Adv. Appl. Prob. **37,** 938–962 (2005)
Printed in Northern Ireland

ON A LOCALIZATION PROPERTY OF WAVELET COEFFICIENTS FOR PROCESSES WITH STATIONARY INCREMENTS, AND APPLICATIONS. I. LOCALIZATION WITH RESPECT TO SHIFT

SERGIO ALBEVERIO,* *Universität Bonn*
SHUJI KAWASAKI,** *Hitotsubashi University*

Abstract

We formulate a localization property of wavelet coefficients for processes with stationary increments, in the estimation problem associated with the processes. A general setting for the estimation is adopted and examples that fit this setting are given. An evaluation of wavelet coefficient decay with respect to shift $k \in \mathbb{N}$ is explicitly derived (only the asymptotic behavior, for large k, was previously known). It is this evaluation that makes it possible to establish the localization property of the wavelet coefficients. In doing so, it turns out that the theory of positive-definite functions plays an important role. As applications, we show that, in the wavelet coefficient domain, estimators that use a simple moment method are nearly as good as maximum likelihood estimators. Moreover, even though the underlying process is long-range dependent and process domain estimates imply the validity of a noncentral limit theorem, for the wavelet coefficient domain estimates we always obtain a central limit theorem with a small prescribed error.

Keywords: Wavelet coefficient; localization; parameter estimation; process with stationary increments; positive-definite function; likelihood; noncentral limit theorem; central limit theorem; long-range dependence

2000 Mathematics Subject Classification: Primary 42C40; 60F05; 60G10; 60G18
Secondary 60G15; 62F10; 62M10

1. Introduction

Let us observe a process $X^T = \{X_t,\ 0 \le t \le T\}$ with stationary increments, and pose the problem of estimating a parameter associated with the process. In this paper, we study the estimation in the wavelet coefficient domain and show a regularization of limit theorems through a localization property of wavelet coefficients, in a sense to be explained below. Therefore, we consider those cases in which the original problems of process domain estimation are well translated into the corresponding problems of wavelet coefficient domain estimation.

Throughout the paper, we assume that X^T is a real-valued process with mean 0 and finite variance, as well as stationary increments. Let ψ be a real-valued wavelet on $\mathbb{R}_+$ satisfying the following assumptions:

(ψ1) ψ has compact support on $W = [0, w]$, for some real $w \ge 1$, and is bounded;

Received 13 December 2004; revision received 12 July 2005.
* Postal address: Institut für Angewandte Mathematik, Universität Bonn, Wegelerstrasse 6, 53115 Bonn, Germany. Email address: albeverio@uni-bonn.de
** Current address: Center for Tsukuba Advanced Research Alliance, The University of Tsukuba, Tennoudai 1-1-1, Tsukuba-shi, Ibaraki-ken, 305-8577, Japan. Email address: kawasaki@wslab.risk.tsukuba.ac.jp

(ψ2) ψ has γth-order vanishing moment for some $\gamma \in \mathbb{N}$, i.e.

$$\int_{\mathbb{R}_+} t^r \psi(t)\,\mathrm{d}t = 0, \qquad r = 0, 1, \dots, \gamma - 1.$$

For each observation length $T > 0$ of sample paths, let $S^T = \{\boldsymbol{s}_j^T,\ j = J_0 + 1, \dots, J\}$ with $\boldsymbol{s}_j^T = \{s_j(k),\ k = 1, \dots, N_{T,j}\}$ be the wavelet coefficient of X^T, defined by

$$s_j(k) = \int_{2^j k}^{2^j(w+k)} \psi_{j,k}(t) X_t\,\mathrm{d}t,$$

where $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$ and J_0 and J are integers such that $J_0 + 1 \le J$. Here, $N_{T,j} = \max\{k\colon 2^j(w+k) \le T\} = \lfloor 2^{-j}T - w \rfloor$ is the maximum number of available wavelet coefficients at scale j up to $t = T$. We remark that if X^T is a process with stationary increments and mean 0, then, for each j, $\{s_j(k), k \in \mathbb{N}\}$ is stationary with respect to k and has mean 0. We also remark that the class of stationary-increment processes contains the class of stationary processes.

The setting of the problem in this paper is as follows. Let a parameter (or statistic) $\xi \in \mathbb{R}$ that we desire to estimate be written as a functional expectation $\xi = \mathrm{E}[\tilde{f}(X_t)]$ or $\xi = \lim_{t\to\infty} \mathrm{E}[\tilde{f}(X_t)]$, for some function $\tilde{f}\colon \mathbb{R} \to \mathbb{R}$. A basic estimator for ξ may be $T^{-1}\int_0^T \tilde{f}(X_t)\,\mathrm{d}t$, which is considered to be a functional $\tilde{f}(X^T)$ of X^T. We then assume that ξ can also be written as $\xi = f(\boldsymbol{\theta})$ for a given $f\colon \mathbb{R}^{\mathcal{J}} \to \mathbb{R}$, $\mathcal{J} = J - J_0$, and is estimated consistently by $\hat{\xi}_T = f(\hat{\boldsymbol{\theta}}_T)$:

$$f(\hat{\boldsymbol{\theta}}_T) \to f(\boldsymbol{\theta}) \quad \text{almost surely (a.s.)} \qquad \text{as } T \to \infty.$$

Here $\hat{\boldsymbol{\theta}}_T = (\hat{\theta}_{T,J_0+1}, \dots, \hat{\theta}_{T,J}) \in \mathbb{R}^{\mathcal{J}}$ converges to $\boldsymbol{\theta} = (\theta_{J_0+1}, \dots, \theta_J) \in \mathbb{R}^{\mathcal{J}}$:

$$\hat{\boldsymbol{\theta}}_T \to \boldsymbol{\theta} \quad \text{a.s.} \qquad \text{as } T \to \infty.$$

Here $\theta_j \in \mathbb{R}$, $j = J_0 + 1, \dots, J$, is assumed to be a functional expectation of the wavelet coefficient at scale j, meaning that $\theta_j = \mathrm{E}[g(s_j(1))]$ for some $g\colon \mathbb{R} \to \mathbb{R}$. As its estimator $\hat{\theta}_{T,j}$, $j = J_0 + 1, \dots, J$, we take in particular the one given by the method of moments, namely

$$\hat{\theta}_{T,j} = \frac{1}{N_{T,j}} \sum_{k=1}^{N_{T,j}} g(s_j(k)), \tag{1}$$

which can be considered a functional $g(\boldsymbol{s}_j^T)$ of $\boldsymbol{s}_j^T$. Thus, $\hat{\boldsymbol{\theta}}_T = \{g(\boldsymbol{s}_j^T),\ j = J_0 + 1, \dots, J\}$. Each θ_j is considered to be a scale component of the desired statistic, and f sums up to produce the desired statistic.

In this setting, we use the terminology *process domain estimator* $\tilde{f}(X^T)$ and *wavelet coefficient domain estimator* $f(\{g(\boldsymbol{s}_j^T)\}) = f \circ g(S^T)$; we do so throughout the paper.

The following examples give cases in which we have this form of estimate.

Example 1. (*Estimation of variance and covariance.*) For a stationary process X^T, let us consider the problem of estimating the variance $\sigma^2 = \mathrm{E}[X_0^2]$ and covariance $r(\tau) = \mathrm{E}[X_\tau X_0]$, $\tau \in [0, T]$. From the Fourier representation of X_t and resolution of unity [17, Theorem 1.1, p. 332], i.e. $\sum_{j\in\mathbb{Z}} |\hat{\psi}(2^j\lambda)|^2 := 1$, where ψ is the Fourier transform of ψ, we have

$$\sigma^2 = \sum_{j\in\mathbb{Z}} 2^{-j}\theta_j = f(\boldsymbol{\theta}), \qquad \theta_j = \sigma_j^2 := \mathrm{E}[s_j^2(0)].$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

In this case, $g(x) = x^2,\ x \in \mathbb{R}$. On the other hand, from Lemma 1, below, we have

$$r(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T X_t X_{t+\tau}\,\mathrm{d}t = \lim_{T\to\infty} \sum_{j=-\infty}^{J_T} \frac{2^{-j}}{N_{T,j}} \sum_{k=1}^{N_{T,j}} s_j^{(X)}(k) s_j^{(\tilde{X})}(k)$$

$$= \lim_{T\to\infty} \sum_{j=-\infty}^{J_T} 2^{-j}\hat{\theta}_{T,j}(\tau) = \lim_{T\to\infty} f(\hat{\boldsymbol{\theta}}_T(\tau)) \quad \text{with } g(x,y) = xy,\ x, y \in \mathbb{R},$$

where $\tilde{X}_t = X_{t+\tau}$, $s_j^{(X)}(k)$ is the wavelet coefficient of X_t, and $J_T = \max\{j \in \mathbb{Z}\colon N_{T,j} \geq 1\}$. In this case,

$$\theta_j = \theta_j(\tau) = \mathrm{E}[s_j^{(X)}(0)s_j^{(\tilde{X})}(0)] = \lim_{T\to\infty} N_{T,j}^{-1} \sum_{k=1}^{N_{T,j}} g(s_j^{(X)}(k), s_j^{(\tilde{X})}(k)) \quad \text{a.s.}$$

and $r(\tau) = f(\boldsymbol{\theta}(\tau)) = \sum_{j\in\mathbb{Z}} 2^{-j}\theta_j(\tau)$.

Lemma 1. *Let X^T and Y^T be ergodic stationary processes with finite variances and let*

$$s_j^{(X^T)}(k) \quad \textit{and} \quad s_j^{(Y^T)}(k)$$

be their wavelet coefficients. Then

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T X_t Y_t\,\mathrm{d}t = \lim_{T\to\infty} \sum_{j=-\infty}^{J_T} \frac{2^{-j}}{N_{T,j}} \sum_{k=1}^{N_{T,j}} s_j^{(X^T)}(k) s_j^{(Y^T)}(k).$$

Example 2. (*Estimation of power spectrum.*) A way of estimating the power spectrum $h(\lambda)$ of a stationary process X^T, for some $\lambda_0 \in \mathbb{R}$, is by using the smoothed periodogram. Then, as is well known (see, e.g. [21, pp. 432–449]), the smoothed periodogram, written in the equivalent form

$$I_T(\lambda_0) = \int_{-T}^{T} \mathrm{e}^{\mathrm{i}\lambda_0\tau}\hat{r}_T(\tau)\,\mathrm{d}\tau, \qquad \hat{r}_T(\tau) = K_T(\tau)\hat{r}_T^0(\tau),$$

consistently estimates $h(\lambda_0)$ as $T \to \infty$. Here K_T is an appropriate δ-approximating sequence in the frequency domain and $\hat{r}_T^0(\tau) = T^{-1}\int_0^T X_t X_{t+\tau}\,\mathrm{d}t$. Let

$$\hat{\theta}_{T,j}^0(\tau) = N_{T,j}^{-1} \sum_{k=1}^{N_{T,j}} s_j^{(X)}(k)s_j^{(\tilde{X})}(k) \quad \text{and} \quad \theta_j^0(\tau) = \mathrm{E}[s_j^{(X)}(0)s_j^{(\tilde{X})}(0)].$$

Since

$$\hat{\theta}_{T,j}(\lambda) = \int_{-T}^{T} \mathrm{e}^{\mathrm{i}\lambda\tau} K_T(\tau)\hat{\theta}_{T,j}^0(\tau)\,\mathrm{d}\tau \to \int_{\mathbb{R}} \mathrm{e}^{\mathrm{i}\tau\lambda}\theta_j^0(\tau)\,\mathrm{d}\tau = \theta_j(\lambda) \quad \text{a.s.} \qquad \text{as } T \to \infty,$$

we may take

$$f(\hat{\boldsymbol{\theta}}_T(\lambda_0)) = \sum_{j=-\infty}^{J_T} 2^{-j}\hat{\theta}_{T,j}(\lambda_0) \to h(\lambda_0) = \sum_{j\in\mathbb{Z}} 2^{-j}\theta_j(\lambda_0) = f(\boldsymbol{\theta}(\lambda_0)) \quad \text{a.s.},$$

according to the arguments of Example 1.

Example 3. (*Estimation of Hermite expansion coefficients.*) When two stationary Gaussian processes X^T and Y^T are observed, let us consider the problem of estimating a functional $f\colon \mathbb{R} \to \mathbb{R}$ in the model $Y_t = f(X_t)$ by estimating the Hermite expansion coefficients $\{c_l\}$ of f, which are such that

$$f(x) = \sum_{l \in \mathbb{N}} c_l H_l(x) \quad \text{in } L^2\left(\mathbb{R}, \mathrm{e}^{-x^2/2} \frac{\mathrm{d}x}{\sqrt{2\pi}}\right).$$

Let $v_l^2 = \mathrm{E}[H_l^2(X_0)]$ and $v_{l,j}^2 = \mathrm{E}[(s_j^{(H_l(X))}(0))^2]$. Since $\mathrm{E}[H_l(X_t)Y_t] = c_l v_l^2$, we have

$$c_l = \frac{\sum_{j \in \mathbb{Z}} 2^{-j}\, \mathrm{E}[s_j^{(H_l(X))}(0) s_j^{(Y)}(0)]}{\sum_{j \in \mathbb{Z}} 2^{-j} v_{l,j}^2} = \frac{\sum_{j \in \mathbb{Z}} 2^{-j} \theta_j^{(1)}(l)}{\sum_{j \in \mathbb{Z}} 2^{-j} \theta_j^{(2)}(l)} = f(\boldsymbol{\theta}(l)),$$

according to the arguments of Example 1. Here $\boldsymbol{\theta}(l) = (\theta_j^{(1)}(l), \theta_j^{(2)}(l))$,

$$\theta_j^{(1)}(l) = \mathrm{E}[s_j^{(H_l(X))}(0) s_j^{(Y)}(0)],$$

and $\theta_j^{(2)}(l) = v_{l,j}^2$, with $\boldsymbol{g}(x, y) = (xy, x^2)$.

