The Annals of Statistics
2008, Vol. 36, No. 4, 1925–1956
DOI: 10.1214/07-AOS527

A WAVELET WHITTLE ESTIMATOR OF THE MEMORY PARAMETER OF A NONSTATIONARY GAUSSIAN TIME SERIES[1]

BY E. MOULINES, F. ROUEFF AND M. S. TAQQU

Télécom Paris/CNRS LTCI and Boston University

We consider a time series $X = \{X_k, k \in \mathbb{Z}\}$ with memory parameter $d_0 \in \mathbb{R}$. This time series is either stationary or can be made stationary after differencing a finite number of times. We study the "local Whittle wavelet estimator" of the memory parameter d_0. This is a wavelet-based semiparametric pseudo-likelihood maximum method estimator. The estimator may depend on a given finite range of scales or on a range which becomes infinite with the sample size. We show that the estimator is consistent and rate optimal if X is a linear process, and is asymptotically normal if X is Gaussian.

1. Introduction.

Let $X \stackrel{\text{def}}{=} \{X_k\}_{k\in\mathbb{Z}}$ be a process, not necessarily stationary or invertible. Denote by ΔX, the first order difference, $(\Delta X)_\ell = X_\ell - X_{\ell-1}$, and by $\Delta^k X$, the kth order difference. Following [9], the process X is said to have memory parameter d_0, $d_0 \in \mathbb{R}$, if for any integer $k > d_0 - 1/2$, $U \stackrel{\text{def}}{=} \Delta^k X$ is covariance stationary with spectral measure

$$\nu_U(d\lambda) = |1 - e^{-i\lambda}|^{2(k-d_0)} \nu^*(d\lambda), \qquad \lambda \in [-\pi, \pi], \tag{1}$$

where ν^* is a nonnegative symmetric measure on $[-\pi, \pi]$ such that, in a neighborhood of the origin, it admits a positive and bounded density. The process X is covariance stationary if and only if $d_0 < 1/2$. When $d_0 > 0$, X is said to exhibit long memory or long-range dependence. The *generalized spectral measure* of X is defined as

$$\nu(d\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{-2d_0} \nu^*(d\lambda), \qquad \lambda \in [-\pi, \pi]. \tag{2}$$

We suppose that we observe $X_1, \ldots, X_n$ and want to estimate the exponent d_0 under the following semiparametric set-up introduced in [15]. Let $\beta \in (0, 2]$, $\gamma > 0$ and $\varepsilon \in (0, \pi]$, and assume that

$$\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon),$$

where $\mathcal{H}(\beta, \gamma, \varepsilon)$ is the class of finite nonnegative symmetric measures on $[-\pi, \pi]$ whose restrictions on $[-\varepsilon, \varepsilon]$ admit a density g, such that, for all $\lambda \in (-\varepsilon, \varepsilon)$,

$$|g(\lambda) - g(0)| \le \gamma g(0) |\lambda|^\beta. \tag{3}$$

Received July 2007; revised July 2007.
[1]Supported in part by NSF Grant DMS-05-05747.
AMS 2000 subject classifications. Primary 62M15, 62M10, 62G05; secondary 62G20, 60G18.
Key words and phrases. Long memory, semiparametric estimation, wavelet analysis.

Since $\varepsilon \leq \pi$, $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$ is only a local condition for λ near 0. For instance, ν^* may contain atoms at frequencies in $(\varepsilon, \pi]$ or have an unbounded density on this domain.

We shall estimate d_0 using the semiparametric *local Whittle wavelet estimator* defined in Section 3. We will show that under suitable conditions, this estimator is consistent (Theorem 3), the convergence rate is optimal (Corollary 4) and it is asymptotically normal (Theorem 5). In Section 4, we discuss how it compares to other estimators.

There are two popular semiparametric estimators for the memory parameter d_0 in the frequency domain:

(1) the Geweke–Porter–Hudak (GPH) estimator introduced in [6] and analyzed in [16], which involves a regression of the log-periodogram on the log of low frequencies;
(2) the local Whittle (Fourier) estimator (or LWF) proposed in [11] and developed in [15], which is based on the Whittle approximation of the Gaussian likelihood, restricted to low frequencies.

Corresponding approaches may be considered in the wavelet domain. By far, the most widely used wavelet estimator is based on the log-regression of the wavelet coefficient variance on the scale index, which was introduced in [1]; see also [14] and [13] for recent developments. A wavelet analog of the LWF, referred to as the *local Whittle wavelet estimator* can also be defined. This estimator was proposed for analyzing noisy data in a parametric context in [23] and was considered by several authors, essentially in a parametric context (see, e.g., [10] and [12]). To our knowledge, its theoretical properties are not known (see the concluding remarks in [22], page 107). The main goal of this paper is to fill this gap in a semiparametric context. The paper is structured as follows. In Section 2, the wavelet analysis of a time series is presented and some results on the dependence structure of the wavelet coefficients are given. The definition and the asymptotic properties of the local Whittle wavelet estimator are given in Section 3: the estimator is shown to be rate optimal under a general condition on the wavelet coefficients, which are satisfied when X is a linear process with four finite moments, and it is shown to be asymptotically normal under the additional condition that X is Gaussian. These results are discussed in Section 4. The proofs can be found in the remaining sections. The linear case is considered in Section 5. The asymptotic behavior of the wavelet Whittle likelihood is studied in Section 6 and weak consistency is studied in Section 7. The proofs of the main results are gathered in Section 8.

2. The wavelet analysis.

The functions $\phi(t)$, $t \in \mathbb{R}$, and $\psi(t)$, $t \in \mathbb{R}$, will denote the father and mother wavelets respectively, and $\hat{\phi}(\xi) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \phi(t) e^{-i\xi t}\,dt$ and $\hat{\psi}(\xi) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \psi(t) e^{-i\xi t}\,dt$ their Fourier transforms. We suppose that ϕ and ψ satisfy the following assumptions:

(W-1) ϕ and ψ are integrable and have compact supports, $\hat{\phi}(0) = \int_{\mathbb{R}} \phi(x)\,dx = 1$ and $\int_{\mathbb{R}} \psi^2(x)\,dx = 1$;

(W-2) there exists $\alpha > 1$ such that $\sup_{\xi \in \mathbb{R}} |\hat{\psi}(\xi)|(1 + |\xi|)^\alpha < \infty$;

(W-3) the function ψ has M vanishing moments, that is, $\int_{\mathbb{R}} t^l \psi(t)\,dt = 0$ for all $l = 0, \dots, M-1$;

(W-4) the function $\sum_{k\in\mathbb{Z}} k^l \phi(\cdot - k)$ is a polynomial of degree l for all $l = 0, \dots, M-1$;

(W-5) d_0, M, α and β are such that $(1+\beta)/2 - \alpha < d_0 \le M$.

Assumption (W-1) implies that $\hat{\phi}$ and $\hat{\psi}$ are everywhere infinitely differentiable. Assumption (W-2) is regarded as a *regularity condition* and assumptions (W-3) and (W-4) are often referred to as *admissibility conditions*. When (W-1) holds, assumptions (W-3) and (W-4) can be expressed in different ways. (W-3) is equivalent to asserting that the first $M-1$ derivative of $\hat{\psi}$ vanish at the origin and hence

$$|\hat{\psi}(\lambda)| = O(|\lambda|^M) \qquad \text{as } \lambda \to 0. \tag{4}$$

And, by [3], Theorem 2.8.1, page 90, (W-4) is equivalent to

$$\sup_{k\neq 0} |\hat{\phi}(\lambda + 2k\pi)| = O(|\lambda|^M) \qquad \text{as } \lambda \to 0. \tag{5}$$

Finally, (W-5) is the constraint on M and α that we will impose on the wavelet-based estimator of the memory parameter d_0 of a process having generalized spectral measure (2) with $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$ for some positive β, γ and ε. Remarks 1 and 7 below provide some insights into (W-5). We may consider nonstationary processes X because the wavelet analysis performs an implicit differentiation of order M. It is perhaps less well known that, in addition, wavelets can be used with noninvertible processes ($d_0 \le -1/2$) due to the regularity condition (W-2). These two properties of the wavelet are, to some extent, similar to the properties of the tapers used in Fourier analysis (see, e.g., [9, 22]).

Adopting the engineering convention that large values of the scale index j correspond to coarse scales (low frequencies), we define the family $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ of translated and dilated functions, $\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k)$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}$. If ϕ and ψ are the scaling and wavelet functions associated with a multiresolution analysis (see [3]), then $\{\psi_{j,k}, j \in \mathbb{Z}, k \in \mathbb{Z}\}$ forms an orthogonal basis in $L^2(\mathbb{R})$. A standard choice are the Daubechies wavelets (DB-M), which are parameterized by the number of their vanishing moments M. The associated scaling and wavelet functions ϕ and ψ satisfy (W-1)–(W-4), where α in (W-2) is a function of M which increases to infinity as M tends to infinity (see [3], Theorem 2.10.1). In this work, however, we neither assume that the pair $\{\phi, \psi\}$ is associated with a multiresolution analysis (MRA), nor that the $\psi_{j,k}$'s form a Riesz basis. Other possible choices are discussed in [14], Section 3.

The wavelet coefficients of the process $X = \{X_\ell, \ell \in \mathbb{Z}\}$ are defined by

$$W_{j,k} \stackrel{\text{def}}{=} \int_{\mathbb{R}} X(t)\psi_{j,k}(t)\,dt, \qquad j \ge 0, k \in \mathbb{Z}, \tag{6}$$

where $X(t) \stackrel{\text{def}}{=} \sum_{k\in\mathbb{Z}} X_k\phi(t-k)$. If (ϕ,ψ) define an MRA, then X_k is identified with the kth approximation coefficient at scale $j=0$ and $W_{j,k}$ are the details coefficients at scale j.

Because translating the functions ϕ or ψ by an integer amounts to translating the sequence $\{W_{j,k}, k\in\mathbb{Z}\}$ by the same integer for all j, we can suppose, without loss of generality, that the supports of ϕ and ψ are included in $[-\mathrm{T},0]$ and $[0,\mathrm{T}]$, respectively, for some integer $\mathrm{T}\ge 1$. Using this convention, it is easily seen that the wavelet coefficient $W_{j,k}$ depends only on the available observations $\{X_1,\dots,X_n\}$ when $j\ge 0$ and $0\le k<n_j$, where, denoting the integer part of x by $[x]$,

$$n_j \stackrel{\text{def}}{=} \max\bigl([2^{-j}(n-\mathrm{T}+1)-\mathrm{T}+1],0\bigr). \tag{7}$$

Suppose that X is a (possibly nonstationary) process with memory parameter d_0 and generalized spectral measure ν. If $M>d_0-1/2$, then $\Delta^M X$ is stationary and hence, by [14], Proposition 1, the sequence of wavelet coefficients $W_{j,\cdot}$ is a stationary process and we can define $\sigma_j^2(\nu) \stackrel{\text{def}}{=} \mathrm{Var}(W_{j,k})$. Our estimator takes advantage of the *scaling* and *weak dependence* properties of the wavelet coefficients, as expressed in the following condition, which will be shown to hold in many cases of interest.

CONDITION 1. There exist $\beta>0$ and $\sigma^2>0$ such that

$$\sup_{j\ge 1} 2^{\beta j}\left|\frac{\sigma_j^2(\nu)}{\sigma^2 2^{2d_0 j}}-1\right|<\infty \tag{8}$$

and

$$\sup_{n\ge 1}\sup_{j=1,\dots,J_n}(1+n_j2^{-2j\beta})^{-1}n_j^{-1}\,\mathrm{Var}\left(\sum_{k=0}^{n_j-1}\frac{W_{j,k}^2}{\sigma_j^2(\nu)}\right)<\infty. \tag{9}$$

Equation (8) states that, up to the multiplicative constant σ^2, the variance $\sigma_j^2(\nu)$ is approximated by $2^{2d_0 j}$ and that the error goes to zero exponentially fast as a function of j. It is a direct consequence of the approximation of the covariance of the wavelet coefficients established in [14]. Equation (9) imposes a bound on the variance of the normalized partial sum of the stationary centered sequence $\{\sigma_j^{-2}(\nu)W_{j,k}^2\}$, which, provided that $n_j 2^{-2j\beta}=O(1)$, is equivalent to what occurs when these variables are independent. We stress that the wavelet coefficients $W_{j,k}$ are, however, **not** independent, nor can they be approximated by independent coefficients; see [14]. Establishing (9) requires additional assumptions on the process X that go beyond its covariance structure since $W_{j,k}^2$ is involved; see Theorem 1, where this property is established for a general class of linear processes. We have isolated relations (8) and (9) because in our semiparametric context, these two relations are sufficient to show that the wavelet Whittle estimator converges to d_0 at the optimal rate (see Theorem 3 below).

Let us recall some definitions and results from [14] which are used here. As noted above, for a given scale j, the process $\{W_{j,k}\}_{k\in\mathbb{Z}}$ is covariance stationary. It will be called the *within-scale* process because all the $W_{j,k}$, $k \in \mathbb{Z}$, share the same j. The situation is more complicated when considering two different scales $j > j'$ because the two-dimensional sequence $\{[W_{j,k}, W_{j',k}]^T\}_{k\in\mathbb{Z}}$ is not stationary, as a consequence of the pyramidal wavelet scheme. A convenient way to define a joint spectral density for wavelet coefficients is to consider the *between-scale* process.

DEFINITION 1. The sequence $\{[W_{j,k}, \mathbf{W}_{j,k}(j-j')^T]^T\}_{k\in\mathbb{Z}}$, where

$$\mathbf{W}_{j,k}(j-j') \stackrel{\text{def}}{=} [W_{j',2^{j-j'}k}, \dots, W_{j',2^{j-j'}k+2^{j-j'}-1}]^T,$$

is called the *between-scale* process at scales $0 \le j' \le j$. $\mathbf{W}_{j,k}(j-j')$ is a $2^{j-j'}$-dimensional vector of wavelet coefficients at scale j'.

Assuming that the generalized spectral measure of X is given by (2) and provided that $M > d_0 - 1/2$, since $\Delta^M X$ is stationary, both the within-scale process and the between-scale process are covariance stationary; see [14]. Let us consider the case $\nu^* \in \mathcal{H}(\beta, \gamma, \pi)$, that is, $\varepsilon = \pi$, so that ν^* admits a density f^* in the space $\mathcal{H}(\beta, \gamma)$ as defined in [14] and ν admits a density $f(\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{-2d_0} f^*(\lambda)$. We denote by $\mathbf{D}_{j,0}(\cdot; f)$ the spectral density of the within-scale process at scale index j and by $\mathbf{D}_{j,j-j'}(\cdot; f)$ the cross spectral density between $\{W_{j,k}\}_{k\in\mathbb{Z}}$ and $\{\mathbf{W}_{j,k}(j-j')\}_{k\in\mathbb{Z}}$ for $j' < j$. It will be convenient to set $u = j - j'$. Theorem 1 in [14] states that, under (W-1)–(W-5), for all $u \ge 0$, there exists $C > 0$ such that for all $\lambda \in (-\pi, \pi)$ and $j \ge u \ge 0$,

$$|\mathbf{D}_{j,u}(\lambda; f) - f^*(0)\mathbf{D}_{\infty,u}(\lambda; d_0)2^{2jd_0}| \le C f^*(0) 2^{(2d_0-\beta)j}, \tag{10}$$

where, for all $u \ge 0$, $d \in (1/2 - \alpha, M]$ and $\lambda \in (-\pi, \pi)$,

$$\mathbf{D}_{\infty,u}(\lambda; d) \stackrel{\text{def}}{=} \sum_{l\in\mathbb{Z}} |\lambda + 2l\pi|^{-2d} \mathbf{e}_u(\lambda + 2l\pi) \overline{\hat{\psi}(\lambda + 2l\pi)} \hat{\psi}\big(2^{-u}(\lambda + 2l\pi)\big), \tag{11}$$

with $\mathbf{e}_u(\xi) \stackrel{\text{def}}{=} 2^{-u/2}[1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T$.

