



HAL
open science

A Wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series

Éric Moulines, François Roueff, Murad S. Taqqu

► **To cite this version:**

Éric Moulines, François Roueff, Murad S. Taqqu. A Wavelet Whittle estimator of the memory parameter of a non-stationary Gaussian time series. 2006. hal-00016446v1

HAL Id: hal-00016446

<https://hal.science/hal-00016446v1>

Preprint submitted on 4 Jan 2006 (v1), last revised 18 Aug 2008 (v4)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A WAVELET WHITTLE ESTIMATOR OF THE MEMORY PARAMETER OF A NON-STATIONARY GAUSSIAN TIME SERIES

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

CNRS LTCI and Boston University

ABSTRACT. We consider discrete-time Gaussian time series with memory parameter $d \in \mathbb{R}$. These time series are either stationary or can be made stationary after differencing a finite number of times. We develop a wavelet-based semiparametric pseudo-likelihood maximum method estimator of the memory parameter d , which can be seen as an extension to the wavelet-transform domain of the Gaussian semi-parametric estimator discussed in Robinson (1995a). The estimator may depend on a given finite range of scales or on a range which become infinite with the sample size. We show that the estimator is, in all cases, consistent and asymptotically normal.

Date: January 4, 2006.

1991 Mathematics Subject Classification. Primary 62M15, 62M10, 62G05. Secondary 62G20, 60G18.

Key words and phrases. long memory; semiparametric estimation; Wavelet analysis.

Murad S. Taqqu would like to thank l'École Normale Supérieure des Télécom-munications in Paris for their hospitality. This research was partially supported by the NSF Grant DMS-0505747 at Boston University.

1. INTRODUCTION AND BASIC ASSUMPTIONS

Let $X \stackrel{\text{def}}{=} \{X_\ell\}_{\ell \in \mathbb{Z}}$ be a Gaussian process, not necessarily stationary. Denote by ΔX the first order difference, $(\Delta X)_\ell = X_\ell - X_{\ell-1}$, and by $\Delta^k X$ the k -th order difference, which is defined iteratively. The process X is said to have memory parameter d , $d \in \mathbb{R}$, if for any integer $k > d - 1/2$, $\Delta^k X$ is stationary with spectral density function

$$f_{\Delta^k X}(\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{2(k-d)} f^*(\lambda) \quad \lambda \in (-\pi, \pi), \quad (1)$$

where f^* is a non-negative symmetric function which is bounded on $(-\pi, \pi)$ and is bounded away from zero in a neighborhood of the origin. Observe the following.

- (1) If $f_{\Delta^k X}(\lambda)$ is integrable, then it is a spectral density. This happens when $k > d - 1/2$, that is, $k \geq [d + 1/2]$. Then $\Delta^k X$ is stationary and so is $\Delta^{k'} X$ for any $k' \geq k$. It is therefore enough to consider in (1) the smallest integer k greater than $d - 1/2$.
- (2) If $d < 1/2$, then X is stationary.
- (3) When $d > 0$, X is said to exhibit long memory or long range dependence (LRD).

The *generalized spectral density* of X is defined as

$$f(\lambda) \stackrel{\text{def}}{=} |1 - e^{-i\lambda}|^{-2d} f^*(\lambda) \quad \lambda \in (-\pi, \pi) \quad (2)$$

The generalized spectral density f characterizes the distribution of $\Delta^k X$ for all $k > d - 1/2$ and it also characterizes the distribution of X up to a (random) polynomial trend of degree $[d - 1/2]$. It is a spectral density only if $d < 1/2$.

There has been considerable interest in long memory processes in the past decades both from a theoretical perspective and a modeling perspective see *e.g.* Baillie (1996) and Doukhan et al. (2003) and the references therein. Here are standard examples.

Examples

- 1) *Fractionally integrated auto-regressive moving-average series (FARIMA)*. This is a mean zero stationary Gaussian time series with spectral density $f(\lambda)$

given by (2), with $d < 1/2$, where f^* is the spectral density of a standard autoregressive moving-average (ARMA) time series $X^{(0)}$. One sometimes writes $X = \Delta^d X^{(0)}$ and regards the index d as the fractional differencing index.

