.0763
	CoF	0.0476	0.0586	0.0669	0.0733	0.0786
WAG	MLE2	0.0727	0.0826	0.0886	0.0937	0.0983
	MLE3	0.0727	0.0827	0.0887	0.0938	0.0984
	CoF	0.0757	0.0893	0.0980	0.1047	0.1104
WMT	MLE2	0.0500	0.0580	0.0635	0.0677	0.0716
	MLE3	0.0502	0.0582	0.0637	0.0680	0.0718
	CoF	0.0518	0.0623	0.0692	0.0742	0.0783
XOM	MLE2	0.0520	0.0619	0.0698	0.0763	0.0810
	MLE3	0.0522	0.0622	0.0702	0.0768	0.0816
	CoF	0.0534	0.0655	0.0752	0.0832	0.0891

B.11. *Robustness check:* $\mu \neq 0$

We also carry out additional simulation studies to compare the performance of alternative methods for estimating the $ARFIMA(0,d,0)$ model. The following two tables report the bias and standard error of three ML estimates. Again, the exact MLE of *mu* outperforms the plug-in MLE, especially when d is very negative. However, the exact MLEs of d and σ perform similarly to the plug-in MLEs.

TABLE VIII

BIAS AND STD OF ALTERNATIVE MLES FOR $ARFIMA(0,d,0)$: $\mu = 1$ AND $\sigma = 1$.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
					$n = 250$					
			d = -0.40			d = -0.30			d = -0.20	
μ	Bias	0.0000	0.0002	0.0000	0.0000	0.0007	0.0004	0.0000	0.0003	0.0000
	Std	0.0000	0.0096	0.0115	0.0000	0.0153	0.0165	0.0000	0.0235	0.0241
d	Bias	-0.0021	-0.0135	-0.0074	-0.0045	-0.0175	-0.0140	-0.0056	-0.0185	-0.0170
	Std	0.0488	0.0502	0.0494	0.0507	0.0534	0.0520	0.0519	0.0545	0.0538
σ	Bias	-0.0029	-0.0051	-0.0042	-0.0024	-0.0047	-0.0043	-0.0022	-0.0044	-0.0042
	Std	0.0445	0.0444	0.0444	0.0418	0.0420	0.0420	0.0454	0.0453	0.0453
			d = -0.10			d = 0.00			d = 0.10	
μ	Bias	0.0000	0.0004	0.0003	0.0000	-0.0006	-0.0006	0.0000	-0.0028	-0.0027
	Std	0.0000	0.0380	0.0384	0.0000	0.0617	0.0616	0.0000	0.1082	0.1086
d	Bias	-0.0044	-0.0174	-0.0168	-0.0056	-0.0194	-0.0193	-0.0045	-0.0193	-0.0193
	Std	0.0525	0.0548	0.0544	0.0498	0.0531	0.0529	0.0495	0.0528	0.0528
σ	Bias	-0.0044	-0.0065	-0.0065	-0.0013	-0.0035	-0.0035	-0.0010	-0.0033	-0.0033
	Std	0.0440	0.0441	0.0441	0.0449	0.0450	0.0450	0.0453	0.0451	0.0451
