Communications for Statistical Applications and Methods
2017, Vol. 24, No. 5, 481–492
https://doi.org/10.5351/CSAM.2017.24.5.481
Print ISSN 2287-7843 / Online ISSN 2383-4757

Convergence rate of a test statistics observed by the longitudinal data with long memory

Yoon Tae Kim[a], Hyun Suk Park[1,a]

[a]Department of Finance and Information Statistics, Hallym University, Korea

Abstract

This paper investigates a convergence rate of a test statistics given by two scale sampling method based on Aït-Sahalia and Jacod (*Annals of Statistics*, **37**, 184–222, 2009). This statistics tests for longitudinal data having the existence of long memory dependence driven by fractional Brownian motion with Hurst parameter $H \in (1/2, 1)$. We obtain an upper bound in the Kolmogorov distance for normal approximation of this test statistic. As a main tool for our works, the recent results in Nourdin and Peccati (*Probability Theory and Related Fields*, **145**, 75–118, 2009; *Annals of Probability*, **37**, 2231–2261, 2009) will be used. These results are obtained by employing techniques based on the combination between Malliavin calculus and Stein's method for normal approximation.

Keywords: Malliavin calculus, multiple stochastic integrals, central limit theorem, Hurst parameter, longitudinal data, fractional Brownian motion

1. Introduction

A fractional Brownian motion $\{B^H, t \geq 0\}$ with Hurst parameter $H \in (0, 1)$ is a centered Gaussian process with the covariance function

$$\mathbb{E}\left[B^H(t)B^H(s)\right] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right), \quad t, s \geq 0.$$

The Hurst parameter $H \in (0, 1)$ characterizes the self-similar behavior of the process. This parameter gives the long-range dependence property of its increments and decides the regularity of the sample paths. Therefore, the problem of properly estimating Hurst parameter H is of the most importance. Many methods to estimate H of $\{B^H, t \geq 0\}$ have been proposed to solve this problem, such as wavelets, k-variations, variograms, maximum likelihood method and spectral methods, some of which can be found in the book by Beran (1994).

This paper investigates a convergence rate of test statistics F_n to see if the error is a Brownian motion or a true fractional Brownian motion in the following longitudinal data:

$$Y(t) = \beta_0 + \beta_1 x(t) + B^H(t), \quad t \in [0, T], \tag{1.1}$$

where $x(t)$ is a non-random function. In terms of the Hurst parameter, this test can be formulated as:

$$H_0 : H = \frac{1}{2} \quad \text{vs.} \quad H_1 : H \neq \frac{1}{2}.$$

[1] Corresponding author: Department of Statistics, Hallym University, 1 Hallimdaehak-gil, Chuncheon 24252, Korea.
E-mail: hspark@hallym.ac.kr

Published 30 September 2017 / journal homepage: http://csam.or.kr

This test statistics F_n, based on the ratio of two realized power variations with different sampling frequencies, has the form:

$$F_n = \frac{\sum_{l=1}^{[T/k\Delta_n]} \left|\Delta_{l,k}^n Y\right|^2}{\sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2}, \tag{1.2}$$

where $\Delta_l^n Y = Y(l\Delta_n) - Y((l-1)\Delta_n)$ and $\Delta_{l,k}^n Y = Y(lk\Delta_n) - Y((l-1)k\Delta_n)$ for determined positive integer k. In the paper Kim and Park (2015), authors prove that

$$\frac{1}{\sqrt{\Delta_n}}\left(F_n - k^{2H-1}\right) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{k^{4H-2}\sigma^2}{T^2}\right), \tag{1.3}$$

where σ^2 is given by

$$\sigma^2 = 2T(k+1)\sum_{j\in\mathbb{Z}} \rho_H(j)^2 - 2k^{-2H}\sum_{l\in\mathbb{Z}}\sum_{j=1}^{k}\left(\sum_{r=1}^{k}\rho_H(lk+r-j)\right)^2. \tag{1.4}$$

Here ρ_H is the covariance function of a fractional Brownian motion expressed as

$$\rho_H(l) = \frac{1}{2}\left(|l+1|^{2H} + |l-1|^{2H} - 2|l|^{2H}\right).$$

From (1.3), we reject H_0 if

$$|F_n - 1| > \sqrt{\Delta_n} z_{\frac{\alpha}{2}} \frac{\sigma}{T},$$

where $\mathbb{P}(Z \geq z_{\alpha/2}) = \alpha/2$ $[Z \sim \mathcal{N}(0,1)]$.

Asymptotic analysis focuses on only describing that the properties (e.g., the central limit theorem (CLT) in our case) of a statistics even when the sample size is finite and similar to the properties when the sample size becomes arbitrarily large. Our main result may give information on how similar the distribution of F_n is with the Gaussian distribution according to sample size.

If the data $\{Y(t)\}$ have the long memory property for each series, i.e., H_0 is rejected, then we may use the model (1.1) for a statistical application. Suppose we observe $\{Y_i(t)\}$ at times $j\Delta_n$, $j = 1, \ldots, [T/\Delta_n]$ and at cross section $i = 1, \ldots, d$. Assume that all series in the longitudinal data have the same Hurst parameter H. For practical purpose, we have to estimate Hurst parameter H first, and then a realization, obtained by the data Y_i, of the estimator $\hat{H}_{\text{ols}}(n,d)$ proposed in this paper is plugged into H in the model (1.1). The estimator $\hat{H}_{\text{ols}}(n,d)$ given above is of the following form:

$$\hat{H}_{\text{ols}}(n,d) = \frac{\sum_{i=1}^{d} \log\left(U_n^{(i)}\right) + d\log k}{d\log k^2},$$

where

$$U_n^{(i)} = \frac{\sum_{l=1}^{[T/k\Delta_n]} \left|\Delta_{l,k}^n Y_i\right|^2}{\sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y_i\right|^2}.$$

The model (1.1) becomes

$$Y_i(t) = (\beta_0 + u_i) + \beta_1 x_i(t) + \epsilon_i(t), \quad i = 1, \ldots, d \text{ and } t \in [0, T], \tag{1.5}$$

where the error term $\epsilon_i(t)$ is a fractional Brownian motion with $\epsilon_i(t+h) - \epsilon_i(t) \sim \mathcal{N}(0, \sigma^2 h^{2\hat{H}_{\text{ols}}(n,d)})$. After that, we may use the usual longitudinal data analysis in order to estimate the linear regression model (1.5).

