Bernoulli **25**(3), 2019, 1870–1900
https://doi.org/10.3150/18-BEJ1039

Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations

MASAAKI FUKASAWA* and TETSUYA TAKABATAKE**

Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, 560-8531, Osaka, Japan. E-mail: *fukasawa@sigmath.es.osaka-u.ac.jp; **takabatake@sigmath.es.osaka-u.ac.jp

This paper proposes feasible asymptotically efficient estimators for a certain class of Gaussian noises with self-similarity and stationarity properties, which includes the fractional Gaussian noises, under high frequency observations. In this setting, the optimal rate of estimation depends on whether either the Hurst or diffusion parameters is known or not. This is due to the singularity of the asymptotic Fisher information matrix for simultaneous estimation of the above two parameters. One of our key ideas is to extend the Whittle estimation method to the situation of high frequency observations. We show that our estimators are asymptotically efficient in Fisher's sense. Further by Monte-Carlo experiments, we examine finite sample performances of our estimators. Finite sample modifications of the asymptotic variances of the estimators are also given, which exhibit almost perfect fits to the numerical results.

Keywords: asymptotic efficiency; fractional Gaussian noises; high frequency observations; local asymptotic normality; Whittle estimation

1. Introduction

Self-similarity and Gaussianity properties of noises in time series data are observed in many fields, for example, hydrology, turbulence, molecular biology and financial economics. Fractional Brownian motion, which is introduced by Kolmogorov [13] and further developed by Mandelbrot and Van Ness [16], is the most fundamental continuous-time model to represent these phenomena. Until now, statistical inference problems of discretely observed fractional Brownian motion have been studied under the large sample asymptotics (cf. Fox and Taqqu [9], Dahlhaus [6], Lieberman *et al.* [14,15], Cohen *et al.* [5]) or the high frequency asymptotics (cf. Coeurjolly [4], Brouste and Iacus [3], Kawai [12], Brouste and Fukasawa [2]). Some of them also discuss the optimality of their estimators or local asymptotic normality (LAN) property under those asymptotics. To the best of our knowledge, however, there is no estimator available so far which is computationally feasible and asymptotically optimal under the high frequency asymptotics, despite that it has become important because of increasing availability of high frequency data thanks to recent developments of information technology.

Now, we review the results of Kawai [12] and Brouste and Fukasawa [2], where they studied the LAN property of the fractional Gaussian noises under high frequency observations. The parameter to be estimated is $(H, \sigma) \in (0, 1] \times (0, \infty)$, where H and σ are the so-called Hurst parameter and diffusion parameter respectively. Kawai [12] obtained a LAN property in a weak sense, where the rate matrix $\bar{\phi}_n(H,\sigma)$ and the asymptotic Fisher information matrix $\mathcal{I}(H,\sigma)$ are given by

$$\bar{\phi}_n(H,\sigma) := \begin{pmatrix} \frac{1}{\sqrt{n}|\log\delta_n|} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix}, \qquad \mathcal{I}(H,\sigma) := \begin{pmatrix} 2 & -\frac{2}{\sigma} \\ -\frac{2}{\sigma} & \frac{2}{\sigma^2} \end{pmatrix}.$$

Here, n is the sample size and δ_n is the length of sampling intervals. The LAN property is only in a weak sense because $\mathcal{I}(H,\sigma)$ is a singular matrix. As a result, the asymptotic lower bounds of risk are derived only in the case that either the Hurst parameter H or the diffusion parameter σ is known. In Brouste and Fukasawa [2], it was shown that the LAN property in the usual sense actually holds even if both of the parameters are unknown. One of their key ideas is using a certain class of non-diagonal rate matrices. The LAN property enables them to determine the optimal rate of estimation for each of H and σ. Then the lower bounds of variance are given as follows; for any estimators $\widehat{H}_n$ and $\widehat{\sigma}_n$,

$$\lim_{C\to\infty}\liminf_{n\to\infty}\sup_{\|\phi_n^{-1}(H_0,\sigma_0)((H,\sigma)^*-(H_0,\sigma_0)^*)\|_{\mathbb{R}^2}<C} nE_{H,\sigma}^{(n)}\big[(\widehat{H}_n-H)^2\big] \geq \mathcal{G}_1(H_0)^{-1},$$

$$\lim_{C\to\infty}\liminf_{n\to\infty}\sup_{\|\phi_n^{-1}(H_0,\sigma_0)((H,\sigma)^*-(H_0,\sigma_0)^*)\|_{\mathbb{R}^2}<C} \frac{n}{\sigma^2(\log\delta_n)^2}E_{H,\sigma}^{(n)}\big[(\widehat{\sigma}_n-\sigma)^2\big] \geq \mathcal{G}_1(H_0)^{-1},$$

for any (H_0,σ_0), where

$$\mathcal{G}_1(H) := \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\partial}{\partial H}\log f_H(\lambda)\right)^2 d\lambda - \frac{1}{2}\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial H}\log f_H(\lambda)\,d\lambda\right)^2 > 0$$

and f_H is the spectral density of the fractional Gaussian noises, which will be given later. It is noteworthy that the optimal rates of estimation are slower than in the case where the other parameter is known.

The LAN property implies the asymptotic efficiency of the maximum likelihood estimator (MLE) in general. The MLE is unfortunately not computationally feasible for the fractional Gaussian noises because the computation of the inverse of the covariance matrices is very heavy. Our main contribution in this context is to construct computationally feasible asymptotically efficient estimators. We deal with the three cases: both the Hurst parameter H and the diffusion parameter σ are unknown, only the Hurst parameter H is known, only the diffusion parameter σ is known. Actually, we work under a more general model of Gaussian noises with self-similarity and stationarity properties, which generalizes the fractional Gaussian noises.

This paper is organized as follows. In Section 3, we introduce the model. In Section 4, we give several examples in our framework. In Section 5, we present and prove our main results. In Section 6, we investigate finite sample performances of our estimators by Monte-Carlo experiments. In the Appendix, we give several lemmas and extensions of the results in Kawai [12], Brouste and Fukasawa [2].

2. Notation and brief review of the Whittle estimator in large sample setting

In this section, we prepare notation used in this paper and briefly review asymptotic properties of the Whittle estimator in large sample setting, that is, the situation that a length of sampling interval $\delta > 0$ is constant and a sample size $n \to \infty$ is considered through this section. In particular, we suppose $\delta \equiv 1$ for notational simplicity.

2.1. Notation

Let $p \in \mathbb{N}$, $\Theta^\dagger$ be a closure of a bounded convex domain in $\mathbb{R}^{p+1}$ and a function $f : \Theta^\dagger \times [-\pi, \pi] \to [-\infty, \infty]$ be sufficiently smooth in the sense specified later. Denote $f_\theta(\lambda) \equiv f(\theta, \lambda)$ and

$$a_q(\theta) := \left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_1}\log f_\theta(\lambda)\,d\lambda, \dots, \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_q}\log f_\theta(\lambda)\,d\lambda\right),$$

$$\mathcal{F}_q(\theta) := \left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_i}\log f_\theta(\lambda)\frac{\partial}{\partial\theta_j}\log f_\theta(\lambda)\,d\lambda\right)_{i,j=1,\dots,q},$$

$$\mathcal{G}_q(\theta) := \left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\partial}{\partial\theta_i}\log g_\theta(\lambda)\frac{\partial}{\partial\theta_j}\log g_\theta(\lambda)\,d\lambda\right)_{i,j=1,\dots,q}$$

for each $q \in \{1, 2, \dots, p+1\}$, where the function $g_\theta(\lambda) \equiv g(\theta, \lambda)$ is defined by

$$g_\theta(\lambda) := \frac{f_\theta(\lambda)}{b(\theta)} \qquad \text{with } b(\theta) := \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_\theta(\lambda)\,d\lambda\right).$$

In particular, denote $a_0(\theta) := 0$ for notational simplicity. Moreover, define

$$I_n(\mathbf{x}_n, \lambda) := \frac{1}{2\pi n}\left|\sum_{j=1}^{n} x_j e^{\sqrt{-1}j\lambda}\right|^2 \qquad \text{with } \mathbf{x}_n = (x_1, \dots, x_n) \in \mathbb{R}^n, \lambda \in [-\pi, \pi].$$

2.2. Asymptotic properties of the Whittle estimator in large sample setting

Let $(\Omega, \mathcal{F}, P)$ be a probability space on which a centered stationary Gaussian sequence $\{X_j\}_{j\in\mathbb{Z}}$ with a spectral density $f(\theta, \lambda)$ is defined, that is, its covariance function is characterized by the function $f : \Theta^\dagger \times [-\pi, \pi] \to [-\infty, \infty]$ as follows:

$$E[X_i X_j] = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\sqrt{-1}(i-j)\lambda} f(\theta, \lambda)\,d\lambda \qquad \text{for any } i, j \in \mathbb{Z}.$$

In the following, denote $f_\theta(\lambda) \equiv f(\theta, \lambda)$ for notational simplicity. Moreover, we impose the following conditions, which are used in many of previous works, see, for example, Fox and Taqqu [9], Dahlhaus [6,7], Lieberman *et al.* [14] and Cohen *et al.* [5], on the spectral density:

(S.1) If θ_1 and θ_2 are distinct elements of $\Theta^\dagger$, a set $\{\lambda \in [-\pi, \pi] : f_{\theta_1}(\lambda) \neq f_{\theta_2}(\lambda)\}$ has a positive Lebesgue measure.

(S.2) $f \in \mathcal{C}^{3,1}(\Theta^\dagger \times [-\pi, \pi] \setminus \{0\})$ and there exists a continuous function $\alpha : \Theta^\dagger \to (-1, 1)$ such that for any $\eta > 0$, there exist positive constants $c_{1,\eta}, c_{2,\eta}$, which only depend on η, such that the following conditions hold for every $(\theta, \lambda) \in \Theta^\dagger \times [-\pi, \pi] \setminus \{0\}$.
   (a) $c_{1,\eta}|\lambda|^{-\alpha(\theta)+\eta} \leq f_\theta(\lambda) \leq c_{2,\eta}|\lambda|^{-\alpha(\theta)-\eta}$.
   (b) For any $l \in \{1, 2, 3\}$ and $k \in \{1, \dots, p+1\}^l$,

$$\left| \frac{\partial^l}{\partial\theta_{k_1} \cdots \partial\theta_{k_l}} f_\theta(\lambda) \right| \leq c_{2,\eta}|\lambda|^{-\alpha(\theta)-\eta},$$

$$\left| \frac{\partial^{l+1}}{\partial\lambda\, \partial\theta_{k_1} \cdots \partial\theta_{k_l}} f_\theta(\lambda) \right| \leq c_{2,\eta}|\lambda|^{-\alpha(\theta)-1-\eta}.$$

(S.3) The matrix $\mathcal{F}_{p+1}(\theta)$ is invertible for any $\theta \in \Theta^\dagger$.

Remark 1. Under the condition (S.2), the spectral density $f_\theta(\lambda)$ and its derivatives are integrable and the differential and integral operators can be freely interchanged.

Remark 2. It is unclear whether the identifiability condition (S.1) implies (S.3). Actually, Taniguchi [20], p.161, and Dzhaparidze [8], p.107, assume the same or equivalent conditions as (S.3) in addition to (S.1). Note that the condition (S.3) is equivalent to the linear independence of a family of functions $\{\partial/\partial\theta_j \log f_\theta(\cdot)\}_{j=1,\dots,p+1}$ for any $\theta \in \Theta^\dagger$. Moreover, its linear independence is equivalent to that of a family of functions $\{\partial/\partial\theta_j \log g_\theta(\cdot)\}_{j=1,\dots,p+1}$ for any $\theta \in \Theta^\dagger$ under the condition (S.2) because $\partial/\partial\theta_j \log g_\theta(\lambda)$ equals $\partial/\partial\theta_j \log f_\theta(\lambda)$ plus a certain constant and these functions are not constant under (S.2). As a result, the condition (S.3) is also equivalent to that the matrices $\mathcal{G}_q(\theta)$ are invertible for any $\theta \in \Theta^\dagger$ and $q \in \{1, 2, \dots, p+1\}$ under (S.2).

Through this section, we consider the following observations

$$\mathbf{X}_n^{(1)} := (X_1, \dots, X_n), \qquad n \in \mathbb{N}, \tag{2.1}$$

and denote by $Q_\theta^{(n)}$ a distribution of $\mathbf{X}_n^{(1)}$. The Whittle likelihood $L^W(\theta)$ and the Whittle estimator $\widehat{\theta}_n^W$ based on observed values $\mathbf{x}_n := \mathbf{X}_n^{(1)}(\omega)$ for a certain $\omega \in \Omega$ are defined in the following way:

$$L_n^W(\theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_\theta(\lambda)\, d\lambda + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(\mathbf{x}_n, \lambda)}{f_\theta(\lambda)}\, d\lambda, \qquad \widehat{\theta}_n^W := \mathop{\arg\min}_{(\theta) \in \Theta \times \Sigma} L_n^W(\theta).$$

The following result is derived from Theorem 5 in Lieberman *et al.* [14] and Theorem 2.4 in Cohen *et al.* [5].

Theorem 1. *Let a true value* θ_0 *be an interior point of the parameter space* $\Theta^\dagger$. *Under* (S.1)–(S.3), *a sequence of the Whittle estimators* $\widehat{\theta}_n^W$ *is asymptotically efficient in Fisher's sense*, *that* *is, it holds that*

$$\mathcal{L}\{\sqrt{n}(\widehat{\theta}_n^W - \theta_0)|Q_{\theta_0}^{(n)}\} \overset{n\to\infty}{\to} \mathcal{N}(0, \mathcal{F}_{p+1}(\theta_0)^{-1}),$$

where $\mathcal{F}_{p+1}(\theta_0)$ is the non-singular asymptotic Fisher information matrix at the point θ_0.