			d = 0.20			d = 0.30			d = 0.40	
μ	Bias	0.0000	0.0029	0.0035	0.0000	-0.0248	-0.0269	0.0000	0.0334	0.0359
	Std	0.0000	0.1948	0.1959	0.0000	0.3532	0.3551	0.0000	0.6428	0.6440
d	Bias	-0.0031	-0.0193	-0.0193	-0.0083	-0.0255	-0.0254	-0.0135	-0.0293	-0.0291
	Std	0.0504	0.0536	0.0537	0.0492	0.0532	0.0532	0.0425	0.0474	0.0474
σ	Bias	-0.0045	-0.0067	-0.0067	-0.0050	-0.0071	-0.0071	-0.0032	-0.0048	-0.0048
	Std	0.0443	0.0445	0.0445	0.0437	0.0438	0.0438	0.0432	0.0433	0.0433
					$n = 1000$					
			d = -0.40			d = -0.30			d = -0.20	
μ	Bias	0.0000	-0.0000	-0.0001	0.0000	0.0001	0.0003	0.0000	0.0004	0.0004
	Std	0.0000	0.0029	0.0035	0.0000	0.0051	0.0054	0.0000	0.0091	0.0093
d	Bias	-0.0014	-0.0052	-0.0029	-0.0001	-0.0040	-0.0030	-0.0010	-0.0048	-0.0045
	Std	0.0249	0.0253	0.0252	0.0249	0.0254	0.0252	0.0246	0.0249	0.0248
σ	Bias	-0.0008	-0.0013	-0.0011	-0.0016	-0.0021	-0.0020	0.0000	-0.0005	-0.0005
	Std	0.0222	0.0222	0.0223	0.0222	0.0222	0.0222	0.0220	0.0221	0.0221
			d = -0.10			d = 0.00			d = 0.10	
μ	Bias	0.0000	0.0001	0.0001	0.0000	-0.0009	-0.0010	0.0000	-0.0026	-0.0025
	Std	0.0000	0.0164	0.0164	0.0000	0.0311	0.0311	0.0000	0.0620	0.0622
d	Bias	-0.0017	-0.0055	-0.0054	-0.0027	-0.0068	-0.0068	-0.0011	-0.0055	-0.0055
	Std	0.0240	0.0246	0.0245	0.0241	0.0248	0.0248	0.0236	0.0243	0.0243
σ	Bias	-0.0006	-0.0011	-0.0011	-0.0006	-0.0011	-0.0011	-0.0009	-0.0014	-0.0014
	Std	0.0226	0.0226	0.0226	0.0229	0.0229	0.0229	0.0227	0.0227	0.0227
			d = 0.20			d = 0.30			d = 0.40	
μ	Bias	0.0000	-0.0092	-0.0098	0.0000	0.0058	0.0047	0.0000	-0.0328	-0.0328
	Std	0.0000	0.1196	0.1204	0.0000	0.2468	0.2505	0.0000	0.5437	0.5466
d	Bias	-0.0023	-0.0067	-0.0067	-0.0014	-0.0056	-0.0056	-0.0044	-0.0087	-0.0087
	Std	0.0256	0.0264	0.0264	0.0244	0.0251	0.0251	0.0226	0.0236	0.0236
σ	Bias	-0.0001	-0.0006	-0.0006	-0.0013	-0.0017	-0.0017	-0.0011	-0.0014	-0.0014
	Std	0.0234	0.0234	0.0234	0.0215	0.0215	0.0215	0.0224	0.0224	0.0224