REMARK 1. The condition (W-5) involves an upper and a lower bound. The lower bound guarantees that the series defined by the right-hand side of (11) omitting the term $l = 0$ converges uniformly for $\lambda \in (\pi, \pi)$. The upper bound guarantees that the term $l = 0$ is bounded at $\lambda = 0$. As a result, $\mathbf{D}_{\infty,u}(\lambda; d)$ is bounded on $\lambda \in (\pi, \pi)$ and, by (10), so is $\mathbf{D}_{j,u}(\lambda; f)$. In particular, the wavelet coefficients are short-range dependent. For details, see the proof of Theorem 1 in [14].

REMARK 2. We stress that (10) may no longer hold if we only assume $\nu^* \in \mathcal{H}(\beta,\gamma,\varepsilon)$ with $\varepsilon < \pi$ since in this case, no condition is imposed on $\nu(d\lambda)$ for $|\lambda| > \varepsilon$ and hence $W_{j,\bullet}$ may not have a density for all j. However, this difficulty can be circumvented by decomposing ν^* as

$$\nu^*(d\lambda) = f^*(\lambda)\,d\lambda + \tilde{\nu}^*(d\lambda), \tag{12}$$

where f^* has support in $[-\varepsilon,\varepsilon]$ and $\tilde{\nu}^*([-\varepsilon,\varepsilon]) = 0$; see the proof of Theorem 1.

Here is a simple interpretation of the bound (10). For any $d \in \mathbb{R}$, $2^{2jd}\mathbf{D}_{\infty,u}(\cdot; d)$ is the spectral density of the wavelet coefficient of the generalized fractional Brownian motion (GFBM) $\{B_{(d)}(\theta)\}$ defined as the Gaussian process indexed by test functions $\theta \in \Theta_{(d)} = \{\theta : \int_{\mathbb{R}} |\xi|^{-2d}|\hat{\theta}(\xi)|^2\,d\xi < \infty\}$ with mean zero and covariance

$$\operatorname{Cov}\big(B_{(d)}(\theta_1), B_{(d)}(\theta_2)\big) = \int_{\mathbb{R}} |\xi|^{-2d}\hat{\theta}_1(\xi)\overline{\hat{\theta}_2(\xi)}\,d\xi. \tag{13}$$

When $d > 1/2$, the condition $\int |\xi|^{-2d}|\hat{\theta}(\xi)|^2\,d\xi < \infty$ requires that $\hat{\theta}(\xi)$ decays sufficiently quickly at the origin and when $d < 0$, it requires that $\hat{\theta}(\xi)$ decreases sufficiently rapidly at infinity. Provided that $d \in (1/2-\alpha, M+1/2)$, the wavelet function ψ and its scaled and translated versions $\psi_{j,k}$ all belong to $\Theta_{(d)}$. Defining the discrete wavelet transform of $B_{(d)}$ as $W_{j,k}^{(d)} \stackrel{\mathrm{def}}{=} B_{(d)}(\psi_{j,k})$, $j \in \mathbb{Z}$, $k \in \mathbb{Z}$ and $\mathbf{W}_{j,k}^{(d)}(u) \stackrel{\mathrm{def}}{=} [W_{j-u,2^u k}^{(d)}, \dots, W_{j-u,2^u k+2^u-1}^{(d)}]$, one obtains

$$\operatorname{Cov}\big(W_{j,k}^{(d)}, \mathbf{W}_{j,k'}^{(d)}(u)\big) = 2^{2dj}\int_{-\pi}^{\pi} \mathbf{D}_{\infty,u}(\lambda; d)e^{i\lambda(k-k')}\,d\lambda; \tag{14}$$

see [14], Remark 5, for more details. Equation (10) shows that the within- and between-scale spectral densities $\mathbf{D}_{j,u}(\lambda;\nu)$ of the process X with memory parameter d may be approximated by the corresponding densities of the wavelet coefficients of the GFBM $B_{(d)}$, with an L^∞-error bounded by $O(2^{(2d_0-\beta)j})$.

The approximation (10) is a crucial step for proving that Condition 1 holds for linear processes. The following theorem is proved in Section 5.

THEOREM 1. *Let X be a process having generalized spectral measure* (2) *with $d_0 \in \mathbb{R}$ and with $\nu^* \in \mathcal{H}(\beta,\gamma,\varepsilon)$ such that $f^*(0) \stackrel{\mathrm{def}}{=} d\nu^*/d\lambda_{|\lambda=0} > 0$, where $\gamma > 0$, $\beta \in (0,2]$ and $\varepsilon \in (0,\pi]$. Then, under* (W-1)–(W-5), *the bound* (8) *holds with $\sigma^2 = f^*(0)\mathrm{K}(d_0)$, where*

$$\mathrm{K}(d) \stackrel{\mathrm{def}}{=} \int_{-\infty}^{\infty} |\xi|^{-2d}|\hat{\psi}(\xi)|^2\,d\xi \qquad \textit{for any } d \in (1/2-\alpha, M+1/2). \tag{15}$$

Suppose, in addition, that there exist an integer $k_0 \le M$ *and a real-valued sequence* $\{a_k\}_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ *such that*

$$(\Delta^{k_0} X)_k = \sum_{t\in\mathbb{Z}} a_{k-t} Z_t, \qquad k \in \mathbb{Z}, \tag{16}$$

where $\{Z_t\}_{t\in\mathbb{Z}}$ *is a weak white noise process such that* $\mathbb{E}[Z_t] = 0$, $\mathbb{E}[Z_t^2] = 1$, $\mathbb{E}[Z_t^4] = \mathbb{E}[Z_1^4] < \infty$ *for all* $t \in \mathbb{Z}$ *and*

$$\operatorname{Cum}(Z_{t_1}, Z_{t_2}, Z_{t_3}, Z_{t_4}) = \begin{cases} \mathbb{E}[Z_1^4] - 3, & \text{if } t_1 = t_2 = t_3 = t_4, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Then, under (W-1)–(W-5), *the bound* (9) *holds and Condition* 1 *is satisfied.*

REMARK 3. Relation (9) does not hold for every long-memory process X, even with arbitrary moment conditions; see [5].

REMARK 4. Any martingale increment process with constant finite fourth moment, as in the assumption A3′ considered in [15], satisfies (17). Another particular case is given by the following corollary, proved in Section 5.

The following result specializes Theorem 1 to a Gaussian process X and shows that at large scales, the wavelet coefficients of X can be approximated by those of a process $\bar{X}$ whose spectral measure $\bar{\nu}$ satisfies the global condition $\bar{\nu} \in \mathcal{H}(\beta, \gamma, \pi)$.

COROLLARY 2. *Let* X *be a Gaussian process having generalized spectral measure* (2) *with* $d_0 \in \mathbb{R}$ *and with* $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$ *such that* $f^*(0) \overset{\text{def}}{=} d\nu^*/d\lambda_{|\lambda=0} > 0$, *where* $\gamma > 0$, $\beta \in (0, 2]$ *and* $\varepsilon \in (0, \pi]$. *Then, under* (W-1)–(W-5), *Condition* 1 *is satisfied with* $\sigma^2 = f^*(0)\mathrm{K}(d_0)$.

There exists, moreover, a Gaussian process $\overline{X}$ *defined on the same probability space as* X *with generalized spectral measure* $\bar{\nu} \in \mathcal{H}(\beta, \gamma, \pi)$ *and wavelet coefficients* $\{\overline{W}_{j,k}\}$ *such that*

$$\begin{aligned} &\sup_{n\ge 1, j\ge 0} \{n_j 2^{j(1+2d_0-2\alpha)} + n_j^2 2^{2j(1-2\alpha)}\}^{-1} \\ &\quad \times \mathbb{E}\left[\left|\sum_{k=0}^{n_j-1} W_{j,k}^2 - \sum_{k=0}^{n_j-1} \overline{W}_{j,k}^2\right|^2\right] < \infty. \end{aligned} \tag{18}$$

3. Asymptotic behavior of the local Whittle wavelet estimator. We first define the estimator. Let $\{c_{j,k}, (j,k) \in \mathcal{I}\}$ be an array of centered independent Gaussian random variables with variance $\operatorname{Var}(c_{j,k}) = \sigma_{j,k}^2$, where $\mathcal{I}$ is a finite set. The negative of its log-likelihood is $(1/2)\sum_{(j,k)\in\mathcal{I}}\{c_{j,k}^2/\sigma_{j,k}^2 + \log(\sigma_{j,k}^2)\}$, up to a constant additive term. Our local Whittle wavelet estimator (LWWE) uses such a contrast process to estimate the memory parameter d_0 by choosing $c_{j,k} = W_{j,k}$. The scaling and weak dependence in Condition 1 then suggest the following *pseudo* negative log-likelihood:

$$\hat{\mathrm{L}}_{\mathcal{I}}(\sigma^2, d) = (1/2) \sum_{(j,k)\in\mathcal{I}} \{W_{j,k}^2/(\sigma^2 2^{2dj}) + \log(\sigma^2 2^{2dj})\}$$
$$= \frac{1}{2\sigma^2} \sum_{(j,k)\in\mathcal{I}} 2^{-2dj} W_{j,k}^2 + \frac{|\mathcal{I}|}{2} \log\bigl(\sigma^2 2^{2\langle\mathcal{I}\rangle d}\bigr),$$

where $|\mathcal{I}|$ denotes the number of elements of the set $\mathcal{I}$ and $\langle\mathcal{I}\rangle$ is defined as the average scale,

$$\langle\mathcal{I}\rangle \stackrel{\text{def}}{=} \frac{1}{|\mathcal{I}|} \sum_{(j,k)\in\mathcal{I}} j. \tag{19}$$

Define $\hat{\sigma}_{\mathcal{I}}^2(d) \stackrel{\text{def}}{=} \mathrm{Argmin}_{\sigma^2>0} \hat{\mathrm{L}}_{\mathcal{I}}(\sigma^2, d) = |\mathcal{I}|^{-1} \sum_{(j,k)\in\mathcal{I}} 2^{-2dj} W_{j,k}^2$. The maximum pseudo-likelihood estimator of the memory parameter is then equal to the minimum of the negative profile log-likelihood (see [21], page 403), $\hat{d}_{\mathcal{I}} \stackrel{\text{def}}{=} \mathrm{Argmin}_{d\in\mathbb{R}} \hat{\mathrm{L}}_{\mathcal{I}}(\hat{\sigma}_{\mathcal{I}}^2(d), d)$, that is,

$$\hat{d}_{\mathcal{I}} = \mathop{\mathrm{Argmin}}_{d\in\mathbb{R}} \tilde{\mathrm{L}}_{\mathcal{I}}(d), \qquad \text{where } \tilde{\mathrm{L}}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log \sum_{(j,k)\in\mathcal{I}} 2^{2d(\langle\mathcal{I}\rangle - j)} W_{j,k}^2. \tag{20}$$

If $\mathcal{I}$ contains at least two different scales, then $\tilde{\mathrm{L}}_{\mathcal{I}}(d) \to \infty$ as $d \to \pm\infty$ and thus $\hat{d}_{\mathcal{I}}$ is finite. The derivative of $\tilde{\mathrm{L}}_{\mathcal{I}}(d)$ vanishes at $d = \hat{d}_{\mathcal{I}}$, that is, $\widehat{\mathbf{S}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) = 0$, where for all $d \in \mathbb{R}$,

$$\widehat{\mathbf{S}}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \sum_{(j,k)\in\mathcal{I}} [j - \langle\mathcal{I}\rangle] 2^{-2jd} W_{j,k}^2. \tag{21}$$

We consider two specific choices for $\mathcal{I}$. For any integers n, j_0 and j_1, $j_0 \le j_1$, the set of all available wavelet coefficients from n observations $X_1, \dots, X_n$ having scale indices between j_0 and j_1 is

$$\mathcal{I}_n(j_0, j_1) \stackrel{\text{def}}{=} \{(j,k) : j_0 \le j \le j_1, 0 \le k < n_j\}, \tag{22}$$

where n_j is given in (7). Consider two sequences, $\{L_n\}$ and $\{U_n\}$, satisfying, for all n,

$$0 \le L_n < U_n \le J_n, \qquad J_n \stackrel{\text{def}}{=} \max\{j : n_j \ge 1\}. \tag{23}$$

The index J_n is the maximal available scale index for the sample size n; L_n and U_n will denote, respectively, the lower and upper scale indices used in the pseudo-likelihood function. The estimator will then be denoted $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$. As shown below, in the semiparametric framework, the lower scale L_n governs the rate of convergence of $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ toward the true memory parameter. There are two possible settings as far as the upper scale U_n is concerned:

(S-1) $U_n - L_n$ is fixed, equal to $\ell > 0$;
(S-2) $U_n \le J_n$ for all n and $U_n - L_n \to \infty$ as $n \to \infty$.

(S-1) corresponds to using a fixed number of scales and (S-2) corresponds to using a number of scales tending to infinity. We will establish the large sample properties of $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ for these two cases.

The following theorem, proved in Section 8, states that under Condition 1, the estimator $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ is consistent.

THEOREM 3 (Rate of convergence). *Assume Condition 1. Let $\{L_n\}$ and $\{U_n\}$ be two sequences satisfying (23) and suppose that, as $n \to \infty$,*

$$L_n^2 (n 2^{-L_n})^{-1/4} + L_n^{-1} \to 0. \tag{24}$$

The estimator $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ defined by (20) and (22) is then consistent with a rate given by

$$\hat{d}_{\mathcal{I}_n(L_n,U_n)} = d_0 + O_{\mathbb{P}}\{(n 2^{-L_n})^{-1/2} + 2^{-\beta L_n}\}. \tag{25}$$

By balancing the two terms in the bound (25), we obtain the optimal rate.

COROLLARY 4 (Optimal rate). *When $n \asymp 2^{(1+2\beta)L_n}$, we obtain the rate*

$$\hat{d}_{\mathcal{I}_n(L_n,U_n)} = d_0 + O_{\mathbb{P}}\big(n^{-\beta/(1+2\beta)}\big). \tag{26}$$

PROOF. By taking $n \asymp 2^{(1+2\beta)L_n}$, the condition $L_n^{-1} + L_n^2 (n 2^{-L_n})^{-1/4} \to 0$ is satisfied and $(nL_n)^{-1/2} \asymp 2^{-\beta L_n} \asymp n^{-\beta/(1+2\beta)}$. This is the minimax rate [7]. □

REMARK 5. Observe that the setting of Theorem 3 includes both cases (S-1) and (S-2). The difference between these settings will appear when computing the limit variance in the Gaussian case; see Theorem 5 below.