3. Model assumption

In this section, we introduce several assumptions for a model considered in this paper, which are inspired from the previous works documented in Section 2.

Assumption 1. Let parameter spaces Θ and Σ be the closures of bounded convex domains in $\mathbb{R}^{p-1} \times (0, 1]$ and $(0, \infty)$, respectively. Moreover, assume true values $\theta_0 \equiv (\psi_0, H_0) \equiv (\psi_0^{(1)}, \dots, \psi_0^{(p-1)}, H_0)$ and σ_0 are interior points of Θ and Σ, respectively. Let the sample size $n \in \mathbb{N}$ and the length of sampling intervals $\delta_n > 0$ satisfy $\inf_{n\in\mathbb{N}} n\delta_n > 0$ and $\delta_n \to 0$ as $n \to \infty$. Let $(\Omega, \mathcal{F}, P)$ be a probability space on which a sequence of observations:

$$\mathbf{X}_n := (X_1^n, \dots, X_n^n), \qquad n \in \mathbb{N},$$

is defined, where $\mathbf{X}_n$ is a n-dimensional centered Gaussian random vector which covariance function is characterized by a spectral density $f^n(\theta, \sigma, \lambda)$, that is, it holds that

$$E[X_i^n X_j^n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\sqrt{-1}(i-j)\lambda} f^n(\theta, \sigma, \lambda)\, d\lambda \qquad \text{for any } i, j = 1, \dots, n.$$

Denote by $P_{\theta,\sigma}^{(n)}$ a distribution of $\mathbf{X}_n$. Moreover, we impose the following condition on the spectral density:

(S.0) $f^n(\theta, \sigma, \lambda)$ is decomposed as $\sigma^2\delta_n^{2H} f(\theta, \lambda)$, where $(\theta, \sigma, \lambda) \mapsto \sigma^2 f(\theta, \lambda)$ satisfies the conditions (S.1)–(S.3) with the parameter space $\Theta^\dagger := \Theta \times \Sigma$.

In the following, denote $f^n(\theta, \sigma, \lambda)$ and $f(\theta, \lambda)$ as $f_{\theta,\sigma}^n(\lambda)$ and $f_\theta(\lambda)$, respectively. Moreover, we sometimes use the notation $\theta_{p+1} \equiv \sigma$ for simplicity.

Remark 3. As far as we consider the asymptotics of high frequency observations, it is natural to regard the data as discrete observations from a continuous-time process. If we consider a stationary stochastic process $X = \{X_t\}_{t\in\mathbb{R}}$ with a spectral density $f_X(\theta, \lambda)$ with $\lambda \in \mathbb{R}$, then the well-known aliasing formula yields that a spectral density of its discrete sample with the length of observations δ_n is given by

$$f^n(\theta, \lambda) := \frac{1}{2\pi\delta_n} \sum_{j\in\mathbb{Z}} f_X\left(\theta, \frac{\lambda + 2\pi j}{\delta_n}\right), \qquad \lambda \in [-\pi, \pi].$$

Moreover, if we consider a centered self-similar stationary Gaussian noise ($\delta_n \equiv 1$) in the sense of Sinai [19], then its spectral density is given by

$$f_{H,\eta}(\lambda) := \eta\{2(1-\cos\lambda)\}\sum_{j\in\mathbb{Z}} \frac{1}{|\lambda + 2\pi j|^{1+2H}}, \qquad \lambda \in [-\pi, \pi], \tag{3.1}$$

for a certain constant $\eta > 0$. Note that certain series are naturally appeared in both cases. As far as we are interested in self-similar stationary Gaussian noises, it suffices to consider stationary Gaussian noises with spectral densities like (3.1). However, we actually work under more slightly general form of spectral densities appeared in (S.0).

Remark 4. In the large sample setting corresponding to $\delta_n \equiv \delta$ and $n \to \infty$, it is not necessary to consider the decomposition of the spectral density in (S.0) as shown in Section 2 because we can include the part $\sigma^2\delta^{2H}$ into the model parameters. This is also mentioned in the seminal paper of Dahlhaus [6], p.1752. In the high frequency setting, however, we cannot do this because the variable δ_n which drives the asymptotics is closely related with the Hurst parameter H. Moreover, the parameter σ and the others play different roles under high frequency observations. In fact, as we see later, the efficient rate of estimator of σ is different from that of the others.

4. Examples

In this section, we give several examples satisfying the assumptions introduced in the previous section. Namely, we treat three models: the fractional Brownian motion, a special case of a fractional Langevin model and its extension.

4.1. Fractional Brownian motion

A centered continuous Gaussian process $\{B_t^H\}_{t\in\mathbb{R}}$ is called a fractional Brownian motion (fBm) with the Hurst parameter H if its covariance function is given by

$$E\left[B_t^H B_s^H\right] := \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right) \qquad \text{for any } s, t \in \mathbb{R}.$$

Such a process actually exists for any $H \in (0, 1]$ from the Kolmogorov's extension and continuity theorems. It is well known that the fBm has the stationary increments and self-similarity properties, that is,

$$\left\{B_{t+k}^H - B_k^H\right\}_{t\in\mathbb{R}} = \left\{B_t^H\right\}_{t\in\mathbb{R}} \quad \text{and} \quad \left\{B_{ct}^H\right\}_{t\in\mathbb{R}} = \left\{c^H B_t^H\right\}_{t\in\mathbb{R}} \tag{4.1}$$

hold in law for any $k \in \mathbb{R}$ and $c > 0$. Consider the following observations:

$$\mathbf{B}_n := \left(\sigma B_{\delta_n}^H, \sigma B_{2\delta_n}^H, \ldots, \sigma B_{n\delta_n}^H\right),$$

where n is the sample size and δ_n is the length of sampling intervals. Denote by $P_{H,\sigma}^{(n)}$ a distribution of a n-dimensional centered Gaussian random vector

$$\mathbf{X}_n := \left(X_1^n, \dots, X_n^n\right) \qquad \text{with } X_j^n := \sigma\left(B_{j\delta_n}^H - B_{(j-1)\delta_n}^H\right).$$

Note that $\mathbf{X}_n$ satisfies all conditions in Assumption 1. Indeed, the properties in (4.1) yield that a covariance function of $\mathbf{X}_n$ is characterized by

$$\begin{aligned} E\left[X_i^n X_j^n\right] &= \frac{\sigma^2 \delta_n^{2H}}{2}\left(|i-j+1|^{2H} - 2|i-j|^{2H} + |i-j-1|^{2H}\right) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\sqrt{-1}(i-j)\lambda} f_{H,\sigma}^n(\lambda)\, d\lambda \end{aligned}$$

for any $i, j = 1, \dots, n$, where $f_{H,\sigma}^n(\lambda) := \sigma^2 \delta_n^{2H} f_H(\lambda)$ and

$$\begin{aligned} f_H(\lambda) &:= C_H\left\{2(1-\cos\lambda)\right\}\sum_{j\in\mathbb{Z}} \frac{1}{|\lambda + 2\pi j|^{1+2H}} \\ &\text{with } C_H := \frac{\Gamma(2H+1)\sin(\pi H)}{2\pi}, \end{aligned} \tag{4.2}$$

see Sinai [19] and Samorodnitsky and Taqqu [18]. Moreover, it is easily shown that the spectral density $f_{H,\sigma}^n$ satisfies the condition (S.0) (cf. Fox and Taqqu [9]).

4.2. Fractional Langevin model

In the context of molecular biology, the movement of particle in homogeneous medium is modeled by the Langevin equation. One of the characteristics of this model is that a mean square displacement of particle linearly grows in time. The particle in inhomogeneous medium, however, does not behave in the same way. Namely, the mean square displacement of particle in this situation grows as a power function in time. This phenomenon is the so-called anomalous diffusion, see, for example, Bouchaud and Georges [1] for more detail. Therefore, we attempt to model this by the following second order stochastic differential equations:

$$\begin{aligned} dZ_t &= Y_t\, dt, \\ dY_t &= -\nabla q(Z_t) - \gamma Y_t\, dt + \sigma\, dB_t^H, \qquad \gamma, \sigma > 0, \end{aligned}$$

where Z and Y represent the position and velocity of particle respectively, q is a certain potential function, γ and σ are the friction and diffusion coefficients respectively, and B^H is the fractional Brownian motion with the Hurst parameter H. Here, we assumed the mass of the particle is equal to 1 for notational simplicity. We call the above equations as a fractional Langevin model named after the Langevin model in the case of $H = 1/2$.

Under the situation of a free particle with no friction term, that is, assume q is a constant function and $\gamma = 0$ in the fractional Langevin model, we consider a statistical inference problem for the unknown parameter (H,σ) based on position data $\{Z_{j\delta_n}\}_{j=1,\dots,n}$, where n is the sample size and δ_n is the length of sampling intervals. In this situation, no velocity data are available so that we consider to substitute numerical derivatives of the position data $\{Z_{j\delta_n}\}_{j=1,\dots,n}$ for the velocity data. Namely, we define a proxy $\{Y_j^n\}_{j=1,\dots,n}$ for the velocity data of particle by

$$Y_j^n := \frac{1}{\delta_n}(Z_{j\delta_n} - Z_{(j-1)\delta_n}) = \frac{1}{\delta_n}\int_{(j-1)\delta_n}^{j\delta_n} Y_t\,dt.$$

Then, we define n-dimensional centered Gaussian random vectors $\mathbf{X}_n := (X_1^n,\dots,X_n^n)$ with $n \in \mathbb{N}$ and denote by $P_{H,\sigma}^{(n)}$ a distribution of $\mathbf{X}_n$, where

$$X_j^n := Y_j^n - Y_{j-1}^n = \frac{\sigma}{\delta_n}\int_{(j-1)\delta_n}^{j\delta_n}\big(B_t^H - B_{t-\delta_n}^H\big)\,dt.$$

Note that $\mathbf{X}_n$ satisfies all conditions in Assumption 1. Indeed, it is easily shown that a covariance function of $\mathbf{X}_n$ is characterized by

$$\begin{aligned} E\big[X_i^n X_j^n\big] &= \frac{\sigma^2\delta_n^{2H}}{2}\big(|i-j+2|^{2H+2} - 4|i-j+1|^{2H+2} \\ &\quad + 6|i-j|^{2H+2} - 4|i-j-1|^{2H+2} + |i-j-2|^{2H+2}\big) \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\sqrt{-1}(i-j)\lambda} f_{H,\sigma}^n(\lambda)\,d\lambda \end{aligned}$$

for any $i,j = 1,\dots,n$, where $f_{H,\sigma}^n(\lambda) := \sigma^2\delta_n^{2H} f_H(\lambda)$ and

$$f_H(\lambda) := C_H\big\{2(1-\cos\lambda)\big\}^2 \sum_{j\in\mathbb{Z}} \frac{1}{|\lambda + 2\pi j|^{3+2H}} \qquad \text{with } C_H := \frac{\Gamma(2H+1)\sin(\pi H)}{2\pi}.$$

Moreover, it is shown that the spectral density $f_{H,\sigma}^n$ satisfies the condition (S.0) in the similar ways to the previous example. Therefore, this example is included in our setting.

More generally, we consider a sequence of n-dimensional centered Gaussian random vectors which covariance functions are characterized by the following spectral density:

$$f_{H,\psi,\sigma}^n(\lambda) := \sigma^2\delta_n^{2H} C_H\big\{2(1-\cos\lambda)\big\}^{\psi+1} \sum_{j\in\mathbb{Z}} \frac{1}{|\lambda + 2\pi j|^{1+2H+2\psi}},$$

where $\psi \in (0,\infty)$ and C_H is given in the above. Note that such Gaussian random vectors exist for any $\psi \in (0,\infty)$ because their covariance matrix are well-defined and positive definite for each $n \in \mathbb{N}$. Moreover, it is easily shown that $f_{H,\psi,\sigma}^n$ satisfies the condition (S.0) in the similar ways to the previous examples. A continuous-time stochastic process behind such a sequence can be formally derived from taking the fractional difference of the order ψ for a fractional integral of the order ψ of the fractional Brownian motion with the Hurst parameter H and appropriately rescaling. However, its rigorous justification is left for future research.

5. Construction of asymptotically efficient estimators

The Whittle estimation method is very useful to estimate the Hurst parameter of a certain classs of stationary Gaussian time series in various aspects, for example, the Whittle estimator enjoys asymptotic efficiency as well as the MLE and can be computed more easily and faster than it because we compute an approximate log-likelihood instead of the exact one which involves the inverse of the covariance matrix. Therefore, we attempt to prove the Whittle estimation method also works well under the high frequency observations. In the following, we consider statistical inference problems of our model in three cases: both parameters (θ, σ) are unknown, only the Hurst parameter H is known, only the diffusion parameter σ is known. In each case, we construct a feasible asymptotically efficient estimator by applying the Whittle estimation method.