Example 4. (*Estimation of Hurst index of fractional Brownian motion.*) Let us consider estimating the Hurst index H, $0 < H < 1$, of fractional Brownian motion. A wavelet-based method for the estimation was proposed in [2] and [1]. The estimator, denoted by $\hat{H}_T$, is given by

$$\hat{H}_T = \sum_{j=J_0+1}^{J} a_{J,J_0}(j) \left[\log_2 \hat{\theta}_{T,j} - \mathcal{J}^{-1} \sum_{j=J_0+1}^{J} \log_2 \hat{\theta}_{T,j}\right] - \frac{1}{2} = f(\hat{\boldsymbol{\theta}}_T),$$

where $a_{J,J_0}(j)$, $j = J_0 + 1, \dots, J$, are the linear least-squares regression coefficients given by

$$a_{J,J_0}(j) = \frac{x_j - \bar{x}_{J,J_0}}{2 \sum_{j=J_0+1}^{J} (x_j - \bar{x}_{J,J_0})^2}$$

with $x_j = j$ and $\bar{x}_{J,J_0} = \mathcal{J}^{-1} \sum_{j=J_0+1}^{J} x_j$, and $g(x) = x^2$, $x \in \mathbb{R}$. In this case, $\theta_j = \sigma_j^2$ and the true H satisfies the above equation with $\hat{\theta}_{T,j}$ replaced by θ_j. The numbers J_0 and J can be finite in this example.

Example 5. (*Estimation of invariant distributions.*) For a given stationary ergodic process X^T, let us consider the problem of estimating the invariant distribution F on $\mathbb{R}$. Let $\mathbf{1}_S(x)$ be the indicator function: $\mathbf{1}_S(x) = 1$ if $x \in S$ and $\mathbf{1}_S(x) = 0$ otherwise. For $a \leq b$, let $c = \{c_{j,k}, k \in \mathbb{N}, j \in \mathbb{Z}\} \subset \mathbb{R}$ be a sequence such that $\mathbf{1}_{[a,b]}(x) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{N}} c_{j,k} \psi_{j,k}(x)$ in $L^2(\mathbb{R})$. Then we can write

$$f(\hat{\boldsymbol{\theta}}_T(c)) = \sum_{j \leq J_T} \sum_{k=1}^{N_{T,j}} c_{j,k} \hat{\theta}_{T,j,k} \to F(\{a \leq x \leq b\}) := \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} c_{j,k} \theta_{j,k} = f(\boldsymbol{\theta}(c)),$$

where

$$\hat{\theta}_{T,j,k} = \frac{1}{T} \int_0^T \psi_{j,k}(t)\, \mathbf{1}_{[a,b]}(X_t)\, \mathrm{d}t \to \int_a^b \psi_{j,k}(x)\, \mathrm{d}F(x) = \theta_{j,k} \quad \text{as } T \to \infty.$$

In this and a companion paper (to appear), we formulate localization properties of the wavelet coefficient S^T with respect to k and j and then propose several applications. In particular, the localization properties provide us with the regularization of certain irregularities (like noncentral limit theorems) in limit theorems. Even if the underlying process is long-range dependent, its wavelet coefficient becomes short-range dependent. As a result, wavelet coefficient domain estimates always satisfy a central limit theorem (CLT). We describe this circumstance below.

In studying the asymptotics of estimates associated with processes having stationary increments, sometimes the limit theorem one obtains is not a CLT but a noncentral limit theorem (NCLT), according to the strength of the long-range dependence. Let $\{X_t\}$ be a stationary-increment Gaussian process with mean increment 0, and let $\{Y_t\}$ be the increment process $Y_t = X_{t+1} - X_t$. We assume ergodicity, with $r_\tau = \mathrm{E}[Y_{s+\tau}Y_s] = O(\tau^{-D})$ as $\tau \to \infty$, for some $0 < D < 1$. Let the parameter to be estimated have the form $A_N^{-1}\sum_{n=1}^N h(Y_n)$, for some function h of Hermite rank $p \geq 1$. Here $A_N \uparrow \infty$ is an appropriate normalization. It is well known (see, e.g. [4], [9], [11], and [23]) that limit theorems hold, with a CLT or an NCLT according to whether $pD \geq 1$ or $pD < 1$, respectively. For example, for estimates of the Hurst index of fractional Brownian motion by variance-type estimators ($p = 2$), we have $D = 2(H-1)$, meaning that $pD = 4(H-1)$. Thus, if $\frac{3}{4} < H < 1$ then the variance-type estimators result in an NCLT.

However, in wavelet coefficient domain estimates,

$$r_j(l) = \mathrm{E}[s_j(l)s_j(0)] = O(l^{-2(H-\gamma)})$$

for each j; hence, the estimator $f(\hat{\boldsymbol{\theta}}_T)$, with j appropriately truncated to satisfy $J_0+1 \leq j \leq J$, obeys the CLT with normalization $A_N = \sqrt{N}$ only if the vanishing moment γ is sufficiently large. The truncation involves a certain error in estimation, but this error can be made as small as desired by letting J_0 be smaller and J larger. If the error is sufficiently small, we can consider only j, $J_0+1 \leq j \leq J$, with a CLT as limit theorem and with scale components of short-range dependence and faster convergence (the larger j is, the slower the convergence becomes, since the number of summands in (1) decreases as $N_{T,j} \sim 2^{-j}T$).

The paper is organized as follows. In Sections 2 and 3 all results are presented. The proofs of all our propositions and theorems are given in Section 4, and Section 5 contains the proofs of all our lemmas. Throughout, $\mathbb{N} = 1, 2, \ldots$ and $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$, while ∂_λ denotes ordinary or distributional derivative with respect to λ.

2. Localization with respect to shifts, and applications

Before stating our theorems, we put forward the following proposition, which gives a general criterion concerning a 'localization property' of stationary sequences. This proposition is used to show the localization property of $\boldsymbol{s}_j^T$ with respect to k, for each j.

Proposition 1. *Let $\mathbf{Y}_N = \{Y_k,\ k = 1, \ldots, N\}$ be a real-valued stationary sequence and $\boldsymbol{\Sigma}_N$ its covariance matrix, i.e. $(\boldsymbol{\Sigma}_N)_{k,k'} = r_{|k-k'|} := \mathrm{cov}[Y_k, Y_{k'}]$, $1 \leq k, k' \leq N$. If there exists a real $\zeta > 1$ such that*

$$\frac{\sum_{k\in\mathbb{N}} r_k^2}{r_0^2} \leq \frac{1}{\zeta^2} \tag{2}$$

then the following inequality holds in the sense of positive definiteness:

$$C_*\boldsymbol{\Lambda}_N \leq \boldsymbol{\Sigma}_N \leq C^*\boldsymbol{\Lambda}_N \quad \textit{for all } N \in \mathbb{N}. \tag{3}$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

Here $C_* = (\zeta - 1)/\zeta$, $C^* = (\zeta + 1)/\zeta$, *and* $\mathbf{\Lambda}_N$ *is the* $N \times N$ *matrix* $\mathrm{diag}(\mathbf{\Sigma}_N)$*, i.e.*

$$(\mathbf{\Lambda}_N)_{k,k'} = \begin{cases} r_0 & \text{if } k = k', \\ 0 & \text{if } k \neq k'. \end{cases}$$

We are interested in the situation where ζ is much larger than 1, in which case the bound in (3) is tighter and $\mathbf{\Sigma}_N$ is considered to be 'close' to the diagonal matrix $\mathbf{\Lambda}_N$, or, equivalently, $\boldsymbol{Y}_N$ is 'close' to an independent sequence.

We will say that a stationary sequence $\boldsymbol{Y}_N$ of the type described in Proposition 1 satisfies the *localization property*.

Now let X^T be a real-valued stationary-increment process with mean 0 and let

$$\boldsymbol{s}_j^T = \{s_j(k),\ k = 1, \ldots, N_{T,j}\}$$

be the set of its wavelet coefficients at scale j. Let $r_j(k) := \mathrm{cov}[s_j(k), s_j(0)]$. Proposition 1 can be applied to $\boldsymbol{s}_j^T$ as follows. We first state the result simply for the case in which $\{X_t\}$ is self-similar as well, in order to make the essence clear, and then for the case in which X^T is not necessarily self-similar but has a regularly varying covariance function. Let $\mathrm{var}[X_1] = 1$ for the sake of simplicity. In the following theorem about self-similar stationary-increment processes, the covariance of X^T is then $\mathrm{cov}[X_s, X_t] = \frac{1}{2}[s^{2H} + t^{2H} - |s - t|^{2H}]$ for some H, $0 < H < 1$ (see, e.g. [12, Theorems 1.3.1 and 3.1.1]).

Theorem 1. *Let* X^T *be a stationary-increment process with* H*-self-similarity. If* $\gamma \geq 2$ *then, for each* j, $r_j(k)$ *satisfies*

$$\frac{r_j(k)}{r_j(0)} \leq (1 + k^2)^{-(\gamma - 1)} \qquad \text{for all } k \in \mathbb{N} \tag{4}$$

and thus satisfies (2) with r_k *replaced by* $r_j(k)$*.*

If we use the fact that $(1 + k^2)^{-2(\gamma-1)} \leq k^{-4(\gamma-1)}$ for $k \geq 4$, then $\sum_{k \in \mathbb{N}} r_j^2(k)/r_j^2(0)$ satisfies

$$\frac{\sum_{k\in\mathbb{N}} r_j^2(k)}{r_j^2(0)} \leq \left(\frac{1}{3}\right)^{2\gamma-3}\left(\frac{4}{5}\right)^{2(\gamma-1)} + \frac{1}{4\gamma - 5}\left(\frac{1}{3}\right)^{4\gamma-5}.$$

If we denote the right-hand side by ζ^{-2}, we have values of ζ for $\gamma = 2, 3, 4$, namely 2.11, 8.10, and 30.4, respectively. Accordingly, (C_*, C^*) takes the values (0.525, 1.48), (0.877, 1.12), and (0.967, 1.03), respectively. Therefore, from the localization point of view, ψ may be hoped to have $\gamma \geq 3$.

We say that S^T has the *localization property with respect to shift* k if (3) holds for each j (ζ does not depend on j for those X^T with self-similarity as in Theorem 1, but it depends on j in general). We call the property k*-localization* for brevity.

Remark 1. In [10], [14], and [24], the asymptotic decay of the wavelet coefficients of fractional Brownian motion, for $0 < H < 1$, was evaluated and found to obey

$$r_j(k) = O(k^{-2(\gamma - H)}) \quad \text{as } k \to \infty. \tag{5}$$

For $\gamma \geq 1$, this decay is faster than that in (4). However, this is just an asymptotic evaluation and does not imply the satisfaction of (2).

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

Now consider the case in which X^T is a process with a regularly varying covariance function. A function $u\colon \mathbb{R}_+ \to (0,\infty)$ is said to be *ρ-regularly varying* [8, p. 18], $\rho \in \mathbb{R}$, if

$$\frac{u(ax)}{u(a)} \to x^\rho, \quad x \in \mathbb{R}_+, \qquad \text{as } a \to \infty.$$

If u is regularly varying then it must be of the form $u(x) = x^\rho \ell(x)$, where $\ell\colon \mathbb{R}_+ \to (0,\infty)$ is a slowly varying function, i.e. one that satisfies $\ell(ax)/\ell(a) \to 1$ as $a \to \infty$. Examples of regularly varying functions include the fractional powers and rational functions. Also, a typical example of a slowly varying function is given by the logarithm.

In the next theorem, we assume that the covariance $r_\tau^X = \mathrm{E}[X_s X_{s+\tau}]$ is $2H$-regularly varying, in the sense that

$$\frac{r_{a\tau}^X}{r_a^X} \to \tau^{2H} \quad \text{as } a \to \infty \qquad \text{or} \qquad r_\tau^X = \tau^{2H}\ell(\tau).$$

This condition is often used by authors in relation to limit theorems of stochastic processes [3], [18], [20], [23]. Here, the choice $\rho = 2H$, $0 < H < 1$, is suggested by the form of the variances of self-similar processes. It will turn out that this form is well suited for the argument below.

Theorem 2. *Let X^T be a stationary-increment process with a regularly varying covariance function. Assume that there exists an absolutely continuous Radon measure $\mu \equiv \mu_{H,\ell}$, $\mu\colon \mathcal{B}((0,\infty)) \to \mathbb{R}_+$, where $\mathcal{B}$ denotes a Borel σ-algebra, such that*

$$\mathrm{E}[(X(\tau) - X(0))^2] = \int_{(0,\infty)} (1 - \mathrm{e}^{-\lambda\tau^2})\,\mathrm{d}\mu(\lambda). \tag{6}$$

Let us denote the density of μ by h, i.e. $\mathrm{d}\mu(\lambda) = h(\lambda)\,\mathrm{d}\lambda$, where $\mathrm{d}\lambda$ is the Lebesgue measure on $(0,\infty)$, and let h satisfy the growth condition

$$h(\lambda) = \begin{cases} O(\lambda^{-\beta}) & \text{as } \lambda \to \infty, \text{ with } \beta > 1, \\ O(\lambda^{-\beta_0}) & \text{as } \lambda \downarrow 0, \text{ with } \beta_0 < 2, \end{cases} \tag{7}$$

and the differential relation

$$3h'(\lambda) + 2h''(\lambda) \geq 0, \qquad \lambda \in (0,\infty). \tag{8}$$

If $\gamma \geq 2$ then the same assertion holds as in Theorem 1.