2) *Fractional Brownian motion (FBM)* It is a continuous-time mean zero Gaussian stationary increment process $\{X(t), t \in \mathbb{R}\}$ which is self-similar with index $H \in (0, 1)$, that is, for all $a > 0$, $X(at) \stackrel{d}{=} a^H X(t)$, $t \in \mathbb{R}$, where $\stackrel{d}{=}$ denotes the equality of the finite-dimensional distributions. The time series $X \stackrel{\text{def}}{=} \{X_\ell\}_{\ell \in \mathbb{Z}}$ is the discrete-time version of FBM. Here $\Delta^k X$ is stationary for $k \geq 1$ and the generalized spectral density f of X is given, up to a multiplicative constant, by

$$f(\lambda) = \sum_{k \in \mathbb{Z}} |\lambda + 2k\pi|^{-(2H+1)} = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \quad \lambda \in (-\pi, \pi).$$

where $d = H + 1/2 \in (1/2, 3/2)$ and

$$f^*(\lambda) = \sum_{k \in \mathbb{Z}} \left| \frac{1 - e^{-i\lambda}}{\lambda + 2k\pi} \right|^{2H+1}, \quad \lambda \in (-\pi, \pi). \quad (3)$$

This function f^* is bounded on $(-\pi, \pi)$. Its behavior at zero frequency is given by

$$f^*(\lambda) = \left| \frac{1 - e^{-i\lambda}}{\lambda} \right|^{2H+1} + |1 - e^{-i\lambda}|^{2H+1} \sum_{k \neq 0} |\lambda + 2k\pi|^{-(2H+1)} = 1 + O(|\lambda|^{2 \wedge (2H+1)}),$$

and hence it is bounded away from zero in a neighborhood of the origin.

3) *Fractional Gaussian noise (FGN)*. It is $X = \Delta X^{(1)}$ where $X^{(1)}$ is the discrete-time FBM introduced above. It is therefore stationary with spectral density

$$f(\lambda) = |1 - e^{-i\lambda}|^2 \sum_{k \in \mathbb{Z}} |\lambda + 2k\pi|^{-(2H+1)} = |1 - e^{-i\lambda}|^{-2d} f^*(\lambda), \quad \lambda \in (-\pi, \pi),$$

where $d = H - 1/2 \in (-1/2, 1/2)$ and $f^*(\lambda)$ is as in (3).

We may consider non-stationary processes X such as FBM because we will use wavelet analysis to estimate d which performs an implicit differentiation of X up to any desired order specified by the number of vanishing moments (denoted M in assumptions (W-3) and (W-4) below) of the wavelet. It is therefore not required to preprocess the data (*e.g.* differentiating and tapering) to estimate

the memory parameter of a possibly nonstationary time-series, in contrast for example with methods based on Fourier analysis.

We suppose that we observe a finite sample of X and want to estimate the exponent d under the following typical smoothness assumptions on f^* . Let $\beta \in (0, 2]$ and $L > 0$. and assume that f^* belongs to the function class $\mathcal{H}(\beta, L)$. $\mathcal{H}(\beta, L)$ is the set of positive even functions g on $[-\pi, \pi]$ such that, for all $\lambda \in [-\pi, \pi]$,

$$|g(\lambda) - g(0)| \leq L g(0) |\lambda|^\beta . \quad (4)$$

This setting is typical in semiparametric estimation, see, *e.g.* , Robinson (1995a).

1.1. Fourier and wavelet estimators. There are two popular semiparametric estimators for the memory parameter d in the frequency domain:

- (1) The GPH estimator introduced in Geweke and Porter-Hudak (1983) and analyzed by Robinson (1995b). It involves a regression of the log-periodogram on the log of low frequencies.
- (2) The local Whittle estimator (or GSE) proposed in Künsch (1987) and developed in Robinson (1995a) is based on the Whittle approximation of the Gaussian likelihood, but adapted to a semiparametric context which focuses on low frequencies.