TABLE IX

BIAS, STD AND RMSE OF ALTERNATIVE MLEs FOR $ARFIMA(0,d,0)$: $\mu = -1$ AND $\sigma = 1$.

		MLE1	MLE2	MLE3	MLE1	MLE2	MLE3	MLE1	MLE2	MLE3
$n = 250$										
		d = -0.40			d = -0.30			d = -0.20		
μ	Bias	0.0000	-0.0005	-0.0006	0.0000	0.0007	0.0007	0.0000	0.0003	0.0003
	Std	0.0000	0.0097	0.0118	0.0000	0.0150	0.0159	0.0000	0.0241	0.0249
d	Bias	-0.0055	-0.0170	-0.0109	-0.0044	-0.0166	-0.0134	-0.0046	-0.0178	-0.0163
	Std	0.0500	0.0518	0.0512	0.0538	0.0558	0.0546	0.0529	0.0555	0.0548
σ	Bias	-0.0031	-0.0054	-0.0045	-0.0032	-0.0054	-0.0050	-0.0008	-0.0031	-0.0029
	Std	0.0452	0.0451	0.0452	0.0462	0.0460	0.0460	0.0439	0.0440	0.0440
		d = -0.10			d = 0.00			d = 0.10		
μ	Bias	0.0000	0.0013	0.0012	0.0000	-0.0017	-0.0016	0.0000	0.0030	0.0028
	Std	0.0000	0.0395	0.0397	0.0000	0.0623	0.0622	0.0000	0.1116	0.1120
d	Bias	-0.0006	-0.0148	-0.0143	-0.0052	-0.0192	-0.0191	-0.0039	-0.0197	-0.0198
	Std	0.0518	0.0557	0.0553	0.0485	0.0525	0.0523	0.0501	0.0538	0.0538
σ	Bias	-0.0014	-0.0038	-0.0037	-0.0006	-0.0029	-0.0029	-0.0027	-0.0051	-0.0051
	Std	0.0436	0.0436	0.0436	0.0446	0.0447	0.0447	0.0454	0.0453	0.0453
		d = 0.20			d = 0.30			d = 0.40		
μ	Bias	0.0000	0.0066	0.0063	0.0000	-0.0070	-0.0048	0.0000	0.0471	0.0471
	Std	0.0000	0.1900	0.1906	0.0000	0.3520	0.3565	0.0000	0.6664	0.6697
d	Bias	-0.0047	-0.0213	-0.0213	-0.0062	-0.0235	-0.0234	-0.0129	-0.0297	-0.0295
	Std	0.0484	0.0535	0.0536	0.0453	0.0498	0.0498	0.0415	0.0464	0.0464
σ	Bias	-0.0028	-0.0051	-0.0051	-0.0044	-0.0065	-0.0065	-0.0021	-0.0038	-0.0037
	Std	0.0444	0.0446	0.0446	0.0446	0.0446	0.0446	0.0446	0.0446	0.0446
$n = 1000$										
		d = -0.40			d = -0.30			d = -0.20		
μ	Bias	0.0000	-0.0000	0.0000	0.0000	-0.0002	-0.0002	0.0000	-0.0003	-0.0003
	Std	0.0000	0.0028	0.0034	0.0000	0.0051	0.0055	0.0000	0.0090	0.0093
d	Bias	-0.0036	-0.0074	-0.0052	-0.0019	-0.0058	-0.0048	-0.0011	-0.0048	-0.0045
	Std	0.0263	0.0270	0.0266	0.0251	0.0256	0.0254	0.0249	0.0252	0.0251
σ	Bias	0.0001	-0.0005	-0.0002	-0.0009	-0.0015	-0.0014	-0.0007	-0.0012	-0.0011
	Std	0.0232	0.0232	0.0232	0.0224	0.0224	0.0224	0.0223	0.0223	0.0223
		d = -0.10			d = 0.00			d = 0.10		
μ	Bias	0.0000	0.0007	0.0007	0.0000	-0.0023	-0.0023	0.0000	-0.0021	-0.0020
	Std	0.0000	0.0169	0.0169	0.0000	0.0314	0.0314	0.0000	0.0608	0.0608
d	Bias	-0.0010	-0.0051	-0.0050	-0.0030	-0.0071	-0.0071	-0.0016	-0.0058	-0.0058
	Std	0.0252	0.0256	0.0256	0.0237	0.0243	0.0243	0.0252	0.0260	0.0260
σ	Bias	-0.0005	-0.0011	-0.0011	-0.0002	-0.0007	-0.0007	-0.0019	-0.0024	-0.0024
	Std	0.0219	0.0219	0.0219	0.0221	0.0220	0.0220	0.0230	0.0229	0.0229
		d = 0.20			d = 0.30			d = 0.40		
μ	Bias	0.0000	-0.0075	-0.0081	0.0000	-0.0035	-0.0022	0.0000	0.0050	0.0069
	Std	0.0000	0.1234	0.1242	0.0000	0.2510	0.2534	0.0000	0.5516	0.5593
d	Bias	-0.0005	-0.0050	-0.0050	-0.0027	-0.0073	-0.0072	-0.0034	-0.0077	-0.0077
	Std	0.0244	0.0251	0.0251	0.0243	0.0253	0.0253	0.0232	0.0243	0.0243
σ	Bias	-0.0012	-0.0017	-0.0017	-0.0004	-0.0008	-0.0008	0.0001	-0.0002	-0.0002
	Std	0.0217	0.0217	0.0217	0.0227	0.0227	0.0227	0.0216	0.0216	0.0216