5.1. Only the Hurst parameter H is known

Assume the true value of the Hurst parameter H_0 is known through this subsection. Under the condition (S.0), we define an estimation function $\bar{L}_N(\psi, \sigma)$ based on observed values $\mathbf{x}_n := \mathbf{X}_n(\omega)$ for a certain $\omega \in \Omega$ as follows:

$$\begin{aligned}\bar{L}_n(\psi, \sigma) &:= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\bigl(\sigma^2 \delta_n^{2H_0} f_\psi(\lambda)\bigr)\, d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\mathbf{x}_n, \lambda)}{\sigma^2 \delta_n^{2H_0} f_\psi(\lambda)}\, d\lambda \\ &= 2H_0 \log \delta_n + \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\bigl(\sigma^2 f_\psi(\lambda)\bigr)\, d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\widetilde{\mathbf{x}}_n, \lambda)}{\sigma^2 f_\psi(\lambda)}\, d\lambda,\end{aligned}$$

where $f_\psi(\lambda) := f_{\psi, H_0}(\lambda)$, $\widetilde{\mathbf{x}}_n := \delta_n^{-H_0}\mathbf{x}_n$. Then we define an estimator for the unknown parameter (ψ, σ) by

$$\begin{aligned}(\bar{\psi}_n, \bar{\sigma}_n) &:= \mathop{\arg\min}_{(\psi, H_0, \sigma) \in \Theta \times \Sigma} \bar{L}_n(\psi, \sigma) \\ &= \mathop{\arg\min}_{(\psi, H_0, \sigma) \in \Theta \times \Sigma} \left\{ \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\bigl(\sigma^2 f_\psi(\lambda)\bigr)\, d\lambda + \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\widetilde{\mathbf{x}}_n, \lambda)}{\sigma^2 f_\psi(\lambda)}\, d\lambda \right\}.\end{aligned}$$

Note that the decomposition of the spectral density in (S.0) and the Gaussianity of the distribution of $\mathbf{X}_n$ yield that the distribution of

$$\widetilde{\mathbf{X}}_n := \bigl(\delta_n^{-H_0} X_1^n, \dots, \delta_n^{-H_0} X_n^n\bigr), \qquad n \in \mathbb{N}, \tag{5.1}$$

is equal to that of $\mathbf{X}_n^{(1)}$ with spectral density $\sigma^2 f_\theta(\lambda)$. As a result, the following result is obtained from Theorem 1 and an easy modification of Theorem 2.4 of Cohen *et al.* [5] even under high frequency observations.

Theorem 2. *Suppose Assumption 1 and H_0 is known. Then $(\bar{\psi}_n, \bar{\sigma}_n)$ is an asymptotically efficient estimator in Fisher's sense, that is, it holds that*

$$\mathcal{L}\left\{ \sqrt{n} \begin{pmatrix} \bar{\psi}_n - \psi_0 \\ \bar{\sigma}_n - \sigma_0 \end{pmatrix} \middle| P_{\theta_0, \sigma_0}^{(n)} \right\} \overset{n \to \infty}{\to} \mathcal{N}\bigl(0, \mathcal{F}(\psi_0, \sigma_0)^{-1}\bigr),$$

where the non-singular asymptotic Fisher information matrix $\mathcal{F}(\psi,\sigma)$ is given by

$$\mathcal{F}(\psi,\sigma) := \left(\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial\theta_i} \log\bigl(\sigma^2 f_\psi(\lambda)\bigr) \frac{\partial}{\partial\theta_j} \log\bigl(\sigma^2 f_\psi(\lambda)\bigr)\, d\lambda \right)_{i,j=1,\dots,p-1,p+1}.$$

5.2. Both parameters (θ,σ) are unknown

In this section, we consider the case that both parameters (θ,σ) are unknown. This case is more difficult than the previous case because we cannot substantially take the appropriate scaling of the data like (5.1). However, we can construct an asymptotically efficient estimator as follows: At first, under the assumption (S.0), we define the reparametrized spectral density $g_\theta(\lambda) \equiv g(\theta,\lambda)$ with $\theta \in \Theta$ by

$$g_\theta(\lambda) := \frac{f_\theta(\lambda)}{b(\theta)} \qquad \text{with } b(\theta) := \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_\theta(\lambda)\, d\lambda \right).$$

Then, g_θ also satisfies the conditions (S.1) and (S.2). Moreover, it is easily shown that the following equality holds for any $\theta \in \Theta$:

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log g_\theta(\lambda)\, d\lambda = 0. \tag{5.2}$$

Then, the Whittle estimation function $\nu_n^2(\theta)$ and the Whittle estimator $\widehat{\theta}_n$ for the reparametrized parameter (ν,θ) with $\nu := \sigma\delta_n^H b(\theta)^{-1/2}$ based on observed values $\mathbf{x}_n := \mathbf{X}_n(\omega)$ for a certain $\omega \in \Omega$ are defined by

$$\nu_n^2(\theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(\mathbf{x}_n,\lambda)}{g_\theta(\lambda)}\, d\lambda, \qquad \widehat{\theta}_n := \operatorname*{arg\,min}_{\theta\in\Theta} \nu_n^2(\theta).$$

Note that the estimator $\widehat{\theta}_n$ also minimizes a rescaled estimation function $\widetilde{\sigma}_n^2(\theta)$ given by

$$\widetilde{\sigma}_n^2(\theta) := \bigl(\delta_n^{2H_0} b(\theta_0)\bigr)^{-1} \nu_n^2(\theta) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(\widetilde{\mathbf{x}}_n,\lambda)}{g_\theta(\lambda)}\, d\lambda, \tag{5.3}$$

where $\widetilde{\mathbf{x}}_n := (\delta_n^{2H_0} b(\theta_0))^{-\frac{1}{2}} \mathbf{x}_n$, and the same argument in (5.1) yields that a distribution of $\widetilde{\mathbf{X}}_n := (\delta_n^{2H_0} b(\theta_0))^{-\frac{1}{2}} \mathbf{X}_n$ is equal to that of $\mathbf{X}_n^{(1)}$ with spectral density $\sigma^2 g_\theta(\lambda)$. In other words, we can formally take an appropriate scaling of the data even if H_0 is unknown. Define $\widetilde{\sigma}_n := (\widetilde{\sigma}_n^2(\widehat{\theta}_n))^{1/2}$. Then, we can regard a random variable $(\widehat{\theta}_n, \widetilde{\sigma}_n)$ as a minimizer of a function $L_n(\theta,\sigma)$ given by

$$L_n(\theta,\sigma) := \log\sigma^2 + \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_n(\widetilde{\mathbf{x}}_n,\lambda)}{\sigma^2 g_\theta(\lambda)}\, d\lambda$$

with respect to (θ,σ) on $\Theta \times \Sigma$. As a result of the above argument, we obtain the following central limit theorem (CLT) from Theorem 1 and elementary computation of its asymptotic covariance matrix using the equality (5.2) and the block matrix.

Lemma 1. *Under Assumption 1, the following CLT holds:*

$$\mathcal{L}\left\{\sqrt{n}\begin{pmatrix}\widehat{\theta}_n-\theta_0\\ \widetilde{\sigma}_n-\sigma_0\end{pmatrix}\Big|P^{(n)}_{\theta_0,\sigma_0}\right\}\overset{n\to\infty}{\to}\mathcal{N}\left(0,\begin{pmatrix}\mathcal{G}_p(\theta_0)^{-1} & 0_{p\times 1}\\ 0_{1\times p} & \frac{\sigma_0^2}{2}\end{pmatrix}\right). \tag{5.4}$$

Note that the random variable $\widetilde{\sigma}_n$ is not a statistic because the true value θ_0 is used in its definition. However, we can construct an estimator $\widehat{\sigma}_n$ for the diffusion parameter σ by substituting the estimators $\widehat{\theta}_n$ and $\widehat{H}_n$ into the true values θ_0 and H_0 in the rescaled function $\widetilde{\sigma}_n$. Namely, we define

$$\widehat{\sigma}_n:=\sqrt{\left(\delta_n^{2\widehat{H}_n}b(\widehat{\theta}_n)\right)^{-1}\nu_n^2(\widehat{\theta}_n)}.$$

In the rest of this section, we show the estimator $(\widehat{\theta}_n,\widehat{\sigma}_n)$ is asymptotically efficient in Fisher's sense. Before showing this claim, we review the definition of the asymptotically efficient estimator in this sense more precisely following by Ibragimov and Has'minski [11], p. 159.

Definition 1. Let $\Theta^\dagger\subset\mathbb{R}^{p+1}$ and a family of measures $\{P^{(n)}_\theta;\theta\in\Theta^\dagger\}$ satisfy the LAN property at a point $\theta_0\in\Theta^\dagger$ as $n\to\infty$, that is, for a certain non-singular $(p+1)\times(p+1)$-matrix $\phi_n(\theta_0)$ and any $u\in\mathbb{R}^{p+1}$, the log-likelihood ratio admits the representation:

$$\log\frac{dP^{(n)}_{\theta_0+\phi_n(\theta_0)u}}{dP^{(n)}_{\theta_0}}=\langle u,\zeta_n(\theta_0)\rangle-\frac{1}{2}\langle\mathcal{J}(\theta_0)u,u\rangle+r_n(\theta_0),$$

where $\mathcal{J}(\theta_0)$ is a non-singular $(p+1)\times(p+1)$-matrix and

$$\zeta_n(\theta_0)\to\mathcal{N}\big(0,\mathcal{J}(\theta_0)\big),\qquad r_n(\theta_0)\to 0,$$

in law under $P^{(n)}_{\theta_0}$ as $n\to\infty$. A sequence of estimators $\widehat{\theta}_n$ is called asymptotically efficient in Fisher's sense at the point θ_0 if it holds that

$$\mathcal{L}\{\phi_n^{-1}(\theta_0)(\widehat{\theta}_n-\theta_0)|P^{(n)}_{\theta_0}\}\overset{n\to\infty}{\to}\mathcal{N}\big(0,\mathcal{J}(\theta_0)^{-1}\big).$$

Here, the matrices $\phi_n(\theta_0)$ and $\mathcal{J}(\theta_0)$ are usually called as a rate and the asymptotic Fisher information matrices, respectively.

Theorem 3. *Suppose Assumption 1 and a rate matrix $\phi_n\equiv\phi_n(\theta_0,\sigma_0)$ satisfies Assumption 2 in Appendix B. Then, the sequence of estimators $(\widehat{\theta}_n,\widehat{\sigma}_n)$ is asymptotically efficient in Fisher's sense, that is,*

$$\mathcal{L}\left\{\phi_n^{-1}\begin{pmatrix}\widehat{\theta}_n-\theta_0\\ \widehat{\sigma}_n-\sigma_0\end{pmatrix}\Big|P^{(n)}_{\theta_0,\sigma_0}\right\}\overset{n\to\infty}{\to}\mathcal{N}\big(0,\mathcal{J}(\theta_0,\sigma_0)^{-1}\big), \tag{5.5}$$

where $\mathcal{J}(\theta,\sigma)$ is given in Appendix B.

Proof. At first, note that

$$\phi_n^{-1}\begin{pmatrix}\widehat{\theta}_n-\theta_0\\ \widehat{\sigma}_n-\sigma_0\end{pmatrix}=\begin{pmatrix}\phi_{n,1}^{-1}(\widehat{\psi}_n-\psi_0)\\ \phi_{n,2}^{-1}\begin{pmatrix}\widehat{H}_n-H_0\\ \widehat{\sigma}_n-\sigma_0\end{pmatrix}\end{pmatrix}, \tag{5.6}$$

where $\phi_{n,1}^{-1}$ and $\phi_{n,2}^{-1}$ are inverse matrices of $\phi_{n,1}$ and $\phi_{n,2}$ respectively, that is,

$$\phi_{n,1}^{-1}=\operatorname{diag}\left(\frac{1}{d_n^{(1)}},\dots,\frac{1}{d_n^{(p-1)}}\right),\qquad \phi_{n,2}^{-1}=\frac{1}{\det(\phi_{n,2})}\begin{pmatrix}\widehat{\beta}_n & -\widehat{\alpha}_n\\ -\beta_n & \alpha_n\end{pmatrix}.$$

Here, $\det(\phi_{n,2})$ is calculated as follows:

$$\det(\phi_{n,2})=\frac{\sigma_0}{n}\left(\sqrt{n}\alpha_n\frac{\sqrt{n}\widehat{\beta}_n}{\sigma_0}-\sqrt{n}\widehat{\alpha}_n\frac{\sqrt{n}\beta_n}{\sigma_0}\right)=\frac{\sigma_0}{n}(\sqrt{n}\alpha_n\widehat{\gamma}_n-\sqrt{n}\widehat{\alpha}_n\gamma_n).$$

Set $\acute{\sigma}_n:=\{(\delta_n^{2H_0}b(\widehat{\theta}_n))^{-1}v_n^2(\widehat{\theta}_n)\}^{1/2}$. At first, the error of $\log\widehat{\sigma}_n-\log\sigma_0$ is expanded as follows by using the delta method and Lemma 6 in the Appendix:

$$\begin{aligned}\log\widehat{\sigma}_n-\log\sigma_0&=(\log\acute{\sigma}_n-\log\sigma_0)-\log\delta_n(\widehat{H}_n-H_0)\\&=\frac{1}{\sigma_0}(\acute{\sigma}_n-\sigma_0)-\log\delta_n(\widehat{H}_n-H_0)+o_{P^{(n)}_{\theta_0,\sigma_0}}\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

In the similar way as the above, we obtain an asymptotic expansion of $\widehat{\sigma}_n-\sigma_0$ as follows:

$$\begin{aligned}\widehat{\sigma}_n-\sigma_0&=\sigma_0(\log\widehat{\sigma}_n-\log\sigma_0)+o_{P^{(n)}_{\theta_0,\sigma_0}}\left(n^{-\frac{1}{2}}\right)\\&=\acute{\sigma}_n-\sigma_0-\sigma_0\log\delta_n(\widehat{H}_n-H_0)+o_{P^{(n)}_{\theta_0,\sigma_0}}\left(n^{-\frac{1}{2}}\right).\end{aligned}\tag{5.7}$$

Therefore, the following asymptotic expansion holds.