We shall now state a central limit theorem for the estimator $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ of d_0, under the additional assumption that X is a Gaussian process. Extensions to non-Gaussian linear processes will be considered in a future work. We denote by $|\cdot|$ the Euclidean norm and define, for all $d \in (1/2 - \alpha, M]$ and $u \in \mathbb{N}$,

$$\mathrm{I}_u(d) \stackrel{\text{def}}{=} \int_{-\pi}^{\pi} |\mathbf{D}_{\infty,u}(\lambda; d)|^2 \, d\lambda = (2\pi)^{-1} \sum_{\tau \in \mathbb{Z}} \mathrm{Cov}^2\big(W_{0,0}^{(d)}, W_{-u,\tau}^{(d)}\big), \tag{27}$$

where we have used (14). We denote, for all integer $\ell \ge 1$,

$$\eta_\ell \stackrel{\text{def}}{=} \sum_{j=0}^{\ell} j \frac{2^{-j}}{2 - 2^{-\ell}} \quad \text{and} \quad \kappa_\ell \stackrel{\text{def}}{=} \sum_{j=0}^{\ell} (j - \eta_\ell)^2 \frac{2^{-j}}{2 - 2^{-\ell}}, \tag{28}$$

$$\mathrm{V}(d_0, \ell) \stackrel{\text{def}}{=} \frac{\pi}{(2-2^{-\ell})\kappa_\ell(\log(2)\mathrm{K}(d_0))^2} \times \Biggl\{ \mathrm{I}_0(d_0) + \frac{2}{\kappa_\ell} \sum_{u=1}^{\ell} \mathrm{I}_u(d_0) 2^{(2d_0-1)u} \times \sum_{i=0}^{\ell-u} \frac{2^{-i}}{2-2^{-\ell}} (i - \eta_\ell)(i + u - \eta_\ell) \Biggr\}, \tag{29}$$

$$\mathrm{V}(d_0, \infty) \stackrel{\text{def}}{=} \frac{\pi}{[2\log(2)\mathrm{K}(d_0)]^2} \Biggl\{ \mathrm{I}_0(d_0) + 2 \sum_{u=1}^{\infty} \mathrm{I}_u(d_0) 2^{(2d_0-1)u} \Biggr\}, \tag{30}$$

where $\mathrm{K}(d)$ is defined in (15). The following theorem is proved in Section 8.

THEOREM 5 (CLT). *Let X be a Gaussian process having generalized spectral measure (2) with $d_0 \in \mathbb{R}$ and $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$ with $\nu^*(-\varepsilon, \varepsilon) > 0$, where $\gamma > 0$, $\beta \in (0, 2]$ and $\varepsilon \in (0, \pi]$. Let $\{L_n\}$ be a sequence such that*

$$L_n^2 (n 2^{-L_n})^{-1/4} + n 2^{-(1+2\beta)L_n} \to 0 \tag{31}$$

and $\{U_n\}$ be a sequence such that either (S-1) or (S-2) holds. Then, under (W-1)–(W-5), we have, as $n \to \infty$,

$$(n 2^{-L_n})^{1/2} \bigl(\hat{d}_{\mathcal{I}_n(L_n, U_n)} - d_0 \bigr) \xrightarrow{\mathcal{L}} \mathcal{N}[0, \mathrm{V}(d_0, \ell)], \tag{32}$$

where $\ell = \lim_{n\to\infty}(U_n - L_n) \in \{1, 2, \dots, \infty\}$.

REMARK 6. The condition (31) is similar to (24), but ensures, in addition, that the bias in (25) is asymptotically negligible.

REMARK 7. The larger the value of β, the smaller the size of the allowed range for d_0 in (W-5) for a given decay exponent α and number M of vanishing moments. Indeed, the range in (W-5) has been chosen so as to obtain a bound on the bias which corresponds to the best possible rate under the condition $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$. If (W-5) is replaced by the weakest condition $d_0 \in (1/2 - \alpha, M]$, which does not depend on β, the same CLT (32) holds, but β in condition (31) must be replaced by $\beta' \in (0, \beta]$. This β' must satisfy $1/2 - \alpha < (1 + \beta')/2 - \alpha < d_0$, that is, $0 < \beta' < 2(d_0 + \alpha) - 1$. When $\beta' < \beta$, one gets a slower rate in (32).

REMARK 8. Relation (32) holds under (S-1), where $\ell < \infty$ and (S-2), where $\ell = \infty$. It follows from (72) and (74) that $\mathrm{V}(d_0, \ell) \to \mathrm{V}(d_0, \infty) < \infty$ as $\ell \to \infty$. Our numerical experiments suggest that in some cases, one may have $\mathrm{V}(d_0, \ell) \le \mathrm{V}(d_0, \ell')$ with $\ell \le \ell'$; see the bottom left panel of Figure 1. In that figure, one indeed notices a bending of the curves for large d, which is more pronounced for small values of M and may be due to a correlation between the wavelet coefficients across scales.

REMARK 9. The most natural choice is $U_n = J_n$, which amounts to using all the available wavelet coefficients with scale index larger than L_n. The case (S-1) is nevertheless of interest. In practice, the number of observations n is finite and the number of available scales $J_n - L_n$ can be small. Since, when n is finite, it is always possible to interpret the estimator $\hat{d}_{\mathcal{I}_n(L_n,J_n)}$ as $\hat{d}_{\mathcal{I}_n(L_n,L_n+\ell)}$ with $\ell = J_n - L_n$, one may approximate the distribution of $(n2^{-L_n})^{1/2}(\hat{d}_{\mathcal{I}_n(L_n,J_n)} - d_0)$ either by $\mathcal{N}(0, \mathrm{V}(d_0, \ell))$ or by $\mathcal{N}(0, \mathrm{V}(d_0, \infty))$. Since the former involves only a single limit, it is likely to provide a better approximation for finite n. Another interesting application involves considering online estimators of d_0: online computation of wavelet coefficients is easier when the number of scales is fixed; see [19].

4. Discussion. The asymptotic variance $\mathrm{V}(d, \ell)$ is defined for all $\ell \in \{1, 2, \ldots, \infty\}$ and all $1/2 + \alpha < d \leq M$ by (29) and (30). Its expression involves the range of scales ℓ and the L^2-norm $\mathrm{I}_u(d_0)$ of the asymptotic spectral density $\mathbf{D}_{\infty,u}(\lambda; d)$ of the wavelet coefficients, both for the "within" scales ($u = 0$) and the "between" scales ($u > 0$). The choice of wavelets does not matter much, as Figure 1 indicates. One can use Daubechies wavelet or Coiflets (for which the scale function also has vanishing moments). What matters is the number of vanishing moments M and the decay exponent α, which both determine the frequency resolution of ψ. For wavelets derived from a multiresolution analysis, M is always known and [3], Remark 2.7.1, page 86, provides a sequence of lower bounds tending to α (we used such lower bounds for the Coiflets used below). For the Daubechies wavelet with M vanishing moments, an analytic formula giving α is available; see [4], equation (7.1.23), page 225 and the table on page 226, and note that our α equals the α of [4] plus 1.

4.1. *The ideal Shannon wavelet case.* The so-called Shannon wavelet ψ_S is such that its Fourier transform $\hat{\psi}_S$ satisfies $|\hat{\psi}_S(\xi)|^2 = 1$ for $|\xi| \in [\pi, 2\pi]$ and is zero otherwise. This wavelet satisfies (W-2)–(W-4) for arbitrary large M and α, but does not have compact support, hence it does not satisfy (W-1). We may not, therefore, choose this wavelet in our analysis. It is of interest, however, because it gives a rough idea of what happens when α and M are large since one can always construct a wavelet ψ satisfying (W-1)–(W-4) which is arbitrarily close to the Shannon wavelet. Using the Shannon wavelet in (11), we get, for all $\lambda \in (-\pi, \pi)$, $\mathbf{D}_{\infty,u}(\lambda; d) = 0$ for $u \geq 1$ and $\mathbf{D}_{\infty,0}(\lambda; d) = (2\pi - |\lambda|)^{-2d}$ so that, for all $d \in \mathbb{R}$, (29) becomes

$$\mathrm{V}(d, \ell) = \frac{\pi g(-4d)}{2(2 - 2^{-\ell})\kappa_\ell \log^2(2) g^2(-2d)} \qquad \text{where } g(x) = \int_\pi^{2\pi} \lambda^x \, d\lambda. \tag{33}$$

This $\mathrm{V}(d, \ell)$ is displayed in Figure 1.

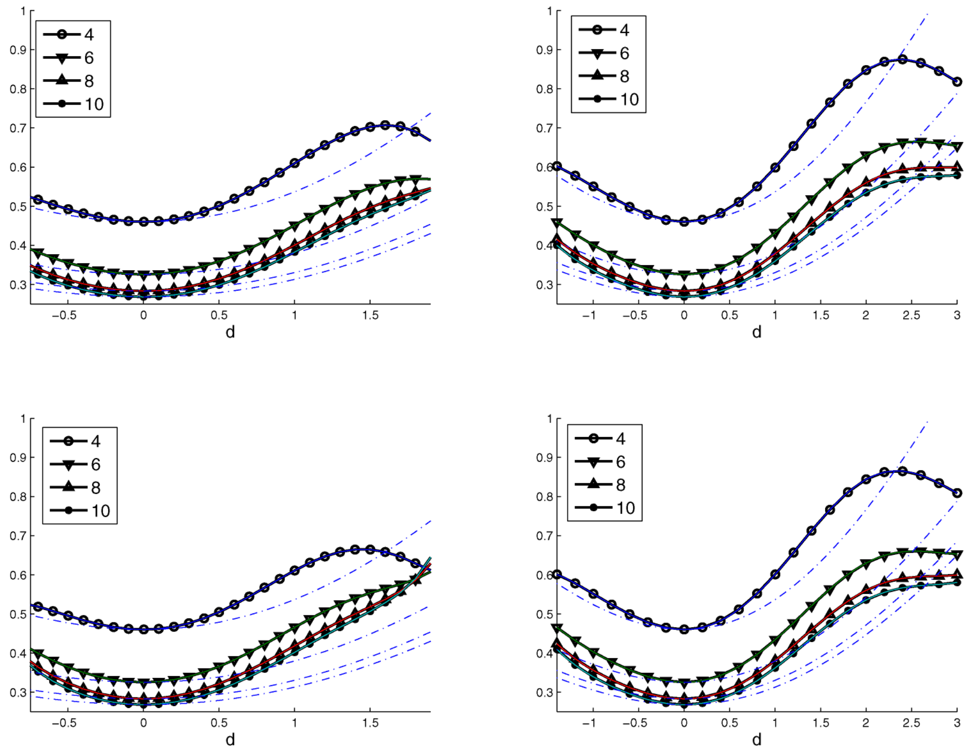

FIG. 1. *Numerical computations of the asymptotic variance* $\mathrm{V}(d,\ell)$ *for the Coiflets and Daubechies wavelets for different values of the number of scales* $\ell = 4, 6, 8, 10$ *and of the number of vanishing moments* $M = 2, 4$. *Top row*: *Coiflets*; *bottom row*: *Daubechies wavelets*; *left column*: $M = 2$; *right column*: $M = 4$. *The dash-dot lines are the asymptotic variances for the Shannon wavelet* [*see* (33)] *with* $\ell = 4, 6, 8, 10$. *For a given* ℓ, *the variances for different orthogonal wavelets coincide at* $d = 0$; *see the comment following* (34). *The right and left columns have different horizontal scales because different values of* M *yield different ranges for* d.

4.2. *Universal lower bound for* $\mathrm{I}_0(d)$. For $\ell = \infty$, using the facts that $\mathrm{I}_0(d) \geq 0$ for $u \geq 1$ and, by the Jensen inequality in (27), $\mathrm{I}_0(d) \geq \mathrm{K}^2(d)/(2\pi)$, we have, for all $1/2 + \alpha < d \leq M$,

$$\mathrm{V}(d, \infty) \geq (8 \log^2(2))^{-1} \simeq 0.2602. \tag{34}$$

This inequality is sharp when $d = 0$ and the wavelet family $\{\psi_{j,k}\}_{j,k}$ forms an orthonormal basis. This is because, in this case, the lower bound $(8\log^2(2))^{-1}$ in (34) equals $\mathrm{V}(0, \infty)$. Indeed, by (13) and Parsevals theorem, the wavelet coefficients $\{B_{(0)}(\psi_{j,k})\}_{j,k}$ are a centered white noise with variance 2π and, by (15) and (27), $\mathrm{K}(0) = 2\pi$ and $\mathrm{I}_u(0) = 2\pi \mathbb{1}(u = 0)$. Then, $\mathrm{V}(0, \ell) = (2(2 - 2^{-\ell})\kappa_\ell \log^2(2))^{-1}$. Since κ_ℓ is increasing with ℓ and tends to 2 as $\ell \to \infty$ (see Lemma 13), $\mathrm{V}(0, \ell) \geq (8 \log^2(2))^{-1} = \mathrm{V}(0, \infty)$. Hence, the lower bound (34) is attained at $d_0 = 0$ if $\{\psi_{j,k}\}_{j,k}$ is an orthonormal basis.

4.3. *Numerical computations.* For a given wavelet ψ, we can compute the variances $\mathrm{V}(d,\ell)$ numerically for any $\ell = 1, 2, \dots, \infty$ and $1/2 + \alpha < d \le M$. It is easily shown that $d \mapsto \mathrm{V}(d,\ell)$ is infinitely differentiable on $1/2 + \alpha < d \le M$ so that interpolation can be used between two different values of d. We compared numerical values of $\mathrm{V}(d,\ell)$ for four different wavelets, with $\ell = 4, 6, 8, 10$, and compared them with the Shannon approximation (33); see Figure 1. We used as wavelets two Daubechies wavelets which have $M = 2$ and $M = 4$ vanishing moments, and $\alpha = 1.3390$ and $\alpha = 1.9125$ decay exponents, respectively, and two so-called Coiflets with the same number of vanishing moments, and $\alpha > 1.6196$ and $\alpha > 1.9834$ decay exponents respectively. For a given number M of vanishing moments, the Coiflet has a larger support than the Daubechies wavelet, resulting in a better decay exponent. The asymptotic variances are different for $M = 2$, in particular, for negative d's, the Coiflet asymptotic variance is closer to that of the Shannon wavelet. The asymptotic variances are very close for $M = 4$.

4.4. *Comparison with Fourier estimators.* Semiparametric Fourier estimators are based on the periodogram. To allow comparison with Fourier estimators, we must first link the normalization factor $n2^{-L_n}$ with the bandwidth parameter m_n (the index of the largest normalized frequency) used by semiparametric Fourier estimators. A Fourier estimator with bandwidth m_n projects the observations $[X_1 \dots X_n]^T$ on the space generated by the vectors $\{\cos(2\pi k \cdot /n), \sin(2\pi k \cdot /n)\}$, $k = 1, \dots, m_n$, whose dimension is $2m_n$; on the other hand, the wavelet coefficients $\{W_{j,k}, j \ge L, k = 0, \dots, n_j - 1\}$ used in the wavelet estimator correspond to a projection on a space whose dimension is at most $\sum_{j=L_n}^{J_n} n_j \sim 2n2^{-L_n}$, where the equivalence holds as $n \to \infty$ and $n2^{-L_n} \to \infty$, by applying (75) with $j_0 = L_n$, $j_1 = J_n$ and $p = 1$. Hence, for m_n or $n2^{-L_n}$ large, it makes sense to consider $n2^{-L_n}$ as an analog of the bandwidth parameter m_n. The maximal scale index U_n is similarly related to the *trimming number* (the index of the smallest normalized frequency), often denoted by l_n (see [16]), that is, $l_n \sim n2^{-U_n}$. We stress that, in absence of trends, there is no need to trim coarsest scales.

With the above notation, the assumption (24) in Theorem 3 becomes $m_n/n + (\log n/m_n)^8 m_n^{-1} \to 0$ and the conclusion (25) is expressed as $\hat{d} = d_0 + O_{\mathbb{P}}(m_n^{-1/2} + (m_n/n)^{\beta})$. The assumption (31) becomes $(\log n/m_n)^8 m_n^{-1} + m_n^{1+2\beta}/n^{2\beta} \to 0$ and the rate of convergence in (32) is $m_n^{1/2}$.