Remark 2. Since $\mathrm{E}[X_s X_t] = \frac{1}{2}\{\mathrm{E}[X_s^2] + \mathrm{E}[X_t^2] - \mathrm{E}[(X_{|s-t|} - X_0)^2]\}$ and $\mathrm{E}[X_s X_t]$ is a positive-definite kernel, we find that $\mathrm{E}[(X_{|s-t|} - X_0)^2]$ is a negative-definite kernel (see [6] for the definition of a negative-definite kernel). On the other hand, $\Psi(s,t) = (s-t)^2$ is a negative-definite kernel [6, Section 3.1.22]. Hence, (6) is equivalent to $V(s,t) := \mathrm{E}[(X_{|s-t|} - X_0)^2]$ being of the form $V = \varphi \circ \Psi$, with $\varphi(z) = \int_0^\infty (1 - \mathrm{e}^{-\lambda z}) h(\lambda)\,\mathrm{d}\lambda$. We remark that a large class of negative-definite kernels can be obtained using φ [6, p. 77] and, thus, that (6) is not too restrictive.

Next, as an application of k-localization, we point out that, in the wavelet coefficient domain, the maximum likelihood estimator (MLE) of θ is 'close' to the estimator given by the simple moment method. Logically, this is a formulation of the localization for a functional $h(\boldsymbol{Y}_N)$ through the localization for $\boldsymbol{Y}_N$ itself. Here the specific functional $h(\boldsymbol{y})$ is the score function $\dot{\ell}(\theta, \boldsymbol{y}) = \partial_\theta \ell(\theta, \boldsymbol{y})$ for $\boldsymbol{y} \in \mathbb{R}^N$, where ℓ is the log-likelihood.

In the sequel, the stationary processes $\boldsymbol{Y}_N$ and stationary-increment processes X^T are assumed to be Gaussian. Thus, $\boldsymbol{Y}_N \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_N)$. We consider the estimator $\hat{\theta}_N^{\mathrm{MM}} \equiv \hat{\theta}_N^{\mathrm{MM}}(\boldsymbol{Y}_N)$ of $\theta = \mathrm{E}[g(Y_0)]$ for a general measurable function $g\colon \mathbb{R} \to \mathbb{R}$ using the moment method, that is

$$\hat{\theta}_N^{\mathrm{MM}} = \frac{1}{N} \sum_{k=1}^{N} g(Y_N(k)),$$

as well as the MLE $\hat{\boldsymbol{\theta}}^*_{\boldsymbol{\Sigma}_N} = \hat{\boldsymbol{\theta}}^*_{\boldsymbol{\Sigma}_N}(\boldsymbol{Y}_N)$, and show that $\mathrm{var}[\hat{\theta}_N^{\mathrm{MM}}]$ is asymptotically almost as small as $\mathrm{var}[\hat{\boldsymbol{\theta}}^*_{\boldsymbol{\Sigma}_N}]$. The log-likelihood function $\ell_{\boldsymbol{\Sigma}_N}(\theta)$ for $\boldsymbol{Y}_N$ is given by

$$\ell_{\boldsymbol{\Sigma}_N}(\theta) = -\tfrac{1}{2} N \log 2\pi - \tfrac{1}{2} \log |\boldsymbol{\Sigma}_N| - \tfrac{1}{2} \boldsymbol{Y}_N^\top \boldsymbol{\Sigma}_N^{-1} \boldsymbol{Y}_N.$$

On the other hand, for the sake of comparison, let us consider the random vector

$$\tilde{\boldsymbol{Y}}_N = \{\tilde{Y}_k,\ k = 1, \dots, N\} \sim \mathcal{N}(0, \boldsymbol{\Lambda}_N)$$

and the MLE $\hat{\boldsymbol{\theta}}^*_{\boldsymbol{\Lambda}_N} = \hat{\boldsymbol{\theta}}_{\boldsymbol{\Lambda}_N}(\tilde{\boldsymbol{Y}}_N)$ associated with the likelihood $\ell_{\boldsymbol{\Lambda}_N}(\theta)$ given by

$$\ell_{\boldsymbol{\Lambda}_N}(\theta) = -\tfrac{1}{2} N \log 2\pi - \tfrac{1}{2} \log |\boldsymbol{\Lambda}_N| - \tfrac{1}{2} \tilde{\boldsymbol{Y}}_N^\top \boldsymbol{\Lambda}_N^{-1} \tilde{\boldsymbol{Y}}_N.$$

As is well known, the MLE $\hat{\theta}^*_{\boldsymbol{\Lambda}_N}$ is given by the moment method:

$$\hat{\theta}^*_{\boldsymbol{\Lambda}_N} = \frac{1}{N} \sum_{k=1}^{N} g(\tilde{Y}_N(k)).$$

Now consider the case in which the CLT holds for the first two estimators, $\hat{\theta}^{\mathrm{MM}}_{\boldsymbol{\Sigma}_N}$ and $\hat{\theta}^*_{\boldsymbol{\Sigma}_N}$: respectively,

$$\sqrt{N}(\hat{\theta}^{\mathrm{MM}}_{\boldsymbol{\Sigma}_N} - \theta) \xrightarrow{\mathrm{w}} \mathcal{N}(0, v^2_{\boldsymbol{\Sigma}}(\theta)) \quad \text{and} \quad \sqrt{N}(\hat{\theta}^*_{\boldsymbol{\Sigma}_N} - \theta) \xrightarrow{\mathrm{w}} \mathcal{N}(0, \mathcal{I}^{-1}_{\boldsymbol{\Sigma}}(\theta))$$

(in the sense of weak convergence) as $N \to \infty$, where $\mathcal{I}_{\boldsymbol{\Sigma}}(\theta) = \lim_{N\to\infty} \mathrm{var}[\dot{\ell}_{\boldsymbol{\Sigma}_N}(\theta)]$ and $v^2_{\boldsymbol{\Sigma}}(\theta)) > 0$. Since the components of $\tilde{\boldsymbol{Y}}_N$ are independent and identically distributed and have finite variance, the CLT for $\hat{\theta}^*_{\boldsymbol{\Lambda}_N}$ follows automatically, meaning that

$$\sqrt{N}(\hat{\theta}^*_{\boldsymbol{\Lambda}_N} - \theta) \xrightarrow{\mathrm{w}} \mathcal{N}(0, \mathcal{I}^{-1}_{\boldsymbol{\Lambda}}(\theta)),$$

with $\mathcal{I}_{\boldsymbol{\Lambda}}(\theta) = \lim_{N\to\infty} \mathrm{var}[\dot{\ell}_{\boldsymbol{\Lambda}_N}(\theta)]$.

In Proposition 2, we give separate evaluations of $v^2_{\boldsymbol{\Sigma}}(\theta)$ using $\mathcal{I}^{-1}_{\boldsymbol{\Lambda}}(\theta)$ and $\mathcal{I}^{-1}_{\boldsymbol{\Sigma}}(\theta)$. Both evaluations are of interest. The latter is done using the former, via

$$\frac{v^2_{\boldsymbol{\Sigma}}(\theta)}{\mathcal{I}^{-1}_{\boldsymbol{\Sigma}}(\theta)} = \frac{\mathcal{I}^{-1}_{\boldsymbol{\Lambda}}(\theta)}{\mathcal{I}^{-1}_{\boldsymbol{\Sigma}}(\theta)} \cdot \frac{v^2_{\boldsymbol{\Sigma}}(\theta)}{\mathcal{I}^{-1}_{\boldsymbol{\Lambda}}(\theta)}. \tag{9}$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

By the Cramér–Rao lower bound (see, e.g. [13]), we have

$$\mathcal{I}_{\mathbf{\Sigma}}^{-1}(\theta) \le v_{\mathbf{\Sigma}}^2(\theta) \tag{10}$$

in general.

We recall that if $\mathrm{var}[g(Y_0)] < \infty$, then g admits the Hermite expansion $g(x) = \sum_{l \ge p} c_l H_l(x)$ in $L^2(\mathbb{R}, \mathrm{e}^{-x^2/2}\mathrm{d}x/\sqrt{2\pi})$, for some Hermite rank $p \ge 1$. Let η be the power spectral density of $\boldsymbol{Y}_N$, meaning that, for $r_k = \mathrm{cov}[Y_N(k), Y_N(0)]$, we have

$$r_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}k\xi}\eta(\xi)\,\mathrm{d}\xi.$$

Proposition 2. *Assume that $\boldsymbol{Y}_N$ is Gaussian and that* $\mathrm{var}[g(Y_0)] < \infty$ *for a g with Hermite rank $p \ge 1$.*

(i) *If there exists a real $\zeta_p > 1$ such that*

$$\frac{\sum_{k\in\mathbb{N}} r_k^{2p}}{r_0^{2p}} \le \frac{1}{\zeta_p^2} \tag{11}$$

and if $r_k = O(k^{-D})$ as $k \to \infty$ for a real $D > 2/p$, then the following inequality holds, where $C_{p*} = (\zeta_p - 1)/\zeta_p$ and $C_p^* = (\zeta_p + 1)/\zeta_p$:

$$C_{p*}\mathcal{I}_{\mathbf{\Lambda}}^{-1}(\theta) \le v_{\mathbf{\Sigma}}^2(\theta) \le C_p^*\mathcal{I}_{\mathbf{\Lambda}}^{-1}(\theta). \tag{12}$$

(ii) *If (11) holds with $p = 1$ then the following inequality holds:*

$$v_{\mathbf{\Sigma}}^2(\theta) \le C_{\mathbf{\Lambda}/\mathbf{\Sigma}} C_p^* \mathcal{I}_{\mathbf{\Sigma}}^{-1}(\theta). \tag{13}$$

Here $C_{\mathbf{\Lambda}/\mathbf{\Sigma}}$, given by

$$C_{\mathbf{\Lambda}/\mathbf{\Sigma}} = \frac{\mathcal{I}_{\mathbf{\Lambda}}^{-1}(\theta)}{\mathcal{I}_{\mathbf{\Sigma}}^{-1}(\theta)} = \bigg\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\eta(\xi)\,\mathrm{d}\xi\bigg\}\bigg\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\eta(\xi)}\,\mathrm{d}\xi\bigg\},$$

satisfies $1 \le C_{\mathbf{\Lambda}/\mathbf{\Sigma}} \le C_*^{-2}$, assuming that $1/\eta(\xi)$ is integrable on $[-\pi, \pi]$.

Thus, as (12) and (13) show, if the parameter $\zeta_p > 1$ can be made sufficiently large then the estimators given by the moment method perform nearly as well as the MLE. This is remarkable, since MLEs require theoretical and computational hardness in solving likelihood equations exactly or even approximately, whereas the moment method does not.

For the lower bound of $v_{\mathbf{\Sigma}}^2(\theta)$ corresponding to (13), we remark that if $C_{p*}C_{\mathbf{\Lambda}/\mathbf{\Sigma}} \ge 1$ then we have $v_{\mathbf{\Sigma}}^2(\theta) \ge C_{\mathbf{\Lambda}/\mathbf{\Sigma}} C_{p*}\mathcal{I}_{\mathbf{\Sigma}}^{-1}(\theta)$ as well. However, the case of $C_{p*}C_{\mathbf{\Lambda}/\mathbf{\Sigma}} \le 1$ is void because of (10). The criterion for $C_{p*}C_{\mathbf{\Lambda}/\mathbf{\Sigma}} \ge 1$ is not clear.

Proposition 2 immediately applies to the wavelet coefficients $\boldsymbol{s}_j^T$. As is well known, the MLE for $f(\boldsymbol{\theta})$ is given by $f(\hat{\boldsymbol{\theta}}_N^*)$ (see, e.g. [19]); in particular, $\hat{\boldsymbol{\theta}}_N^* = (\hat{\theta}_{N,J_0+1}^*, \dots, \hat{\theta}_{N,J}^*)$, with $\hat{\theta}_{N,j}^*$ an MLE for each $j = J_0 + 1, \dots, J$. Thus, we have only to consider the MLE for an arbitrarily fixed j below.