$$\begin{aligned}\phi_{n,2}^{-1}\begin{pmatrix}\widehat{H}_n-H_0\\ \widehat{\sigma}_n-\sigma_0\end{pmatrix}&=\frac{1}{\det(\phi_{n,2})}\begin{pmatrix}\widehat{\beta}_n(\widehat{H}_n-H_0)-\widehat{\alpha}_n(\widehat{\sigma}_n-\sigma_0)\\ -\beta_n(\widehat{H}_n-H_0)+\alpha_n(\widehat{\sigma}_n-\sigma_0)\end{pmatrix}\\&=\frac{\sigma_0}{n\det(\phi_{n,2})}\begin{pmatrix}\widehat{\gamma}_n\sqrt{n}(\widehat{H}_n-H_0)-\sqrt{n}\widehat{\alpha}_n\dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n-\sigma_0)\\ -\gamma_n\sqrt{n}(\widehat{H}_n-H_0)+\sqrt{n}\alpha_n\dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n-\sigma_0)\end{pmatrix}+o_{P^{(n)}_{\theta_0,\sigma_0}}(1)\\&=\frac{\sigma_0}{n\det(\phi_{n,2})}\begin{pmatrix}\widehat{\gamma}_n & -\sqrt{n}\widehat{\alpha}_n\\ -\gamma_n & \sqrt{n}\alpha_n\end{pmatrix}\begin{pmatrix}\sqrt{n}(\widehat{H}_n-H_0)\\ \dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n-\sigma_0)\end{pmatrix}+o_{P^{(n)}_{\theta_0,\sigma_0}}(1)\\&=\left(E_n^*\right)^{-1}\begin{pmatrix}\sqrt{n}(\widehat{H}_n-H_0)\\ \dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n-\sigma_0)\end{pmatrix}+o_{P^{(n)}_{\theta_0,\sigma_0}}(1),\end{aligned}$$

where

$$E_n \equiv E_n(\theta_0,\sigma_0) := \begin{pmatrix} \sqrt{n}\alpha_n & \gamma_n \\ \sqrt{n}\widehat{\alpha}_n & \widehat{\gamma}_n \end{pmatrix}.$$

Note that (5.7) and $\widehat{\alpha}_n/\det(\phi_{n,2}) = O(\sqrt{n})$, $\alpha_n/\det(\phi_{n,2}) = O(\sqrt{n})$ are used in the second equality. From this asymptotic expansion and (5.6), it holds that

$$\phi_n^{-1}\begin{pmatrix} \widehat{\theta}_n - \theta_0 \\ \widehat{\sigma}_n - \sigma_0 \end{pmatrix} = \begin{pmatrix} \sqrt{n}\phi_{n,1} & 0_{p-1\times 2} \\ 0_{2\times p-1} & E_n^* \end{pmatrix}^{-1} \begin{pmatrix} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ \dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n - \sigma_0) \end{pmatrix} + o_{P^{(n)}_{\theta_0,\sigma_0}}(1).$$

Then, from Lemma 6 in the Appendix, $E_n \to E$, $\sqrt{n}\phi_{n,1} \to D$ in matrix norm as $n \to \infty$ and the continuous mapping theorem, it converges to a normal distribution with mean vector $0_{p+1\times 1}$ and covariance matrix given by the inverse matrix of

$$\begin{pmatrix} D & 0_{p-1\times 2} \\ 0_{2\times p-1} & E^* \end{pmatrix}^* \begin{pmatrix} \mathcal{F}_p(\theta) & a_p(\theta) \\ a_p(\theta)^* & 2 \end{pmatrix} \begin{pmatrix} D & 0_{p-1\times 2} \\ 0_{2\times p-1} & E^* \end{pmatrix}. \tag{5.8}$$

Therefore, (5.5) follows from elementary computation of (5.8) by using the block matrix because D is a diagonal matrix. Moreover, the asymptotic efficiency of $\{(\widehat{\theta}_n, \widehat{\sigma}_n)\}_{n\in\mathbb{N}}$ also follows from (5.5) and Theorem 6 in the Appendix. □

5.3. Only the diffusion parameter σ is known

Our purpose in this subsection is to construct an asymptotically efficient estimator for the unknown parameter $\theta = (\psi, H)$ in Fisher's sense in the case that the true value of the diffusion parameter σ_0 is known. Namely, we construct an estimator $(\widehat{\psi}_n, \widehat{H}_n)$ for the parameter (ψ, H) satisfying the following central limit theorem:

$$\mathcal{L}\left\{\begin{pmatrix} \sqrt{n}(\widehat{\psi}_n - \psi_0) \\ \sqrt{n}|\log\delta_n|(\widehat{H}_n - H_0) \end{pmatrix} \middle| P^{(n)}_{\theta_0,\sigma_0}\right\} \overset{n\to\infty}{\to} \mathcal{N}\big(0, \mathcal{I}(\theta_0)^{-1}\big), \tag{5.9}$$

where the non-singular asymptotic Fisher information $\mathcal{I}(\theta)$ is given by

$$\mathcal{I}(\theta) := \begin{pmatrix} \mathcal{G}_{p-1}(\theta)^{-1} & -\frac{1}{2}\mathcal{G}_{p-1}(\theta)^{-1}a_{p-1}(\theta) \\ -\frac{1}{2}a_{p-1}(\theta)^*\mathcal{G}_{p-1}(\theta)^{-1} & \frac{1}{2} + \frac{1}{4}a_{p-1}(\theta)^*\mathcal{G}_{p-1}(\theta)^{-1}a_{p-1}(\theta) \end{pmatrix}^{-1}.$$

In fact, the relation (5.9) means the asymptotic efficiency in Fisher's sense, which follows from Lemma 5 and Theorem 5 in the Appendix. Note that the optimal asymptotic covariance matrix $\mathcal{I}(\theta)^{-1}$ is independent of the estimation error of the Hurst parameter as well as that of the diffusion parameter.

At first, we work under a strong assumption that a value $b(\theta_0)$ is known. Then, we define an estimator $\widehat{H}_n^{\text{zero}}$ as a solution of the following equation with respect to H:

$$\log \frac{\nu_n^2(\widehat{\theta}_n)}{\delta_n^{2H} b(\theta_0)} - \log \sigma_0^2 = 0,$$

where $\widehat{\theta}_n \equiv (\widehat{\psi}_n, \widehat{H}_n) \equiv (\widehat{\psi}_n^{(1)}, \dots, \widehat{\psi}_n^{(p-1)}, \widehat{H}_n)$ is the Whittle estimator defined in Section 5.2. Note that the estimator $\widehat{H}_n^{\text{zero}}$ is given by

$$\widehat{H}_n^{\text{zero}} := \frac{1}{2|\log \delta_n|} \{\log b(\theta_0) - \log \nu_n^2(\widehat{\theta}_n) + \log \sigma_0^2\}.$$

Then, we obtain the following result.

Lemma 2. *Suppose Assumption 1 and $b(\theta_0)$ is known. Then, the following CLT holds*:

$$\mathcal{L}\left\{\begin{pmatrix} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ \sqrt{n}|\log \delta_n|(\widehat{H}_n^{\text{zero}} - H_0) \end{pmatrix} \Big| P_{\theta_0,\sigma_0}^{(n)} \right\} \overset{n\to\infty}{\to} \mathcal{N}(0, \mathcal{A}^{(0)}(\theta_0)),$$

where

$$\mathcal{A}^{(0)}(\theta) := \begin{pmatrix} I_p & 0_{p\times 1} \\ 0_{1\times p} & -\frac{1}{\sigma} \end{pmatrix} \begin{pmatrix} \mathcal{G}_p(\theta)^{-1} & 0_{p\times 1} \\ 0_{1\times p} & \frac{\sigma^2}{2} \end{pmatrix} \begin{pmatrix} I_p & 0_{p\times 1} \\ 0_{1\times p} & -\frac{1}{\sigma} \end{pmatrix}^* = \begin{pmatrix} \mathcal{G}_p(\theta)^{-1} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

In particular, the estimator $(\widehat{\psi}_n, \widehat{H}_n^{\text{zero}})$ enjoys the asymptotic normality property.

Proof. At first, the definition of $\widehat{H}_n^{\text{zero}}$, Lemma 1 and the delta method yield that

$$\begin{aligned} \sqrt{n}|\log \delta_n|(\widehat{H}_n^{\text{zero}} - H_0) &= -\frac{\sqrt{n}}{2}\left\{\log \frac{\nu_n^2(\widehat{\theta}_n)}{\delta_n^{2H_0} b(\theta_0)} - \log \sigma_0^2\right\} \\ &= -\frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0) + o_{P_{\theta_0,\sigma_0}^{(n)}}(1) \qquad \text{as } n \to \infty. \end{aligned} \tag{5.10}$$

Then, the following asymptotic expansion holds as $n \to \infty$:

$$\begin{pmatrix} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ \sqrt{n}|\log \delta_n|(\widehat{H}_n^{\text{zero}} - H_0) \end{pmatrix} = \begin{pmatrix} I_p & 0_{p\times 1} \\ 0_{1\times p} & -\frac{1}{\sigma_0} \end{pmatrix} \cdot \sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta_0 \\ \widetilde{\sigma}_n - \sigma_0 \end{pmatrix} + o_{P_{\theta_0,\sigma_0}^{(n)}}(1).$$

The first claim follows from Lemma 1 and the continuous mapping theorem. Moreover, the second claim also follows from the first one and the continuous mapping theorem. □

The assumption that $b(\theta_0)$ is known is too strict; however Lemma 2 has an important implication that a precise estimate of $b(\theta_0)$ leads to an efficient estimation of H. In fact, the convergence rate $\sqrt{n}|\log\delta_n|$ and the asymptotic variance 1/2 of H in Lemma 2 are the best possible for fractional Gaussian noise model with σ known; see Kawai [12].

Now, we remove the assumption that $b(\theta_0)$ is known. Consider to substitute the Whittle estimator $\widehat{\theta}_n$ to the true value θ_0 in $b(\theta_0)$. Namely, we define an estimator by

$$\widehat{H}_n^{\mathrm{one}} := \frac{1}{2|\log\delta_n|}\bigl\{\log b(\widehat{\theta}_n) - \log\nu_n^2(\widehat{\theta}_n) + \log\sigma_0^2\bigr\}.$$

The following result is proved in the same way as the proof of Lemma 2 by using Lemma 6 in the Appendix instead of Lemma 1.

Lemma 3. *Under Assumption* 1, *the following CLT holds*:

$$\mathcal{L}\left\{\begin{pmatrix}\sqrt{n}(\widehat{\theta}_n - \theta_0)\\ \sqrt{n}|\log\delta_n|\bigl(\widehat{H}_n^{\mathrm{one}} - H_0\bigr)\end{pmatrix}\Bigg| P_{\theta_0,\sigma_0}^{(n)}\right\} \overset{n\to\infty}{\to} \mathcal{N}\bigl(0, \mathcal{A}^{(1)}(\theta_0)\bigr),$$

where $\mathcal{A}^{(1)}(\theta)$ *is given in Lemma* 6 *in the* Appendix. *In particular, the estimator* $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{one}})$ *enjoys the asymptotic normality property.*

Although $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{one}})$ is rate-efficient, unfortunately, it is not asymptotically efficient; compare $\mathcal{A}^{(1)}(\theta)$ with $\mathcal{I}(\theta)^{-1}$. This is due to that $\widehat{\theta}_n$ that is substituted to $b(\theta_0)$ includes an estimate of H which is not rate-efficient. Therefore, its asymptotic variance contributes to that of $\widehat{H}_n^{\mathrm{one}}$. However, we can construct an asymptotically efficient estimator by using the estimator $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{one}})$. Namely, we define an estimator by

$$\widehat{H}_n^{\mathrm{two}} := \frac{1}{2|\log\delta_n|}\bigl\{\log b\bigl(\widehat{\theta}_n^{\mathrm{one}}\bigr) - \log\nu_n^2(\widehat{\theta}_n) + \log\sigma_0^2\bigr\},$$

where $\widehat{\theta}_n^{\mathrm{one}} := (\widehat{\psi}_n, \widehat{H}_n^{\mathrm{one}})$. Then, we can prove the asymptotic efficiency of the estimator $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{two}})$ as follows.

Theorem 4. *Under Assumption* 1, *the estimator* $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{two}})$ *is asymptotically efficient in Fisher's sense*, *that is*, *the estimator* $(\widehat{\psi}_n, \widehat{H}_n^{\mathrm{two}})$ *satisfies the CLT* (5.9).