The most efficient Fourier estimator is the local Whittle (Fourier) estimator studied in [15]; provided that

(1) the process $\{X_k\}$ is stationary and has spectral $f(\lambda) = |1 - e^{-i\lambda}|^{-2d_0} f^\star(\lambda)$ with $d_0 \in (-1/2, 1/2)$ and $f^*(\lambda) = f^*(0) + O(|\lambda|^\beta)$ as $\lambda \to 0$,

(2) the process $\{X_k\}$ is linear and causal, $X_k = \sum_{j=0}^{\infty} a_j Z_{k-j}$, where $\{Z_k\}$ is a martingale increment sequence satisfying $\mathbb{E}[Z_k^2 \mid \mathcal{F}_{k-1}] = 1$ a.s., $\mathbb{E}[Z_k^3 | \mathcal{F}_{k-1}] = \mu^3$ a.s. and $\mathbb{E}[Z_k^4] = \mathbb{E}[Z_1^4]$, where $\mathcal{F}_k = \sigma(Z_{k-l}, l \ge 0)$ and $a(\lambda) \overset{\text{def}}{=} \sum_{k=0}^{\infty} a_k e^{-ik\lambda}$

is differentiable in a neighborhood $(0,\delta)$ of the origin and $|da/d\lambda(\lambda)| = O(|a(\lambda)|/\lambda)$ as $\lambda \to 0^+$ (see A2′)

(3) $m_n^{-1} + (\log m_n)^2 m_n^{1+2\beta}/n^{2\beta} \to 0$ (see A4′) ,

then $m_n^{1/2}(\hat{d}_{m_n} - d_0)$ is asymptotically zero-mean Gaussian with variance $1/4$. This asymptotic variance is smaller than (but very close to) our lower bound in (34) and comparable to the asymptotic variance obtained numerically for the Daubechies wavelet with two vanishing moments; see the left-hand panel in Figure 1. Also, note that while the asymptotic variance of the Fourier estimators is a constant, the asymptotic variances of the wavelet estimators depend on d_0 (see Figure 1). In practice, one estimates the limiting variance $\mathrm{V}(d_0,\ell)$ by $\mathrm{V}(\hat{d},\ell)$ in order to construct asymptotic confidence intervals. The continuity of $\mathrm{V}(\cdot,\ell)$ and the consistency of $\hat{d}$ justify this procedure.

We would like to stress, however, that the wavelet estimator has some distinctive advantages. From a theoretical standpoint, for a given β, the wavelet estimator is rate optimal, that is, for $\beta \in (0,2]$, the rate is $n^{\beta/1+2\beta}$ (see Corollary 4) and the CLT is obtained for any rate $o(n^{\beta/1+2\beta})$. For the local Whittle Fourier estimator, the best rate of convergence is $O((n/\log^2(n))^{\beta/1+2\beta})$ and the CLT is obtained for any rate $o((n/\log^2(n))^{\beta/1+2\beta})$. This means that for any given β, the wavelet estimator has a faster rate of convergence and can therefore yield, for an appropriate admissible choice of the finest scale, shorter confidence intervals. Another advantage of the wavelet Whittle estimator over this estimator is that the optimal rate of convergence is shown to hold for $\nu^\star \in \mathcal{H}(\beta,\gamma,\varepsilon)$ without any further regularity assumption, such as the density f^* of ν^* having to be differentiable in a neighborhood of zero, with a given growth of the logarithmic derivative. To the best of our knowledge, the GPH estimator is the only Fourier estimator which has been shown, in a Gaussian context, to achieve the rate $O(n^{\beta/(1+2\beta)})$ (see [7]); its asymptotic variance is $\pi^2/24 \simeq 0.4112$. It is larger than the lower bound (34) and larger than the asymptotic variance obtained by using standard Daubechies wavelets with $\ell \geq 6$ on the range $(-1/2, 1/2)$ of d_0 allowed for the GPH estimator (see Figure 1). When pooling frequencies, the asymptotic variance of the GPH estimator improves and tends to $1/4$ (the local Whittle Fourier asymptotic variance) as the number of pooled frequencies tends to infinity; see [16].

Thus far, we have compared our local Whittle wavelet estimator with the local Whittle Fourier (LWF) and GPH estimators in the context of a stationary and invertible process X, that is, for $d_0 \in (-1/2, 1/2)$. As already mentioned, the wavelet estimators can be used for arbitrarily large ranges of the parameter d_0 by appropriately choosing the wavelet so that (W-5) holds. There are two main ways of adapting the LWF estimator to larger ranges of d: differentiating and tapering the data (see [22]) or, as promoted by [20], modifying the local Whittle likelihood, yielding the so-called exact local Whittle Fourier (ELWF) estimator. The theoretical analysis of these methods is performed under the same set of assumptions as in [15], so the same comments on the nonoptimality of the rate and on the restriction on $f^\star$ apply. Also, note that the model considered by [20] for X differs from the model of integrated processes defined by (16) and is not time-shift invariant; see their equation (1). In addition, their estimator is not invariant under the addition of a constant in the data, a drawback which is not easily dealt with; see their Remark 2. The asymptotic variance of the ELWF estimator has been shown to be $1/4$, the same as the LFW estimator, provided that the range (Δ_1, Δ_2) for d_0 is of width $\Delta_2 - \Delta_1 \le 9/2$. The asymptotic variance of our local Whittle wavelet estimator with eight scales, using the Daubechies wavelet with $M = 4$ zero moments, is at most 0.6 on a range of same width; see the left-hand panel in Figure 1. Again, this comparison does not take into account the logarithmic factor in the rate of convergence imposed by the conditions on the bandwidth m_n. Concerning the asymptotic variances of tapered Fourier estimators, increasing the allowed range for d_0 means increasing the taper order (see [8] and [17]), which, as already explained, inflates the asymptotic variance of the estimates. In contrast, for the wavelet methods, by increasing the number of vanishing moments M of, say, a Daubechies wavelet, the allowed range for d_0 is arbitrarily large while the asymptotic variance converges to the ideal Shannon wavelet case, derived in (33); the numerical values are displayed in Figure 1 for different values of the number of scales ℓ. The figure shows that larger values of ℓ tend to yield a smaller asymptotic variance. One should thus choose the largest possible M and the maximal number of scales. This prescription cannot be applied to a small sample because increasing the support of the wavelet decreases the number of available scales. The Daubechies wavelets with $M = 2$ to $M = 4$ are commonly used in practice.

From a practical standpoint, the wavelet estimator is computationally more efficient than the aforementioned Fourier estimators. Using the fast pyramidal algorithm, the wavelet transform coefficients are computed in $O(n)$ operations. The function $d \mapsto \tilde{\mathrm{L}}_{\mathcal{I}}(d)$ can be minimized using the Newton algorithm [2], Chapter 9.5, whose convergence is guaranteed because $\tilde{\mathrm{L}}_{\mathcal{I}}(d)$ is convex in d. The complexity of the minimization procedure is related to the computational cost of evaluation of the function $\tilde{\mathrm{L}}_{\mathcal{I}}$ and its two first derivatives. Assume that these functions need to be evaluated at p distinct values $d_1, \dots, d_p$. We first compute the empirical variance of the wavelet coefficients $n_j^{-1} \sum_{k=0}^{n_j-1} W_{j,k}^2$ for the scales $j \in \{L_n, \dots, U_n\}$, which does not depend on d and requires $O(n)$ operations. For $\mathcal{I} = \mathcal{I}_n(L_n, U_n)$, $\tilde{\mathrm{L}}_{\mathcal{I}}$ and all of its derivatives are linear combinations of these $U_n - L_n + 1 = O(\log(n))$ empirical variances with weights depending on d. The total complexity for computing the wavelet Whittle estimator in an algorithm involving p iterations is thus $O(n + p\log(n))$. The local Whittle Fourier (LWF) contrast being convex, the same Newton algorithm converges, but the complexity is slightly higher. The computation of the Fourier coefficients requires $O(n\log(\boldsymbol{n}))$ operations. The number of terms in the LWF contrast function (see [15], page 1633) is of order m_n [which is typically of order $O(n^\gamma)$, where $\gamma \in (0, 1/1 + 2\beta)$], so the evaluation of the LWF contrast function (and its derivatives) for p distinct values of the memory parameter $d_1, \dots, d_p$ requires $O(pm_n)$ operations. The overall complexity of computing the LWF estimator in a Newton algorithm involving p steps is therefore $O(n\log(n) + pm_n)$. Differentiating and tapering the data only adds $O(n)$ operations, so the same complexity applies in this case. The ELWF estimator is much more computationally demanding and is impractical for large data sets: for each value of the memory coefficient d at which the pseudo-likelihood function is evaluated, the algorithm calls for the fractional integration or differentiation of the observations, namely, $(\Delta^d X)_k, k = 1, \dots, n$, and the computation of the Fourier transform of $\{(\Delta^d X)_1, \dots, (\Delta^d X)_n\}$. In this context, $(\Delta^d X)_k \stackrel{\text{def}}{=} \sum_{l=0}^{k} \frac{(-d)_l}{l!} X_{k-l}, k = 1, \dots, n$, where $(x)_0 = 1$ and $(x)_k = x(x+1)\cdots(x+k-1)$ for $k \geq 1$ denote the Pochhammer symbols. The complexity of this procedure is thus $O(n^2 + n\log(n))$. The complexity for p function evaluations, therefore, is $O(p(n^2 + n\log(n)))$. The convexity of the criterion is not assured, so a minimization algorithm can possibly be trapped in a local minimum. These drawbacks make the ELWF estimator impractical for large data sets, say of size $10^6 - 10^7$, as encountered in teletraffic analysis or high-frequency financial data.

5. Condition 1 holds for linear and Gaussian processes.

PROOF OF THEOREM 1. For any scale index $j \in \mathbb{N}$, define by $\{h_{j,l}\}_{l\in\mathbb{Z}}$ the sequence $h_{j,l} \stackrel{\text{def}}{=} 2^{-j/2} \int_{-\infty}^{\infty} \phi(t+l)\psi(2^{-j}t)\,dt$ and by $H_j(\lambda) \stackrel{\text{def}}{=} \sum_{l\in\mathbb{Z}} h_{j,l} e^{-i\lambda l}$ its associated discrete-time Fourier transform. Since ϕ and ψ are compactly supported, $\{h_{j,l}\}$ has a finite number of nonzero coefficients. As shown by [14], Relation 13, for any sequence $\{x_l\}_{l\in\mathbb{Z}}$, the discrete wavelet transform coefficients at scale j are given by $W^x_{j,k} = \sum_{l\in\mathbb{Z}} x_l h_{j,2^j k - l}$. In addition, it follows from [14], Relation 16, that $H_j(\lambda) = (1 - e^{-i\lambda})^M \tilde{H}_j(\lambda)$, where $\tilde{H}_j(\lambda)$ is a trigonometric polynomial, that is, $\tilde{H}_j(\lambda) = \sum_{l\in\mathbb{Z}} \tilde{h}_{j,l} e^{-i\lambda l}$, where $\{\tilde{h}_{j,l}\}$ has a finite number of nonzero coefficients.

Define $\overline{\nu}$ and $\tilde{\nu}$ as the restrictions of ν on $[-\varepsilon, \varepsilon]$ and on its complementary set, respectively. These definitions imply that

$$\sigma_j^2(\nu) = \sigma_j^2(\overline{\nu}) + \sigma_j^2(\tilde{\nu}). \tag{35}$$

Since $\nu^* \in \mathcal{H}(\beta, \gamma, \varepsilon)$, the corresponding decomposition for ν^* reads as in (12), so $\overline{\nu}$ admits a density $f(\lambda) = |1 - e^{-i\lambda}|^{-2d_0} f^*(\lambda)$ on $\lambda \in [-\pi, \pi]$, where $f^*(\lambda) = 0$ for $\lambda \notin [-\varepsilon, \varepsilon]$ and $|f^*(\lambda) - f^*(0)| \leq \gamma f^*(0)|\lambda|^\beta$ on $\lambda \in [-\varepsilon, \varepsilon]$. Hence, (10) holds: by [14], Theorem 1, there exists a constant C such that for all $j \geq 0$ and $\lambda \in (-\pi, \pi)$,

$$|\mathbf{D}_{j,u}(\lambda; f) - f^*(0)\mathbf{D}_{\infty,u}(\lambda; d_0)2^{2jd_0}| \leq C f^*(0)\bar{\gamma} 2^{(2d_0-\beta)j}. \tag{36}$$

Recall that $\mathbf{D}_{j,0}(\lambda; f)$ is the spectral density of a stationary series with variance $\sigma_j^2(\overline{\nu}) = \int_{-\pi}^{\pi} \mathbf{D}_{j,0}(\lambda; f)\,d\lambda$. Similarly, by (14) and (15), $\mathbf{D}_{\infty,0}(\lambda; d_0)$ is the spectral density of a stationary series with variance $\mathrm{K}(d_0)$. Thus, after integration on $\lambda \in (-\pi, \pi)$, (36) with $u = 0$ yields

$$|\sigma_j^2(\overline{\nu}) - f^*(0)\mathrm{K}(d_0)2^{2jd_0}| \le 2\pi C f^*(0)\bar{\gamma}2^{(2d_0-\beta)j}. \tag{37}$$

By [14], Proposition 9, there exists a constant C such that $|H_j(\lambda)| \le C2^{j(M+1/2)} \times |\lambda|^M(1+2^j|\lambda|)^{-\alpha-M}$ for any $\lambda \in [-\pi, +\pi]$, which implies that

$$\begin{aligned}\sigma_j^2(\tilde{\nu}) &= 2\int_\varepsilon^\pi |H_j(\lambda)|^2\nu(d\lambda) \le C2^{(1+2M)j}\int_\varepsilon^\pi \lambda^{2M}(1+2^j\lambda)^{-2\alpha-2M}\nu(d\lambda)\\ &\le C\pi^{2M}2^{(1+2M)j}(1+\varepsilon 2^j)^{-2\alpha-2M}\nu([\varepsilon,\pi])\\ &= O(2^{j(1-2\alpha)}) = o(2^{j(2d_0-\beta)}),\end{aligned} \tag{38}$$

since, by (W-5), $1-2\alpha-2d_0+\beta < 0$. Relations (35), (37) and (38) prove (8).

We now consider (9). We have, for all $j \ge 0$ and $n \ge 1$, (see [18], Theorem 2, page 34),

$$\begin{aligned}\mathrm{Var}\left(\sum_{k=0}^{n_j-1} W_{j,k}^2\right) &= \sum_{\tau=-n_j+1}^{n_j-1}(n_j-|\tau|)\,\mathrm{Cov}(W_{j,0}^2, W_{j,\tau}^2)\\ &= \sum_{\tau=-n_j+1}^{n_j-1}(n_j-|\tau|)[2\,\mathrm{Cov}^2(W_{j,0}, W_{j,\tau})\\ &\qquad + \mathrm{Cum}(W_{j,0}, W_{j,0}, W_{j,\tau}, W_{j,\tau})].\end{aligned} \tag{39}$$

Using (16), since $M \ge k_0$, we may write

$$W_{j,k} = \sum_{t\in\mathbb{Z}} \tilde{h}_{j,2^jk-t}(\Delta^M X)_t = \sum_{t\in\mathbb{Z}} b_{j,2^jk-t}Z_t, \tag{40}$$

where $b_{j,\cdot} \stackrel{\text{def}}{=} \tilde{h}_{j,\cdot} \star (\Delta^{M-k_0}a)$ belongs to $\ell^2(\mathbb{Z})$. By (17), we thus obtain

$$\mathrm{Cum}(W_{j,0}, W_{j,0}, W_{j,\tau}, W_{j,\tau}) = (\mathbb{E}[Z_1^4]-3)\sum_{t\in\mathbb{Z}} b_{j,t}^2 b_{j,2^j\tau-t}^2,$$

which, in turns, implies that

$$\begin{aligned}\sum_{\tau\in\mathbb{Z}} |\,\mathrm{Cum}(W_{j,0}, W_{j,0}, W_{j,\tau}, W_{j,\tau})| &= |\mathbb{E}[Z_1^4]-3| \sum_{t,\tau\in\mathbb{Z}} b_{j,t}^2 b_{j,2^j\tau-t}^2\\ &\le |\mathbb{E}[Z_1^4]-3|\sigma_j^4(\nu)\end{aligned} \tag{41}$$

since, by (40), $\sum_t b_{j,t}^2 = \sigma_j^2(\nu)$.