By setting $\boldsymbol{Y}_N = \boldsymbol{s}_j^T$, $r_k = r_j(k)$, and $N = N_{T,j}$, we have the following corollary, which is an easy consequence of Theorem 2 and Proposition 2 except for the relation of p, H, and γ. We remark that the spectral density η_j of $\boldsymbol{s}_j$ is such that

$$r_j(k) = \frac{2^j}{2\pi}\int_{-2^{-j}\pi}^{2^{-j}\pi} \mathrm{e}^{\mathrm{i}2^j k\xi}\eta_j(\xi)\,\mathrm{d}\xi.$$

Corollary 1. *Let X^T be Gaussian, have stationary increments, and have a $2H$-regularly varying covariance function. Suppose that* $\mathrm{var}[g(X_0)] < \infty$ *for a g with Hermite rank $p \geq 1$, and that X^T satisfies the assumptions of Theorem 2. If $\gamma \geq 2$, then we have the same conclusion as in Proposition 2, with $\boldsymbol{Y}_N$ and $\eta(\xi)$ replaced by $\boldsymbol{s}_j^T$ and $\eta_j(2^j\xi)$, respectively.*

3. Central limit theorem for $\hat{\boldsymbol{\theta}}_T$ and $f(\hat{\boldsymbol{\theta}}_T)$

Now let us consider the central limit theorem for $\hat{\boldsymbol{\theta}}_T$ and $f(\hat{\boldsymbol{\theta}}_T)$ using multiple scales $j = J_0 + 1, \ldots, J$ simultaneously for integers $J_0 + 1 \leq J$. Let us assume that $\{X_t\}$ is Gaussian and that $\mathrm{E}[g(X_0)] = 0$ and $\mathrm{var}[g(X_0)] < \infty$. Thus, g has Hermite rank p for some $p \geq 1$ and a Hermite expansion of the form $g(x) = \sum_{n \geq p} c_l H_l(x)$ in $L^2(\mathbb{R}, \mathrm{e}^{-x^2/2}\mathrm{d}x/\sqrt{2\pi})$. Here, we assume the covariance function $r_\tau^X := \mathrm{cov}[X_0, X_\tau]$ to be smoothly varying, which property is a little stronger than the regular variation one. A function $\varphi\colon \mathbb{R}_+ \to (0, \infty)$ of ρ-regular variation is said to be *smoothly varying* (or *ρ-smoothly varying*) [8, Section 1.8.1] if

$$\frac{x^i \partial^i \varphi(x)}{\varphi(x)} \to \prod_{\nu=0}^{i} (\rho - \nu) \quad \text{as } x \to \infty, \qquad i \in \mathbb{N}.$$

Theorem 3. *Let X^T be a stationary-increment process with a $2H$-smoothly varying covariance function. If $\gamma \geq 2$ then*

$$\sqrt{N_{T,J}}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}) \xrightarrow{\mathrm{w}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_J). \tag{14}$$

The (j, j')th component of $\boldsymbol{\Sigma}_J$, with $j \leq j'$, is given by

$$(\boldsymbol{\Sigma}_J)_{j,j'} = \frac{1}{d_j} \sum_{\nu \in \mathbb{Z}} r_{j,j'}(\nu), \qquad r_{j,j'}(\nu) = \mathrm{cov}[g(s_j(k)), g(s_{j'}(k'))]\Big|_{k - 2^{j'-j}k' = \nu}, \tag{15}$$

or

$$(\boldsymbol{\Sigma}_J)_{j,j'} = \sum_{\nu \in \mathbb{Z}} r_{j,j'}^{(Y)}(\nu), \qquad r_{j,j'}^{(Y)}(\nu) = \mathrm{cov}[Y_j(n), Y_{j'}(n')]\Big|_{n - n' = \nu}, \tag{16}$$

where $d_j = 2^{J-j}$ and $Y_j(n) = d_j^{-1} \sum_{k=1}^{d_j} g(s_j(d_j n + k))$. Moreover, if

$$\boldsymbol{\partial f} := (\partial f/\partial x_{J_0+1}, \ldots, \partial f/\partial x_J)^\top$$

is continuous in a neighborhood of $\boldsymbol{\theta}$, then

$$\sqrt{T}(f(\hat{\boldsymbol{\theta}}_T) - f(\boldsymbol{\theta})) \xrightarrow{\mathrm{w}} \mathcal{N}(0, v^2(\boldsymbol{\theta})) \quad \text{as } T \to \infty, \tag{17}$$

with $v^2(\boldsymbol{\theta}) = 2^J \boldsymbol{\partial f}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}_J \boldsymbol{\partial f}(\boldsymbol{\theta})$.

In order to understand the implications of this CLT in the wavelet coefficient domain, let us recall the NCLT obtained in the process domain limit theorem and compare them. In these respective domains, let $\{X_t\}$ and $\{X_n\}$ be stationary Gaussian processes with long-range dependence such that their covariance is regularly varying when the time lag is large. Let A_N be an appropriate normalization, with $A_N \to \infty$ as $x \to \infty$. Then the nonlinear functional of the form

$$\frac{1}{A_N} \int_0^{Nt} f(X_s)\,\mathrm{d}s \quad \text{or} \quad \frac{1}{A_N} \sum_{n=0}^{\lfloor Nt \rfloor} f(X_n)$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

converges weakly with Skorokhod's topology in $D[0,1]$ (the metric space of right-continuous paths with left limits) to the Hermite processes given by the multiple Wiener–Itô integrals

$$Z_t = C_q \int_{\mathbb{R}^p}' \int_0^t \prod_{l=1}^{p} (s - y_j)^{-(1+q)/2} \mathbf{1}(\{y_j \le s\}) \, \mathrm{d}s \, \mathrm{d}B(y_1) \cdots \mathrm{d}B(y_p),$$

for some q such that $0 < q < 1/p$, where B is the standard Brownian motion, $\int_{\mathbb{R}^p}'$ stands for the integral over $\mathbb{R}^p$ that avoids $\{y_i = y_j\}$, $i \neq j$, $\mathbf{1}$ is the indicator function, and $C_q > 0$ is a constant depending only on q (see [11], [12], [20], and [23]). Thus, the parameter estimation in the above CLT corresponds to weak convergence of the marginal distribution for $t = 1$. We remark that Hermite processes have finite variances.

When we wish to use a statistical procedure like hypothesis testing, an evaluation of percentiles of the limit distribution associated with the interval estimation is necessary. In the case of process domain estimations, we must evaluate the percentile of Z_t (or Z_1), which might be difficult.

On the other hand, in the case of a wavelet-based estimation, we can perform the interval estimation using the above CLT as follows. Let $z_{a/2} > 0$ be the percentile corresponding to $100a\%$ confidence, for some a, $0 < a < 1$, and let us write $f \equiv f_{J_0,J}$ and $v^2(\boldsymbol{\theta}) \equiv v^2_{J_0,J}(\boldsymbol{\theta})$ to make the dependence on J and J_0 clear. Note that $f_{-\infty,\infty}$ corresponds to exact estimation. Then, according to the CLT,

$$\lim_{T\to\infty} \Pr(\{|f_{J_0,J}(\hat{\boldsymbol{\theta}}_T) - f_{J_0,J}(\boldsymbol{\theta})| \le z_{a/2} v_{J,J_0}(\boldsymbol{\theta}) T^{-1/2}\}) = 1 - a.$$

For an arbitrarily small $\varepsilon > 0$, we can choose large J and $|J_0|$ such that $|f_{J_0,J}(\boldsymbol{\theta}) - f_{-\infty,\infty}(\boldsymbol{\theta})| \le \varepsilon$. It then turns out that the truncated estimation satisfies

$$\lim_{T\to\infty} \Pr(\{|f_{J_0,J}(\hat{\boldsymbol{\theta}}_T) - f_{-\infty,\infty}(\boldsymbol{\theta})| \le \varepsilon + z_{a/2} v_{J,J_0}(\boldsymbol{\theta}) T^{-1/2}\}) \ge 1 - a. \tag{18}$$

For a fixed a, if we would like to make ε smaller we have to make J and $|J_0|$ larger, but this makes $v_{J,J_0}(\boldsymbol{\theta})$ larger. Thus, given this trade-off, an appropriate criterion for choosing J and J_0 given ε may be of future interest. We note, however, that if ε is large compared with $z_{a/2} v_{J,J_0}(\boldsymbol{\theta}) T^{-1/2}$, then taking T larger in (18) is meaningless. Thus, ε should be kept smaller than the latter term, which means that ε is to be determined as a function of T.

By truncating the scales with $J < \infty$, we can exclude the lower-frequency components contained in the original process, which cause long-range dependence and imply an NCLT. As a result, by allowing a small error ε, we can perform the estimation with only the short-range-dependent components, and obtain a CLT.

The CLT (14) itself is obtained as a result of a variant of k-localization. Here we use an asymptotic estimate rather than the evaluation of the form (4); see Remark 3, below.

Lemma 2. *For each pair* (k, k'), $1 \le k \le d_j$, $1 \le k' \le d_{j'}$, *we have*

$$\mathrm{E}[s_j(d_j n + k) s_{j'}(k')] = O(n^{-2(\gamma - H)} \ell(n)) \quad \text{as } n \to \infty.$$

There are many known results about limit theorems for functionals of general stationary Gaussian vectors (see [3], [4], [18], and [20], for example). However, the important point here is how to adapt the wavelet coefficients to those general results, because $(s_{J_0+1}(k), \ldots, s_J(k))$ is not a stationary vector sequence with respect to k. See [5] for the case of Hurst index estimation,

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

which shows how complicated it is to show the validity of the CLT. We can overcome this difficulty by reconstructing the wavelet coefficients S^T in such a way that the resulting sequence is a stationary vector and the asymptotic covariance matrix $\mathbf{\Sigma}_J$ is unchanged.

Remark 3. In the case that $j' > j$, it might be possible to consider an expression similar to (4) as the sufficient condition for Theorem 3, instead of Lemma 2. In the CLT, however, it is better to have a more precise asymptotic evaluation, as given in Lemma 2. In fact, to obtain the CLT, it turns out that $2p(\gamma - H) > 1$, i.e. $\gamma > H + 1/2p$, is a sufficient condition. Thus, according to the value of (p, H), there are cases in which the minimum γ needed for the CLT to hold is $\gamma = 1$. On the other hand, by the evaluation for integer arguments, $2p(\gamma - 1) > 1$, i.e. $\gamma > 1 + 1/2p$, is a sufficient condition irrespective of the value of H. Thus, with the latter evaluation we can conclude only that $\gamma = 2$ is sufficient for the CLT for all $H,\ 0 < H < 1$.

We also note that a covariance of the reduced form (15) or (16) is obtained if $\gamma > H + 1/p$. There are indeed cases in which $\gamma = 1$ is sufficient for the CLT to hold but $\gamma = 2$ is sufficient to have a covariance of reduced form. The condition $\gamma \geq 2$ is sufficient for both to hold.

4. Proofs of propositions and theorems

In this section, we give proofs of the propositions and theorems presented above. Proofs of the lemmas are all given in the next section.

4.1. Proof of Proposition 1

To prove (3), we verify that the determinant of each principal minor of $C^*\mathbf{\Lambda}_N - \mathbf{\Sigma}_N$ and $\mathbf{\Sigma}_N - C_*\mathbf{\Lambda}_N$ is nonnegative, which is a necessary and sufficient condition for the square matrices to be nonnegative definite [6, Theorem 3.1.16]. To this end, we follow the induction procedure. Let $g_n,\ n = 1, 2, \dots, N$ denote the determinant of the $n \times n$ principal minor: $g_n = |C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n|$. The elements of $C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n$ are

$$(C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n)_{i,j} = \begin{cases} r_0/\zeta & \text{if } i = j, \\ -r_{|i-j|} & \text{if } i \neq j. \end{cases}$$

First, we have $g_2 = (r_0/\zeta)^2 - r_1^2$. The condition in Proposition 1 implies that $g_2 \geq 0$. Next, let us assume that $g_n \geq 0$ for some $n,\ 2 \leq n \leq N - 1$. Then $C^*\mathbf{\Lambda}_{n+1} - \mathbf{\Sigma}_{n+1}$ is of the form

$$C^*\mathbf{\Lambda}_{n+1} - \mathbf{\Sigma}_{n+1} = \begin{bmatrix} r_0/\zeta & -\boldsymbol{r}_n^\top \\ -\boldsymbol{r}_n & C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n \end{bmatrix},$$

where $\boldsymbol{r}_N = (r_1, r_2, \dots, r_N)^\top$. Let $\boldsymbol{x}_n \in \mathbb{R}^n$ be the solution of the linear equation

$$(C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n)\boldsymbol{x}_n = \boldsymbol{r}_n.$$

Since sweeping out does not change the value of the determinant, we have

$$g_{n+1} = \begin{vmatrix} A_n & -\boldsymbol{r}_n^\top \\ \mathbf{0}_n & C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n \end{vmatrix},$$

where the scalar A_n is given by

$$A_n = r_0/\zeta - \boldsymbol{r}_n^\top \boldsymbol{x}_n = r_0/\zeta - \boldsymbol{r}_n^\top (C^*\mathbf{\Lambda}_n - \mathbf{\Sigma}_n)^{-1}\boldsymbol{r}_n \tag{19}$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

and $\mathbf{0}_n$ is the n-dimensional zero vector. Hence, we have $g_{n+1} = A_n g_n$ and it is enough to show that $A_n \geq 0$.