Proof. In the same way as (5.10), the following asymptotic expansion holds as $n\to\infty$:

$$\begin{aligned}&\sqrt{n}|\log\delta_n|\bigl(\widehat{H}_n^{\mathrm{two}} - H_0\bigr)\\ &\quad= -\frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0) + \frac{\sqrt{n}}{2}\bigl(\log b\bigl(\widehat{\theta}_n^{\mathrm{one}}\bigr) - \log b(\theta_0)\bigr) + o_{P_{\theta_0,\sigma_0}^{(n)}}(1).\end{aligned} \tag{5.11}$$

Moreover, the second term in the right-hand side of (5.11) is expanded as follows:

$$\begin{aligned}\frac{\sqrt{n}}{2}\bigl(\log b\bigl(\widehat{\theta}_n^{\mathrm{one}}\bigr) - \log b(\theta_0)\bigr) &= -\frac{1}{2}a_p(\theta_0)^*\sqrt{n}\bigl(\widehat{\theta}_n^{\mathrm{one}} - \theta_0\bigr) + o_{P_{\theta_0,\sigma_0}^{(n)}}(1)\\ &= -\frac{1}{2}a_{p-1}(\theta_0)^*\sqrt{n}(\widehat{\psi}_n - \psi_0) + o_{P_{\theta_0,\sigma_0}^{(n)}}(1)\end{aligned} \tag{5.12}$$

as $n \to \infty$, where the $\sqrt{n}|\log \delta_n|$-consistency of the estimator $\widehat{H}_n^{\mathrm{one}}$ is used in the second equality. Therefore, we obtain the following asymptotic expansion from (5.11) and (5.12):

$$\begin{pmatrix} \sqrt{n}(\widehat{\psi}_n - \psi_0) \\ \sqrt{n}|\log \delta_n|(\widehat{H}_n^{\mathrm{two}} - H_0) \end{pmatrix} = \begin{pmatrix} I_{p-1} & 0_{p\times 1} \\ -\frac{1}{2} a_{p-1}(\theta_0)^* & -\frac{1}{\sigma_0} \end{pmatrix} \cdot \sqrt{n} \begin{pmatrix} \widehat{\psi}_n - \psi_0 \\ \widetilde{\sigma}_n - \sigma_0 \end{pmatrix} + o_{P^{(n)}_{\theta_0,\sigma_0}}(1)$$

as $n \to \infty$. As a result, the conclusion also follows from the same way as the last part of the proof of Lemma 6 in the Appendix. □

6. Simulation studies

6.1. Implementation and simulation studies of Whittle estimator

In this part, we explain how to implement the Whittle estimator for the Hurst and diffusion parameters and show its finite sample performance by the Monte-Carlo experiments. Although we perform simulation studies only for the case of the fGn, an extension to the other models shown in Section 4 is actually straightforward. Therefore, our implementation procedure is applicable for most of interesting self-similar stationary Gaussian noises appeared in practice, also see Remark 3 in Section 3. To the best of our knowledge, no explicit form of the Whittle likelihood for the spectral density of fGn is known so far. Therefore, some kinds of approximations for this likelihood are needed. In practical, we approximate this likelihood by the Riemann sum at the points of the Fourier frequencies $\lambda_j^n := 2\pi j/n,\ j = 1, \dots, n$ as follows:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{I_n(\lambda)}{g_H(\lambda)}\,d\lambda = \frac{1}{\pi}\int_{0}^{\pi} \frac{I_n(\lambda)}{g_H(\lambda)}\,d\lambda \approx \frac{1}{n}\sum_{j=1}^{\lfloor n/2 \rfloor} \frac{I_n(\lambda_j^n)}{g_H(\lambda_j^n)} \qquad \text{as } n \to \infty. \tag{6.1}$$

Then, the sum of (6.1) is effectively calculated by the well-known FFT (Fast Fourier Transform) algorithm. In particular, a calculation speed by the FFT algorithm is fastest when the number of data n is equal to 2^m for a certain $m \in \mathbb{N}$. Note that the symmetric property of the periodogram $I_n(\lambda)$ and the normalized spectral density $g_H(\lambda)$ with respect to $\lambda \in [-\pi, \pi]$ are used in the first equality of (6.1). Finally, estimated values of $\widehat{H}_n$ are obtained by substituting normalized Paxson's approximation spectral density, see Paxson [17] and a supplementary article Fukasawa and Takabatake [10] for more detail, for g_H and minimizing the sum in (6.1).

In our simulation studies, we use FGN package of R to generate fractional Gaussian noises and fix several parameters as follows: the true value of the diffusion parameter $\sigma_0 = 0.5$, the time horizon $T = 1$, the number of the approximations of the spectral density $k = 200$, the number of repetition of the Monte-Carlo experiments $m = 1000$. We vary the Hurst parameter H_0 and the length of the observation interval δ_n.

First, we examine the Whittle estimator $\widehat{H}_n$ constructed in Section 5.2. Table 1 reports the rescaled bias and MSE (Mean Square Error) of $\widehat{H}_n$:

$$\mathrm{Bias}_H(H_0, \delta_n) := \frac{1}{m}\sum_{j=1}^{m} \sqrt{n}(\widehat{H}_{n,j} - H_0),$$

$$\mathrm{MSE}_H(H_0, \delta_n) := \frac{1}{m}\sum_{j=1}^{m}\left\{\sqrt{n}(\widehat{H}_{n,j} - H_0)\right\}^2, \qquad \delta_n := 1/n,$$

where $\widehat{H}_{n,j}$ is the jth estimated value of $\widehat{H}_n$ in the Monte-Carlo experiments. Table 1 suggests that $\widehat{H}_n$ has a negative bias for all cases if δ_n is not so small. However, this bias seems to vanish for all cases as δ_n becomes smaller. In Figure 1, red and blue lines superimposed on the histogram represent the probability density of the theoretical asymptotic error distribution provided by Theorem 3 and that of a normal distribution with a mean $\mathrm{Bias}_H(H_0, \delta_n)$ and a variance $\mathrm{MSE}_H(H_0, \delta_n)$, respectively. These figures confirm that the Whittle estimator $\widehat{H}_n$ has finite sample performances which are consistent to the asymptotic results in Theorem 3.

Next, we examine the Whittle estimator $\widehat{\sigma}_n$ constructed in Section 5.2. Table 2 reports the rescaled bias and MSE of $\widehat{\sigma}_n$:

$$\mathrm{Bias}_\sigma(H_0, \delta_n) := \frac{1}{m}\sum_{j=1}^{m}\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_{n,j} - \sigma_0),$$

$$\mathrm{MSE}_\sigma(H_0, \delta_n) := \frac{1}{m}\sum_{j=1}^{m}\left\{\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_{n,j} - \sigma_0)\right\}^2, \qquad \delta_n := 1/n,$$

where $\widehat{\sigma}_{n,j}$ is the jth estimated value of $\widehat{\sigma}_n$ in the Monte-Carlo experiments.

In Figure 2, green and blue lines superimposed on the histogram represent the probability density of the theoretical asymptotic error distribution provided by Theorem 3 and that of a normal distribution with a mean $\mathrm{Bias}_\sigma(H_0, \delta_n)$ and a variance $\mathrm{MSE}_\sigma(H_0, \delta_n)$ respectively. In contrast to the case of the Whittle estimator $\widehat{H}_n$, the blue line is distinguishable from the green one for each H_0 even if δ_n is quite small. If δ_n is sufficiently small, however, they seem to fit well the red one, which represents the probability density of a centered normal distribution with a variance:

$$V_\sigma(H_0, \delta_n) := \frac{\sigma_0^2}{2|\log\delta_n|^2} + \sigma_0^2\left\{1 - \frac{1}{2|\log\delta_n|}\partial_H \log b(H_0)\right\}^2\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\partial_H \log g_{H_0}(\lambda)\right]^2 d\lambda\right\}^{-1}. \tag{6.2}$$

This variance, derived in Appendix C based on Theorem 3, converges to the limit variance as $n \to \infty$ but very slowly because $|\log\delta_n|$ very slowly diverges, for example $|\log(1/2^{18})| \approx 12.477$. Moreover, Figure 3 suggests that all values of the MSE in Table 2 fits well the values of $V_\sigma(H_0, 2^{-18})$, and those values are a bit smaller than the limit variances when H_0 is small and those are quite larger than them when H_0 is large. Note that the smallness (resp. largeness) of the values of (6.2) is caused from the positiveness (resp. negativeness) of values of $\partial_H \log b(H_0)$ and those size, see also Figure 4 and Table 3. Furthermore, we can also see that this phenomena happens even for the other asymptotically efficient estimators under high frequency observations which include the MLE given by Brouste and Fukasawa [2], see Appendix C for more detail. Therefore, careful treatment at this point is necessary for high frequency data analysis.

Table 1. Table of rescaled bias and MSE of the Whittle estimator $\widehat{H}_n$ with several values of δ_n and H_0

Table of $\widehat{H}_n$		$\delta_n = 1/2^8$	$\delta_n = 1/2^{13}$	$\delta_n = 1/2^{18}$
$H_0 = 0.1$	Bias	-0.0109	0.02458	0.01158
	MSE	0.1538	0.12914	0.12909
$H_0 = 0.2$	Bias	-0.0922	-0.01550	-0.01569
	MSE	0.2994	0.22613	0.23672
$H_0 = 0.3$	Bias	-0.1139	0.00238	0.02617
	MSE	0.3392	0.29208	0.30897
$H_0 = 0.4$	Bias	-0.1902	-0.01717	-0.00525
	MSE	0.4158	0.35315	0.36804
$H_0 = 0.5$	Bias	-0.2231	-0.07692	0.00596
	MSE	0.4899	0.41390	0.38944
$H_0 = 0.6$	Bias	-0.2854	-0.08426	-0.05027
	MSE	0.5568	0.44233	0.42982
$H_0 = 0.7$	Bias	-0.2304	-0.06622	-0.04082
	MSE	0.5435	0.42367	0.42112
$H_0 = 0.8$	Bias	-0.2475	-0.01685	-0.04263
	MSE	0.5771	0.45629	0.44602
$H_0 = 0.9$	Bias	-0.1861	0.02320	0.02074
	MSE	0.5700	0.47100	0.45968

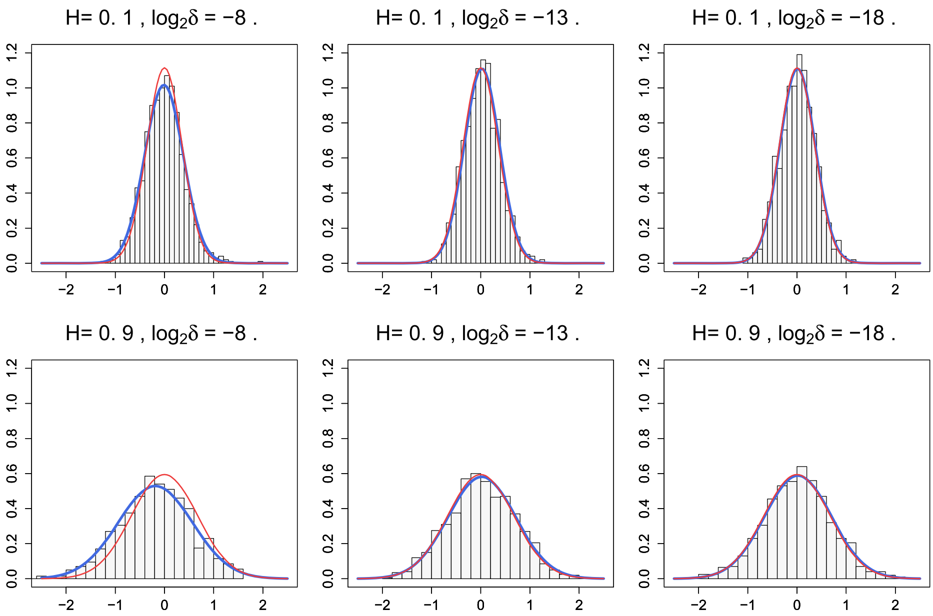

Figure 1. Histograms of $\{\sqrt{n}(\widehat{H}_{n,j} - H_0)\}_{j=1,\dots,m}$ and plot of probability densities of several normal distributions with different values of δ_n and H_0.

Table 2. Table of rescaled bias and MSE of the Whittle estimator $\widehat{\sigma}_n$ with several values of δ_n and H_0

Table of $\widehat{\sigma}_n$		$\delta_n = 1/2^8$	$\delta_n = 1/2^{13}$	$\delta_n = 1/2^{18}$
$H_0 = 0.1$	Bias	0.00851	0.016242	0.006291
	MSE	0.03294	0.027818	0.027987
$H_0 = 0.2$	Bias	−0.02560	−0.002493	−0.006068
	MSE	0.06293	0.050319	0.053721
$H_0 = 0.3$	Bias	−0.02707	0.008957	0.015442
	MSE	0.07307	0.068760	0.071989
$H_0 = 0.4$	Bias	−0.05837	0.000172	0.000279
	MSE	0.08863	0.086517	0.089748
$H_0 = 0.5$	Bias	−0.07483	−0.026713	0.004175
	MSE	0.10484	0.103572	0.098181
$H_0 = 0.6$	Bias	−0.10024	−0.032532	−0.022897
	MSE	0.12959	0.115973	0.113178
$H_0 = 0.7$	Bias	−0.07076	−0.024223	−0.018831
	MSE	0.15596	0.125084	0.118769
$H_0 = 0.8$	Bias	−0.06555	0.006731	−0.021478
	MSE	0.21589	0.165668	0.143015
$H_0 = 0.9$	Bias	0.12212	0.052899	0.019976
	MSE	0.99965	0.269497	0.207527

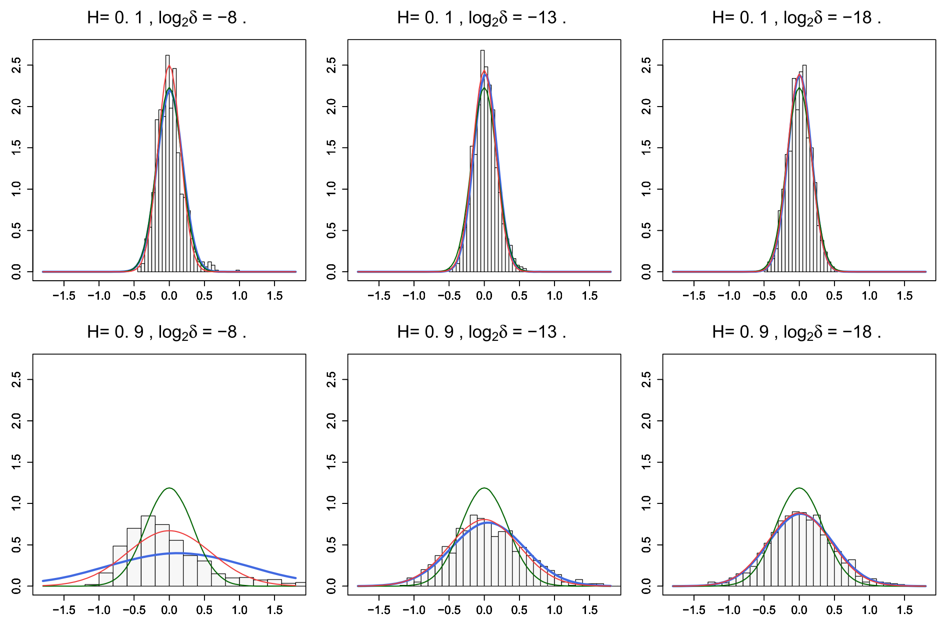

Figure 2. Histograms of $\{\sqrt{n}/|\log \delta_n|(\widehat{\sigma}_{n,j} - \sigma_0)\}_{j=1,\dots,m}$ and plot of probability densities of several normal distributions with different values of δ_n and H_0.