The corresponding approaches in the time/scale (wavelet) domain are:

- (1) Regression of the logarithm of the wavelet coefficient variance on the scale index, which has been introduced in Abry and Veitch (1998). See Bardet et al. (2000) for a treatment in a semiparametric context assuming observations in continuous time and Moulines et al. (2005) in discrete time.
- (2) Approximation of the Gaussian likelihood in the wavelet domain which has been proposed by Wornell and Oppenheim (1992), for noisy data (see also Kaplan and Kuo (1993), McCoy and Walden (1996) and the references therein). Although this estimator was the first wavelet estimator introduced in the context of long memory, to our knowledge, its theoretical properties have never been established rigorously (see the concluding remarks in (Velasco, 1999, p. 107)).

We want to study what we will call the *Wavelet Whittle estimator* which is the wavelet parallel to the local Whittle estimator. Working in the semiparametric setting where $f^* \in \mathcal{H}(\beta, L)$ is regarded a nuisance function, we show that the Wavelet Whittle estimator is consistent and asymptotically normal. We intend to compare this estimator to the ones mentioned above in a subsequent work.

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

- (W-1) ϕ and ψ are integrable and have compact supports, $\widehat{\phi}(0) = \int_{\mathbb{R}} \phi(x)dx = 1$ and $\int_{\mathbb{R}} \psi^2(x)dx = 1$.
- (W-2) There exists $\alpha > 1$ such that $\sup_{\xi \in \mathbb{R}} |\widehat{\psi}(\xi)| (1 + |\xi|)^\alpha < \infty$,
- (W-3) The function ψ has M vanishing moments, *i.e.* $\int_{\mathbb{R}} t^l \psi(t) dt = 0$ 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 $l = 0, \dots, M - 1$.

Assumption (W-1) implies that $\widehat{\phi}$ and $\widehat{\psi}$ are everywhere infinitely differentiable. 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 $\widehat{\psi}$ vanish at the origin and hence

$$|\widehat{\psi}(\lambda)| = O(|\lambda|^M) \quad \text{as } \lambda \rightarrow 0. \quad (5)$$

And, by (Cohen, 2003, Theorem 2.8.1, Page 90), (W-4) is equivalent to

$$\sup_{k \neq 0} |\widehat{\phi}(\lambda + 2k\pi)| = O(|\lambda|^M) \quad \text{as } \lambda \rightarrow 0. \quad (6)$$

We define the rescaled and shifted wavelets as

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k), j \in \mathbb{Z}, k \in \mathbb{Z}, \quad (7)$$

adopting the signal processing convention that large values of j correspond to low frequencies. Here j denote the scale index and k the position index. If ϕ and ψ are the scaling and the wavelet functions associated to a multiresolution analysis (see Cohen (2003)), 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 vanishing moments M (see *e.g.* Craigmile and Percival (2005)). 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 (Cohen, 2003, Chapter ?)).

In this work, however, we do not assume that the $\psi_{j,k}$ in (7) form an orthogonal basis. We may choose any ϕ and ψ as long as (W-1)-(W-4) are satisfied. For example, we may set, for some positive integer N ,

$$\phi(x) = \mathbb{1}_{[0,1]}^{*N}(x) \quad \text{and} \quad \psi(x) = \frac{d^N}{dx^N} \mathbb{1}_{[0,1]}^{*2N}(x),$$

where $\mathbb{1}_{[0,1]}$ is the indicator function of $[0, 1]$ and $\mathbb{1}_{[0,1]}^{*N}$ denotes the N -order self-convolution of $\mathbb{1}_{[0,1]}$. Then

$$|\widehat{\phi}(\xi)| = |2 \sin(\xi/2)/\xi|^N \quad \text{and} \quad \widehat{\psi}(\xi) = |\xi|^N |2 \sin(\xi/2)/\xi|^{2N}.$$

Using (5) and (6), one easily checks that (W-1)-(W-4) are satisfied with M and α both equal to N . Of course the $\{\psi_{j,k}\}$ are not orthonormal for this choice of ψ .