The main tool for the proof of a Berry-Esseen bound is the combination of Stein's method and Malliavin calculus as well as the result in Nourdin and Peccati (2009a, 2009b). Recently, Berry-Esseen bounds for various statistics for estimators of parameters , involved in stochastic differential equations and stochastic partial differential equations, have been much studied (Kim and Park, 2016, 2017a, 2017b).

2. Preliminaries

In this section, we briefly review some facts about Malliavin calculus for Gaussian processes. For a more detailed reference, see Nualart (2006). Suppose that $\mathfrak{H}$ is a real separable Hilbert space with scalar product denoted by $\langle \cdot, \cdot \rangle_{\mathfrak{H}}$. Let $X = \{X(h), h \in \mathfrak{H}\}$ be an isonormal Gaussian process, that is a centered Gaussian family of random variables such that $\mathbb{E}[X(h)X(g)] = \langle h, g \rangle_{\mathfrak{H}}$. If $X = B^H$, then

$$\mathbb{E}\left[B^H(t)B^H(s)\right] = \langle \mathbf{1}_{[0,s]}, \mathbf{1}_{[0,t]} \rangle_{\mathfrak{H}} = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right).$$

For every $q \geq 1$, let $\mathcal{H}_q$ be the q^{th} Wiener chaos of X, that is the closed linear subspace of $\mathbb{L}^2(\Omega)$ generated by $\{H_q(X(h)) : h \in \mathfrak{H}, \|h\|_{\mathfrak{H}} = 1\}$, where H_q is the q^{th} Hermite polynomial. We define a linear isometric mapping $I_q : \mathfrak{H}^{\odot q} \to \mathcal{H}_q$ by $I_q(h^{\otimes q}) = H_q(X(h))$, where $\mathfrak{H}^{\odot n}$ is the symmetric tensor product. The following duality formula holds

$$\mathbb{E}[F I_q(h)] = \mathbb{E}\left[\langle D^q F, h \rangle_{\mathfrak{H}^{\otimes q}}\right], \tag{2.1}$$

for any element $h \in \mathfrak{H}^{\odot q}$ and any random variable $F \in \mathbb{D}^{q,2}$. Here

$$\|F\|_{q,2}^2 = \mathbb{E}\left[F^2\right] + \sum_{k=1}^{q} \mathbb{E}\left[\left\|D^k F\right\|_{\mathfrak{H}^{\otimes k}}^2\right],$$

where D^k is the iterative Malliavin derivative. The linear isometric mapping I_q satisfies $I_q(f) = I_q(\tilde{f})$ and

$$\mathbb{E}\left[I_p(f) I_q(g)\right] = \begin{cases} 0, & \text{if } p \neq q, \\ p! \left\langle \tilde{f}, \tilde{g} \right\rangle_{\mathfrak{H}}, & \text{if } p = q, \end{cases} \tag{2.2}$$

where $\tilde{f}$ denotes the symmetrization of f.

If $f \in \mathfrak{H}^{\odot p}$, the Malliavin derivative of the multiple stochastic integrals is given by

$$D_z I_q(f_q) = q I_{q-1}\left(f_q(\cdot, z)\right), \quad \text{for } z \in [0,1]^2. \tag{2.3}$$

Let $\{e_l, l \geq 1\}$ be a complete orthonormal system in $\mathfrak{H}$.

If $f \in \mathfrak{H}^{\odot p}$ and $g \in \mathfrak{H}^{\odot q}$, the contraction $f \otimes_r g$, $1 \le r \le p \wedge q$, is the element of $\mathbb{H}^{\otimes(p+q-2r)}$ defined by

$$f \otimes_r g = \sum_{l_1,\dots,l_r=1}^{\infty} \langle f, e_{l_1} \otimes \cdots \otimes e_{l_r} \rangle_{\mathfrak{H}^{\otimes r}} \otimes \langle g, e_{l_1} \otimes \cdots \otimes e_{l_r} \rangle_{\mathfrak{H}^{\otimes r}}. \tag{2.4}$$

Notice that the tensor product $f \otimes g$ and the contraction $f \otimes_r g$, $1 \le r \le p \wedge q$ are not necessarily symmetric even though f and g are symmetric. We will denote their symmetrizations by $f \tilde{\otimes} g$ and $f \tilde{\otimes}_r g$, respectively. The following formula for the product of the multiple stochastic integrals will be frequently used to prove the main result in this paper:

Proposition 1. *Let $f \in \mathfrak{H}^{\odot p}$ and $g \in \mathfrak{H}^{\odot q}$ be two symmetric functions. Then*

$$I_p(f)I_q(g) = \sum_{r=0}^{p\wedge q} r! \binom{p}{r}\binom{q}{r} I_{p+q-2r}(f \otimes_r g). \tag{2.5}$$