Since $\boldsymbol{\Sigma}_n \boldsymbol{\Lambda}_n^{-1}$ is positive definite, it can be diagonalized by an orthogonal matrix $\boldsymbol{B}_n$, so that $\boldsymbol{\Sigma}_n \boldsymbol{\Lambda}_n^{-1} = \boldsymbol{B}_n^\top \boldsymbol{V}_n \boldsymbol{B}_n$ for a diagonal matrix $\boldsymbol{V}_n = \operatorname{diag}(v_{n,1}, \dots, v_{n,n})$ with $v_{n,i} \in \mathbb{R}_+$. Then, on the right-hand side of (19), $(C^* \boldsymbol{\Lambda}_n - \boldsymbol{\Sigma}_n)^{-1}$ can be written as

$$(C^* \boldsymbol{\Lambda}_n - \boldsymbol{\Sigma}_n)^{-1} = \frac{1}{C^*} \boldsymbol{\Lambda}_n^{-1/2} \left(\boldsymbol{I} - \frac{\boldsymbol{\Sigma}_n \boldsymbol{\Lambda}_N^{-1}}{C^*} \right)^{-1} \boldsymbol{\Lambda}_n^{-1/2}$$

$$= \frac{1}{C^*} \boldsymbol{\Lambda}_n^{-1/2} \sum_{l=0}^{\infty} \frac{1}{(C^*)^l} \boldsymbol{B}_n^\top \boldsymbol{V}_n^l \boldsymbol{B}_n \boldsymbol{\Lambda}_n^{-1/2}.$$

Therefore, A_n becomes

$$A_n = \frac{r_0}{\zeta} - \frac{1}{C^*} \sum_{l=0}^{\infty} \left(\frac{1}{C^*} \right)^l \boldsymbol{r}_n^\top \boldsymbol{\Lambda}_n^{-1/2} \boldsymbol{B}_n^\top \boldsymbol{V}_n^l \boldsymbol{B}_n \boldsymbol{\Lambda}_n^{-1/2} \boldsymbol{r}_n$$

$$= \frac{r_0}{\zeta} - \frac{1}{C^*} \sum_{l=0}^{\infty} \left(\frac{1}{C^*} \right)^l \| \boldsymbol{V}_n^{l/2} \boldsymbol{B}_n \boldsymbol{\Lambda}_n^{-1/2} \boldsymbol{r}_n \|^2,$$

where $\|\boldsymbol{x}_n\|^2 = \sum_{i=1}^n x_i^2$ for a vector $\boldsymbol{x}_n = (x_1, \dots, x_n)$. Let $v_n^* = \max_{1 \leq i \leq n} v_{n,i}$ and $v^* = \sup_{n \in \mathbb{N}} v_n^*$; then $v_n^* \leq 1$. In fact, $v_n^* \uparrow v^*$ by [16, p. 65, Equations (9) and (10)], while in the present case $v^* = \max_{-\pi \leq \xi \leq \pi} \eta(\xi)/r_0 = \eta(0)/r_0 = 1$ by [15, Theorem 4.2.1, p. 154]. Therefore, $C^* - v_n^* \geq (1+\zeta)/\zeta - 1 = \zeta^{-1}$. Since $\| \boldsymbol{V}_n^{l/2} \boldsymbol{B}_n \boldsymbol{\Lambda}_n^{-1/2} \boldsymbol{r}_n \|^2 \leq ((v_n^*)^l / \sigma^2) \|\boldsymbol{r}_n\|^2$, we have

$$A_n \geq r_0 \left[\frac{1}{\zeta} - \frac{\|\boldsymbol{r}_n\|^2}{C^* r_0^2} \sum_{l=0}^{\infty} \left(\frac{v_n^*}{C^*} \right)^l \right] = r_0 \left[\frac{1}{\zeta} - \frac{\|\boldsymbol{r}_n\|^2}{r_0^2} \frac{1}{C^* - v_n^*} \right] \geq r_0 \left[\frac{1}{\zeta} - \frac{\zeta \|\boldsymbol{r}_n\|^2}{r_0^2} \right].$$

Thus, we obtain

$$A_n \geq r_0 \zeta \left[\frac{1}{\zeta^2} - \frac{1}{r_0^2} \sum_{i \in \mathbb{N}} r_i^2 \right] \geq 0$$

by condition (2).

4.2. Proof of Theorem 1

Since $r_j(k),\ k \in \mathbb{N}_0$, can be written as

$$r_j(k) = 2^{(2H+1)j} \left(-\frac{1}{2} \right) \iint_{W^2} \psi(s)\psi(t) |s - t + k|^{2H} \, \mathrm{d}s \, \mathrm{d}t = 2^{(2H+1)j} r_0(k), \tag{20}$$

as is well known (see, e.g. [14]), to prove (4) it is enough to consider $r_0(k)$, i.e. we will show that $\sum_{k \in \mathbb{N}} r_0^2(k)/r_0^2(0) \leq \zeta^{-2}$. By a formula given in [6, Corollary 3.2.10], the term

$$|s - t + k|^{2H} = \{(s - t + k)^2\}^H$$

can be written as

$$|s - t + k|^{2H} = c_H \int_{(0,\infty)} (1 - \mathrm{e}^{-\lambda (s-t+k)^2}) \, \mathrm{d}\tilde{\mu}_H(\lambda), \tag{21}$$

where $\mathrm{d}\tilde{\mu}_H(\lambda) = \lambda^{-(1+H)}\,\mathrm{d}\lambda$ and c_H is an appropriate positive constant. By changing variable to $\xi = \sqrt{\lambda}$ and substituting the Fourier integral

$$\mathrm{e}^{-\xi^2(s-t+k)^2} = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(s-t+k)^2}} \exp\left\{-\frac{x^2}{2(s-t+k)^2}\right\} \mathrm{e}^{\mathrm{i}\xi x}\,\mathrm{d}x$$

into (21), we have

$$|s-t+k|^{2H} = C_H \int_{(0,\infty)} (1-\mathrm{e}^{\mathrm{i}\lambda(s-t+k)})\,\mathrm{d}\mu_H(\lambda),$$

where

$$C_H = \frac{c_H}{2}\int_{\mathbb{R}} |x|^{2H} \frac{\mathrm{e}^{-x^2/2}\,\mathrm{d}x}{\sqrt{2\pi}} \quad \text{and} \quad \mathrm{d}\mu_H(\lambda) = \lambda^{-(1+2H)}\,\mathrm{d}\lambda.$$

We note the identity [22, p. 232]

$$\mathrm{e}^{\mathrm{i}\lambda k} = \frac{(I-\partial_\lambda^2)^m}{(1+k^2)^m}\mathrm{e}^{\mathrm{i}\lambda k}, \qquad m \in \mathbb{N},$$

where I is the identity operator. Let us first consider the case $m = 1$. By Lemma 3, in (20) $r_0(k) \geq 0$ for all $k \in \mathbb{N}_0$. Hence, by integration by parts, we have

$$\begin{aligned} 0 \leq r_0(k) &= C_H \int_{(0,\infty)} \mathrm{e}^{\mathrm{i}\lambda k}\left[\iint_{W^2} \psi(s)\psi(t)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\,\mathrm{d}s\,\mathrm{d}t\right]\mathrm{d}\mu_H(\lambda) \\ &= \frac{C_H}{1+k^2}\int_{(0,\infty)} \mathrm{e}^{\mathrm{i}\lambda k}\left[\iint_{W^2} \psi(s)\psi(t)(I-\partial_\lambda^2)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\lambda^{-(1+2H)}\,\mathrm{d}s\,\mathrm{d}t\right]\mathrm{d}\lambda \qquad (22) \end{aligned}$$

if $\gamma > H+1$. We can write

$$(I-\partial_\lambda^2)[\mathrm{e}^{\mathrm{i}\lambda(s-t)}\lambda^{-(1+2H)}] = (1-\varphi(s,t))\mathrm{e}^{\mathrm{i}\lambda(s-t)}\lambda^{-(3+2H)},$$

where

$$\varphi(s,t) = -\lambda^2(s-t)^2 - 2(1+2H)\mathrm{i}\lambda(s-t) + (1+2H)(2+2H).$$

We then have

$$\iint_{W^2} \psi(s)\psi(t)(I-\partial_\lambda^2)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\lambda^{-(1+2H)}\,\mathrm{d}s\,\mathrm{d}t \leq |\hat{\psi}(\lambda)|^2\lambda^{-(1+2H)},$$

since $\varphi(s,t)$ is a positive-definite kernel (see [6, Sections 3.1.9 and 3.1.10]) and so is $\mathrm{e}^{\mathrm{i}\lambda(s-t)}$; hence, $\varphi(s,t)\mathrm{e}^{\mathrm{i}\lambda(s-t)}$ is positive definite, by the fact that the product of two positive-definite kernels is again positive definite [6, Section 3.1.12]. Therefore, we have

$$r_0(k) \leq \frac{C_H}{1+k^2}\int_{(0,\infty)} |\hat{\psi}(\lambda)|^2\,\mathrm{d}\mu_H(\lambda) = \frac{r_0(0)}{1+k^2}$$

(see the expression for $r_0(0)$ in (22)).

For the case $m \geq 2$, we have

$$0 \leq r_0(k) = \frac{C_H}{(1+k^2)^m}\int_{(0,\infty)} \{(I-\partial_\lambda^2)^m\mathrm{e}^{\mathrm{i}\lambda k}\}\left[\iint_{W^2} \psi(s)\psi(t)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\lambda^{-(1+2H)}\,\mathrm{d}s\,\mathrm{d}t\right]\mathrm{d}\lambda,$$

where

$$(I-\partial_\lambda^2)^m e^{i\lambda k} = \left\{ I - \frac{(1+k^2)^m - 1}{k^2}\partial_\lambda^2 \right\} e^{i\lambda k}. \tag{23}$$

Hence, we obtain

$$r_0(k) = \frac{C_H}{(1+k^2)^m} \times \int_{(0,\infty)} e^{i\lambda k}\left[\iint_{W^2} \psi(s)\psi(t)\left\{ I - \frac{(1+k^2)^m - 1}{k^2}\partial_\lambda^2 \right\} e^{i\lambda(s-t)}\lambda^{-(1+2H)}\, ds\, dt\right] d\lambda \tag{24}$$

if $\gamma > H + m$, and, thus, $r_0(k) \le r_0(0)/(1+k^2)^m$ by the same reason as in the $m = 1$ case.

Finally we claim that the largest possible value of m is $\gamma - 1$. This follows from an integrability argument for (24) with respect to λ. Assumption ($\psi 1$) implies that $\hat{\psi}$ is an entire function, meaning that it has a Taylor expansion for all $\lambda \in \mathbb{R}_+$. Also, assumption ($\psi 2$) implies that $\partial_\lambda^n \hat{\psi}(0) = 0$, $n = 0, \dots, \gamma - 1$. Thus, the Taylor expansion is of the form $\hat{\psi}(\lambda) = \sum_{n=\gamma}^{\infty} a_n \lambda^n$ for some $\{a_n\} \subset \mathbb{C}$. Among the terms contained in

$$(I-\partial_\lambda^2)^m |\hat{\psi}(\lambda)|^2 \lambda^{-(1+2H)} = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \sum_{l=0}^{2i} \binom{2i}{l} \left\{ \prod_{\nu=0}^{l-1} 1 + 2H + \nu \right\} \times \sum_{n=0}^{\infty} b_{n+2\gamma} \left\{ \prod_{\nu=0}^{2i-l-1} n + 2\gamma - \nu \right\} \lambda^{n+2\gamma-(2i+1+2H)},$$

the one with the 'worst' singularity as $\lambda \downarrow 0$ has $n = 0$ and $i = m$. Hence, the assertion holds if $2\gamma - (1 + 2H + 2m) > -1$, i.e. $\gamma > m + H$. Since $m \ge 1$ and $0 < H < 1$, we conclude that if $\gamma \ge 2$ then (4) holds with $m = \lfloor \gamma - H \rfloor = \gamma - 1$.

Lemma 3. *If $\{X_t\}$ has stationary increments and a regularly varying covariance function, then $r_j(k) \ge 0$ for all $k \in \mathbb{N}$.*

Remark 4. A more precise evaluation of $r_j(k)/r_j(0)$ may be possible if we evaluate the term associated with ∂_λ^2 in (23) rather than neglecting it. Also, in order to have the decay rate given in (5), certain additional steps, like taking fractional derivatives of the Fourier transform $\hat{\psi}$, might be necessary.

4.3. Proof of Theorem 2

Since

$$E[X_s X_t] = \tfrac{1}{2}\{E[X_s^2] + E[X_t^2] - E[(X_{|s-t|} - X_0)^2]\},$$

that $E[X_s X_t]$ has regular variation as a function of $|s-t|$ implies that $E[(X_{|s-t|} - X_0)^2]$ has regular variation. Also, that $E[X_s X_t]$ is a positive-definite kernel implies that $E[(X_{|s-t|} - X_0)^2]$ is a negative-definite kernel. It is clear that if $\tilde{\ell}$ is a slowly varying function then $\ell(\tau) = \tilde{\ell}(\tau^{1/2})$ is slowly varying as well. Hence, as a function of $(s-t)^2$, we can write

$$E[(X_{|s-t|} - X_0)^2] = \Psi((s-t)^2), \qquad \Psi(x) = x^H \ell(x) \quad \text{for } x \in \mathbb{R}_+,$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

where $0 < H < 1$ and ℓ is a slowly varying function. From the assumption of the theorem, the negative-definite kernel $\mathrm{E}[(X_{|s-t|} - X_0)^2]$ has the representation

$$\mathrm{E}[(X_{|s-t|} - X_0)^2] = \int_{(0,\infty)} (1 - \mathrm{e}^{-\lambda(s-t)^2}) h_{H,\ell}(\lambda)\,\mathrm{d}\lambda.$$

By a calculation similar to that in the proof of Theorem 1 (cf. (20)), the covariance $r_j(k)$ can be written as

$$\begin{aligned} r_j(k) &= 2^j\left(-\frac{1}{2}\right)\iint_{W^2} \psi(s)\psi(t)|2^j(s-t+k)|^{2H}\ell(2^{2j}(s-t+k)^2)\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_{(0,\infty)} \mathrm{e}^{\mathrm{i}\lambda k}\left[\iint_{W^2} \psi(s)\psi(t)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\,\mathrm{d}s\,\mathrm{d}t\right]\eta_j(\lambda)\,\mathrm{d}\lambda, \end{aligned}$$

where $\eta_j \equiv \eta_{H,\ell,j}$ is defined by

$$\eta_j(\lambda) = 2^j\lambda\int_{\mathbb{R}_+} \frac{1}{\sqrt{2\pi}} x^{-2}\exp\{-2^{-(2j+1)}x^2\} h_{H,\ell}\left(\left(\frac{\lambda}{x}\right)^2\right)\mathrm{d}x, \qquad \lambda \in (0,\infty),$$

which exists for each $\lambda \in (0,\infty)$ by assumption (7) and Lemma 4, below. For the same reason as in the proof of Theorem 1, we have only to consider the case $m = 1$. We have

$$r_j(k) = \frac{2^j}{1+k^2}\int_{(0,\infty)} \mathrm{e}^{\mathrm{i}\lambda k}\left[\iint_{W^2} \psi(s)\psi(t)(I - \partial_\lambda^2)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\eta_j(\lambda)\,\mathrm{d}s\,\mathrm{d}t\right]\mathrm{d}\lambda, \tag{25}$$

where $(I - \partial_\lambda^2)\mathrm{e}^{\mathrm{i}\lambda(s-t)}\eta_j(\lambda) = (\eta_j(\lambda) - \varphi(s,t))\mathrm{e}^{\mathrm{i}\lambda(s-t)}$ with

$$\varphi(s,t) = -(s-t)^2\eta_j(\lambda) + \mathrm{i}(s-t)\partial_\lambda\eta_j(\lambda) + \partial_\lambda^2\eta_j(\lambda).$$

In order to neglect the term associated with ∂_λ^2 in (25), so that we can evaluate $r_j(k)$ using only terms associated with I, it is sufficient that $\varphi(s,t)$ is a positive-definite kernel, for which in turn $\partial_\lambda^2\eta_j(\lambda) \geq 0$ is a sufficient condition. It follows that this is satisfied if

$$\partial_\lambda^2\left[\lambda h_{H,\ell}\left(\left(\frac{\lambda}{x}\right)^2\right)\right] = \frac{2\lambda}{x^2}\left\{3h'\left(\left(\frac{\lambda}{x}\right)^2\right) + \frac{2\lambda^2}{x^2}h''\left(\left(\frac{\lambda}{x}\right)^2\right)\right\} \geq 0,$$

i.e. if $3h'(\lambda) + 2\lambda h''(\lambda) \geq 0$.