To any real valued sequence $x = \{x_k, k \in \mathbb{Z}\}$, we associate the functions

$$\mathbf{x}_n(t) \stackrel{\text{def}}{=} \sum_{k=1}^n x_k \phi(t-k) \quad \text{and} \quad \mathbf{x}(t) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} x_k \phi(t-k) \quad (8)$$

by using the father wavelet ϕ as interpolator. The (details) wavelet coefficients are then defined as

$$W_{j,k}^{\mathbf{x}} \stackrel{\text{def}}{=} \int_{\mathbb{R}} \mathbf{x}(t) \psi_{j,k}(t) dt \quad j \geq 0, k \in \mathbb{Z}. \quad (9)$$

We can suppose without loss of generality that ϕ has support in $(-T, 0)$ and ψ has support in $(0, T)$ for some integer $T \geq 1$. Since, for all $k \leq 0$ and all $k > n$, $\phi(\cdot - k)$ has its support outside $[0, n - T + 1]$, we have $\mathbf{x}_n(t) = \mathbf{x}(t)$ for all $t \in [0, n - T + 1]$. Moreover, since $\psi_{j,k}$ has support in $2^j(k, k + T)$, we have

$$W_{j,k}^{\mathbf{x}} = W_{j,k}^{\mathbf{x}_n} = \int_{\mathbb{R}} \mathbf{x}_n(t) \psi_{j,k}(t) dt, \quad (10)$$

for all indices (j, k) such that $j \geq 0$ and $0 \leq k < n_j$, where, denoting the integer part of x by $[x]$,

$$n_j \stackrel{\text{def}}{=} [2^{-j}(n - T + 1) - T + 1] \vee 0 \quad (11)$$

is the number of available wavelet coefficients at scale index j . We will drop the superscript \mathbf{x} or \mathbf{x}_n when there is no danger of confusion. In the following sections, we will perform a wavelet analysis of a random process $X = \{X_\ell, \ell \in \mathbb{Z}\}$, in which case we will write $W_{j,k}^{\mathbf{X}}$ or simply $W_{j,k}$ when no ambiguity may occur.

1.3. The Wavelet Whittle estimator. Consider an array of centered independent Gaussian random variables $\{c_{j,k}, (j, k) \in \mathcal{I}\}$ with variance $\mathbb{E}c_{j,k}^2$, where \mathcal{I} is a finite set. Its log-likelihood, up to a negative multiplicative constant and an additive constant, may be expressed as

$$\sum_{(j,k) \in \mathcal{I}} \{c_{j,k}^2 / \mathbb{E}c_{j,k}^2 + \log(\mathbb{E}c_{j,k}^2)\} .$$

Our Wavelet Whittle estimator uses such a contrast process to estimate the memory parameter d . The idea is to identify $\{c_{j,k}, (j, k) \in \mathcal{I}\}$ with the wavelet coefficients $\{W_{j,k}^{\mathbf{X}}, (j, k) \in \mathcal{I}\}$. This is because the wavelet coefficients $W_{j,k}^{\mathbf{X}}$ are weakly correlated. Pseudo-maximum likelihood estimation consists in choosing an effective approximation of the likelihood. We will show that this approximation of the likelihood has indeed good properties for the estimation of the memory parameter d in a semiparametric framework if we set

$$\mathbb{E}c_{j,k}^2 \stackrel{\text{def}}{=} \sigma^2 2^{2jd} , \quad (12)$$

since, as we show in Moulines et al. (2005), one has $\mathbb{E}W_{j,k}^2 = \sigma^2 2^{2dj} \{1 + O(2^{-\beta j})\}$ as $j \rightarrow \infty$, under the smoothness assumption $f^* \in \mathcal{H}(\beta, L)$, see (4).