Now we introduce the *infinitesimal generator L of the Ornstein-Uhlenbeck semigroup* and the relation of the operator L with the operators D and δ (see Subsection 1.4 in Nualart (2006) for more details). Let $F \in L^2(\Omega)$ be a square integrable random variable. For each $n \ge 1$, we will denote by $\mathbb{J}_n : L^2(\Omega) \to \mathbb{H}_n$ the orthogonal projection on the n^{th} Wiener chaos $\mathcal{H}_n$. The operator L is defined through the projection operator $\mathbb{J}_n$, $n = 0, 1, 2 \dots$, as $L = \sum_{n=0}^{\infty} -n\mathbb{J}_n F$, and is called the *infinitesimal generator of the Ornstein-Uhlenbeck semigroup*. The relationship between the operator D, δ and L is given as: $\delta DF = -LF$, that is, for $F \in L^2(\Omega)$ the statement $F \in \mathrm{Dom}(L)$ is equivalent to $F \in \mathrm{Dom}(\delta D)$ (i.e. $F \in \mathbb{D}^{1,2}$ and $DF \in \mathrm{Dom}(\delta)$), and in this case $\delta DF = -LF$. We also define the operator L^{-1}, which is the *pseudo-inverse* of L, as $L^{-1}F = \sum_{n=1}^{\infty} \mathbb{J}_n(F)/n$. Note that L^{-1} is an operator with values in $\mathbb{D}^{2,2}$ and $LL^{-1}F = F - \mathbb{E}[F]$ for all $F \in L^2(\Omega)$.

3. Main results

In this section, we investigate a convergence rate of CLT in (1.3). First recall that for every $z \in \mathbb{R}$, the function

$$\begin{aligned} f_z(x) &= e^{\frac{x^2}{2}} \int_{-\infty}^{x} \left\{ \mathbf{1}_{(-\infty,z]}(u) - \Phi(z) \right\} e^{-\frac{u^2}{2}} du \\ &= \begin{cases} \sqrt{2\pi} e^{\frac{x^2}{2}} \Phi(x)\{1 - \Phi(z)\}, & \text{if } x \le z, \\ \sqrt{2\pi} e^{\frac{x^2}{2}} \Phi(z)\{1 - \Phi(x)\}, & \text{if } x > z \end{cases} \end{aligned} \tag{3.1}$$

is a solution to the following Stein equation such that $\|f_z\|_\infty \le \sqrt{2\pi}/4$ and $\|f_z'\|_\infty \le 1$:

$$f'(x) - xf(x) = \mathbf{1}_{(-\infty,z]}(x) - \Phi(z), \tag{3.2}$$

The derivative of f_z is given by

$$f_z'(x) = \begin{cases} \{1 - \Phi(z)\} \left\{ 1 + \sqrt{2\pi} x e^{\frac{x^2}{2}} \Phi(x) \right\}, & \text{if } x \le z, \\ \Phi(z) \left[-1 + \sqrt{2\pi} x e^{\frac{x^2}{2}} \{1 - \Phi(x)\} \right], & \text{if } x > z. \end{cases} \tag{3.3}$$

We use the following lemma, given by Michael and Pfanzagl (1971), to prove our main result.

Lemma 1. *Let $(\Omega, \mathfrak{F}, \mathbb{P})$ be a probability space and G_n and V_n be a $\mathfrak{F}$-measurable function such that $V_n > 0$ a.s. for all n. Then for any $\epsilon > 0$, we have*

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{G_n}{V_n} \le z\right) - \Pr(Z \le z) \right| = \sup_{z\in\mathbb{R}} |\mathbb{P}(U_n \le z) - \mathbb{P}(Z \le z)| + \mathbb{P}(|V_n - 1| \ge \epsilon) + \epsilon. \tag{3.4}$$

Now we obtain the Berry-Esseen bound of the test statistics F_n given in (1.2).

Theorem 1. *Suppose that $|x(t) - x(s)| \le c|t - s|$ for $c > 0$, where $x(t)$ is given in the equation (1.5). If $H > 1/2$, there exists a constant $c > 0$ such that, for sufficiently large n,*

$$\sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{T}{\sqrt{\Delta_n} k^{2H-1}\sigma}\left(F_n - k^{2H-1}\right) \le z\right) - \mathbb{P}(Z \le z) \right| \le c \min\left\{\Delta_n^{\frac{3-4H}{2}}, \Delta_n^{\frac{1-H}{2}}\right\}. \tag{3.5}$$

where σ^2 is given in (1.4).

Proof: Throughout this proof, c stands for an absolute constant with possibly different values in different places. Using Lemma 1, we have that for any $0 < \epsilon < 1$,

$$\begin{aligned}
&\sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{T}{\sqrt{\Delta_n} k^{2H-1}\sigma}\left(F_n - k^{2H-1}\right) \le z\right) - \Pr(Z \le z) \right| \\
&= \sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{T}{\sqrt{\Delta_n} k^{2H-1}\sigma} \frac{\sum_{l=1}^{[T/k\Delta_n]} \left|\Delta_{l,k}^n Y\right|^2 - k^{2H-1} \sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2}{\sum_{l=1}^{[T/\Delta_n]} \left|\Delta_{l,k}^n Y\right|^2} \le z \right) - \mathbb{P}(Z \le z) \right| \\
&= \sup_{z\in\mathbb{R}} \left| \mathbb{P}\left(\frac{\Delta_n^{1-2H}}{\sqrt{\Delta_n} k^{2H-1}\sigma} \left\{ \sum_{l=1}^{[T/k\Delta_n]} \left|\Delta_{l,k}^n Y\right|^2 - k^{2H-1} \sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2 \right\} \le z \right) - \mathbb{P}(Z \le z) \right| \\
&\quad + \mathbb{P}\left(\left| \frac{\Delta_n^{1-2H}}{T} \sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2 - 1 \right| \ge \epsilon \right) + \epsilon.
\end{aligned} \tag{3.6}$$