We shall now bound $\sum_{\tau=-n_j+1}^{n_j-1}\mathrm{Cov}^2(W_{j,0}, W_{j,\tau})$. One can define uncorrelated wavelet coefficients $\{\overline{W}_{j,k}\}$ and $\{\tilde{W}_{j,k}\}$, associated with the generalized spectral measures $\overline{\nu}$ and $\tilde{\nu}$, respectively and such that $W_{j,k} = \overline{W}_{j,k} + \tilde{W}_{j,k}$ for all $j \ge 0$ and

$k \in \mathbb{Z}$. Therefore, $\mathrm{Cov}^2(W_{j,0}, W_{j,\tau}) = \mathrm{Cov}^2(\overline{W}_{j,0}, \overline{W}_{j,0}) + \mathrm{Cov}^2(\tilde{W}_{j,0}, \tilde{W}_{j,\tau}) + 2\,\mathrm{Cov}(\overline{W}_{j,0}, \overline{W}_{j,\tau})\,\mathrm{Cov}(\tilde{W}_{j,0}, \tilde{W}_{j,\tau})$. By (8), $\sigma_j^2(\nu) \asymp 2^{2jd_0}$. Therefore, by (36) and using [14], Proposition 3, equation (30), for all $j \geq 0$, $\{\sigma_j^{-1}(\nu)\overline{W}_{j,k}, k \in \mathbb{Z}\}$ is a stationary process whose spectral density is bounded above by a constant independent of j. Parsevals theorem implies that $\sup_{j\geq 1} \sigma_j^{-4}(\nu) \sum_{\tau\in\mathbb{Z}} \mathrm{Cov}^2(\overline{W}_{j,0}, \overline{W}_{j,\tau}) < \infty$, hence

$$\sup_{n\geq 1} \sup_{j=1,\dots,J_n} n_j^{-1}\sigma_j^{-4}(\nu) \sum_{\tau=-n_j+1}^{n_j-1} (n_j - |\tau|)\,\mathrm{Cov}^2(\overline{W}_{j,0}, \overline{W}_{j,\tau}) < \infty. \tag{42}$$

Now, consider $\{\tilde{W}_{j,k}\}$. The Cauchy–Schwarz inequality and the stationarity of the within-scale process imply that $\mathrm{Cov}^2(\tilde{W}_{j,0}, \tilde{W}_{j,\tau}) \leq \mathrm{Var}^2(\tilde{W}_{j,0}) = \sigma_j^4(\tilde{\nu}) = O(2^{2j(1-2\alpha)})$, by (38), and since $\sigma_j^2(\nu) \asymp 2^{2jd_0}$, we get

$$\sup_{n\geq 1} \sup_{j=1,\dots,J_n} \frac{2^{2j(2\alpha+2d_0-1)}}{n_j^2\sigma_j^4(\nu)} \sum_{\tau=-n_j+1}^{n_j-1} (n_j - |\tau|)\,\mathrm{Cov}^2(\tilde{W}_{j,0}, \tilde{W}_{j,\tau}) < \infty. \tag{43}$$

Finally, using the fact that, for any $j \geq 1$, $\mathbf{D}_{j,0}(\lambda; f)$ is the spectral density of the process $\{\overline{W}_{j,k}\}$ and denoting by $\tilde{\nu}_j$ the spectral measure of $\{\tilde{W}_{j,k}\}_{k\in\mathbb{Z}}$, it is straightforward to show that

$$\begin{aligned} A(n,j) &\stackrel{\text{def}}{=} \sum_{\tau=-n_j+1}^{n_j-1} (n_j - |\tau|)\,\mathrm{Cov}(\overline{W}_{j,0}, \overline{W}_{j,\tau})\,\mathrm{Cov}(\tilde{W}_{j,0}, \tilde{W}_{j,\tau}) \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \mathbf{D}_{j,0}(\lambda'; f) \left|\sum_{k=0}^{n_j-1} e^{ik(\lambda+\lambda')}\right|^2 \tilde{\nu}_j(d\lambda)\,d\lambda' \\ &\leq 2\pi n_j \sigma_j^2(\tilde{\nu}) \|\mathbf{D}_{j,0}(\cdot; f)\|_\infty. \end{aligned}$$

This implies that $A(n,j) \geq 0$ and using (38), (36) and $\sigma_j^2(\nu) \asymp 2^{2jd_0}$, we get

$$\sup_{n\geq 1} \sup_{j=1,\dots,J_n} \frac{2^{j(2\alpha+2d_0-1)}}{n_j\sigma_j^4(\nu)} |A(n,j)| < \infty. \tag{44}$$

Using the fact that $W_{j,k} = \overline{W}_{j,k} + \tilde{W}_{j,k}$ and $\overline{W}_{j,k}$ and $\tilde{W}_{j,k}$ are uncorrelated, (39), (41), (42), (43), (44) and $1 - 2\alpha - 2d_0 < -\beta < 0$ yield (9). □

REMARK 10. If $\varepsilon = \pi$ in the assumptions of Theorem 1, then, in the above proof, $\tilde{W}_{j,k} = 0$ for all (j,k), so not only (9) holds, but also the stronger relation

$$\sup_{n\geq 1} \sup_{j=1,\dots,J_n} n_j^{-1}\,\mathrm{Var}\left(\sum_{k=0}^{n_j-1} \frac{W_{j,k}^2}{\sigma_j^2(\nu)}\right) < \infty. \tag{45}$$

PROOF OF COROLLARY 2. Condition 1 holds because Theorem 1 applies to a Gaussian process. Moreover, since its fourth order cumulants are zero, the relation $W_{j,k}^2 = \overline{W}_{j,k}^2 + \tilde{W}_{j,k}^2 + 2\overline{W}_{j,k}\tilde{W}_{j,k}$, (43) and (44) yield

$$\operatorname{Var}\left(\sum_{k=0}^{n_j-1}(W_{j,k}^2 - \overline{W}_{j,k}^2)\right) \le C\left[\frac{n_j^2\sigma_j^4(\nu)}{2^{2j(2\alpha+2d_0-1)}} + \frac{n_j\sigma_j^4(\nu)}{2^{j(2\alpha+2d_0-1)}}\right],$$

where C is a positive constant. Since $\overline{W}_{j,k}$ and $\tilde{W}_{j,k}$ are uncorrelated, $\mathbb{E}[W_{j,k}^2 - \overline{W}_{j,k}^2] = \sigma_j^2(\tilde{\nu})$, hence the last display, $\sigma_j^2(\nu) \asymp 2^{2jd_0}$ and (38) yield (18). □

6. Asymptotic behavior of the contrast process. We decompose the contrast (20) into a sum of a (deterministic) function of d and a random process indexed by d,

$$\tilde{\mathrm{L}}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \mathrm{L}_{\mathcal{I}}(d) + \mathrm{E}_{\mathcal{I}}(d) + \log\bigl(|\mathcal{I}|\sigma^2 2^{2d_0\langle\mathcal{I}\rangle}\bigr), \tag{46}$$

where the log term does not depend on d (and thus may be discarded) and

$$\mathrm{L}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log\left(\frac{1}{|\mathcal{I}|}\sum_{(j,k)\in\mathcal{I}} 2^{2(d_0-d)j}\right) - \frac{1}{|\mathcal{I}|}\sum_{(j,k)\in\mathcal{I}}\log\bigl(2^{2(d_0-d)j}\bigr), \tag{47}$$

$$\mathrm{E}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log\left[1 + \sum_{(j,k)\in\mathcal{I}}\frac{2^{2(d_0-d)j}}{\sum_{\mathcal{I}} 2^{2(d_0-d)j}}\left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1\right)\right], \tag{48}$$

with σ^2 defined in (8).

PROPOSITION 6. *For any finite and nonempty set $\mathcal{I} \subset \mathbb{N}\times\mathbb{Z}$, the function $d \to \mathrm{L}_{\mathcal{I}}(d)$ is nonnegative, convex and vanishes at $d = d_0$. Moreover, for any sequence $\{L_n\}$ such that $n2^{-L_n} \to \infty$ as $n \to \infty$, and for any constants $d_{\min}$ and $d_{\max}$ in $\mathbb{R}$ satisfying $d_0 - 1/2 < d_{\min} \le d_{\max}$,*

$$\liminf_{n\to\infty}\ \inf_{d\in[d_{\min},d_{\max}]}\ \inf_{j_1=L_n+1,\dots,J_n} \ddot{\mathrm{L}}_{\mathcal{I}_n(L_n,j_1)}(d) > 0, \tag{49}$$

where $\mathcal{I}_n$ is defined in (22) and $\ddot{\mathrm{L}}_{\mathcal{I}}$ denotes the second derivative of $\mathrm{L}_{\mathcal{I}}$.

PROOF. By concavity of the log function, $\mathrm{L}_{\mathcal{I}}(d) \ge 0$ and is zero if $d = d_0$. If $\mathcal{I} = \mathcal{I}_n(L_n, j_1)$ with $j_1 \ge L_n + 1$, one can compute $\ddot{\mathrm{L}}_{\mathcal{I}}(d)$ and show that it can be expressed as $\ddot{\mathrm{L}}_{\mathcal{I}}(d) = (2\log(2))^2 \operatorname{Var}(N)$, where N is an integer-valued random variable such that $\mathbb{P}(N = j) = 2^{2(d_0-d)j}n_j / \sum_{j=L_n}^{j_1} 2^{2(d_0-d)j}n_j$ for $j \ge 0$. Let $d \ge d_{\min} > d_0 - 1/2$. Then,

$$\mathbb{P}(N = L_n) \ge \bigl(1 - 2^{2(d_0-d_{\min})-1}\bigr)\{1 - \mathrm{T}2^{L_n}(n - \mathrm{T} + 1)^{-1}\}.$$

Since $n2^{-L_n} \to \infty$, the term between the brackets tends to 1 as $n \to \infty$. Hence, for n large enough, we have $\inf_{d \ge d_{\min}} \mathbb{P}(N = L_n) \ge (1 - 2^{2(d_0 - d_{\min})-1})/2$. Similarly, one finds, for n large enough, $\inf_{d \in [d_{\min}, d_{\max}]} \mathbb{P}(N = L_n + 1) \ge (1 - 2^{2(d_0 - d_{\min})-1}) 2^{2(d_0 - d_{\max})-1}/2$. Hence,

$$\begin{aligned}
\inf_{d \in [d_{\min}, d_{\max}]} \operatorname{Var}(N) &\ge \{L_n - \mathbb{E}(N)\}^2 \mathbb{P}(N = L_n) \\
&\quad + \{L_n + 1 - \mathbb{E}(N)\}^2 \mathbb{P}(N = L_n + 1) \\
&\ge \bigl(1 - 2^{2(d_0 - d_{\min})-1}\bigr) 2^{2(d_0 - d_{\max})-2} \\
&\quad \times \bigl(\{L_n - \mathbb{E}(N)\}^2 + \{L_n + 1 - \mathbb{E}(N)\}^2\bigr) \\
&\ge \bigl(1 - 2^{2(d_0 - d_{\min})-1}\bigr) 2^{2(d_0 - d_{\max})-4},
\end{aligned}$$

where the last inequality is obtained by observing that either $\mathbb{E}(N) - L_n \ge 1/2$ or $L_n + 1 - \mathbb{E}(N) < 1/2$. □

We now show that the random component $\mathrm{E}_{\mathcal{J}}(d)$ of the contrast (46) tends to 0 uniformly in d. For all $\rho > 0$, $q \ge 0$ and $\delta \in \mathbb{R}$, define the set of real-valued sequences

$$\mathcal{B}(\rho, q, \delta) \stackrel{\text{def}}{=} \bigl\{\{\mu_j\}_{j \ge 0} : |\mu_j| \le \rho(1 + j^q) 2^{j\delta} \text{ for all } j \ge 0\bigr\}. \tag{50}$$

Define, for any $n \ge 1$, any sequence $\boldsymbol{\mu} \stackrel{\text{def}}{=} \{\mu_j\}_{j \ge 0}$ and $0 \le j_0 \le j_1 \le J_n$,

$$\tilde{\mathbf{S}}_{n, j_0, j_1}(\boldsymbol{\mu}) \stackrel{\text{def}}{=} \sum_{j=j_0}^{j_1} \mu_{j-j_0} \sum_{k=0}^{n_j - 1} \biggl[\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1\biggr]. \tag{51}$$

PROPOSITION 7. *Under Condition 1, for any $q \ge 0$ and $\delta < 1$, there exists $C > 0$ such that for all $\rho \ge 0$, $n \ge 1$ and $j_0 = 1, \dots, J_n$,*

$$\begin{aligned}
&\Bigl\{\mathbb{E} \sup_{\boldsymbol{\mu} \in \mathcal{B}(\rho, q, \delta)} \sup_{j_1 = j_0, \dots, J_n} |\tilde{\mathbf{S}}_{n, j_0, j_1}(\boldsymbol{\mu})|^2\Bigr\}^{1/2} \\
&\qquad \le C \rho n 2^{-j_0} [H_{q,\delta}(n 2^{-j_0}) + 2^{-\beta j_0}],
\end{aligned} \tag{52}$$

where, for all $x \ge 0$, $H_{q,\delta}(x) \stackrel{\text{def}}{=} \begin{cases} x^{-1/2}, & \textit{if } \delta < 1/2, \\ \log^{q+1}(2 + x) x^{-1/2}, & \textit{if } \delta = 1/2, \\ \log^q(2 + x) x^{\delta - 1}, & \textit{if } \delta > 1/2. \end{cases}$

PROOF. We set $\rho = 1$ without loss of generality. We write

$$\tilde{\mathbf{S}}_{n, j_0, j_1}(\boldsymbol{\mu}) = \sum_{j=j_0}^{j_1} \frac{\sigma_j^2(\nu)}{\sigma^2 2^{2d_0 j}} \mu_{j-j_0} \sum_{k=0}^{n_j - 1} \biggl[\frac{W_{j,k}^2}{\sigma_j^2(\nu)} - 1\biggr] + \sum_{j=j_0}^{j_1} n_j \mu_{j-j_0} \biggl[\frac{\sigma_j^2(\nu)}{\sigma^2 2^{2d_0 j}} - 1\biggr]$$

and denote the two terms of the right-hand side of this equality as $\tilde{\mathbf{S}}^{(0)}_{n,j_0,j_1}(\boldsymbol{\mu})$ and $\tilde{\mathbf{S}}^{(1)}_{n,j_0,j_1}(\boldsymbol{\mu})$, respectively. By (8), $C_1 \stackrel{\text{def}}{=} \sup_{j\geq 0} 2^{\beta j}|\sigma_j^2(\nu)/(\sigma^2 2^{2d_0 j}) - 1| < \infty$, which implies $\sup_{j\geq 0}|\sigma_j^2(\nu)/(\sigma^2 2^{2d_0 j})| \leq 1 + C_1$. Hence, if $\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)$, then

$$|\tilde{\mathbf{S}}^{(0)}_{n,j_0,j_1}(\boldsymbol{\mu})| \leq (1 + C_1) \sum_{j=j_0}^{J_n} (1 + (j - j_0)^q) 2^{(j-j_0)\delta} \left| \sum_{k=0}^{n_j - 1} \left(\frac{W_{j,k}^2}{\sigma_j^2(\nu)} - 1 \right) \right|.$$

Using the Minkowski inequality and $n_j \leq n2^{-j}$, (9) implies that there exists a constant C_2 such that

$$\begin{aligned} &\left\{ \mathbb{E}\left[\sup_{\boldsymbol{\mu}\in\mathcal{B}(1,q,\delta)} \sup_{j_1 = j_0,\ldots,J_n} |\tilde{\mathbf{S}}^{(0)}_{n,j_0,j_1}(\boldsymbol{\mu})|^2 \right] \right\}^{1/2} \\ &\qquad \leq (1 + C_1) C_2 \sum_{j=j_0}^{J_n} (1 + (j - j_0)^q) 2^{(j-j_0)\delta} [(n2^{-j})^{1/2} + n2^{-(1+\beta)j}]. \end{aligned} \tag{53}$$