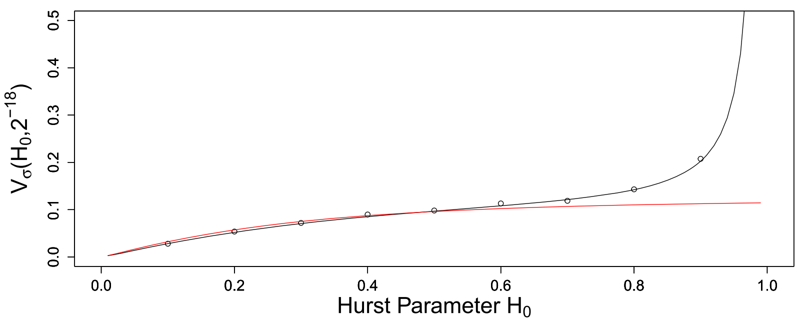

Figure 3. This figure represents a plot of $V_\sigma(H_0, 2^{-18})$ (black) and the theoretical asymptotic variance (red) with respect to H_0. Moreover, the values of MSE in Table 2 are superimposed on them.

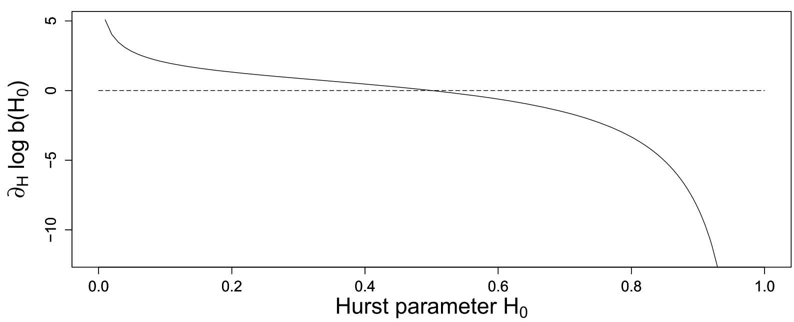

Figure 4. The solid line in this figure represents a plot of $\partial_H \log b(H_0)$ with respect to H_0.

Table 3. Values of $\partial_H \log b(H_0)$ with several different values of H_0 are summarized

$H_0 = 0.1$	$H_0 = 0.2$	$H_0 = 0.3$	$H_0 = 0.4$	$H_0 = 0.5$	$H_0 = 0.6$	$H_0 = 0.7$	$H_0 = 0.8$	$H_0 = 0.9$
2.026	1.321	0.874	0.464	−0.000	−0.613	−1.554	−3.331	−8.449

6.2. Simulation studies of estimator $\widehat{H}_n^{\text{two}}$

Table 4 reports the rescaled bias and MSE of $\widehat{H}_n^{\text{two}}$:

$$\mathrm{Bias}_2(H_0, \delta_n) := \frac{1}{m} \sum_{j=1}^{m} \sqrt{n} |\log \delta_n| \big(\widehat{H}_{n,j}^{\text{two}} - H_0\big),$$

$$\mathrm{MSE}_2(H_0,\delta_n) := \frac{1}{m}\sum_{j=1}^{m}\left\{\sqrt{n}|\log\delta_n|\left(\widehat{H}^{\mathrm{two}}_{n,j}-H_0\right)\right\}^2, \qquad \delta_n := 1/n,$$

where $\widehat{H}^{\mathrm{two}}_{n,j}$ be the jth estimated value of $\widehat{H}^{\mathrm{two}}_{n}$ in the Monte-Carlo experiments.

In Figure 5, green and blue lines superimposed on the histogram represent the probability density of the theoretical asymptotic error distribution provided by Theorem 4 and that of a normal distribution with a mean $\mathrm{Bias}_2(H_0,\delta_n)$ and a variance $\mathrm{MSE}_2(H_0,\delta_n)$ respectively. Figure 5 also suggests that the blue line is distinguishable from the green one for each H_0 even if δ_n is quite small. If δ_n is sufficiently small, however, they seem to fit well the red one, which represents the probability density of a centered normal distribution with a variance:

$$\begin{aligned} V_2(H_0,\delta_n) := {} & \frac{1}{2}\left(1+\frac{1}{2|\log\delta_n|}\partial_H \log b(H_0)\right)^2 \\ & + \frac{1}{|\log\delta_n|^2}\left(\frac{1}{2}\partial_H \log b(H_0)\right)^4\left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\partial_H \log g_{H_0}(\lambda)\right]^2 d\lambda\right)^{-1}. \end{aligned} \tag{6.3}$$

This variance, derived in Appendix C, converges to the limit variance as $n\to\infty$ but very slowly. Interestingly, Figure 6 suggests that all values of the MSE in Table 4 fits well the values of $V_2(H_0,2^{-18})$.

In the rest of this section, we attempt to answer the following two questions by using the approximation formula (6.3):

1. Why values of $\mathrm{MSE}_2(H_0,\delta_n)$ in Table 4 is larger (resp. smaller) than those of the theoretical asymptotic variance when $H_0 < 1/2$ (resp. $H_0 > 1/2$ except for $H_0 = 0.9$)?
2. Why the value of $\mathrm{MSE}_2(H_0,\delta_n)$ when $H_0 = 0.9$ is extremely large?

At first, we consider the first question. The phenomena of the first question seems to be caused from the effect of the first term of (6.3). Indeed, $(2|\log\delta_n|)^{-1}\partial_H \log b(H_0)$ in this term takes relatively large positive (resp. negative) value in these cases because $|\log\delta_n|$ very slowly diverges and $\partial_H \log b(H_0)$ takes relatively large positive (resp. negative) value, see also Figure 4 and Table 3. Note that the second term of (6.3) in these cases has almost no influence to the value of (6.3) if δ_n is quite small.

Next, we consider the second question. Contrast to the previous one, the phenomena of the second question seems to be caused from the effect of the second term of (6.3). Indeed, $|\log\delta_n|^{-2}(\frac{1}{2}\partial_H \log b(H_0))^4$ in the second term of (6.3) takes extremely large positive value when $H_0 = 0.9$ because $|\log\delta_n|$ slowly diverges and $\partial_H \log b(H_0)$ extremely large positive value, see also Figure 4 and Table 3. Then, the positive effect of the value of $\mathrm{MSE}_2(H_0,\delta_n)$ caused from the first term of (6.3) is relatively larger than the negative one caused from the second term of (6.3), which is mentioned in the answer of the previous question. Furthermore, we can also see that this phenomena happens even for the other asymptotically efficient estimator in this setting. See Appendix C for more detail. Therefore, careful treatment at this point is also necessary for high frequency data analysis.

Table 4. Table of rescaled bias and MSE of the estimator $\widehat{H}_n^{\text{two}}$ with deferent values of δ_n and H_0

Table of $\widehat{H}_n^{\text{two}}$		$\delta_n = 1/2^8$	$\delta_n = 1/2^{13}$	$\delta_n = 1/2^{18}$
$H_0 = 0.1$	Bias	−0.0056	−0.04572	−0.00617
	MSE	0.6912	0.67937	0.60245
$H_0 = 0.2$	Bias	0.0994	0.00822	−0.00135
	MSE	0.6400	0.58856	0.57736
$H_0 = 0.3$	Bias	0.0226	−0.01569	−0.02641
	MSE	0.5861	0.52786	0.55571
$H_0 = 0.4$	Bias	0.0248	0.00459	−0.01623
	MSE	0.5307	0.52183	0.51395
$H_0 = 0.5$	Bias	0.0814	−0.01789	0.03082
	MSE	0.5470	0.47506	0.51824
$H_0 = 0.6$	Bias	0.0556	0.01956	−0.00108
	MSE	0.4966	0.46402	0.46335
$H_0 = 0.7$	Bias	0.0403	0.02279	0.00173
	MSE	0.3873	0.41353	0.42827
$H_0 = 0.8$	Bias	−0.0752	0.01975	0.01357
	MSE	0.3480	0.37197	0.41573
$H_0 = 0.9$	Bias	−0.4978	0.05361	0.03233
	MSE	6.1539	2.00298	1.15611

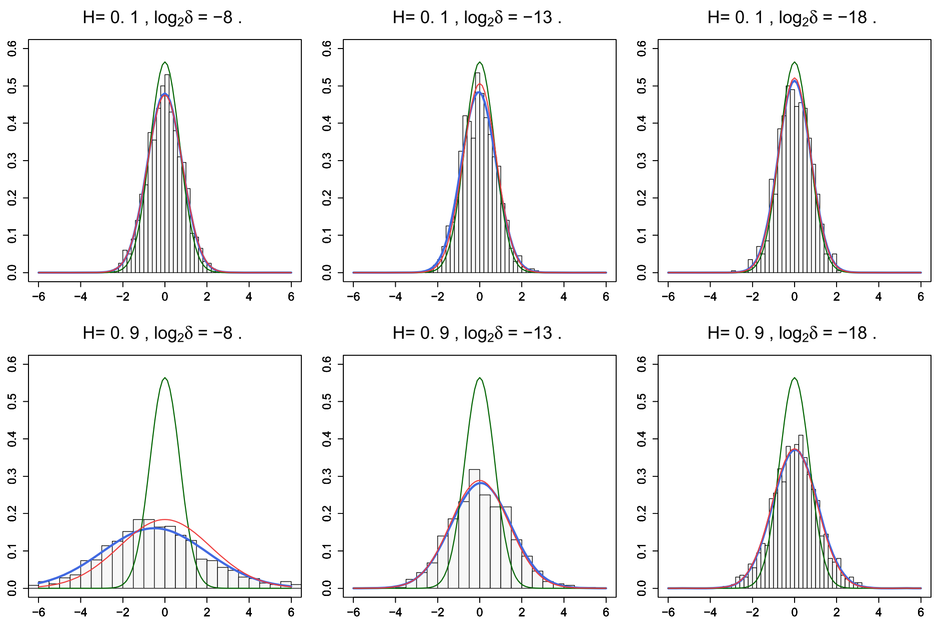

Figure 5. Histograms of $\{\sqrt{n}|\log\delta_n|(\widehat{H}_{n,j}^{\text{two}} - H_0)\}_{j=1,\dots,m}$ and plot of probability densities of several normal distributions with different values of δ_n and H_0.

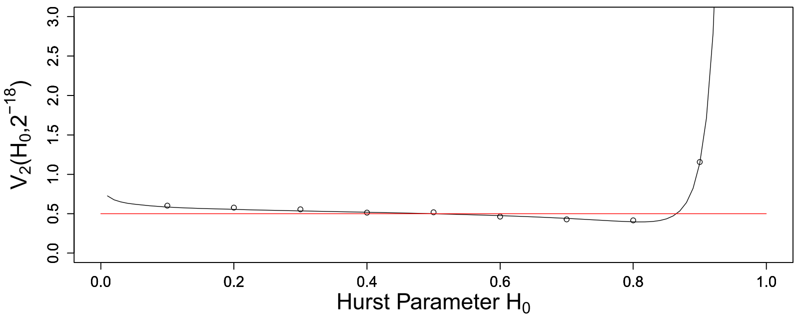

Figure 6. This figure represents a plot of $V_2(H_0, 2^{-18})$ (black) and the theoretical asymptotic variance (red) with respect to H_0. Moreover, the values of MSE in Table 4 are superimposed on them.

Appendix A: Preliminary lemmas

In this appendix, we show several lemmas used in the proof of main results.