First consider the second term in (3.6). We write

$$\begin{aligned}
\sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2 &= \sum_{l=1}^{[T/\Delta_n]} \left\{ \beta_1^2 \left(\Delta_l^n x\right)^2 + 2\beta_1 \left(\Delta_l^n x\right)\left(\Delta_l^n B^H\right) + \left(\Delta_l^n B^H\right)^2 \right\} \\
&:= A_1^n + A_2^n + A_3^n.
\end{aligned} \tag{3.7}$$

Therefore,

$$\mathbb{P}\left(\left| \frac{\Delta_n^{1-2H}}{T} \sum_{l=1}^{[T/\Delta_n]} \left|\Delta_l^n Y\right|^2 - 1 \right| \ge \epsilon \right) \le \frac{1}{\epsilon} \left\{ \frac{\Delta_n^{1-2H}}{T} \left|A_1^n\right| + \frac{\Delta_n^{1-2H}}{T} \mathbb{E}\left[\left|A_2^n\right|\right] + \mathbb{E}\left[\left| \frac{\Delta_n^{1-2H}}{T} A_3^n - 1 \right|\right] \right\}. \tag{3.8}$$

By the assumption on $x(t)$, the first term in (3.8) can be estimated as

$$\frac{\Delta_n^{1-2H}}{T} \left|A_1^n\right| \le c\beta_1^2 \frac{\Delta_n^{1-2H}}{T} \left[\frac{T}{\Delta_n}\right] \Delta_n^2 \le c\Delta_n^{2(1-H)}. \tag{3.9}$$

By the Cauchy-Schwartz inequality, we get

$$\begin{aligned}\frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|A_2^n\right|\right] &\le c\beta_1\frac{\Delta_n^{1-2H}}{T}\Delta_n\sum_{l=1}^{[T/\Delta_n]}\mathbb{E}\left[\left|\Delta_l^n B^H\right|\right]\\ &\le c\beta_1\frac{\Delta_n^{1-2H}}{T}\Delta_n\sum_{l=1}^{[T/\Delta_n]}\sqrt{\mathbb{E}\left[\left(\Delta_l^n B^H\right)^2\right]}\\ &\le c\Delta_n^{1-H}. \end{aligned} \tag{3.10}$$

As for the third term in (3.8), we estimate

$$\begin{aligned}\mathbb{E}\left[\left|\frac{\Delta_n^{1-2H}}{T}A_3^n - 1\right|\right] &= \frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|A_3^n - \frac{T}{\Delta_n}\Delta_n^{2H}\right|\right]\\ &\le \frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|A_3^n - \left[\frac{T}{\Delta_n}\right]\Delta_n^{2H}\right|\right] + \Delta_n^{2H}\left(\frac{T}{\Delta_n} - \left[\frac{T}{\Delta_n}\right]\right). \end{aligned} \tag{3.11}$$

By using the computation of $\mathbb{V}\mathrm{ar}(\sum_{l=1}^{[T/\Delta_n]}(\Delta_l^n B^H)^2)$ in Kim and Park (2015), the first term in (3.11) can be bounded

$$\begin{aligned}\frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|A_3^n - \left[\frac{T}{\Delta_n}\right]\Delta_n^{2H}\right|\right] &= \frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|\sum_{l=1}^{[T/\Delta_n]}\left(\Delta_l^n B^H\right)^2 - \sum_{l=1}^{[T/\Delta_n]}\mathbb{E}\left[\left(\Delta_l^n B^H\right)^2\right]\right|\right]\\ &\le \frac{\Delta_n^{1-2H}}{T}\sqrt{\mathbb{V}\mathrm{ar}\left(\sum_{l=1}^{[T/\Delta_n]}\left(\Delta_l^n B^H\right)^2\right)}\\ &\le \frac{\Delta_n^{1-2H}}{T}\sqrt{\left[\frac{T}{\Delta_n}\right]}\Delta_n^{2H}\sqrt{\left[\frac{T}{\Delta_n}\right]^{2H} - \left(\left[\frac{T}{\Delta_n}\right]-1\right)^{2H}}. \end{aligned} \tag{3.12}$$

By the mean value theorem, we have $[T/\Delta_n]^{2H} - ([T/\Delta_n]-1)^{2H} \le 2H([T/\Delta_n])^{2H-1}$. This inequality proves that the right-hand side of (3.12) can be estimated as

$$\frac{\Delta_n^{1-2H}}{T}\mathbb{E}\left[\left|A_3^n - \left[\frac{T}{\Delta_n}\right]\Delta_n^{2H}\right|\right] \le c\Delta_n^{1-H}. \tag{3.13}$$

From (3.11) and (3.13), it follows that

$$\mathbb{E}\left[\left|\frac{\Delta_n^{1-2H}}{T}A_3^n - 1\right|\right] \le c\left(\Delta_n^{1-H} + \Delta_n^{2H}\right). \tag{3.14}$$

By combining the above estimates (3.9), (3.10), and (3.14), we obtain that for every $\epsilon > 0$,

$$\mathbb{P}\left(\left|\frac{\Delta_n^{1-2H}}{T}\sum_{l=1}^{[T/\Delta_n]}\left|\Delta_l^n Y\right|^2 - 1\right| \ge \epsilon\right) \le c\frac{\Delta_n^{1-H}}{\epsilon} \tag{3.15}$$