The sum over the first term is $O(n2^{-j_0} H_{q,\delta}(n2^{-j_0}))$ since $J_n - j_0 \asymp \log_2 n + \log_2 2^{-j_0} = \log_2(n2^{-j_0})$. The sum over the second term is $O(n2^{-(1+\beta)j_0})$ since $\delta < 1$ and $1 + \beta > 1$, so (53) is $O((n2^{-j_0})\{H_{q,\delta}(n2^{-j_0}) + 2^{-\beta j_0}\})$ since $2^{J_n} \asymp n$. Now, by the definition of C_1 above and since $n_j \leq n2^{-j}$, we get

$$\sup_{\boldsymbol{\mu}\in\mathcal{B}(1,q,\delta)} \sup_{j_1 = j_0,\ldots,J_n} |\tilde{\mathbf{S}}^{(1)}_{n,j_0,j_1}(\boldsymbol{\mu})| \leq C_1 n \sum_{j=j_0}^{J_n} (1 + (j - j_0)^q) 2^{(j-j_0)\delta} 2^{-j(1+\beta)},$$

which is $O(n2^{-(1+\beta)j_0})$. The two last displays yield (52). □

COROLLARY 8. *Let $\{L_n\}$ be a sequence such that $L_n^{-1} + (n2^{-L_n})^{-1} \to 0$ as $n \to \infty$ and let $\mathrm{E}_{\mathbf{I}}(d)$ be defined as in* (48). *Condition* 1 *then implies that as $n \to \infty$*:

(a) *for any $\ell \geq 0$,*

$$\sup_{d\in\mathbb{R}} |\mathrm{E}_{\mathbf{I}_n(L_n, L_n+\ell)}(d)| = O_{\mathbb{P}}((n2^{-L_n})^{-1/2} + 2^{-\beta L_n});$$

(b) *for all $d_{\min} > d_0 - 1/2$, setting $\delta = 2(d_0 - d_{\min})$,*

$$\sup_{d\geq d_{\min}} \sup_{j_1 = L_n,\ldots,J_n} |\mathrm{E}_{\mathbf{I}_n(L_n, j_1)}(d)| = O_{\mathbb{P}}(H_{0,\delta}(n2^{-L_n}) + 2^{-\beta L_n}).$$

PROOF. The definitions (48) and (51) imply that, for $0 \leq j_0 \leq j_1 \leq J_n$,

$$\mathrm{E}_{\mathbf{I}_n(j_0,j_1)}(d) = \log[1 + (n2^{-j_0})^{-1} \tilde{\mathbf{S}}_{n,j_0,j_1}[\boldsymbol{\mu}(d, j_0, j_1)]]$$

with $\boldsymbol{\mu}(d, j_0, j_1)$ is the sequence $\{\mu_j(d, j_0, j_1)\}_{j\geq 0}$ defined by

$$\mu_j(d, j_0, j_1) \stackrel{\text{def}}{=} n2^{-j_0} \frac{2^{2(d_0-d)(j+j_0)}}{\sum_{j'=j_0}^{j_1} 2^{2(d_0-d)j'} n_{j'}} \mathbb{1}(j \leq j_1 - j_0). \tag{54}$$

The bounds (a) and (b) then follow from Proposition 7, the Markov inequality and the following bounds.

Part (a). In this case, we apply Proposition 7 with $\delta = 0$. Indeed, using the fact that $\mu_j(d, L_n, L_n + \ell) n_{L_n+j} \leq n2^{-L_n}$ for all $j = 0, \ldots, \ell$ and is zero otherwise, we have that $\mu_j \leq n2^{-L_n}/n_{L_n+\ell} \to 2^\ell$ as $n \to \infty$, since $n2^{-L_n} \to \infty$. Then, for large enough n, $\boldsymbol{\mu}(d, L_n, L_n + \ell) \in \mathcal{B}(2^{\ell+1}, 0, 0)$ for all $d \in \mathbb{R}$.

Part (b). Here, we still apply Proposition 7, but with $\delta = 2(d_0 - d_{\min}) < 1$, implying that $H_{0,\delta}(n2^{-L_n}) \to 0$. Indeed, since the denominator of the ratio appearing in (54) is at least $2^{2(d_0-d)L_n} n_{L_n}$, we have $\sup_{j_1 \geq L_n} \sup_{d \geq d_{\min}} |\mu_j(d, L_n, j_1)| \leq n2^{-L_n} n_{L_n}^{-1} 2^{\delta j}$. Since $n2^{-L_n} \sim n_{L_n}$ as $n \to \infty$, we get that, for large enough n, $\boldsymbol{\mu}(d, L_n, j_1) \in \mathcal{B}(2, 0, \delta)$ for all $d \geq d_{\min}$ and $j_1 \geq L_n$. □

7. Weak consistency. We now establish a preliminary result on the consistency of $\hat{d}$. It does not provide an optimal rate, but it will be used in the proof of Theorem 3, which provides the optimal rate. By the definition of $\hat{d}$ and (46), we have

$$0 \geq \tilde{\mathrm{L}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) - \tilde{\mathrm{L}}_{\mathcal{I}}(d_0) = \mathrm{L}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) + \mathrm{E}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) - \mathrm{E}_{\mathcal{I}}(d_0). \tag{55}$$

The basic idea for proving consistency is to show that (1) the function $d \mapsto \tilde{\mathrm{L}}(d)$ behaves as $(d - d_0)^2$ up to a multiplicative positive constant and (2) the function $d \mapsto \mathrm{E}(d)$ tends to zero in probability, uniformly in d. Proposition 6 will prove (1) and Corollary 8 will yield (2).

PROPOSITION 9 (Weak consistency). *Let $\{L_n\}$ be a sequence such that $L_n^{-1} + (n2^{-L_n})^{-1} \to 0$ as $n \to \infty$. Condition 1 implies that as $n \to \infty$,*

$$\sup_{j_1 = L_n+1, \ldots, J_n} \left|\hat{d}_{\mathcal{I}_n(L_n, j_1)} - d_0\right| = O_{\mathbb{P}}\{(n2^{-L_n})^{-1/4} + 2^{-\beta L_n/2}\}. \tag{56}$$

PROOF. The proof proceeds in four steps.

Step 1. For any positive integer ℓ, $|\hat{d}_{\mathcal{I}_n(L_n, L_n+\ell)} - d_0| = o_{\mathbb{P}}(1)$.

Step 2. There exists $d_{\min} \in (d_0 - 1/2, d_0)$ such that, as $n \to \infty$,

$$\mathbb{P}\left\{\inf_{j_1 = L_n+2, \ldots, J_n} \hat{d}_{\mathcal{I}_n(L_n, j_1)} \leq d_{\min}\right\} \to 0.$$

Combining this with Step 1 yields $\mathbb{P}\{\inf_{j_1 = L_n+1, \ldots, J_n} \hat{d}_{\mathcal{I}_n(L_n, j_1)} \leq d_{\min}\} \to 0$.

Step 3. For any $d_{\max} > d_0$, as $n \to \infty$, $\mathbb{P}\{\sup_{j_1 = L_n+1, \ldots, J_n} \hat{d}_{\mathcal{I}_n(L_n, j_1)} \geq d_{\max}\} \to 0$.

Step 4. Define $H_{0,\delta}$ as in Proposition 7. For all $d_{\min} \in (d_0 - 1/2, d_0)$ and $d_{\max} > d_0$, setting $\delta = 2(d_0 - d_{\min})$, we have

$$\sup_{j_1 = L_n+1,\dots,J_n} \left[\mathbb{1}_{[d_{\min},d_{\max}]}(\hat{d}_{\mathcal{I}_n(L_n,j_1)})(\hat{d}_{\mathcal{I}_n(L_n,j_1)} - d_0)^2\right] = O_{\mathbb{P}}(H_{0,\delta}(n2^{-L_n}) + 2^{-\beta L_n}).$$

Before proving these four steps, let us briefly explain how they yield (56). First, observe that they imply that $\sup_{j_1=L_n+1,\dots,J_n} |\hat{d}_{\mathcal{I}_n(L_n,j_1)} - d_0| = o_{\mathbb{P}}(1)$. Then, applying Step 4 again with $d_{\min} \in (d_0 - 1/4, d_0)$, so that $H_{0,\delta}(x) = x^{-1/2}$, we obtain (56). □

PROOF OF STEP 1. Using standard arguments for contrast estimation (similar to those detailed is Step 3 and Step 4 below), this step is a direct consequence of Proposition 6 and Corollary 8(a). □

PROOF OF STEP 2. Using (20), we have, for all $d \in \mathbb{R}$,

$$\tilde{\mathrm{L}}_{\mathcal{I}}(d) - \tilde{\mathrm{L}}_{\mathcal{I}}(d_0) = \log\left(\sum_{(j,k)\in\mathcal{I}} 2^{2(d-d_0)(\langle\mathcal{I}\rangle - j)} \frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}}\right) - \log\left(\sum_{(j,k)\in\mathcal{I}} \frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}}\right).$$

For some $d_{\min} \in (d_0 - 1/2, d_0)$ to be specified later, we set

$$w_{\mathcal{I},j}(d) \overset{\text{def}}{=} 2^{2(j-\langle\mathcal{I}\rangle)(d_0-d)} \mathbb{1}\{j \le \langle\mathcal{I}\rangle\} + 2^{2(j-\langle\mathcal{I}\rangle)(d_0-d_{\min})} \mathbb{1}\{j > \langle\mathcal{I}\rangle\}, \tag{57}$$

so that for all j and $d \le d_{\min}$, $w_{\mathcal{I},j}(d) \le 2^{2(d-d_0)(\langle\mathcal{I}\rangle - j)}$. We further obtain, for all $d \le d_{\min}$,

$$\tilde{\mathrm{L}}_{\mathcal{I}}(d) - \tilde{\mathrm{L}}_{\mathcal{I}}(d_0) \ge \log \frac{\Sigma_{\mathcal{I}}(d) + A_{\mathcal{I}}(d)}{1 + B_{\mathcal{I}}}, \tag{58}$$

where $\Sigma_{\mathcal{I}}(d) \overset{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k)\in\mathcal{I}} w_{\mathcal{I},j}(d)$, $A_{\mathcal{I}}(d) \overset{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k)\in\mathcal{I}} w_{\mathcal{I},j}(d) \times (\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1)$ and $B_{\mathcal{I}} \overset{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k)\in\mathcal{I}} (\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1)$. We will show that $d_{\min} \in (d_0 - 1/2, d_0)$ may be chosen in such a way that

$$\liminf_{n\to\infty} \inf_{d \le d_{\min}} \inf_{j_1 = L_n+2,\dots,J_n} \Sigma_{\mathcal{I}_n(L_n,j_1)}(d) > 1, \tag{59}$$

$$\sup_{j_1 = L_n+2,\dots,J_n} \left(\sup_{d \le d_{\min}} |A_{\mathcal{I}_n(L_n,j_1)}(d)| + |B_{\mathcal{I}_n(L_n,j_1)}|\right) = o_{\mathbb{P}}(1). \tag{60}$$

By (55), $\tilde{\mathrm{L}}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) \le \tilde{\mathrm{L}}_{\mathcal{I}}(d_0)$. Then, $\inf_{j_1=L_n+2,\dots,J_n} \hat{d}_{\mathcal{I}_n(L_n,j_1)} \le d_{\min}$ would imply that there exists $j_1 = L_n + 2, \dots, J_n$ such that $\inf_{d \le d_{\min}} \tilde{\mathrm{L}}_{\mathcal{I}_n(L_n,j_1)}(d) - \tilde{\mathrm{L}}(d_0) \le 0$, an event whose probability tends to zero as a consequence of (58)–(60). Hence, these equations yield Step 2. It thus remains to show that (59) and (60) hold. By Lemma 13, since $n2^{-L_n} \to \infty$, we have, for n large enough,

$$\sup_{j_1=L_n,\dots,J_n} \langle \mathcal{I}_n(L_n, j_1)\rangle < L_n + 1. \tag{61}$$

Using $w_{\mathcal{I}_n(L_n,j_1),L_n}(d) \geq 0$ and, for n large enough, $w_{\mathcal{I}_n(L_n,j_1),j}(d) \geq 2^{2(j-(L_n+1))(d_0-d_{\min})}$, for $j \geq L_n + 1$, we get, for all $d \leq d_{\min} < d_0$ and $j_1 = L_n + 2, \dots, J_n$,

$$\Sigma_{\mathcal{I}_n(L_n,j_1)}(d) \geq \frac{2^{-2(L_n+1)(d_0-d_{\min})}}{|\mathcal{I}_n(L_n, J_n)|} \sum_{j=L_n+1}^{L_n+2} 2^{2j(d_0-d_{\min})} n_j.$$

Since $n2^{-L_n} \to \infty$, using Lemma 13, $n \asymp 2^{J_n}$ and the fact that $2(d_0 - d_{\min}) - 1 < 0$, straightforward computations give that the LHS in the previous display is asymptotically equivalent to $(1 - 2^{\{2(d_0-d_{\min})-1\}2})/(4 - 2^{2(d_0-d_{\min})+1})$. There are values of $d_{\min} \in (d_0 - 1/2, d_0)$ such that this ratio is strictly larger than 1. For such a choice and for n large enough, (59) holds.

We now check (60). Observing that, for $\mathcal{I}_n \stackrel{\text{def}}{=} \mathcal{I}_n(L_n, j_1)$ and using the notation (51), $A_{\mathcal{I}_n}(d) = |\mathcal{I}_n|^{-1}|\tilde{\mathbf{S}}_{n,L_n,j_1}(\{w_{\mathcal{I}_n,L_n+j}(d)\})|$ and $B_{\mathcal{I}_n} = |\mathcal{I}_n|^{-1}\tilde{\mathbf{S}}_{n,L_n,j_1}(\mathbb{1})$, the bound (60) follows from $|\mathcal{I}_n| \geq n_{L_n} \sim n2^{-L_n}$ and Proposition 7 since, for all $d \leq d_{\min}$ and $j \geq 0$, $w_{\mathcal{I}_n,L_n+j}(d) \leq 2^{2(L_n+j-\langle \mathcal{I}_n\rangle)(d_0-d_{\min})} \leq 2^{2j(d_0-d_{\min})}$, which shows that $\{w_{\mathcal{I}_n,L_n+j}(d)\}_{j\geq 0}$ belongs to $\mathcal{B}(1, 0, \delta)$ with $\delta = 2(d_0 - d_{\min})) < 1$. □

PROOF OF STEP 3. By (55), $\mathrm{L}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) \leq \mathrm{E}_{\mathcal{I}}(d_0) - \mathrm{E}_{\mathcal{I}}(\hat{d}_{\mathcal{I}})$, so, for any $d_{\max} \geq d_0$, one has $\inf_{d\geq d_{\max}} \mathrm{L}_{\mathcal{I}}(d) \leq 2\sup_{d\geq d_0} |\mathrm{E}_{\mathcal{I}}(d)|$ on the event $\{\hat{d}_{\mathcal{I}} \geq d_{\max}\}$. By Proposition 6, there exists $c > 0$ such that, for n large enough, $\mathrm{L}_{\mathcal{I}_n(L_n,j_1)}(d) \geq c$ uniformly for $d \geq d_{\max}$ and $j_1 = L_n + 1, \dots, J_n$. Thus, for n large enough,

$$\mathbb{P}\Big\{\sup_{j_1=L_n+1,\dots,J_n} \hat{d}_{\mathcal{I}_n(L_n,j_1)} \geq d_{\max}\Big\} \leq \mathbb{P}\Big\{2\sup_{d\geq d_0}\sup_{j_1=L_n+1,\dots,J_n} \big|\mathrm{E}_{\mathcal{I}_n(L_n,j_1)}(d)\big| \geq c\Big\},$$

which tends to 0 as $n \to \infty$, by Corollary 8(b). □

PROOF OF STEP 4. Equation (55) implies that $\mathbb{1}_{[d_{\min},d_{\max}]}(\hat{d}_{\mathcal{I}})\mathrm{L}_{\mathcal{I}}(\hat{d}_{\mathcal{I}}) \leq 2\sup_{d\geq d_{\min}} |\mathrm{E}_{\mathcal{I}}(d)|$. Let c denote the liminf in the left-hand side of (49) when $d_{\min} = d_{\min}$ and $d_{\max} = d_{\max}$. Proposition 6 and a second order Taylor expansion of $\mathrm{L}_{\mathcal{I}}$ around d_0 give that, for n large enough, for all $j_1 = L_n + 1, \dots, J_n$ and $d \in [d_{\min}, d_{\max}]$, $\mathrm{L}_{\mathcal{I}_n(L_n,j_1)}(d) \geq (c/4)(d - d_0)^2$. Hence, for n large enough,

$$\sup_{j_1=L_n+1,\dots,J_n} \big[\mathbb{1}_{[d_{\min},d_{\max}]}\big(\hat{d}_{\mathcal{I}_n(L_n,j_1)}\big)\big(\hat{d}_{\mathcal{I}_n(L_n,j_1)} - d_0\big)^2\big] \leq \frac{8}{c}\sup_{d\geq d_{\min}} \big|\mathrm{E}_{\mathcal{I}_n(L_n,j_1)}(d)\big|.$$

Corollary 8(b) then yields Step 4. □

REMARK 11. Proposition 9 implies that if $L_n \le U_n \le J_n$ with $L_n^{-1} + (n2^{-L_n})^{-1} \to 0$ as $n \to \infty$, then $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$ is a consistent estimator of d_0. While the rate provided by (56) is not optimal, it will be used to derive the optimal rates of convergence (Theorem 3).