Lemma 4. *Let* $f : \Theta^\dagger \times [-\pi, \pi] \to [-\infty, \infty]$ *satisfy* (S.2). *Then, the matrices* $\mathcal{F}_q(\theta)$ *and* $\mathcal{G}_q(\theta)$ *are connected with the following relation*:

$$\mathcal{G}_q(\theta) = \mathcal{F}_q(\theta) - \frac{1}{2} a_q(\theta) a_q(\theta)^* \qquad \text{for each } q = 1, \dots, p+1 \text{ and } \theta \in \Theta^\dagger.$$

Proof. Fix $q \in \{1, \dots, p+1\}$. Note that the following equality holds for any $\theta \in \Theta^\dagger$:

$$\frac{\partial}{\partial \theta_j} \log g_\theta(\lambda) = \frac{\partial}{\partial \theta_j} \log f_\theta(\lambda) - \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_j} \log f_\theta(\lambda) \, d\lambda. \tag{A.1}$$

Then, it easily follows from (A.1) that (i, j)-component of the matrix $\mathcal{G}_q(\theta)$ is equal to

$$\begin{aligned} &\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f_\theta(\lambda) \frac{\partial}{\partial \theta_j} \log f_\theta(\lambda) \, d\lambda \\ &\quad - \frac{1}{2} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_i} \log f_\theta(\lambda) \, d\lambda \right) \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_j} \log f_\theta(\lambda) \, d\lambda \right) \end{aligned}$$

for any $\theta \in \Theta^\dagger$. Therefore, the conclusion follows. □

Lemma 5. *Let* $\Theta^\dagger := \Theta \times \Sigma$ *given in Assumption* 1 *and* $f : \Theta^\dagger \times [-\pi, \pi] \to [-\infty, \infty]$ *satisfy* (S.2)–(S.3). *Then, the following equality holds for each* $q = 1, \dots, p$ *and* $\theta \in \Theta^\dagger$:

$$\begin{pmatrix} \mathcal{F}_q(\theta) & a_q(\theta) \\ a_q(\theta)^* & 2 \end{pmatrix}^{-1} = \begin{pmatrix} \mathcal{G}_q(\theta)^{-1} & -\frac{1}{2}\mathcal{G}_q(\theta)^{-1} a_q(\theta) \\ -\frac{1}{2} a_q(\theta)^* \mathcal{G}_q(\theta)^{-1} & \frac{1}{2} + \frac{1}{4} a_q(\theta)^* \mathcal{G}_q(\theta)^{-1} a_q(\theta) \end{pmatrix}.$$

Proof. In order to obtain the conclusion, it suffices to prove that

$$\begin{pmatrix} \mathcal{F}_q(\theta) & a_q(\theta) \\ a_q(\theta)^* & 2 \end{pmatrix} \begin{pmatrix} \mathcal{G}_q(\theta)^{-1} & -\frac{1}{2}\mathcal{G}_q(\theta)^{-1}a_q(\theta) \\ -\frac{1}{2}a_q(\theta)^*\mathcal{G}_q(\theta)^{-1} & \frac{1}{2}+\frac{1}{4}a_q(\theta)^*\mathcal{G}_q(\theta)^{-1}a_q(\theta) \end{pmatrix} = I_q \tag{A.2}$$

holds for each $q = 1, \dots, p$ and $\theta \in \Theta^\dagger$, where I_q is the unit matrix of size q. The equality (A.2) easily follows from elementary computations using the block matrix and the following equality derived from Lemma 4:

$$\mathcal{F}_q(\theta)\mathcal{G}_q(\theta)^{-1} = I_q + \frac{1}{2}a_q(\theta)a_q(\theta)^*\mathcal{G}_q(\theta)^{-1}$$

for each $q = 1, \dots, p$ and $\theta \in \Theta^\dagger$. Therefore, we finish the proof. □

Lemma 6. *Under Assumption 1, the following CLT holds.*

$$\mathcal{L}\left\{ \begin{pmatrix} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ \frac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n - \sigma_0) \end{pmatrix} \Big| P^{(n)}_{\theta_0,\sigma_0} \right\} \overset{n\to\infty}{\to} \mathcal{N}\big(0, \mathcal{A}^{(1)}(\theta_0)\big),$$

where $\acute{\sigma}_n := \{(\delta_n^{2H_0} b(\widehat{\theta}_n))^{-1}\nu_n^2(\widehat{\theta}_n)\}^{1/2}$ *and*

$$\begin{aligned} \mathcal{A}^{(1)}(\theta) &:= \begin{pmatrix} I_p & 0_{p\times 1} \\ -\frac{1}{2}a_p(\theta)^* & \frac{1}{\sigma} \end{pmatrix} \begin{pmatrix} \mathcal{G}_p(\theta)^{-1} & 0_{p\times 1} \\ 0_{1\times p} & \frac{\sigma^2}{2} \end{pmatrix} \begin{pmatrix} I_p & 0_{p\times 1} \\ -\frac{1}{2}a_p(\theta)^* & \frac{1}{\sigma} \end{pmatrix}^* \\ &= \begin{pmatrix} \mathcal{G}_p(\theta)^{-1} & -\frac{1}{2}\mathcal{G}_p(\theta)^{-1}a_p(\theta) \\ -\frac{1}{2}a_p(\theta)^*\mathcal{G}_p(\theta)^{-1} & \frac{1}{2}+\frac{1}{4}a_p(\theta)^*\mathcal{G}_p(\theta)^{-1}a_p(\theta) \end{pmatrix} = \begin{pmatrix} \mathcal{F}_p(\theta) & a_p(\theta) \\ a_p(\theta)^* & 2 \end{pmatrix}^{-1}. \end{aligned}$$

Proof. At first, we obtain the following asymptotic expansion from Lemma 1 and the delta method:

$$\begin{aligned} \sqrt{n}(\log\acute{\sigma}_n - \log\sigma_0) &= \sqrt{n}(\log\widetilde{\sigma}_n - \log\sigma_0) - \frac{\sqrt{n}}{2}\big(\log b(\widehat{\theta}_n) - \log b(\theta_0)\big) \\ &= \frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0) - \frac{\sqrt{n}}{2}\big(\log b(\widehat{\theta}_n) - \log b(\theta_0)\big) + o_{P^{(n)}_{\theta_0,\sigma_0}}(1) \end{aligned} \tag{A.3}$$

as $n \to \infty$. Moreover, the second term in (A.3) is expanded by

$$\begin{aligned} \log b(\widehat{\theta}_n) - \log b(\theta_0) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \log f_{\widehat{\theta}_n}(\lambda)\,d\lambda - \frac{1}{2\pi}\int_{-\pi}^{\pi}\log f_{\theta_0}(\lambda)\,d\lambda \\ &= a_p(\theta_0)^*(\widehat{\theta}_n - \theta_0) + o_{P^{(n)}_{\theta_0,\sigma_0}}\big(n^{-\frac{1}{2}}\big) \qquad \text{as } n\to\infty, \end{aligned} \tag{A.4}$$

where Lemma 1 and the delta method are used again in the second equality. Then, we also obtain the following asymptotic expansion from (A.3), (A.4) and the delta method:

$$\begin{pmatrix} \sqrt{n}(\widehat{\theta}_n - \theta_0) \\ \dfrac{\sqrt{n}}{\sigma_0}(\acute{\sigma}_n - \sigma_0) \end{pmatrix} = \begin{pmatrix} I_p & 0_{p\times 1} \\ -\dfrac{1}{2}a_p(\theta_0)^* & \dfrac{1}{\sigma_0} \end{pmatrix} \cdot \sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta_0 \\ \widetilde{\sigma}_n - \sigma_0 \end{pmatrix} + o_{P^{(n)}_{\theta_0,\sigma_0}}(1) \qquad \text{as } n \to \infty.$$

The conclusion follows from Lemma 1, Lemma 5 and the continuous mapping theorem. □

Appendix B: LAN property under high frequency observations

In this appendix, we show several extensions of the results in Kawai [12] and Brouste and Fukasawa [2] into our model framework without proof. These results can be proved in the similar argument. At first, we show the extension of the result in Kawai [12]. Set a rate matrix $\bar{\phi}_n(\theta, \sigma)$ as follows.

$$\bar{\phi}_n \equiv \bar{\phi}_n(\theta, \sigma) := \begin{pmatrix} \dfrac{1}{\sqrt{n}} I_{p-1} & 0_{p-1\times 1} & 0_{p-1\times 1} \\ 0_{1\times p-1} & \dfrac{1}{\sqrt{n}\log \delta_n} & 0 \\ 0_{1\times p-1} & 0 & \dfrac{1}{\sqrt{n}} \end{pmatrix}.$$

Then, we obtain the following weak LAN property.

Theorem 5. *Suppose Assumption 1. The family of measures $\{P^{(n)}_{\theta,\sigma}; (\theta, \sigma) \in \Theta \times \Sigma\}$ is LAN at any points $(\theta_0, \sigma_0) \in \Theta \times \Sigma$ for the rate matrix $\bar{\phi}_n(\theta_0, \sigma_0)$ in a weak sense, that is, the log-likelihood ratio admits the following representation for any $u \in \mathbb{R}^{p+1}$:*

$$\log \frac{dP^{(n)}_{(\theta_0,\sigma_0)+\bar{\phi}_n(\theta_0,\sigma_0)u}}{dP^{(n)}_{\theta_0,\sigma_0}} = \langle u, \bar{\zeta}_n(\theta_0, \sigma_0)\rangle - \frac{1}{2}\langle \mathcal{I}(\theta_0, \sigma_0)u, u\rangle + \bar{r}_n(\theta_0, \sigma_0),$$

where

$$\bar{\zeta}_n(\theta_0, \sigma_0) \to \mathcal{N}\big(0, \mathcal{I}(\theta_0, \sigma_0)\big), \qquad \bar{r}_n(\theta_0, \sigma_0) \to 0,$$

in law under $P^{(n)}_{\theta_0,\sigma_0}$ as $n \to \infty$ and the matrix $\mathcal{I}(\theta, \sigma)$ is given by

$$\begin{pmatrix} \mathcal{F}_{p-1}(\theta) & a_{p-1}(\theta) & \dfrac{1}{\sigma} a_{p-1}(\theta) \\ a_{p-1}(\theta)^* & 2 & \dfrac{2}{\sigma} \\ \dfrac{1}{\sigma} a_{p-1}(\theta)^* & \dfrac{2}{\sigma} & \dfrac{2}{\sigma^2} \end{pmatrix}.$$

In particular, the asymptotic Fisher information matrix $\mathcal{I}(\theta,\sigma)$ is singular unless either the Hurst parameter H or the diffusion parameter σ is known.

Next, we show the extension of the result in Brouste and Fukasawa [2]. Here, we introduce a certain class of rate matrices.

Assumption 2. Consider a matrix

$$\phi_n \equiv \phi_n(\theta,\sigma) := \begin{pmatrix} \phi_{n,1} & 0_{p-1\times 2} \\ 0_{2\times p-1} & \phi_{n,2} \end{pmatrix},$$

where

$$\phi_{n,1} \equiv \phi_{n,1}(\theta,\sigma) := \operatorname{diag}\bigl(d_n^{(1)},\dots,d_n^{(p-1)}\bigr), \qquad \phi_{n,2} \equiv \phi_{n,2}(\theta,\sigma) := \begin{pmatrix} \alpha_n & \widehat{\alpha}_n \\ \beta_n & \widehat{\beta}_n \end{pmatrix},$$

with the following properties:

1. $|\phi_{n,1}| = d_n^{(1)}\cdots d_n^{(p-1)} \neq 0$ and $|\phi_{n,2}| = \alpha_n\widehat{\beta}_n - \widehat{\alpha}_n\beta_n \neq 0$.
2. $\alpha_n\sqrt{n} \to \alpha$ for some $\alpha \in \mathbb{R}$.
3. $\widehat{\alpha}_n\sqrt{n} \to \widehat{\alpha}$ for some $\widehat{\alpha} \in \mathbb{R}$.
4. For $j = 1,\dots,p-1$, $d_n^{(j)}\sqrt{n} \to d^{(j)}$ for some $d^{(j)} \in \mathbb{R}\setminus\{0\}$.
5. $\gamma_n := \alpha_n\sqrt{n}\log\delta_n + \beta_n\sqrt{n}\sigma^{-1} \to \gamma$ for some $\gamma \in \mathbb{R}$.
6. $\widehat{\gamma}_n := \widehat{\alpha}_n\sqrt{n}\log\delta_n + \widehat{\beta}_n\sqrt{n}\sigma^{-1} \to \widehat{\gamma}$ for some $\widehat{\gamma} \in \mathbb{R}$.
7. $d^{(1)}\cdots d^{(p-1)} \neq 0$ and $\alpha\widehat{\gamma} - \widehat{\alpha}\gamma \neq 0$.

Then, we obtain the following LAN property.

Theorem 6. *Suppose Assumption 1. The family of measures $\{P_{\theta,\sigma}^{(n)}; (\theta,\sigma) \in \Theta\times\Sigma\}$ is LAN at any points $(\theta_0,\sigma_0) \in \Theta\times\Sigma$ for the rate matrix $\phi_n(\theta_0,\sigma_0)$ satisfying Assumption 2, that is, the log-likelihood ratio admits the following representation for any $u \in \mathbb{R}^{p+1}$:*

$$\log \frac{dP^{(n)}_{(\theta_0,\sigma_0)+\phi_n(\theta_0,\sigma_0)u}}{dP^{(n)}_{\theta_0,\sigma_0}} = \bigl\langle u, \zeta_n(\theta_0,\sigma_0)\bigr\rangle - \frac{1}{2}\bigl\langle \mathcal{J}(\theta_0,\sigma_0)u, u\bigr\rangle + r_n(\theta_0,\sigma_0),$$

where

$$\zeta_n(\theta_0,\sigma_0) \to \mathcal{N}\bigl(0,\mathcal{J}(\theta_0,\sigma_0)\bigr), \qquad r_n(\theta_0,\sigma_0) \to 0,$$

in law under $P^{(n)}_{\theta_0,\sigma_0}$ as $n\to\infty$ and the matrix $\mathcal{J}(\theta,\sigma)$ is given by

$$\begin{pmatrix} D & 0_{p-1\times 2} \\ 0_{2\times p-1} & E \end{pmatrix} \begin{pmatrix} \mathcal{F}_p(\theta) & a_p(\theta) \\ a_p(\theta)^* & 2 \end{pmatrix} \begin{pmatrix} D & 0_{p-1\times 2} \\ 0_{2\times p-1} & E \end{pmatrix}^*,$$

where

$$D \equiv D(\theta,\sigma) := \mathrm{diag}\big(d^{(1)},\dots,d^{(p-1)}\big), \qquad E \equiv E(\theta,\sigma) := \begin{pmatrix} \alpha & \gamma \\ \widehat{\alpha} & \widehat{\gamma} \end{pmatrix}.$$

In particular, the matrix $\mathcal{J}(\theta,\sigma)$ is nondegenerate.

Appendix C: Approximation formula of theoretical asymptotic variance

In this section, finite sample modifications of the asymptotic variances of asymptotically efficient estimators given in Section 5 is derived for the 2-dimensional parameter (H,σ).