To determine the largest possible value of m, we consider

$$\partial_\lambda^i\eta_j(\lambda) = \int_{\mathbb{R}_+} \frac{1}{\sqrt{2\pi}} x^{-2}\exp\{-2^{-(2j+1)}x^2\}\left\{\partial_\lambda^{i-1}h_j\left(\left(\frac{\lambda}{x}\right)^2\right) + \lambda\partial_\lambda^i h_j\left(\left(\frac{\lambda}{x}\right)^2\right)\right\}\mathrm{d}x.$$

Since $\partial_\lambda^i h_j((\lambda/x)^2) = O(x^{2\beta-2})$ for all i as $x \downarrow 0$, it turns out that the growth assumption (7) implies that this integral is finite for all i. Also, by Lemma 4, below, we have

$$\partial_\lambda^i h_j((\lambda/x)^2) = O(\lambda^{-2(1+H)-i}) \quad \text{as } \lambda \downarrow 0,$$

which leads to the sufficient condition for the finiteness of the integral in (25) with respect to λ, namely

$$\begin{aligned} 2\gamma - 2(1+H) - 2m + 1 > -1 &\Leftrightarrow m < \gamma - H \\ &\Leftrightarrow m = \lfloor \gamma - H \rfloor = \gamma - 1. \end{aligned}$$

Lemma 4. *Under condition (7), β_0 is equal to $1 + H$.*

Remark 5. It is easily seen that the function $\lambda^{-(1+2H)}$ in Theorem 1, as a special case of $\eta_j(\lambda)$, satisfies condition (8).

4.4. Proof of Proposition 2

(i) Let $v^2_{\mathbf{\Sigma}_N}(\theta) = \mathrm{var}[\sqrt{N}(\hat\theta^{\mathrm{MM}}_{\mathbf{\Sigma}_N} - \theta)]$. We then have

$$\begin{aligned} v^2_{\mathbf{\Sigma}_N}(\theta) &= \sum_{l\ge p} c_l^2\biggl(\mathrm{var}[H_l(Y_0)] + \frac{2}{N}\sum_{k=1}^{N-1}(N-k)\,\mathrm{cov}[H_l(Y_k), H_l(Y_0)]\biggr)\\ &\to \sum_{l\ge p} c_l^2\biggl(\mathrm{var}[H_l(Y_0)] + 2\sum_{k\in\mathbb{N}}\mathrm{cov}[H_l(Y_k), H_l(Y_0)]\biggr)\\ &= v^2_{\mathbf{\Sigma}}(\theta) \quad \text{as } N\to\infty, \end{aligned} \tag{26}$$

by the decay condition $r_k = O(k^{-D})$, $D > 2/p$. By applying Proposition 1 to $\{H_l(Y_k)\}$ instead of $\{Y_k\}$, for each $l \ge p$, we obtain the following statement, where $(\mathbf{\Sigma}_N^{(l)})_{k,k'} = \mathrm{cov}[H_l(Y_k), H_l(Y_{k'})]$, $C_{l*} = (\zeta_l - 1)/\zeta_l$, and $C_l^* = (\zeta_l + 1)/\zeta_l$:

$$\text{if}\quad \frac{\sum_{k\in\mathbb{N}} r_k^{2l}}{r_0^{2l}} \le \frac{1}{\zeta_l^2}\quad \text{for some } \zeta_l > 1,\quad \text{then}\quad C_{l*}\mathbf{\Lambda}_N^l \le \mathbf{\Sigma}_N^{(l)} \le C_l^*\mathbf{\Lambda}_N^l\quad \text{for all } N\in\mathbb{N}.$$

Since $\sum_{k\in\mathbb{N}} r_k^{2l}/r_0^{2l} \le \sum_{k\in\mathbb{N}} r_k^{2p}/r_0^{2p}$ implies that $C_{p*} \le C_{l*}$ and $C_l^* \le C_p^*$ for $l \ge p$, and since

$$v^2_{\mathbf{\Sigma}_N}(\theta) = \boldsymbol{x}_N^\top \mathbf{\Sigma}_N^{(l)}\boldsymbol{x}_N \quad \text{with } \boldsymbol{x}_N = (N^{-1/2}, \dots, N^{-1/2}) \in \mathbb{R}^N,$$

for (26) condition (11) implies that

$$C_{p*}\sum_{l\ge p} c_l^2\,\mathrm{var}[H_l(Y_0)] \le v^2_{\mathbf{\Sigma}}(\theta) \le C_p^*\sum_{l\ge p} c_l^2\,\mathrm{var}[H_l(Y_0)].$$

Assertion (12) then follows from $\mathcal{I}_{\mathbf{\Lambda}}^{-1}(\theta) = \mathrm{var}[g(Y_0)] = \sum_{l\ge p} c_l^2\,\mathrm{var}[H_l(Y_0)]$.

(ii) In view of (9), it remains to evaluate $C_{\mathbf{\Lambda}/\mathbf{\Sigma}}$. We evaluate it separately in Proposition 3, below.

Let us symbolically denote $\mathcal{I}_{\mathbf{\Sigma}_N}(\theta) = \mathrm{var}[\dot\ell_{\mathbf{\Sigma}_N}(\theta)]$ and $\mathcal{I}_{\mathbf{\Lambda}_N}^{-1}(\theta) = \mathrm{var}[\dot\ell_{\mathbf{\Lambda}_N}(\theta)]$ (although $\mathcal{I}_{\mathbf{\Sigma}_N}(\theta) \to \mathcal{I}_{\mathbf{\Sigma}}(\theta)$ as $N\to\infty$, we have $\mathcal{I}_{\mathbf{\Sigma}_N}^{-1}(\theta) \ne \mathrm{var}[\sqrt{N}(\hat\theta^*_{\mathbf{\Sigma}_N} - \theta)]$ for finite N).

Proposition 3. *Under the same assumptions as in Proposition 2 and condition (2), the following inequalities hold, where C^* and C_* are as given in (3):*

$$(C^*)^{-2}\mathcal{I}_{\mathbf{\Lambda}_N}^{-1}(\theta) \le \mathcal{I}_{\mathbf{\Sigma}_N}^{-1}(\theta) \le C_*^{-2}\mathcal{I}_{\mathbf{\Lambda}_N}^{-1}(\theta) \quad \textit{for all } N\in\mathbb{N}.$$

Moreover, if $1/\eta(\xi)$ is integrable on $[-\pi, \pi]$ then

$$\lim_{N\to\infty}\frac{\mathcal{I}_{\mathbf{\Lambda}_N}^{-1}(\theta)}{\mathcal{I}_{\mathbf{\Sigma}_N}^{-1}(\theta)} = \biggl\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\eta(\xi)\,\mathrm{d}\xi\biggr\}\biggl\{\frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{1}{\eta(\xi)}\,\mathrm{d}\xi\biggr\} \ge 1. \tag{27}$$

4.5. Proof of Proposition 3

First we prove that $C_{\mathbf{\Lambda}/\mathbf{\Sigma}} \le C_*^{-2}$. By Lemma 5, θ is a smooth function of r_0. Hence, we can write

$$\frac{\mathrm{d}\ell_{\mathbf{\Sigma}_N}(\theta)}{\mathrm{d}\theta} = \frac{\mathrm{d}r_0}{\mathrm{d}\theta}\cdot\frac{\mathrm{d}\ell_{\mathbf{\Sigma}_N}(\theta)}{\mathrm{d}r_0} \quad \text{and} \quad \frac{\mathrm{d}\ell_{\mathbf{\Lambda}_N}(\theta)}{\mathrm{d}\theta} = \frac{\mathrm{d}r_0}{\mathrm{d}\theta}\cdot\frac{\mathrm{d}\ell_{\mathbf{\Lambda}_N}(\theta)}{\mathrm{d}r_0}.$$

Therefore, we will show that

$$\frac{\mathrm{d}\ell_{\mathbf{\Lambda}_N}(\theta)}{\mathrm{d}r_0}\left(\frac{\mathrm{d}\ell_{\mathbf{\Sigma}_N}(\theta)}{\mathrm{d}r_0}\right)^{-1} \le C_*^{-2},$$

which can be calculated in general no matter what $\mathrm{d}r_0/\mathrm{d}\theta$ is.

To this end, it suffices to show that

$$0 \le (C^*)^{-1}\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Lambda}_N| \le \frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Sigma}_N| \le C_*^{-1}\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Lambda}_N|, \tag{28}$$

$$(C^*)^{-2}\boldsymbol{y}_N^\top\frac{\mathrm{d}}{\mathrm{d}r_0}\mathbf{\Lambda}_N^{-1}\boldsymbol{y}_N \le \boldsymbol{y}_N^\top\frac{\mathrm{d}}{\mathrm{d}r_0}\mathbf{\Sigma}_N^{-1}\boldsymbol{y}_N \le C_*^{-2}\boldsymbol{y}_N^\top\frac{\mathrm{d}}{\mathrm{d}r_0}\mathbf{\Lambda}_N^{-1}\boldsymbol{y}_N \le 0, \qquad \boldsymbol{y}_N \in \mathbb{R}^N. \tag{29}$$

In fact, from (28) and (29) we have

$$C_*\left|\frac{\mathrm{d}}{\mathrm{d}r_0}\ell_{\mathbf{\Sigma}_N}\right| \le \left|\frac{\mathrm{d}}{\mathrm{d}r_0}\ell_{\mathbf{\Lambda}_N}\right| \quad \text{a.s.} \tag{30}$$

and, hence,

$$\mathcal{I}_{\mathbf{\Lambda}_N}^{-1}(\theta) \le C_*^{-2}(\theta)\mathcal{I}_{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}).$$

From (30), we separate cases to obtain $|(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Sigma}_N}| \le (C^*)^{-1}|(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Lambda}_N}|$ if $(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Sigma}_N} \ge 0$ and $|(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Sigma}_N}| \le C_*^{-1}|(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Lambda}_N}|$ if $(\mathrm{d}/\mathrm{d}r_0)\ell_{\mathbf{\Sigma}_N} \le 0$.