Let us set

$$U_n = \frac{\Delta_n^{1-2H}}{\sqrt{\Delta_n}k^{2H-1}\sigma}\left\{\sum_{l=1}^{[T/k\Delta_n]}\left|\Delta_{l,k}^n Y\right|^2 - k^{2H-1}\sum_{l=1}^{[T/\Delta_n]}\left|\Delta_l^n Y\right|^2\right\}.$$

Using the multiplication formula of multiple stochastic integral in (2.5) yields

$$(k\Delta_n)^{1-2H}\sum_{l=1}^{[T/k\Delta_n]}\left|\Delta_{l,k}^n B^H\right|^2 = I_2(f_{n,k,2}) + (k\Delta_n)k\left[\frac{T}{k\Delta_n}\right], \tag{3.16}$$

$$\Delta_n^{1-2H}\sum_{l=1}^{[T/\Delta_n]}\left|\Delta_l^n B^H\right|^2 = I_2(f_{n,1,2}) + \Delta_n\left[\frac{T}{\Delta_n}\right], \tag{3.17}$$

where the kernels $f_{n,k,2}$ are given by

$$f_{n,k,2} = (k\Delta_n)^{1-2H}\sum_{l=1}^{[T/k\Delta_n]}\mathbf{1}^{\otimes 2}_{[(l-1)k\Delta_n, lk\Delta_n]}.$$

Also we define a kernel $f_{n,k,1}$ as:

$$f_{n,k,1} = 2\beta_1(k\Delta_n)^{1-2H}\sum_{l=1}^{[T/k\Delta_n]}\left(\Delta_{l,k}^n x\right)\mathbf{1}_{[(l-1)k\Delta_n, lk\Delta_n]}.$$

Hence it follows from (3.16) and (3.17) that

$$U_n = \frac{1}{\sqrt{\Delta_n}\sigma}\left[I_1\left(f_{n,k,1} - f_{n,1,1}\right) + I_2\left(f_{n,k,2} - f_{n,1,2}\right)\right] + \frac{1}{\sqrt{\Delta_n}\sigma}\mathbb{E}[U_n]. \tag{3.18}$$

Here $\mathbb{E}[U_n]$ is given by

$$\mathbb{E}[U_n] = (k\Delta_n)^{1-2H}\sum_{l=1}^{[T/k\Delta_n]}\beta_1^2\left(\Delta_{l,k}^n x\right)^2 - \Delta_n^{1-2H}\sum_{l=1}^{[T/\Delta_n]}\beta_1^2\left(\Delta_l^n x\right)^2 + (k\Delta_n)\left[\frac{T}{k\Delta_n}\right] - \Delta_n\left[\frac{T}{\Delta_n}\right].$$

Applying Lemma 2.3 in Nourdin and Peccati (2009b) to the first term of the right-hand side (3.6), we have, using $\|f_z\|_\infty \le \sqrt{2\pi}/4$ and $\|f_z'\|_\infty \le 1$, that

$$\begin{aligned}\sup_{z\in\mathbb{R}}|\mathbb{P}(U_n \le z) - \mathbb{P}(Z \le z)| &= \left|\mathbb{E}\left[f_z'(U_n) - U_n f_z(U_n)\right]\right| \\ &= \left|\mathbb{E}\left[f_z'(U_n)\left(1 - \langle DU_n, -DL^{-1}U_n\rangle_{\mathfrak{H}}\right)\right] - \frac{1}{\sqrt{\Delta_n}\sigma}\mathbb{E}[U_n]\mathbb{E}\left[f_z'(U_n)\right]\right| \\ &\le \sqrt{\mathbb{E}\left[(1 - \langle DU_n, -DL^{-1}U_n\rangle_{\mathfrak{H}})^2\right]} + \frac{\sqrt{2\pi}}{4\sqrt{\Delta_n}\sigma}\left|\mathbb{E}[U_n]\right|. \end{aligned} \tag{3.19}$$

For simplicity, we set

$$g_{n,1} = \frac{f_{n,k,1} - f_{n,1,1}}{\sqrt{\Delta_n}\sigma} \quad \text{and} \quad g_{n,2} = \frac{f_{n,k,2} - f_{n,1,2}}{\sqrt{\Delta_n}\sigma}.$$

Following the proof of Proposition 3.7 in Nourdin and Peccati (2009b), we estimate

$$\begin{aligned}&\mathbb{E}\left[(1 - \langle DU_n, -DL^{-1}U_n\rangle_{\mathfrak{H}})^2\right] \\ &= \mathbb{E}\left[(1 - \langle g_{n,1} + 2I_1(g_{n,2}), g_{n,1} + I_1(g_{n,2})\rangle_{\mathfrak{H}})^2\right] \\ &= \mathbb{E}\left[(1 - \|g_{n,1}\|^2_{\mathfrak{H}} - 3\langle g_{n,1}, I_1(g_{n,2})\rangle_{\mathfrak{H}} - 2\langle I_1(g_{n,2}), I_1(g_{n,2})\rangle_{\mathfrak{H}})^2\right] \\ &\le 2\left(1 - \|g_{n,1}\|^2_{\mathfrak{H}} - 2\|g_{n,2}\|^2_{\mathfrak{H}^{\otimes 2}}\right)^2 + 18\|g_{n,1}\otimes_1 g_{n,2}\|^2_{\mathfrak{H}} + 16\|g_{n,2}\otimes_1 g_{n,2}\|^2_{\mathfrak{H}^{\otimes 2}}. \end{aligned} \tag{3.20}$$