8. Proofs of Theorems 3 and 5.

NOTATIONAL CONVENTION. In the following, $\{L_n\}$ and $\{U_n\}$ are two sequences satisfying (23). The only difference between the two following settings (S-1) (where $U_n - L_n$ is fixed) and (S-2) (where $U_n - L_n \to \infty$) lies in the computations of the asymptotic variances in Theorem 5 (CLT). Hence, we shall hereafter write L, U, $\mathcal{I}_n$, $\hat{d}_n$, $\widehat{\mathbf{S}}_n$ and $\tilde{\mathbf{S}}_n$ for L_n, U_n, $\mathcal{I}_n(L_n, U_n)$, $\hat{d}_{\mathcal{I}_n(L_n,U_n)}$, $\widehat{\mathbf{S}}_{\mathcal{I}_n(L_n,U_n)}$ and $\tilde{\mathbf{S}}_{n,L_n,U_n}$, respectively.

We will use the explicit notation when the distinction between these two cases (S-1) and (S-2) is necessary, namely, when computing the limiting variances in the proof of Theorem 5.

PROOF OF THEOREM 3. Since $\widehat{\mathbf{S}}_n(\hat{d}_n) = 0$ [see (21)] a Taylor expansion of $\widehat{\mathbf{S}}_n$ around $d = \hat{d}_n$ yields

$$\widehat{\mathbf{S}}_n(d_0) = 2\log(2)(\hat{d}_n - d_0) \sum_{(j,k)\in\mathcal{I}_n} (j - \langle\mathcal{I}_n\rangle) j 2^{-2j\tilde{d}_n} W_{j,k}^2 \tag{62}$$

for some $\tilde{d}_n$ between d_0 and $\hat{d}_n$. The proof of Theorem 3 now consists of bounding $\widehat{\mathbf{S}}_n(d_0)$ from above and showing that $\sum_{\mathcal{I}_n} (j - \langle\mathcal{I}_n\rangle) j 2^{-2j\tilde{d}_n} W_{j,k}^2$, appropriately normalized, has a strictly positive limit.

By the definitions of $\widehat{\mathbf{S}}_n$ [see (21)], $\tilde{\mathbf{S}}_n$ [see (51)] and $\langle\mathcal{I}_n\rangle$ [see (19)], we have $\widehat{\mathbf{S}}_n(d_0) = \tilde{\mathbf{S}}_n(\sigma^2\{j + L - \langle\mathcal{I}_n\rangle\}_{j\ge 0})$. Since $L \le \langle\mathcal{I}_n\rangle \le L + 1$ for n large enough [see (61)] the sequence $\sigma^2\{j + L - \langle\mathcal{I}_n\rangle\}_{j\ge 0}$ belongs to $\mathcal{B}(\sigma^2, 1, 0)$ [see (50)], and Proposition 7, together with the Markov inequality, yields, as $n \to \infty$,

$$\begin{aligned} \widehat{\mathbf{S}}_n(d_0) &= n2^{-L} O_{\mathbb{P}}\big(H_{1,0}(n2^{-L}) + 2^{-\beta L}\big) \\ &= n2^{-L} O_{\mathbb{P}}\big((n2^{-L})^{-1/2} + 2^{-\beta L}\big), \end{aligned} \tag{63}$$

which is the desired upper bound.

We shall now show that the sum in (62) multiplied by $n2^{-L}$ has a strictly positive lower bound. Applying Proposition 9, we have

$$|\tilde{d}_n - d_0| \le |\hat{d}_n - d_0| = O_{\mathbb{P}}\big((n2^{-L})^{-1/4} + 2^{-\beta L/2}\big).$$

Using the fact that $|2^{2j(d_0-\tilde{d}_n)}-1|\le 2^{2j|d_0-\tilde{d}_n|}-1\le 2\log(2)j|d_0-\tilde{d}_n|2^{2j|d_0-\tilde{d}_n|}$, we have that, on the event $\{|d_0-\tilde{d}_n|\le 1/4\}$,

$$\left|\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2\tilde{d}_n j}}-\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2d_0 j}}\right| \le 2\log(2)|d_0-\tilde{d}_n|2^{2L|d_0-\tilde{d}_n|}\sum_{(j,k)\in\mathcal{I}_n}|j-\langle\mathcal{I}_n\rangle|j^2\frac{W_{j,k}^2}{2^{2d_0 j}}2^{(j-L)/2}.$$

Using (8), (61), $j^2=(j-L)^2+2(j-L)L+L^2$ and $n_j\le n2^{-j}$, there is a constant $C>0$ such that

$$\mathbb{E}\sum_{(j,k)\in\mathcal{I}_n}|j-\langle\mathcal{I}_n\rangle|j^2\frac{W_{j,k}^2}{2^{2d_0 j}}2^{(j-L)/2} \le Cn2^{-L}\sum_{j=L}^{U}|j-\langle\mathcal{I}_n\rangle|j^2 2^{-(j-L)/2}=O(L^2n2^{-L}).$$

Hence, since $L^2(n2^{-L})^{-1/4}\to 0$, the last three displays yield, as $n\to\infty$,

$$\left|\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2\tilde{d}_n j}}-\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2d_0 j}}\right|=o_{\mathbb{P}}(n2^{-L}). \tag{64}$$

We now write

$$\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2d_0 j}} =\sigma^2\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}}-1\right)+\sigma^2\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j.$$

With the notation (51), the first term on the right-hand side is $\tilde{\mathbf{S}}_n(\boldsymbol{\mu})$, where $\boldsymbol{\mu}$ is the sequence $\sigma^2\{(j+L-\langle\mathcal{I}_n\rangle)(j+L)\}_{j\ge 0}$. In view of (61), $(j+L-\langle\mathcal{I}_n\rangle)(j+L)\le j^2+jL$, so the sequence $\boldsymbol{\mu}$ is the sum of two sequences belonging to $\mathcal{B}(\sigma^2,2,0)$ and $\mathcal{B}(\sigma^2 L,1,0)$, respectively. Applying Proposition 7 together with the Markov inequality, we get that our $\tilde{\mathbf{S}}_n(\boldsymbol{\mu})=n2^{-L}O_{\mathbb{P}}(H_{0,0}(n2^{-L})+LH_{1,0}(n2^{-L}))=n2^{-L}o_{\mathbb{P}}(1)$ since $L(n2^{-L})^{-1/2}\to 0$. Moreover, by Lemma 13, $\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\sim(n2^{-L})(2-2^{-(U-L)})\kappa_{U-L}$ as $n\to\infty$. Hence,

$$\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2d_0 j}}=(n2^{-L})\{(2-2^{-(U-L)})\kappa_{U-L}+o_{\mathbb{P}}(1)\},$$

and (64) and the previous display yield

$$\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)j\frac{W_{j,k}^2}{2^{2\tilde{d}_n j}} = (n2^{-L})\{\sigma^2(2-2^{-(U-L)})\kappa_{U-L}+o_{\mathbb{P}}(1)\}. \tag{65}$$

Since $\kappa_\ell > 0$ for all $\ell \geq 1$ and $\kappa_\ell \to 2$ as $\ell \to \infty$ (see Lemma 13), and since we assumed $U - L \geq 1$, the sequence $(2-2^{-(U-L)})\kappa_{U-L}$ is bounded below by a positive constant, so (62), (63) and (65) imply (25). □

PROOF OF THEOREM 5. Define $f^*(0) \stackrel{\text{def}}{=} d\nu^*/d\lambda_{|\lambda=0}$. Since $\nu^* \in \mathcal{H}(\beta,\gamma,\varepsilon)$ and $\nu^*(-\varepsilon,\varepsilon) > 0$, we have $f^*(0) > 0$. Without loss of generality, we set $f^*(0) = 1$. By Corollary 2, conditions (8) and (9) hold with $\sigma^2 = \mathrm{K}(d_0)$. Moreover, (31) implies that $L^{-1} + L^2(n2^{-L})^{-1/4} \to 0$, so we may apply (65), which, with (62), gives

$$(n2^{-L})^{1/2}(\hat{d}_n - d_0) = \frac{(n2^{-L})^{-1/2}\widehat{\mathbf{S}}_n(d_0)}{2\log(2)\sigma^2(2-2^{-(U-L)})\kappa_{U-L}}(1+o_{\mathbb{P}}(1)). \tag{66}$$

Define $\overline{\mathbf{S}}_n$ as $\widehat{\mathbf{S}}_n$ in (21), but with the wavelet coefficients $\overline{W}_{j,k}$ defined in Corollary 2 replacing the wavelet coefficients $W_{j,k}$. Let us write

$$\widehat{\mathbf{S}}_n(d_0) = (\widehat{\mathbf{S}}_n(d_0) - \overline{\mathbf{S}}_n(d_0)) + \mathbb{E}_f[\overline{\mathbf{S}}_n(d_0)] + (\overline{\mathbf{S}}_n(d_0) - \mathbb{E}_f[\overline{\mathbf{S}}_n(d_0)]). \tag{67}$$

By Corollary 2, using Minkowski's and Markov's inequalities, (61), $n_j \leq n2^{-j}$ and $d_0 + \alpha > (1+\beta)/2$, we obtain, as $n \to \infty$,

$$\widehat{\mathbf{S}}_n(d_0) - \overline{\mathbf{S}}_n(d_0) = o_{\mathbb{P}}((n2^{-L})^{1/2}).$$

Since $\sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle) = 0$ and $\mathbb{E}_f[W_{j,k}^2] = \sigma_j^2(\nu)$, we may write

$$\begin{aligned}\mathbb{E}_f[\widehat{\mathbf{S}}_n(d_0)] &= \sum_{(j,k)\in\mathcal{I}_n}(j-\langle\mathcal{I}_n\rangle)(2^{-2d_0 j}\sigma_j^2(\nu) - \sigma^2) \\ &= O(n2^{-(1+\beta)L}) = o((n2^{-L})^{1/2}),\end{aligned}$$

where the O-term follows from (8), (61) and $n_j \leq n2^{-j}$ and the o-term follows from (31). Using (66), (67) and the two last displays, we finally get that

$$(n2^{-L})^{1/2}(\hat{d}_n - d_0) = \frac{(n2^{-L})^{-1/2}(\overline{\mathbf{S}}_n(d_0) - \mathbb{E}_f[\overline{\mathbf{S}}_n(d_0)])}{2\log(2)\sigma^2(2-2^{-(U-L)})\kappa_{U-L}}(1+o_{\mathbb{P}}(1)).$$

Because $\overline{f}(\lambda) = |1-e^{-i\lambda}|^{-2d_0}[f^*\mathbb{1}_{[-\epsilon,\epsilon]}](\lambda)$ and $f^*\mathbb{1}_{[-\epsilon,\epsilon]} \in \mathcal{H}(\beta,\gamma',\pi)$ for some $\gamma' > 0$, we may apply Proposition 10 below to determine the asymptotic behavior of $\overline{\mathbf{S}}_n(d_0) - \mathbb{E}_f[\overline{\mathbf{S}}_n(d_0)]$ as $n \to \infty$. Since $\sigma^2 = f^*(0)\mathrm{K}(d_0)$ (Theorem 1), this yields the result and completes the proof. □

The following proposition provides a CLT when the condition on ν^* is global, namely $\nu^* \in \mathcal{H}(\beta, \gamma, \pi)$. It covers the cases (S-1), where $U - L \to \ell < \infty$ and (S-2), where $U - L \to \infty$.

PROPOSITION 10. *Let X be a Gaussian process having generalized spectral measure (2) with $d_0 \in \mathbb{R}$ and $\nu^* \in \mathcal{H}(\beta, \gamma, \pi)$, with $f^*(0) \stackrel{\text{def}}{=} d\nu^*/d\lambda_{|\lambda=0} > 0$, where $\gamma > 0$ and $\beta \in (0, 2]$. Let L and U be two sequences satisfying (23) and suppose that $L^{-1} + (n2^{-L})^{-1} \to 0$ and $U - L \to \ell \in \{1, 2, \ldots, \infty\}$ as $n \to \infty$. Then, as $n \to \infty$,*

$$\text{(68)} \qquad \frac{(n2^{-L})^{-1/2}\{\widehat{\mathbf{S}}_{\mathcal{I}_n(L,U)}(d_0) - \mathbb{E}_f[\widehat{\mathbf{S}}_{\mathcal{I}_n(L,U)}(d_0)]\}}{2\log(2) f^*(0)\mathrm{K}(d_0)(2 - 2^{-(U-L)})\kappa_{U-L}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathrm{V}(d_0, \ell)),$$

where κ_k is defined in (28) and $\mathrm{V}(d_0, \ell)$ in (29) for $\ell < \infty$ and $\mathrm{V}(d_0, \infty)$ in (30).

PROOF. We take $f^*(0) = 1$, without loss of generality. As $n \to \infty$, since $U - L \to \ell$, we have $\kappa_{U-L} \to \kappa_\ell$, by setting, in the special case where $\ell = \infty$, $\kappa_\infty = 2$; see Lemma 13. This gives the deterministic limit of the denominator in (68). The limit distribution of the numerator is obtained by applying Lemma 12 below. Let A_n and Γ_n be the square matrices indexed by the pairs $(j, k), (j, k) \in \mathcal{I}_n \times \mathcal{I}_n$ (in lexicographic order) and defined as follows:

(1) A_n is the diagonal matrix such that $[A_n]_{(j,k),(j,k)} = (n2^{-L})^{-1/2}\operatorname{sign}(j - \langle \mathcal{I}_n \rangle)$ for all $(j, k) \in \mathcal{I}_n$;
(2) Γ_n is the covariance matrix of the vector $[|j - \langle \mathcal{I}_n \rangle|^{1/2} 2^{-d_0 j} W_{j,k}]_{(j,k) \in \mathcal{I}_n}$.