To show (28), we write

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Sigma}_N| &= \frac{(\mathrm{d}/\mathrm{d}r_0)|\mathbf{\Sigma}_N|}{|\mathbf{\Sigma}_N|} = \frac{(\mathrm{d}/\mathrm{d}r_0)|\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}\mathbf{\Lambda}_N|}{|\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}\mathbf{\Lambda}_N|} \\ &= \frac{|\mathbf{\Lambda}_N|(\mathrm{d}/\mathrm{d}r_0)|\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}| + |\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}|(\mathrm{d}/\mathrm{d}r_0)|\mathbf{\Lambda}_N|}{|\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}||\mathbf{\Lambda}_N|}.\end{aligned}$$

Since $|\mathbf{\Lambda}_N| = (r_0/N)(\mathrm{d}/\mathrm{d}r_0)|\mathbf{\Lambda}_N|$, we have

$$\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Sigma}_N| = \left[1 + \frac{r_0}{N}\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Sigma}_N\mathbf{\Lambda}_N^{-1}|\right]\frac{\mathrm{d}}{\mathrm{d}r_0}\log|\mathbf{\Lambda}_N|.$$

Define the matrices $\boldsymbol{Q}_N$, $\boldsymbol{R}_N$, and $\boldsymbol{U}_N$ by

$$(\boldsymbol{Q}_N)_{i,j} = (\boldsymbol{R}_N - r_0\boldsymbol{I}_N)_{i,j} = \begin{cases} 0 & \text{if } i = j, \\ r_{|i-j|} & \text{if } i \ne j, \end{cases}$$

and $\boldsymbol{U}_N = \operatorname{diag}(u_{N,1}, \dots, u_{N,N})$, where $\{u_{N,i}, 1 \le i \le N\}$ are the eigenvalues of $\boldsymbol{Q}_N$, meaning that $\boldsymbol{Q}_N = \boldsymbol{T}_N^\top \boldsymbol{U}_N \boldsymbol{T}_N$ for some orthogonal matrix $\boldsymbol{T}_N$. Then $|\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1}|$ is given by

$$|\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1}| = |(\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1})^{-1}|^{-1} = \left|\left(\boldsymbol{I}_N + \frac{\boldsymbol{Q}_N}{r_0}\right)^{-1}\right|^{-1} = \left|\boldsymbol{T}_N^\top \sum_{i \in \mathbb{N}_0} \left(-\frac{\boldsymbol{U}_N}{r_0}\right)^i \boldsymbol{T}_N\right|^{-1}$$

$$= \left|\operatorname{diag}\left(\frac{r_0}{r_0 + u_{N,l}}\right)_{1 \le l \le N}\right|^{-1} = \prod_{l=1}^{N} \frac{r_0 + u_{N,l}}{r_0}.$$

Hence,

$$\frac{\mathrm{d}}{\mathrm{d}r_0} \log |\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1}| = -\sum_{l=1}^{N} \frac{u_{N,l}}{r_0(r_0 + u_{N,l})}$$

and we have

$$\frac{\mathrm{d}}{\mathrm{d}r_0} \log |\boldsymbol{\Sigma}_N| = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{r_0}{r_0 + u_{N,i}}\right] \frac{\mathrm{d}}{\mathrm{d}r_0} \log |\boldsymbol{\Lambda}_N|. \tag{31}$$

Let $\{\rho_{N,i}, 1 \le i \le N\}$ be the eigenvalues of $\boldsymbol{R}_N$. From the eigenequation

$$0 = |u_{N,i} \boldsymbol{I}_N - \boldsymbol{Q}_N| = |(r_0 + u_{N,i}) \boldsymbol{I}_N - \boldsymbol{R}_N|,$$

it follows that $r_0 + u_{N,i} = \rho_{N,i} \ge 0$.

For the right-hand side of (28), i.e. $|(\mathrm{d}/\mathrm{d}r_0)\ell_{\boldsymbol{\Sigma}_N}| \le C_*^{-1} |(\mathrm{d}/\mathrm{d}r_0)\ell_{\boldsymbol{\Lambda}_N}|$, by setting $\rho_{*N} = \min_{1 \le i \le N} \rho_{N,i}$ we have

$$\frac{\mathrm{d}}{\mathrm{d}r_0} \log |\boldsymbol{\Sigma}_N| \le \frac{r_0}{\rho_{*N}} \frac{\mathrm{d}}{\mathrm{d}r_0} \log |\boldsymbol{\Lambda}_N|.$$

Here, the lower bound in (3),

$$\boldsymbol{\Sigma}_N - C_* \boldsymbol{\Lambda}_N = \boldsymbol{\Sigma}_N - C_* r_0 \boldsymbol{I}_N \ge 0,$$

implies that $C_* r_0 \le \rho_{*N}$, i.e. $r_0/\rho_{*N} \le C_*^{-1}$. The converse inequality in (28) is obtained similarly (using $(C^*)^{-1} \le r_0/\rho_N^*$ in (31), with $\rho_N^* = \max_{1 \le i \le N} \rho_{N,i}$). The positiveness in (28) follows from $(\mathrm{d}/\mathrm{d}r_0) \log |\boldsymbol{\Lambda}_N| = N/r_0 \ge 0$. Hence, the proof of (28) is complete.

We next show (29). Since $\boldsymbol{\Sigma}_N^{-1} = (\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1} \boldsymbol{\Lambda}_N)^{-1}$ and $\boldsymbol{\Lambda}_N^{-1} = -r_0 (\mathrm{d}/\mathrm{d}r_0) \boldsymbol{\Lambda}_N^{-1}$, by using $(\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1})^{-1} = \boldsymbol{T}_N^\top \operatorname{diag}(r_0/(r_0 + u_{N,i}))_{1 \le i \le N} \boldsymbol{T}_N$ we have

$$\frac{\mathrm{d}}{\mathrm{d}r_0} \boldsymbol{\Sigma}_N^{-1} = \left(\frac{\mathrm{d}}{\mathrm{d}r_0} \boldsymbol{\Lambda}_N^{-1}\right) \left[(\boldsymbol{\Sigma}_n \boldsymbol{\Lambda}_N^{-1})^{-1} - r_0 \frac{\mathrm{d}}{\mathrm{d}r_0} (\boldsymbol{\Sigma}_N \boldsymbol{\Lambda}_N^{-1})^{-1}\right]$$

$$= \left(\frac{\mathrm{d}}{\mathrm{d}r_0} \boldsymbol{\Lambda}_N^{-1}\right) \boldsymbol{T}_N^\top \operatorname{diag}\left(\left(\frac{r_0}{r_0 + u_{N,i}}\right)^2\right)_{1 \le i \le n} \boldsymbol{T}_N.$$

By the same arguments as in the proof of (28), we recover (29).

Finally, for (27), by using the theorem of [16, p. 64] we obtain

$$\lim_{n \to \infty} \frac{1}{N} \sum_{l=1}^{N} \frac{r_0}{\rho_{N,l}} = \frac{r_0}{2\pi} \int_{-\pi}^{\pi} \frac{1}{\eta(\xi)} \,\mathrm{d}\xi. \tag{32}$$

Since $r_0 = (2\pi)^{-1} \int_{-\pi}^{\pi} \eta(\xi) \,\mathrm{d}\xi$, by Jensen's inequality the right-hand side of (32) is larger than or equal to 1. This proves that $C_{\boldsymbol{\Lambda}/\boldsymbol{\Sigma}} \ge 1$.

Lemma 5. *If* $\operatorname{var}[g(s_j(0))] < \infty$ *then* θ_j *is a smooth function of* σ_j^2.

4.6. Proof of Corollary 1

By considering Lemma 2 in light of Theorem 3, in particular with $2^J n = l$, $k = 0$, and $k' = 0$, we obtain $r_j(l) = Cl^{-2(\gamma-H)}\ell(l)(1 + O(l^{-1}))$, as $l \to \infty$, in (36). Therefore, for an arbitrarily small $\varepsilon > 0$, $2p(\gamma - H - \varepsilon) > 1$ holds by the condition of Corollary 1. Thus, we have part (i) of Proposition 2. On the other hand, as in the case of Theorem 2, $\gamma \geq 2$ implies (2), which leads to part (ii) of Proposition 2.

4.7. Proof of Theorem 3

For the wavelet coefficient S^T, let us consider a sequence of random vectors

$$\boldsymbol{s} = \{\boldsymbol{s}(n) \in \mathbb{R}^d,\ n \in \mathbb{N}_0\}, \qquad d = \sum_{j=J_0+1}^{J} d_j, \qquad d_j = 2^{J-j},$$

defined by

$$\boldsymbol{s}(n) = (\boldsymbol{s}_{J_0+1}(n), \ldots, \boldsymbol{s}_J(n)),$$

where each $\boldsymbol{s}_j$, $j = J_0 + 1, \ldots, J$, denotes the following subvector in $\mathbb{R}^{d_j}$:

$$\boldsymbol{s}_j(n) = (s_j(d_j n + 1), \ldots, s_j(d_j n + k), \ldots, s_j(d_j(n + 1))).$$

It is easily seen that $\{\boldsymbol{s}_n\}$ is a stationary random vector in the usual sense, i.e. the covariance matrix $\mathrm{E}[\boldsymbol{s}_j^\top(m+n)\boldsymbol{s}_{j'}(m)]$ does not depend on $m \in \mathbb{N}_0$ for each pair (j, j').

Now let us take the sequence $\{\boldsymbol{Y}(n),\ n = 1, \ldots, N_{T,J}\}$ of vectors

$$\boldsymbol{Y}(n) = (Y_{J_0+1}(n), \ldots, Y_J(n)),$$

where

$$Y_j(n) = \frac{1}{d_j}\sum_{k=1}^{d_j} g(s_j(d_j n + k)).$$

Then $\{\boldsymbol{Y}(n)\}$ is vector stationary, i.e. $\mathrm{cov}[Y_j(m), Y_{j'}(m+n)]$ does not depend on $m \in \mathbb{N}_0$ for each pair (j, j'). Let $\hat{\boldsymbol{\theta}}_T^{\boldsymbol{Y}} = (\hat{\theta}_{T,J_0+1}^{\boldsymbol{Y}}, \ldots, \hat{\theta}_{T,J}^{\boldsymbol{Y}})$ be defined by

$$\hat{\theta}_{T,j}^{\boldsymbol{Y}} = \frac{1}{N_{T,J}}\sum_{n=1}^{N_{T,J}} Y_j(n), \qquad j = J_0 + 1, \ldots, J.$$

As is easily seen, $\hat{\boldsymbol{\theta}}_T^{\boldsymbol{Y}} \to \boldsymbol{\theta}$ a.s. and $\mathrm{E}[\hat{\boldsymbol{\theta}}_T^{\boldsymbol{Y}}] = \boldsymbol{\theta}$. Also,

$$\Pr(\{|\sqrt{N_{T,J}}(\hat{\boldsymbol{\theta}}_T^{\boldsymbol{Y}} - \boldsymbol{\theta}) - \sqrt{N_{T,J}}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})| > \varepsilon\}) \to 0 \quad \text{as } T \to \infty,$$

for all $\varepsilon > 0$. Thus, by [7, Theorem 3.1, p. 27], to show that the desired CLT (14) holds it is equivalent to show that the alternative CLT

$$\sqrt{N_{T,J}}(\hat{\boldsymbol{\theta}}_T^{\boldsymbol{Y}} - \boldsymbol{\theta}) \xrightarrow{\mathrm{w}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_J) \tag{33}$$

holds.

By [3, Theorem 4], to prove (33) it is sufficient to show that

$$\sum_{n\in\mathbb{N}} |\mathrm{cov}[Y_j(n)Y_{j'}(0)]|^p < \infty \quad \text{and} \quad \sum_{n\in\mathbb{N}} |\mathrm{cov}[Y_j(0)Y_{j'}(n)]|^p < \infty.$$

For the first condition, it then suffices to show that

$$\sum_{n\in\mathbb{N}} |\,\mathrm{E}[s_j(d_j n + k)s_{j'}(d_{j'}\cdot 0 + k')]|^p < \infty$$

for each k and k' and, for the second condition, that the same inequality holds with j and j' exchanged. From Lemma 2 these convergence statements hold if $2(\gamma - H)p > 1$, i.e. $\gamma > H + 1/2p$.

Equation (15) is obtained from

$$\begin{aligned}\operatorname{cov}[\sqrt{N_{T,J}}\hat{\theta}_{T,j}, \sqrt{N_{T,J}}\hat{\theta}_{T,j'}]&\\ = \frac{r_{j,j'}(0)}{d_j} + \frac{1}{d_j N_{T,j'}} \sum_{k=1}^{N_{T,j'}-1} &\{(N_{T,j'} - \lfloor 2^{-(j'-j)}k\rfloor) r_{j,j'}(k)\\ &+ (N_{T,j'} - \lceil 2^{-(j'-j)}k\rceil) r_{j,j'}(-k)\}\end{aligned}$$

and Lemma 2. In fact, the terms $N_{T,j}^{-1}\sum_{k=1}^{N_{T,j}-1} k r_{j,j'}(\pm k)$ tend to 0 as $T \to \infty$ since they are $O((N_{T,j})^{-2(\gamma-H)p+1})$, whereas $\gamma \geq 2$ implies that $-2(\gamma - H)p + 1 < -1$. Equation (16) is obtained similarly. Finally, (17) follows from Slutsky's theorem [13].

5. Proofs of lemmas

5.1. Proof of Lemma 1

Let the wavelet expansions of $X_t\,\mathbf{1}_{[0,T]}(t)$ and $Y_t\,\mathbf{1}_{[0,T]}(t)$ be

$$\sum_{j\in\mathbb{Z}}\sum_k c_{j,k}^{(X^T)}\psi_{j,k}(t) \quad \text{and} \quad \sum_{j\in\mathbb{Z}}\sum_k c_{j,k}^{(Y^T)}\psi_{j,k}(t),$$

respectively. Since $\operatorname{supp}(\psi_{j,k}) = [2^j k, 2^j(k+w)]$, $c_{j,k}^{(X^T)}$ and $c_{j,k}^{(Y^T)}$ all vanish for $2^j k \geq T$, i.e. $k \geq \lceil 2^{-j}T\rceil$, or for $2^j(k+w) \leq 0$, i.e. $k \leq -w$. From Parseval's equality for L^2-functions and their wavelet coefficients, it follows that

$$\begin{aligned}\lim_{T\to\infty}\frac{1}{T}\int_0^T X_t Y_t\,\mathrm{d}t &= \lim_{T\to\infty}\frac{1}{T}\sum_{j\in\mathbb{Z}}\sum_{k=-\lfloor w\rfloor}^{\lceil 2^{-j}T\rceil} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)}\\ &= \lim_{T\to\infty}\frac{1}{T}\sum_{j\in\mathbb{Z}}\sum_{k=1}^{\lceil 2^{-j}T\rceil} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} + \lim_{T\to\infty}\frac{1}{T}\sum_{j\in\mathbb{Z}}\sum_{k=-\lfloor w\rfloor}^{0} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)}.\end{aligned}$$

Since $2^{-j}T/N_{T,j} \to 1$ as $T \to \infty$, the first term on the right-hand side is the same as

$$\begin{aligned}\sum_{j\in\mathbb{Z}}\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{\lceil 2^{-j}T\rceil} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} &= \sum_{j\in\mathbb{Z}}\lim_{T\to\infty}\frac{2^{-j}}{N_{T,j}}\sum_{k=1}^{N_{T,j}} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)}\\ &= \lim_{T\to\infty}\sum_{j\in\mathbb{Z}}\frac{2^{-j}}{N_{T,j}}\sum_{k=1}^{N_{T,j}} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)}.\end{aligned}$$

Here, the change of

$$\lim_{T\to\infty}\sum_{j\in\mathbb{Z}}\frac{1}{T}\sum_{k=1}^{\lceil 2^{-j}T\rceil} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} \quad \text{into} \quad \sum_{j\in\mathbb{Z}}\lim_{T\to\infty}\frac{1}{T}\sum_{k=1}^{\lceil 2^{-j}T\rceil} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)}$$

is allowed by dominated convergence.