C.1. Case of Whittle estimator $\widehat{\sigma}_n$

A finite sample modification of the asymptotic variance of the Whittle estimator for the diffusion parameter is given in this subsection. At first note that, if the length of the observation interval δ_n is quite small, Lemma 1 yields that the joint distribution of $\sqrt{n}(\widehat{H}_n - H_0)$ and $\sqrt{n}(\widetilde{\sigma}_n - \sigma_0)$ under $P^{(n)}_{H_0,\sigma_0}$ is approximated by that of centered normal random variables Z_1 and Z_2 given by

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim \mathcal{N}\left(0, \mathrm{diag}\left\{\left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\big[\partial_H \log g_{H_0}(\lambda)\big]^2 d\lambda\right)^{-1}, \frac{\sigma_0^2}{2}\right\}\right), \tag{C.1}$$

where Z_1 and Z_2 are independent because the joint distribution of (Z_1, Z_2) is a bivariate normal random variable without correlation.

We derive a finite sample modification of the asymptotic variance of $\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_n - \sigma_0)$. At first, $\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_n - \sigma_0)$ is asymptotically expanded as follows by using Lemma 1, Theorem 3 and the delta method:

$$\begin{aligned}
&\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_n - \sigma_0) \\
&\quad = \frac{\sqrt{n}}{|\log\delta_n|}\sigma_0(\log\widehat{\sigma}_n - \log\sigma_0) + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{|\log\delta_n|}{\sqrt{n}}\right) \\
&\quad = \frac{\sqrt{n}}{|\log\delta_n|}\Big\{(\log\widetilde{\sigma}_n - \log\sigma_0) - \log\delta_n(\widehat{H}_n - H_0) \\
&\qquad - \frac{1}{2}\big(\log b(\widehat{H}_n) - \log b(H_0)\big)\Big\} \\
&\qquad + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{|\log\delta_n|}{\sqrt{n}}\right)
\end{aligned} \tag{C.2}$$

$$= \frac{\sqrt{n}}{\sigma_0 |\log \delta_n|}(\widetilde{\sigma}_n - \sigma_0) + \sqrt{n}(\widehat{H}_n - H_0) - \frac{1}{2|\log \delta_n|}\partial_H \log b(H_0)\sqrt{n}(\widehat{H}_n - H_0) + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{|\log \delta_n|}{\sqrt{n}}\right)$$

as $n \to \infty$. Then, the asymptotic variance of $\frac{\sqrt{n}}{|\log \delta_n|}(\widehat{\sigma}_n - \sigma_0)$ is approximately derived from ignoring the remainder term of (C.2) and using the asymptotic independence of $\sqrt{n}(\widehat{H}_n - H_0)$ and $\sqrt{n}(\widetilde{\sigma}_n - \sigma_0)$ under $P^{(n)}_{H_0,\sigma_0}$ and (C.1) as follows:

$$\begin{aligned}
&\mathrm{Var}^{(n)}_{H_0,\sigma_0}\left[\frac{\sqrt{n}}{|\log \delta_n|}(\widehat{\sigma}_n - \sigma_0)\right] \\
&\approx \frac{1}{|\log \delta_n|^2}\mathrm{Var}^{(n)}_{H_0,\sigma_0}\left[\sqrt{n}(\widetilde{\sigma}_n - \sigma_0)\right] \\
&\quad + \sigma_0^2\left\{1 - \frac{1}{2|\log \delta_n|}\partial_H \log b(H_0)\right\}^2 \mathrm{Var}^{(n)}_{H_0,\sigma_0}\left[\sqrt{n}(\widehat{H}_n - H_0)\right] \\
&\approx \frac{\sigma_0^2}{2|\log \delta_n|^2} + \sigma_0^2\left\{1 - \frac{1}{2|\log \delta_n|}\partial_H \log b(H_0)\right\}^2 \left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[\partial_H \log g_{H_0}(\lambda)\right]^2 d\lambda\right\}^{-1}.
\end{aligned}$$

C.2. Case of two-step estimator $\widehat{H}_n^{\mathrm{two}}$

At first, an asymptotic expansion of $\sqrt{n}|\log \delta_n|(\widehat{H}_n^{\mathrm{one}} - H_0)$ is obtained in the similar way as the proof of Lemma 3 and Lemma 6 in the Appendix as follows:

$$\begin{aligned}
&\sqrt{n}|\log \delta_n|\big(\widehat{H}_n^{\mathrm{one}} - H_0\big) \\
&= -\frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0) + \frac{1}{2}\partial_H \log b(H_0)\sqrt{n}(\widehat{H}_n - H_0) \\
&\quad + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{1}{\sqrt{n}}\right) \qquad \text{as } n \to \infty.
\end{aligned} \tag{C.3}$$

A similar argument in the above yields that

$$\begin{aligned}
&\sqrt{n}|\log \delta_n|\big(\widehat{H}_n^{\mathrm{two}} - H_0\big) \\
&= -\frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0) + \frac{1}{2|\log \delta_n|}\partial_H \log b(H_0)\sqrt{n}|\log \delta_n|\big(\widehat{H}_n^{\mathrm{one}} - H_0\big) \\
&\quad + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{1}{\sqrt{n}}\right) \qquad\qquad\qquad (\mathrm{C}.4) \\
&= -\left(1 + \frac{1}{2|\log \delta_n|}\partial_H \log b(H_0)\right)\frac{\sqrt{n}}{\sigma_0}(\widetilde{\sigma}_n - \sigma_0)
\end{aligned}$$

$$+\frac{1}{|\log\delta_n|}\left(\frac{1}{2}\partial_H \log b(H_0)\right)^2 \cdot \sqrt{n}(\widehat{H}_n - H_0) + O_{P^{(n)}_{H_0,\sigma_0}}\left(\frac{1}{\sqrt{n}}\right)$$

as $n \to \infty$. Note that the delta method and the above CLT are used in the third equality, and (C.3) is used in the fourth one. Therefore, an approximate variance of $\frac{\sqrt{n}}{|\log\delta_n|}(\widehat{\sigma}_n - \sigma_0)$ is derived from ignoring the remainder term of (C.4) and using the asymptotic independence of $\sqrt{n}(\widehat{H}_n - H_0)$ and $\sqrt{n}(\widetilde{\sigma}_n - \sigma_0)$ under $P^{(n)}_{H_0,\sigma_0}$ and (C.1) as follows:

$$\begin{aligned}
&\mathrm{Var}^{(n)}_{H_0,\sigma_0}\big[\sqrt{n}|\log\delta_n|\big(\widehat{H}^{\mathrm{two}}_n - H_0\big)\big] \\
&\approx \left(1+\frac{1}{2|\log\delta_n|}\partial_H \log b(H_0)\right)^2 \frac{1}{\sigma_0^2}\mathrm{Var}^{(n)}_{H_0,\sigma_0}\big[\sqrt{n}(\widetilde{\sigma}_n - \sigma_0)\big] \\
&\quad + \frac{1}{|\log\delta_n|^2}\left(\frac{1}{2}\partial_H \log b(H_0)\right)^4 \mathrm{Var}^{(n)}_{H_0,\sigma_0}\big[\sqrt{n}(\widehat{H}_n - H_0)\big] \\
&\approx \frac{1}{2}\left(1+\frac{1}{2|\log\delta_n|}\partial_H \log b(H_0)\right)^2 \\
&\quad + \frac{1}{|\log\delta_n|^2}\left(\frac{1}{2}\partial_H \log b(H_0)\right)^4 \left(\frac{1}{4\pi}\int_{-\pi}^{\pi}\big[\partial_H \log g_{H_0}(\lambda)\big]^2 d\lambda\right)^{-1}.
\end{aligned}$$

C.3. Additional remark for maximum likelihood estimator

In this subsection, we additionally remark about the generality of the finite sample modifications of asymptotic variances derived in the previous two subsection. Actually, the same argument discussed in the above holds true for more general class of asymptotically efficient estimators under high frequency observations in each setting. In the rest of this section, we briefly summarize them. First, we consider the case that both parameters (θ, σ) are unknown. It is easily shown that for any asymptotically efficient estimator $(\widehat{\theta}_n, \widehat{\sigma}_n) \equiv (\widehat{\psi}_n, \widehat{H}_n, \widehat{\sigma}_n)$ under high frequency observations in this case, that is, it satisfies (5.5) in Theorem 3, a random variable $(\widehat{\theta}_n, \widetilde{\sigma}_n)$ defined by

$$\widetilde{\sigma}_n := \frac{\delta_n^{\widehat{H}_n} b(\widehat{\theta}_n)}{\delta_n^{H_0} b(\theta_0)}\widehat{\sigma}_n, \qquad H \in (0,1], n \in \mathbb{N},$$

also satisfies (5.4) in Lemma 1. Therefore, the same argument in Appendix C.1 holds true. Next, we consider the case that the diffusion parameter σ is known. Define one-step and two-step estimators for the Hurst parameter H defined by

$$\widehat{H}^{\mathrm{one}}_n := \frac{1}{2|\log\delta_n|}\big\{\log b(\widehat{\theta}_n) - \log\big(b(\widehat{\theta}_n)\delta_n^{\widehat{H}_n}\widehat{\sigma}_n\big) + \log\sigma_0^2\big\},$$

$$\widehat{H}^{\mathrm{two}}_n := \frac{1}{2|\log\delta_n|}\big\{\log b\big(\widehat{\theta}^{\mathrm{one}}_n\big) - \log\big(b(\widehat{\theta}_n)\delta_n^{\widehat{H}_n}\widehat{\sigma}_n\big) + \log\sigma_0^2\big\},$$

where $(\widehat{\theta}_n, \widehat{\sigma}_n) \equiv (\widehat{\psi}_n, \widehat{H}_n, \widehat{\sigma}_n)$ is any asymptotically efficient estimator in the previous case and $\widehat{\theta}_n^{\text{one}} := (\widehat{\psi}_n, \widehat{H}_n^{\text{one}})$. Therefore, the same argument in . C.2 also holds true. As a result, the finite sample modifications of the asymptotic variances are also applicable for the MLE given by Brouste and Fukasawa [2] and the one-step and two-step estimators based on the MLE. These results have an important implication that all asymptotically efficient estimators under high frequency observations also suffer from the same problem of finite sample efficiency loss documented in Section 6 and they explain the reason why their phenomena happen.

Acknowledgements

The authors are grateful to anonymous referees for fruitful comments and suggestion for improvements. The second author was supported by JSPS KAKENHI Grant Number 17J04605.

Supplementary Material

Supplement to “Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations” (DOI: 10.3150/18-BEJ1039SUPP; .pdf). We explain how to implement spectral densities of self-similar stationary Gaussian noises and their derivatives with respect to parameters for more detail. This procedure is applicable for all examples shown in Section 4 of the original article.

References

[1] Bouchaud, J.-P. and Georges, A. (1990). Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **195** 127–293. MR1081295

[2] Brouste, A. and Fukasawa, M. (2018). Local asymptotic normality property for fractional Gaussian noise under high frequency observations. *Ann. Statist.* To appear.

[3] Brouste, A. and Iacus, S.M. (2013). Parameter estimation for the discretely observed fractional Ornstein–Uhlenbeck process and the Yuima R package. *Comput. Statist.* **28** 1529–1547. MR3120827

[4] Coeurjolly, J.-F. (2001). Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.* **4** 199–227. MR1856174

[5] Cohen, S., Gamboa, F., Lacaux, C. and Loubes, J.-M. (2013). LAN property for some fractional type Brownian motion. *ALEA Lat. Am. J. Probab. Math. Stat.* **10** 91–106. MR3083920

[6] Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766. MR1026311

[7] Dahlhaus, R. (2006). Correction: “Efficient parameter estimation for self-similar processes” [Ann. Statist. **17** (1989), no. 4, 1749–1766; MR1026311]. *Ann. Statist.* **34** 1045–1047. MR2283403

[8] Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. New York: Springer. MR0812272

[9] Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532. MR0840512

[10] Fukasawa, M. and Takabatake, T. (2019). Supplement to “Asymptotically efficient estimators for self-similar stationary Gaussian noises under high frequency observations.” DOI:10.3150/18-BEJ1039SUPP.

[11] Ibragimov, I.A. and Has'minskiĭ, R.Z. (1981). *Statistical Estimation*: *Asymptotic Theory*. *Applications of Mathematics* **16**. Berlin: Springer. MR0620321

[12] Kawai, R. (2013). Fisher information for fractional Brownian motion under high-frequency discrete sampling. *Comm*. *Statist*. *Theory Methods* **42** 1628–1636. MR3041490

[13] Kolmogoroff, A.N. (1940). Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C*. *R*. (*Doklady*) *Acad*. *Sci*. *URSS* (*N*.*S*.) **26** 115–118. MR0003441

[14] Lieberman, O., Rosemarin, R. and Rousseau, J. (2009). Asymptotic theory for maximum likelihood estimation in stationary fractional Gaussian processes, under short-, long- and intermediate memory. *CiteSeerX*.

[15] Lieberman, O., Rosemarin, R. and Rousseau, J. (2012). Asymptotic theory for maximum likelihood estimation of the memory parameter in stationary Gaussian processes. *Econometric Theory* **28** 457–470. MR2913638

[16] Mandelbrot, B.B. and Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev*. **10** 422–437. MR0242239

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

[18] Samorodnitsky, G. and Taqqu, M.S. (1994). *Stable Non-Gaussian Random Processes*: *Stochastic Models with Infinite Variance*. New York: Chapman & Hall. MR1280932

[19] Sinai, Y.G. (1976). Self-similar probability distributions. *Theory Probab*. *Appl*. **21** 64–80.

[20] Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann*. *Statist*. **11** 157–169. MR0684873

Received August 2017 and revised March 2018