Let $\rho(A)$ denote the spectral radius of the square matrix A, that is, the maximum of the absolute value of its eigenvalues. Of course, $\rho[A_n] = (n2^{-L})^{-1/2}$. Moreover, $\rho[\Gamma_n] \le \sum_{j=L}^{U} \rho[\Gamma_{n,j}]$, where $\Gamma_{n,j}$ is the covariance matrix of the vector $[|j - \langle \mathcal{I}_n \rangle|^{1/2} 2^{-d_0 j} W_{j,k}]_{k=0,\ldots,n_j-1}$. Since $\{W_{j,k}\}_{k \in \mathbb{Z}}$ is a stationary time series, by Lemma 11,

$$\rho[\Gamma_{n,j}] \le |j - \langle \mathcal{I}_n \rangle| 2^{-2d_0 j} 2\pi \sup_{\lambda \in (-\pi,\pi)} \mathbf{D}_{j,0}(\lambda; \nu).$$

From (10), since $\mathbf{D}_{\infty,0}(\cdot; d_0)$ is bounded on $(-\pi, \pi)$, we get, for a constant C not depending on n, $\rho[\Gamma_n] \le C \sum_{j=L}^{U} |j - \langle \mathcal{I}_n \rangle|$. By (61), the latter sum is $O((U - L)^2)$. Hence, as $n \to \infty$, since $U - L \le J_n - L = O(\log(n2^{-L}))$, we have $\rho[A_n]\rho[\Gamma_n] = O((n2^{-L})^{-1/2}(U - L)^2) \to 0$, so the conditions of Lemma 12 are met, provided that $(n2^{-L})^{-1} \operatorname{Var}(\widehat{\mathbf{S}}_n(d_0))$ has a finite limit.

To conclude the proof, we need to compute this limit. In [14], Proposition 2, it is shown that for all $u = 0, 1, \ldots,$ as $j \to \infty$ and $n_j \to \infty$,

$$\text{(69)} \qquad c_n(j, u) \stackrel{\text{def}}{=} 2^{-4d_0 j} n_{j-u} \operatorname{Cov}(\hat{\sigma}_j^2, \hat{\sigma}_{j-u}2) \to 4\pi \mathrm{I}_u(d_0),$$

where $\mathrm{I}_u(d)$ is defined in (27) and $\hat{\sigma}_j^2 \stackrel{\text{def}}{=} \frac{1}{n_j}\sum_{k=0}^{n_j-1} W_{j,k}^2$. Since $\widehat{\mathbf{S}}_n(d_0) = \sum_{j=L}^{U}(j - \langle \mathcal{I}_n \rangle)2^{-2jd_0} n_j \hat{\sigma}_j^2$, we obtain

$$
\begin{aligned}
&(n2^{-L})^{-1}\operatorname{Var}(\widehat{\mathbf{S}}_n(d_0)) \\
&\quad = \sum_{i=0}^{U-L} (i + L - \langle \mathcal{I}_n \rangle)^2 2^{-i} \frac{n_{L+i}}{n2^{-(L+i)}} c_n(L+i, 0) \\
&\qquad + 2\sum_{i=1}^{U-L}\sum_{u=1}^{i} (i + L - \langle \mathcal{I}_n \rangle)(i - u + L - \langle \mathcal{I}_n \rangle) \\
&\qquad\qquad \times 2^{2d_0 u - i} \frac{n_{L+i}}{n2^{-(L+i)}} c_n(L+i, u).
\end{aligned} \tag{70}
$$

By the Cauchy–Schwarz inequality, (45), (8) and $n_{j-u} \asymp n_j 2^{-u}$ imply that $|c_n(j,u)| \le C 2^{-2d_0 u + u/2}$, where C is a positive constant. Using this bound, (61) and $n_j \le n2^{-j}$ for bounding the terms of the two series in the right-hand side of (70) yields the following convergent series: $\sum_{i=0}^{\infty}(i+1)^2 2^{-i}$ and $\sum_{i=1}^{\infty}\sum_{u=1}^{i}(i+1)(i-u+1)2^{-i+u/2}$. Using the assumptions on U and L, we have $n_{L+i} \sim n2^{-(L+i)}$ for any $i \ge 0$ and by Lemma 13, $\langle \mathcal{I}_n \rangle - L \to \eta_\ell$ as $n \to \infty$. Hence, by dominated convergence, (70) and (69) finally give that, as $n \to \infty$, $(n2^{-L})^{-1}\operatorname{Var}(\widehat{\mathbf{S}}_n(d_0))$ converges to

$$
4\pi\left[\mathrm{I}_0(d_0)\kappa_l(2 - 2^{-\ell}) + 2\sum_{1 \le u \le i \le \ell} (i - \eta_\ell)(i - \eta_\ell - u)2^{2d_0 u - i}\mathrm{I}_u(d_0)\right], \tag{71}
$$

where in the case $\ell = \infty$, we have set $2^{-\infty} = 0$, $\eta_\infty = 1$ and $\kappa_\infty = 2$. Note that the above bound on $|c_n(j,u)|$ and (69) imply that as $u \to \infty$,

$$
\mathrm{I}_u(d_0) = O(2^{-2d_0 u + u/2}), \tag{72}
$$

which confirms that the series in (71) is convergent for $\ell = \infty$. Finally, dividing this variance by the squared limit of the denominator in (68), we get the limit variance in (68), namely (29) and (30). □

The following lemmas were used in the proof of Proposition 10.

LEMMA 11. *Let* $\{\xi_\ell, \ell \in \mathbb{Z}\}$ *be a stationary process with spectral density* g *and let* Γ_n *be the covariance matrix of* $[\xi_1, \ldots, \xi_n]$. *Then,* $\rho(\Gamma_n) \le 2\pi \|g\|_\infty$.

LEMMA 12. *Let* $\{\xi_n, n \ge 1\}$ *be a sequence of Gaussian vectors with zero mean and covariance* Γ_n. *Let* $(A_n)_{n\ge1}$ *be a sequence of deterministic symmetric matrices such that* $\lim_{n\to\infty}\operatorname{Var}(\xi_n^T A_n \xi_n) = \sigma^2 \in [0,\infty)$. *Assume that* $\lim_{n\to\infty}[\rho(A_n)\rho(\Gamma_n)] = 0$. *Then,* $\xi_n^T A_n \xi_n - \mathbb{E}[\xi_n^T A_n \xi_n] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$.

PROOF. The result is obvious if $\sigma = 0$, hence we may assume $\sigma > 0$. Let $n \geq 1$, k_n be the rank of Γ_n and Q_n denote an $n \times k_n$ full-rank matrix such that $Q_n Q_n^T = \Gamma_n$. Let $\zeta_n \sim \mathcal{N}(0, I_{k_n})$, where I_k is the identity matrix of size $k \times k$. Then, for any $k_n \times k_n$ unitary matrix U_n, $U_n \zeta_n \sim \mathcal{N}(0, I_{k_n})$ and hence $Q_n U_n \zeta_n$ has the same distribution as ξ_n. Moreover, since A_n is symmetric, so is $Q_n^T A_n Q_n$. Choose U_n to be a unitary matrix such that $\Lambda_n \stackrel{\text{def}}{=} U_n^T (Q_n^T A_n Q_n) U_n$ is a diagonal matrix. Thus, $\zeta_n^T \Lambda_n \zeta_n = (Q_n U_n \zeta_n)^T A_n (Q_n U_n \zeta_n)$ has the same distribution as $\xi_n^T A_n \xi_n$. Since Λ_n is diagonal, $\zeta_n^T \Lambda_n \zeta_n$ is a sum of independent r.v.'s of the form $\sum_{k=1}^{k_n} \lambda_{k,n} \zeta_{k,n}^2$, where $(\zeta_{1,n}, \dots, \zeta_{k_n,n})$ are independent centered unit-variance Gaussian r.v.'s and $\lambda_{k,n}$ are the diagonal entries of Λ_n. Note that $\sum_{k=1}^{k_n} \lambda_{k,n} = \mathbb{E}[\xi_n^T A_n \xi_n]$. To check the asymptotic normality, we verify that the Lindeberg conditions hold for the sum of centered independent r.v.'s: $\xi_n^T A_n \xi_n - \mathbb{E}[\xi_n^T A_n \xi_n] = \sum_{k=1}^{k_n} \lambda_{k,n} (\zeta_{k,n}^2 - 1)$. Under the stated assumptions,

$$\sum_{k=1}^{k_n} \lambda_{k,n}^2 \mathbb{E}(\zeta_{k,n}^2 - 1)^2 = \operatorname{Var}(\xi_n^T A_n \xi_n) \to \sigma^2 \qquad \text{as } n \to \infty$$

and $\rho(\Lambda_n) = \rho(Q_n^T A_n Q_n) \leq \rho(A_n) \sup_{\|x\|=1} \|Q_n x\|^2 = \rho(A_n)\rho(\Gamma_n) \to 0$. Since $\rho(\Lambda_n) = \max_{1 \leq k \leq k_n} |\lambda_{k,n}|$, for all $\epsilon > 0$,

$$\begin{aligned}
&\sum_{k=1}^{k_n} \lambda_{k,n}^2 \mathbb{E}\big[(\zeta_{k,n}^2 - 1)^2 \mathbb{1}\big(|\lambda_{k,n}(\zeta_{k,n}^2 - 1)| \geq \epsilon\big)\big] \\
&\qquad \leq \left(\sum_{k=1}^{k_n} \lambda_{k,n}^2\right) \mathbb{E}\big[(\zeta_{1,n}^2 - 1)^2 \mathbb{1}\big(\rho(\Lambda_n)|\zeta_{1,n}^2 - 1| \geq \epsilon\big)\big] \to 0 \qquad \text{as } n \to \infty.
\end{aligned}$$

Hence, the Lindeberg conditions hold provided $\sigma > 0$. □

LEMMA 13. *Let $p, \ell \geq 0$, η_ℓ and κ_ℓ be defined as in* (28), $\langle \mathcal{I} \rangle$ *as in* (19) *and*

$$\mathcal{J}(\mathcal{I}) \stackrel{\text{def}}{=} |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} (j - \langle \mathcal{I} \rangle)^2 = |\mathcal{I}|^{-1} \sum_{(j,k) \in \mathcal{I}} j(j - \langle \mathcal{I} \rangle).$$

We have

$$\eta_\ell = \frac{1 - 2^{-\ell}(1 + \ell/2)}{1 - 2^{-(\ell+1)}} \in (0, 1), \qquad \lim_{\ell \to \infty} \eta_\ell = 1, \qquad \lim_{\ell \to \infty} \kappa_\ell = 2, \tag{73}$$

$$\textit{for all } u \geq 0 \qquad \lim_{\ell \to \infty} \frac{1}{\kappa_\ell} \sum_{i=0}^{\ell-u} \frac{2^{-i}}{2 - 2^{-\ell}} (i - \eta_\ell)(i + u - \eta_\ell) = 1 \tag{74}$$

and for all $n \geq 1$ and $0 \leq j_0 \leq j_1 \leq J_n$,

$$\left| \sum_{j=j_0}^{j_1} (j - j_0)^p n_j - n 2^{-j_0} \sum_{i=0}^{j_1 - j_0} i^p 2^{-i} \right| \leq 2(\mathrm{T} - 1)(j_1 - j_0)^{p+1}. \tag{75}$$

Moreover, if $0 \le L_n \le J_n$ *with* $n2^{-L_n} \to \infty$ *as* $n \to \infty$, *then*

$$\sup_{j_1=L_n,\dots,J_n} \big||\mathcal{I}_n(L_n, j_1)| - n2^{-L_n}(2 - 2^{-(j_1-L_n)})\big| = O(\log(n2^{-L_n})),$$

$$\sup_{j_1=L_n,\dots,J_n} |\langle \mathcal{I}_n(L_n, j_1)\rangle - L_n - \eta_{j_1-L_n}| = O(\log^2(n2^{-L_n})(n2^{-L_n})^{-1}),$$

$$\sup_{j_1=L_n,\dots,J_n} |\mathcal{J}[\mathcal{I}_n(L_n, j_1)] - \kappa_{j_1-L_n}| = O(\log^3(n2^{-L_n})(n2^{-L_n})^{-1}).$$

Acknowledgments. We would like to thank the referees and the Associate Editor for their helpful comments. Murad S. Taqqu would like to thank l'École Normale Supérieure des Télecommunications in Paris for their hospitality.

REFERENCES

[1] ABRY, P. and VEITCH, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inform. Theory* **44** 2–15. MR1486645

[2] BOYD, S. and VANDENBERGHE, L. (2004). *Convex Optimization*. Cambridge Univ. Press. MR2061575

[3] COHEN, A. (2003). *Numerical Analysis of Wavelet Methods*. North-Holland, Amsterdam. MR1990555

[4] DAUBECHIES, I. (1992). *Ten Lectures on Wavelets*. SIAM, Philadelphia. MR1162107

[5] FAŸ, G., ROUEFF, F. and SOULIER, P. (2007). Estimation of the memory parameter of the infinite-source Poisson process. *Bernoulli*. **13** 473–491. MR2331260

[6] GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–238. MR0738585

[7] GIRAITIS, L., ROBINSON, P. M. and SAMAROV, A. (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* **18** 49–61. MR1437741

[8] HURVICH, C. M., MOULINES, E. and SOULIER, P. (2002). The FEXP estimator for potentially nonstationary linear time series. *Stoch. Proc. App.* **97** 307–340. MR1875337

[9] HURVICH, C. M. and RAY, B. K. (1995). Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. *J. Time Ser. Anal.* **16** 17–41. MR1323616

[10] KAPLAN, L. M. and KUO, C.-C. J. (1993). Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis. *IEEE Trans. Signal Process.* **41** 3554–3562.

[11] KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Probability Theory and Applications. Proc. World Congr. Bernoulli Soc.* **1** 67–74. VNU Sci. Press, Utrecht. MR1092336

[12] MCCOY, E. J. and WALDEN, A. T. (1996). Wavelet analysis and synthesis of stationary long-memory processes. *J. Comput. Graph. Statist.* **5** 26–56. MR1380851

[13] MOULINES, E., ROUEFF, F. and TAQQU, M. S. (2006). Central Limit Theorem for the log-regression wavelet estimation of the memory parameter in the Gaussian semi-parametric context. *Fractals* **15** 301–313.

[14] MOULINES, E., ROUEFF, F. and TAQQU, M. S. (2007). On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. *J. Time Ser. Anal.* **28**. MR2345656

[15] ROBINSON, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. MR1370301

[16] ROBINSON, P. M. (1995). Log-periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072. MR1345214

[17] ROBINSON, P. M. and HENRY, M. (2003). Higher-order kernel semiparametric M-estimation of long memory. *J. Econometrics* **114** 1–27. MR1962371

[18] ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser, Boston. MR0885090

[19] ROUGHAN, M., VEITCH, D. and ABRY, P. (2000). Real-time estimation of the parameters of long-range dependence. *IEEE/ACM Transactions on Networking* **8** 467–478.

[20] SHIMOTSU, K. and PHILLIPS, P. C. B. (2005). Exact local Whittle estimation of fractional integration. *Ann. Statist.* **33** 1890–1933. MR2166565

[21] VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press. MR1652247

[22] VELASCO, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* **20** 87–127. MR1678573

[23] WORNELL, G. W. and OPPENHEIM, A. V. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.* **40** 611–623.

E. MOULINES
F. ROUEFF
E.N.S.T. 46
RUE BARRAULT
75634 PARIS CÉDEX 13
FRANCE
E-MAIL: moulines@tsi.enst.fr
roueff@tsi.enst.fr

M. S. TAQQU
DEPARTMENT OF MATHEMATICS AND STATISTICS
BOSTON UNIVERSITY
BOSTON, MASSACHUSETTS 02215
USA
E-MAIL: murad@math.bu.edu