Hence, it suffices to show that the second term tends to 0. Since the second term tends to a finite limit by Schwarz's inequality, for all $\varepsilon > 0$ there exists a j_0 such that

$$\frac{1}{T}\sum_{|j|>j_0}\sum_{k=-\lfloor w\rfloor}^{0} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} \le \varepsilon.$$

Thus,

$$\begin{aligned}\frac{1}{T}\sum_{j\in\mathbb{Z}}\sum_{k=-\lfloor w\rfloor}^{0} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} &\le \frac{1}{T}\sum_{|j|\le j_0}\sum_{k=-\lfloor w\rfloor}^{0} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} + \varepsilon \\ &= \frac{1}{T}\sum_{|j|\le j_0}\sum_{k=-\lfloor w\rfloor}^{0} c_{j,k}^{(X^T)}c_{j,k}^{(Y^T)} + \varepsilon.\end{aligned}$$

The first term on the right-hand side of this tends to 0 as $T\to\infty$, which completes the proof.

5.2. Proof of Lemma 2

We have

$$\begin{aligned}&\mathrm{E}[s_j(d_jn+k)s_{j'}(k')] \\ &\quad = 2^{(j+j')/2}\bigg(-\frac{1}{2}\bigg)\iint_{W^2}\psi(s)\psi(t)\,\mathrm{E}[\{X(2^j(s+k+d_jn)) - X(2^{j'}(t+k'))\}^2]\,\mathrm{d}s\,\mathrm{d}t.\end{aligned} \tag{34}$$

From the general form of regularly varying functions, we can write

$$\begin{aligned}&\mathrm{E}[\{X(2^j(s+k+d_jn)) - X(2^{j'}(t+k'))\}^2] \\ &\quad = (2^Jn)^{2H}\ell(2^Jn)\bigg|1+\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg|^{2H} \\ &\qquad \times \ell\bigg(2^Jn\bigg|1+\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg|\bigg)\frac{1}{\ell(2^Jn)}.\end{aligned} \tag{35}$$

Taylor expansions of the last two factors on the right-hand side of (35) yield

$$\bigg|1+\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg|^{2H} = \sum_{l\in\mathbb{N}_0}\binom{2H}{l}\bigg(\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg)^l$$

and

$$\begin{aligned}&\ell\bigg(2^Jn\bigg|1+\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg|\bigg)\frac{1}{\ell(2^Jn)} \\ &\quad = 1+\sum_{i\in\mathbb{N}}\frac{1}{i!}\frac{(2^Jn)^i\ell^{(i)}(2^Jn)}{\ell(2^Jn)}\bigg(\frac{2^j(s+k)-2^{j'}(t+k')}{2^Jn}\bigg)^i.\end{aligned}$$

https://doi.org/10.1239/aap/1134587748 Published online by Cambridge University Press

We observe that a slowly varying function is 0-regularly varying and that the smooth variation property of ℓ implies that [8, Equation 1.8.1, p. 44], for $i \geq 1$,

$$\frac{(2^J n)^i \ell^{(i)}(2^J n)}{\ell(2^J n)} \to \prod_{\nu=0}^{i-1}(0-\nu) = 0 \quad \text{as } n \to \infty.$$

Since those terms with $0 \leq l \leq 2\gamma - 1$ in the product of the two terms in the integral (34) vanish by assumption ($\psi 2$), we have

$$\mathrm{E}[s_j(d_j n + k)s_{j'}(k')] = C(2^J n)^{-2(\gamma - H)}\ell(2^J n)(1 + O(n^{-1})) \quad \text{as } n \to \infty, \tag{36}$$

where

$$C = 2^{3(j+j')/2}\frac{(-1)^{\gamma+1}}{2}\bigg(\int_W \psi(t)t^\gamma\,\mathrm{d}t\bigg)^2 \Bigg\{\sum_{i=0}^{2\gamma}\binom{2H}{2\gamma - i}\frac{1}{i!}\frac{(2^J n)^i \ell^{(i)}(2^J n)}{\ell(2^J n)}\Bigg\}.$$

5.3. Proof of Lemma 3

It is enough to prove that the kernel $\Psi(s,t) = |s-t+k|^{2H}$ is a negative-definite-type kernel for $k \in \mathbb{N}$. The kernel $\Psi_0(s,t) = (s-t+k)^2$ is of negative-definite type [6, Section 3.1.22] (but is not a true negative-definite kernel). In general, for a negative-definite kernel $\tilde{\Psi}$, its power $\tilde{\Psi}^H$ is also negative definite for $0 < H < 1$ [6, Section 3.2.10]. We show that this claim is also true for a negative-definite-type kernel.

To this end, we must verify that [6, Theorem 3.2.9, p. 77] holds not only for negative-definite kernels, but also for negative-definite-type kernels. Following the proof of [6, Theorem 3.2.9], it turns out that we are done if $\mathrm{e}^{-\lambda\tilde{\Psi}(s,t)}$ is positive definite. For this, we check the 'only if' part of [6, Theorem 3.2.2, p. 74] and we have to prove, in particular, that $\tilde{\Psi}$ *is of negative-definite type only if* $\mathrm{e}^{-\lambda\tilde{\Psi}}$ *is of positive-definite type*.

Following the proof of [6, Theorem 3.2.2], we take the parameters Ψ and x_0 there to be $\Psi(s,t) = (s-t+k)^2$ and $x_0 = k$. Then $\Psi(s,x_0) = s^2$, $\Psi(t,x_0) = t^2$, $\Psi(x_0,x_0) = k^2$, and

$$\varphi(s,t) = \Psi(s,x_0) + \Psi(t,x_0) - \Psi(s,t) - \Psi(x_0,x_0) = 2st - 2k(s-t) - 2k^2,$$

which is of positive-definite type in the following sense: for all $\{c_i,\ i = 1,\dots,n\} \subset \mathbb{C}$ with $\sum_{i=1}^n c_i = 0$,

$$\sum_{i=1}^n\sum_{j=1}^n c_i\bar{c}_j\varphi(s_i,s_j) = 2\bigg|\sum_{i=1}^n c_i s_i\bigg|^2 \geq 0.$$

Therefore, $\mathrm{e}^{-\Psi(s,x_0)}\mathrm{e}^{-\Psi(t,x_0)}$ is positive definite (see [6, Section 3.1.9]), $\mathrm{e}^{\varphi(s,t)}$ is of positive-definite type (see [6, Section 3.1.14]), and

$$\mathrm{e}^{-\Psi(s,t)} = \mathrm{e}^{\varphi(s,t)}\mathrm{e}^{-\Psi(s,x_0)}\mathrm{e}^{-\Psi(t,x_0)}\mathrm{e}^{-\Psi(x_0,x_0)}$$

is of positive-definite type (see [6, Section 3.1.12]). It is clear that, for claims 3.1.9, 3.1.14, and 3.1.12 of [6], the positive definiteness property can be replaced by the positive definiteness-type property.

5.4. Proof of Lemma 4

By the argument for the slowly varying part ℓ in the proof of Theorem 2, we have

$$|ax|^{2H}\ell((ax)^2) = \int_0^\infty (1 - \mathrm{e}^{-\lambda(ax)^2})h(\lambda)\,\mathrm{d}\lambda = \int_0^\infty (1 - \mathrm{e}^{-\lambda})h\Big(\frac{\lambda}{(ax)^2}\Big)\frac{\mathrm{d}\lambda}{(ax)^2}.$$

For an arbitrary $\varepsilon > 0$, we can take a large $z \equiv z_\varepsilon > 0$ such that

$$\int_0^\infty (1 - \mathrm{e}^{-\lambda})h\Big(\frac{\lambda}{(ax)^2}\Big)\frac{\mathrm{d}\lambda}{(ax)^2} = \int_0^z (1 - \mathrm{e}^{-\lambda a^2})h\Big(\frac{\lambda}{x^2}\Big)\frac{\mathrm{d}\lambda}{x^2} + \varepsilon.$$

Thus, if we let x become larger, we can write

$$\frac{|ax|^{2H}\ell((ax)^2)}{|x|^{2H}\ell(x^2)} = \frac{\int_0^z (1-\mathrm{e}^{-\lambda})(\lambda/(ax)^2)^{-\beta_0}\,\mathrm{d}\lambda/a^2 + \varepsilon'}{\int_0^z (1-\mathrm{e}^{-\lambda})(\lambda/x^2)^{-\beta_0}\,\mathrm{d}\lambda + \varepsilon''},$$

which approaches a^{2H} as $x \to \infty$. Letting $\varepsilon > 0$ become smaller and, thus, z and then x become larger, we conclude that $\beta_0 = 1 + H$.

5.5. Proof of Lemma 5

The assumption of this lemma implies that g can be expanded in Hermite polynomials as follows (the series converges in $L^2(\mathbb{R}, \mathrm{e}^{-x^2/2}\mathrm{d}x/\sqrt{2\pi})$):

$$g(x) = \sum_{l\in\mathbb{N}} c_l H_l(x).$$

Here, without loss of generality, we assume that $c_0 = 0$. Then, since $s_j(k) \sim \mathcal{N}(0, \sigma_j^2)$, by writing $H_{2l}(x) = \sum_{m=0}^{l} h_{l,m}x^{2m}$ we have

$$\mathrm{E}[H_{2l}(s_j(0))] = \sum_{m=0}^{l} (2m-1)!!\, h_{l,m}\sigma_j^{2m}$$

and, hence,

$$\theta_j = \sum_{l\in\mathbb{N}} c_{2l}\,\mathrm{E}[H_{2l}(s_j(0))] = \sum_{m\in\mathbb{N}} a_m\sigma_j^{2m}$$

with $a_m = (2m-1)!!\sum_{l=m}^{\infty} c_{2l}h_{l,m}$. Therefore, θ_j is a smooth function of σ_j^2.

Acknowledgement

The authors are sincerely grateful to Professor Minoru W. Yoshida of the University of Electro-Communications, Tokyo, for helpful discussions.

References

[1] Abry, P. and Veitch, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inf. Theory* **44,** 2–15.

[2] Abry, P., Gonçalvès, P. and Flandrin, P. (1995). Wavelets, spectrum analysis and $1/f$ processes. In *Wavelets and Statistics* (Lecture Notes Statist. **103**), Springer, New York, pp. 15–30.

[3] Arcones, M. A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Prob.* **22,** 2242–2274.

[4] Arcones, M. A. (2000). Distributional limit theorems over a stationary Gaussian sequence of random vectors. *Stoch. Process. Appl.* **88,** 135–159.
[5] Bardet, J.-M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long-range dependent processes. *Statist. Inference Stoch. Process.* **3,** 85–99.
[6] Berg, C., Christensen, J. P. R. and Ressel, P. (1984). *Harmonic Analysis on Semigroups. Theory of Positive Definite and Related Functions* (Graduate Texts Math. **100**). Springer, New York.
[7] Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.
[8] Bingham, N. H., Goldie, C. M. and Teugels, J. T. (1987). *Regular Variation*. Cambridge University Press.
[9] Breuer, P. and Major, P. (1983). Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.* **13,** 425–441.
[10] Dijkerman, R. W. and Mazumdar, R. R. (1994). On the correlation structure of the wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inf. Theory* **40,** 1609–1612.
[11] Dobrushin, R. L. and Major, P. (1979). Non-central limit theorems for nonlinear functionals of Gaussian fields. *Z. Wahrscheinlichkeitsth.* **50,** 27–52.
[12] Embrechts, P. and Maejima, M. (2002). *Selfsimilar Processes*. Princeton University Press.
[13] Ferguson, T. S. (2002). *A Course in Large Sample Theory*. Chapman and Hall, London.
[14] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inf. Theory* **38,** 910–917.
[15] Fuller, W. A. (1996). *Introduction to Statistical Time Series*, 2nd edn. John Wiley, New York.
[16] Grenander, U. and Szegö, G. (1984). *Toeplitz Forms and Their Applications*, 2nd edn. Chelsea, New York.
[17] Hernández, E. and Weiss, G. (1996). *A First Course on Wavelets*. CRC Press, Boca Raton, FL.
[18] Ho, H. C. and Sun, T. C. (1987). A central limit theorem for noninstantaneous filters of a stationary Gaussian process. *J. Multivariate Anal.* **22,** 144–155.
[19] Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York.
[20] Major, P. (1981). *Multiple Wiener–Itô Integrals* (Lecture Notes Math. **849**). Springer, Berlin.
[21] Priestley, M. B. (1981). *Spectral Analysis and Time Series*, Vol. 1. Academic Press, London.
[22] Stein, E. M. (1993). *Harmonic Analysis, Real-Variable Methods, Orthogonality and Oscillatory Integrals*. Princeton University Press.
[23] Taqqu, M. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrscheinlichkeitsth.* **50,** 53–83.
[24] Tewfik, A. H. and Kim, M. (1992). Correlation structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inf. Theory* **38,** 904–909.