Obviously, the first term can be estimated as

$$
\begin{aligned}
&\left(1-\|g_{n,1}\|_{\mathfrak{H}}^2-2\|g_{n,2}\|_{\mathfrak{H}^{\otimes 2}}^2\right)^2 \\
&\leq 2\|g_{n,1}\|_{\mathfrak{H}}^4+2\left(1-2\|g_{n,2}\|_{\mathfrak{H}^{\otimes 2}}^2\right)^2 \\
&\leq \frac{4}{\Delta_n^2\sigma^4}\left(\|f_{n,k,1}\|_{\mathfrak{H}}^4+\|f_{n,1,1}\|_{\mathfrak{H}}^4\right)+\frac{2}{\sigma^4}\left\{\sigma^2-\frac{2}{\Delta_n}\left(\|f_{n,k,2}\|_{\mathfrak{H}^{\otimes 2}}^2-2\langle f_{n,k,2},f_{n,1,2}\rangle_{\mathfrak{H}^{\otimes 2}}+\|f_{n,1,2}\|_{\mathfrak{H}^{\otimes 2}}^2\right)\right\}^2 \\
&:= B_1^n+B_2^n.
\end{aligned}
$$

We note that

$$
|\rho_H(l)| = H(2H-1)|l|^{2H-2}+o\left(|l|^{2H-2}\right) \quad \text{as } |l|\to\infty. \tag{3.21}
$$

For sufficiently large n, we estimate, from (3.21),

$$
\begin{aligned}
\sum_{l,l'=1}^{[T/k\Delta_n]}|\rho_H(l-l')| &\leq \sum_{|j|<[T/\Delta_n]}\left(\left[\frac{T}{\Delta_n}\right]-|j|\right)\left||j'+1|^{2H}+|j'-1|^{2H}-2|j|^{2H}\right| \\
&\leq \frac{T}{k}\sum_{j=1}^{[T/k\Delta_n]}\Delta_n^{-1}j^{2H-2}\leq \frac{T}{k}\Delta_n^{-1}\left\{1+\left(\left[\frac{T}{k\Delta_n}\right]\right)^{2H-1}\right\} \\
&\leq c\left(\Delta_n^{-1}+\Delta_n^{-2H}\right).
\end{aligned} \tag{3.22}
$$

Direct computation and the estimate (3.22) give

$$
\begin{aligned}
B_1^n &\leq c\frac{1}{\Delta_n^2\sigma^4}\Delta_n^{8-4H}\left[\left\{k\sum_{l,l'=1}^{[T/k\Delta_n]}\rho_H(l-l')\right\}^2+\left\{\sum_{l,l'=1}^{[T/\Delta_n]}\rho_H(l-l')\right\}^2\right] \\
&\leq c\left(\Delta_n^{4-4H}+\Delta_n^{6-8H}\right).
\end{aligned} \tag{3.23}
$$

As for the term B_2^n, we compute the three terms in B_2^n

$$
\begin{aligned}
\Delta_n^{-1}\|f_{n,k,2}\|_{\mathfrak{H}^{\otimes 2}}^2 &= \Delta_n^{-1}(k\Delta_n)^{2-4H}\sum_{l,l'=1}^{[T/k\Delta_n]}\left\langle \mathbf{1}^{\otimes 2}_{[(l-1)k\Delta_n,lk\Delta_n]},\mathbf{1}^{\otimes 2}_{[(l'-1)k\Delta_n,l'k\Delta_n]}\right\rangle_{\mathfrak{H}^{\otimes 2}} \\
&= \Delta_n^{-1}(k\Delta_n)^2\sum_{l,l'=1}^{[T/k\Delta_n]}\rho_H(l-l')^2 \\
&= \Delta_n^{-1}k^2\sum_{|j|<[T/k\Delta_n]}\Delta_n\left(\left[\frac{T}{k\Delta_n}\right]-|j|\right)\rho_H(j)^2.
\end{aligned} \tag{3.24}
$$

Using a similar argument as for the first term in (3.24) yields

$$\begin{aligned}
\Delta_n^{-1}\langle f_{n,k,2}, f_{n,1,2}\rangle_{\mathfrak{H}^{\otimes 2}} &= k^{1-2H}\Delta_n^{1-4H}\sum_{l=1}^{[T/k\Delta_n]}\sum_{l'=1}^{[T/\Delta_n]}\left\langle \mathbf{1}^{\otimes 2}_{[(l-1)k\Delta_n, lk\Delta_n]}, \mathbf{1}^{\otimes 2}_{[(l'-1)\Delta_n, l'\Delta_n]}\right\rangle_{\mathbb{H}^{\otimes 2}} \\
&= k^{1-2H}\Delta_n\sum_{l=1}^{[T/k\Delta_n]}\sum_{l'=1}^{[T/k\Delta_n]}\sum_{j=1}^{k}\left(\left\langle \mathbf{1}_{[(l-1)k, lk]}, \mathbf{1}_{[(l'-1)k+j-1,(l'-1)k+j]}\right\rangle_{\mathbb{H}}\right)^2 \\
&= k^{1-2H}\Delta_n\sum_{l=1}^{[T/k\Delta_n]}\sum_{l'=1}^{[T/k\Delta_n]}\sum_{j=1}^{k}\left(\sum_{r=1}^{k}\rho_H((l-l')k+r-j)\right)^2 \\
&= k^{1-2H}\Delta_n\sum_{|l|<[T/k\Delta_n]}\sum_{j=1}^{k}([T/k\Delta_n]-|l|)\left(\sum_{r=1}^{k}\rho_H(lk+r-j)\right)^2. \qquad (3.25)
\end{aligned}$$

Substituting $k=1$ into k in (3.24), we have

$$\Delta_n^{-1}\|f_{n,1,2}\|^2_{\mathfrak{H}^{\otimes 2}} = \Delta_n^{-1}\sum_{|j|<[T/\Delta_n]}\Delta_n\left(\left[\frac{T}{\Delta_n}\right]-|j|\right)\rho_H(j)^2. \qquad (3.26)$$

Combining the above results (3.24), (3.25), and (3.26), we obtain

$$\begin{aligned}
B_2^n \le \frac{2}{\sigma^4}\Bigg[& 2\left|kT\sum_{j\in\mathbb{Z}}\rho_H(j)^2 - k^2\sum_{|j|<[T/k\Delta_n]}\Delta_n\left(\left[\frac{T}{k\Delta_n}\right]-|j|\right)\rho_H(j)^2\right|^2 \\
&+ 4\left|k^{-2H}\sum_{l\in\mathbb{Z}}\sum_{j=1}^{k}\left(\sum_{r=1}^{k}\rho_H(lk+r-j)\right)^2 - k^{1-2H}\Delta_n\sum_{|l|<[T/k\Delta_n]}\sum_{j=1}^{k}\left(\left[\frac{T}{k\Delta_n}\right]-|l|\right)\left(\sum_{r=1}^{k}\rho_H(lk+r-j)\right)^2\right|^2 \\
&+ 2\left|T\sum_{j\in\mathbb{Z}}\rho_H(j)^2 - \sum_{|j|<[T/\Delta_n]}\Delta_n\left(\left[\frac{T}{\Delta_n}\right]-|j|\right)\rho_H(j)^2\right|^2 \\
:=\ & B_{21}^n + B_{22}^n + B_{23}^n.
\end{aligned}$$

Obviously, for sufficiently large n, we have

$$\begin{aligned}
B_{21}^n &\le \frac{4}{\sigma^4}\left[kT\sum_{|j|\ge[T/k\Delta_n]}\rho_H(j)^2 + kT\sum_{|j|<[T/k\Delta_n]}\left\{1-\frac{k\Delta_n}{T}\left[\frac{T}{k\Delta_n}\right]\right\}\rho_H(j)^2 + k^2\Delta_n\sum_{|j|<[T/k\Delta_n]}|j|\rho_H(j)^2\right]^2 \\
&\le c\left\{\Delta_n^{3-4H} + \Delta_n\left(1+\Delta_n^{3-4H}\right) + \Delta_n\left(1+\Delta_n^{2-4H}\right)\right\}^2 \\
&\le c\Delta_n^{2(3-4H)}. \qquad (3.27)
\end{aligned}$$

By a similar estimate as for the term B_{21}^n in (3.27), we get $B_{22}^n \le c\Delta_n^{2(3-4H)}$ and $B_{23}^n \le c\Delta_n^{2(3-4H)}$. Thus we have

$$B_2^n \le c\Delta_n^{2(3-4H)}. \qquad (3.28)$$

As for the second and third terms in (3.20), observe that $\|g_{n,1} \otimes_1 g_{n,2}\|^2_{\mathfrak{H}} \leq c\|f_{n,k,1} \otimes_1 f_{n,k,2}\|^2_{\mathfrak{H}}$. First write

$$
\begin{aligned}
f_{n,k,1} \otimes_1 f_{n,k,2} &= \frac{2\beta_1 (k\Delta_n)^{2-4H}}{\Delta_n \sigma^2} \sum_{l,l'=1}^{[T/k\Delta_n]} \left(\Delta^n_{l,k} x\right) \langle \mathbf{1}_{[(l-1)k\Delta_n, lk\Delta_n]}, \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]} \rangle_{\mathfrak{H}} \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]} \\
&= \frac{2\beta_1 (k\Delta_n)^{2-2H}}{\Delta_n \sigma^2} \sum_{l,l'=1}^{[T/k\Delta_n]} \left(\Delta^n_{l,k} x\right) \rho_H(l-l') \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]}. \qquad (3.29)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\left\| f_{n,k,1} \otimes_1 f_{n,k,2} \right\|^2_{\mathfrak{H}} &= \frac{4\beta_1^2 (k\Delta_n)^4}{\Delta_n^2 \sigma^4} \sum_{l,l',j,j'=1}^{[T/k\Delta_n]} \left(\Delta^n_{l,k} x\right)\left(\Delta^n_{j,k} x\right) \rho_H (l-l') \rho_H (j-j') \rho_H (l'-j') \\
&\leq c\Delta_n^4 \sum_{l,l',j,j'=1}^{[T/k\Delta_n]} \rho_H (l-l') \rho_H (j-j') \rho_H (l'-j'). \qquad (3.30)
\end{aligned}
$$

For the sum in (3.30), we decompose as follows

$$
\sum_{l>l'>j>j'}^{[T/k\Delta_n]} + \sum_{l>l'>j'>j}^{[T/k\Delta_n]} + \cdots . \qquad (3.31)
$$

For the first term, we have

$$
\begin{aligned}
\Delta_n^4 \sum_{l>l'>j>j'}^{[T/k\Delta_n]} \rho_H (l-l') \rho_H (j-j') \rho_H (l'-j') &\leq \Delta_n^4 \sum_{l>l'>j>j'}^{[T/k\Delta_n]} (l-l')^{2H-2} (j-j')^{4H-4} \\
&\leq \Delta_n^2 \sum_{l=1}^{[T/k\Delta_n]} l^{2H-2} \sum_{j=1}^{[T/k\Delta_n]} j^{4H-4} \\
&\leq c\Delta_n^2 \left\{ \left(1 + \left[\frac{T}{k\Delta_n} \right]^{2H-1} \right) \left(1 + \left[\frac{T}{k\Delta_n} \right]^{4H-3} \right) \right\} \\
&\leq c\Delta_n^{3-2H}. \qquad (3.32)
\end{aligned}
$$

Obviously, the same bound also holds for the other terms in (3.31). As for the last term in (3.20), observe that $\|g_{n,2} \otimes_1 g_{n,2}\|^2_{\mathfrak{H}} \leq c\|f_{n,k,2} \otimes_1 f_{n,k,2}\|^2_{\mathfrak{H}}$.

$$
\begin{aligned}
f_{n,k,2} \otimes_1 f_{n,k,2} &= (k\Delta_n)^{2-4H} \sum_{l,l'=1}^{[T/k\Delta_n]} \langle \mathbf{1}_{[(l-1)k\Delta_n, lk\Delta_n]}, \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]} \rangle_{\mathbb{H}} \times \mathbf{1}_{[(l-1)k\Delta_n, lk\Delta_n]} \tilde{\otimes} \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]} \\
&= (k\Delta_n)^{2-2H} \sum_{l,l'=1}^{[T/k\Delta_n]} \rho_H (l-l') \mathbf{1}_{[(l-1)k\Delta_n, lk\Delta_n]} \tilde{\otimes} \mathbf{1}_{[(l'-1)k\Delta_n, l'k\Delta_n]}. \qquad (3.33)
\end{aligned}
$$

Let us set $\rho_{n,H}(j) = |\rho_{n,H}(j)| \mathbf{1}_{\{|j| \leq [T/k\Delta_n]\}}$. By using the arguments in the quadratic variation of the fractional Brownian motion studied by Nourdin (2013), we obtain, from (3.33),

$$\begin{aligned}\left\|f_{n,k,2}\otimes_1 f_{n,k,2}\right\|^2_{\mathbb{H}^{\otimes 2}} &= \frac{4\beta_1^2(k\Delta_n)^4}{\Delta_n^2\sigma^4}\sum_{l,l',j,j'=1}^{[T/k\Delta_n]}\rho_H\left(l-l'\right)\rho_H(j-j')\rho_H(l-j)\rho_H(l'-j')\\ &\le c\Delta_n^2\sum_{l,j'=1}^{[T/k\Delta_n]}\sum_{j,l'\in\mathbb{Z}}\rho_{n,H}(l-l')\rho_{n,H}(j-j')\rho_{n,H}(l-j)\rho_{n,H}(l'-j')\\ &\le c\Delta_n\sum_{l\in\mathbb{Z}}(\rho_{n,H}*\rho_{n,H})(l)^2\le k^4\Delta_n\left(\sum_{|l|\le[T/k\Delta_n]}|\rho_H(l)|^{\frac{4}{3}}\right)^3\\ &\le c\Delta_n\left(1+\left[\frac{T}{k\Delta_n}\right]^{\frac{8H-5}{3}}\right)^3\le c\Delta_n^{2(3-4H)}.\end{aligned}\tag{3.34}$$

The last term in (3.19) can easily estimated as

$$\frac{\sqrt{2\pi}}{4\sqrt{\Delta_n}\sigma}\left|\mathbb{E}[U_n]\right|\le c\left(\Delta_n^{\frac{3-4H}{2}}+\sqrt{\Delta_n}\right).\tag{3.35}$$

By combining all these bounds (3.15), (3.23), (3.28), (3.32), (3.34), and (3.35), together with $\epsilon = \Delta_n^{(1-H)/2}$, the proof of Theorem is now completed. □

Remark 1. We are not sure that the upper bound, obtained in Theorem 1, is an optimal bound in the following sense: the bound $\varphi(F_n)$ is *optimal* for the sequence $\{F_n\}$ with respect to some distance d if there exist constants $0 < c < C < \infty$, independent of n, such that

$$c\varphi(F_n)\le d(F_n,Z)\le C\varphi(F_n),\quad \text{for all } n\ge 1.\tag{3.36}$$

An optimal rates of convergence in the *Kolmogorov distance* will be derived in future studies; therefore, we should develop the techniques to find an lower bound.

Acknowledgements

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014-S1A5B6A02048942).

References

Aït-Sahalia Y and Jacod J (2009). Testing for jumps in a discretely observed process. *The Annals of Statistics*, **37**, 184–222.

Beran J (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, London.

Kim YT and Park HS (2015). Estimation of Hurst parameter in the longitudinal data with long memory, *Communications for Statistical Applications and Methods*, **22**, 295–304.

Kim YT and Park HS (2016). Berry-Esseen Type bound of a sequence $\{\frac{X_N}{Y_N}\}$ and its application, *Journal of the Korean Statistical Society*, **45**, 544–556.

Kim YT and Park HS (2017a). Optimal Berry-Esseen bound for statistical estimations and its application to SPDE, *Journal of Multivariate Analysis*, **155**, 284–304.

Kim YT and Park HS (2017b). Optimal Berry-Esseen bound for an estimator of parameter in the Ornstein-Uhlenbeck process, *Journal of the Korean Statistical Society*, **46**, 413–425.

Michel R and Pfanzagl J (1971). The accuracy of the normal approximation for minimum contrast estimates, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **18**, 73–84.

Nourdin I (2013). *Lectures on Gaussian approximations with Malliavin calculus, Séminaire de Probabilités*, **45**, 3–89.

Nourdin I and Peccati G (2009a). Stein's method on Wiener chaos, *Probability Theory and Related Fields*, **145**, 75–118.

Nourdin I and Peccati G (2009b). Stein's method and exact Berry-Esseen asymptotics for functionals for functionals of Gaussian fields, *The Annals of Probability*, **37**, 2231–2261.

Nualart D (2006). *Malliavin Calculus and Related Topics* (2nd ed), Springer, Berlin.

Received May 25, 2017; Revised August 28, 2017; Accepted August 29, 2017