Applying the contrast process to the wavelet coefficients, we obtain, for all $\sigma^2 > 0$ and $d \in \mathbb{R}$,

$$\begin{aligned} \widehat{\mathcal{L}}_{\mathcal{I}}(\sigma^2, d) &= \sum_{(j,k) \in \mathcal{I}} \{W_{j,k}^2 / (\sigma^2 2^{2dj}) + \log(\sigma^2 2^{2dj})\} \\ &= \frac{1}{\sigma^2} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} W_{j,k}^2 + (\#\mathcal{I}) \log(\sigma^2) + 2 \log(2)d \sum_{(j,k) \in \mathcal{I}} j , \end{aligned}$$

where $\#\mathcal{I}$ denotes the number of elements of the set \mathcal{I} . The Wavelet Whittle pseudo-maximum likelihood estimator is then defined as

$$(\widehat{d}_{\mathcal{I}}, \widehat{\sigma}_{\mathcal{I}}^2) \stackrel{\text{def}}{=} \underset{d \in \mathbb{R}, \sigma^2 > 0}{\text{Argmin}} \widehat{\mathcal{L}}_{\mathcal{I}}(\sigma^2, d).$$

For any given value of the memory parameter d , the maximum estimator of the scale coefficient is given by

$$\widehat{\sigma}_{\mathcal{I}}^2(d) \stackrel{\text{def}}{=} \underset{\sigma^2 > 0}{\text{Argmin}} \widehat{L}_{\mathcal{I}}(\sigma^2, d) = \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} 2^{-2dj} W_{j,k}^2 .$$

We define the profile likelihood $\widetilde{L}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \widehat{L}_{\mathcal{I}}(\widehat{\sigma}_{\mathcal{I}}^2(d), d)$, which may be rewritten as

$$\widetilde{L}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \frac{2 \log(2)d}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} j + \log \left(\sum_{(j,k) \in \mathcal{I}} 2^{-2dj} W_{j,k}^2 \right) \quad (13)$$

$$= \log \left(\sum_{(j,k) \in \mathcal{I}} 2^{2d(\eta(\mathcal{I})-j)} W_{j,k}^2 \right) , \quad (14)$$

where

$$\eta(\mathcal{I}) \stackrel{\text{def}}{=} \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} j . \quad (15)$$

The pseudo maximum likelihood estimator of the memory parameter is equal to the minimum of the profile likelihood $\widehat{d}_{\mathcal{I}} = \underset{d \in \mathbb{R}}{\text{Argmin}} \widetilde{L}_{\mathcal{I}}(d)$. Provided that \mathcal{I} contains at least two different scales, then

$$\min\{j : (j, k) \in \mathcal{I} \text{ for some } k\} < \eta(\mathcal{I}) < \max\{j : (j, k) \in \mathcal{I} \text{ for some } k\} ,$$

so that there are indices j for which $2d(\eta(\mathcal{I}) - j)$ will be positive and others for which it will be negative and hence $\widetilde{L}_{\mathcal{I}}(d)$ tends to ∞ as d tend to $-\infty$ and $+\infty$. Thus, $\widehat{d}_{\mathcal{I}}$ is finite and the derivative of $\widetilde{L}_{\mathcal{I}}(d)$ vanishes at $d = \widehat{d}_{\mathcal{I}}$, *i.e.* $\widehat{\mathbf{S}}_{\mathcal{I}}(\widehat{d}_{\mathcal{I}}) = 0$ where, for all $d \in \mathbb{R}$,

$$\widehat{\mathbf{S}}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \sum_{(j,k) \in \mathcal{I}} [j - \eta(\mathcal{I})] 2^{-2jd} W_{j,k}^2 . \quad (16)$$

We consider two specific choices for \mathcal{I} . For any integers n , J_0 and J_1 satisfying

$$0 \leq J_0 < J_1 \leq J \stackrel{\text{def}}{=} \max\{j : n_j \geq 1\} = \lceil \log_2\{(n - T + 1)/T\} \rceil , \quad (17)$$

we define $\mathcal{I}_n(J_0, J_1) \stackrel{\text{def}}{=} \{(j, k) : J_0 \leq j \leq J_1 \text{ and } 0 \leq k < n_j\}$, where n_j is defined in (11). From now on, J_0 and J_1 are sequences indexed by n but, for notational simplicity, the dependence of J_0 , J_1 and J in n is implicit. The index J is the maximal available scale index corresponding to the sample size n and J_0 and J_1 denote the lower and upper scale indices used by the estimator. As explained above, the estimator $\widehat{d}_{\mathcal{I}}$ is well defined as soon as \mathcal{I} contains two different scales

which, for $\mathcal{I} = \mathcal{I}_n(J_0, J_1)$, is equivalent to having (17). As we will see below, in the semiparametric framework, J_0 governs the rate of convergence of $\widehat{d}_{\mathcal{I}}$ towards the true memory parameter. There are two possible settings as far as J_1 is concerned:

(C-1) $J_1 - J_0$ is fixed, equal to $\ell > 0$, so that $\mathcal{I} = \mathcal{I}_n(J_0, J_0 + \ell)$,

(C-2) $J_1 = J$ so that $\mathcal{I} = \mathcal{I}_n(J_0, J)$.

We will study the large sample properties of the estimators $\widehat{d}_{\mathcal{I}_n(J_0, J_0 + \ell)}$ and $\widehat{d}_{\mathcal{I}_n(J_0, J)}$ as n tends to infinity for appropriate choices of J_0 with the goal of establishing consistency and asymptotic normality. See the comments following Theorem 4 below for a discussion about these two cases. In our results, we will always assume that $J_0 \geq 0$ and $n2^{-J_0} \rightarrow \infty$. Note that in both cases (C-1) and (C-2), this implies (17) for n large enough. The rates will depend on n and on the scale index J_0 . For Fourier estimators, the bounds are generally expressed as functions of n and a *bandwidth* parameter m , equal to the number of discrete Fourier frequencies used to construct the estimator. To allow comparison, we use n and m to express our results, where m is defined as the number of wavelet coefficients used by $\widehat{d}_{\mathcal{I}_n(J_0, J_0 + \ell)}$ or $\widehat{d}_{\mathcal{I}_n(J_0, J)}$, given by $m = \sum_{j=J_0}^{J_0 + \ell} n_j$ and $m = \sum_{j=J_0}^J n_j$, respectively. From Lemma 11, we have that $n2^{-J_0} \rightarrow \infty$ is equivalent to $m \rightarrow \infty$ and, when $n2^{-J_0} \rightarrow \infty$,

$$m \sim \begin{cases} (2 - 2^{-\ell})n2^{-J_0} & \text{if } m = \sum_{j=J_0}^{J_0 + \ell} n_j, \\ 2n2^{-J_0} & \text{if } m = \sum_{j=J_0}^J n_j. \end{cases} \quad (18)$$

The paper is structured as follows. We state the main results in Section 2. The asymptotic behavior of the contrast process is described in Section 3. The proof of consistency is given in Section 4. Rates of convergence are studied in Section 5. Useful approximations to the spectral density of the wavelet coefficients are derived in Section 6. Asymptotic normality is established in Section 7. The appendix contains auxiliary results which are used in the proofs.

2. RESULTS

We suppose that the wavelet coefficients are those of a Gaussian process with a generalized spectral density f of the form (2) and we denote by d_0 the true

value of the memory parameter,

$$f(\lambda) = |1 - e^{-i\lambda}|^{-2d_0} f^*(\lambda), \quad \lambda \in (-\pi, \pi). \quad (19)$$

We denote by \mathbb{P}_f the probability distribution of $\{X_\ell, \ell \in \mathbb{Z}\}$ corresponding to f according to the above model and by \mathbb{E}_f its associated expectation. If the number of vanishing moments M in (W-3) is large enough then, for all $j \geq 0$, the process of wavelet coefficients $\{W_{j,k}, k \in \mathbb{Z}\}$ is stationary and we denote

$$\sigma_j^2(f) \stackrel{\text{def}}{=} \text{Var}_f(W_{j,k}). \quad (20)$$

The two following basic conditions, which hold under appropriate conditions on f^* (Theorem 1) will be used to establish the consistency and the rate of convergence of $\widehat{d}_{\mathcal{I}}$.

- There exist $\beta > 0$ and $\sigma^2 > 0$ such that, for all $j \geq 0$,

$$\sup_{j \geq 0} 2^{\beta j} \left| \frac{\sigma_j^2(f)}{\sigma^2 2^{2d_0 j}} - 1 \right| < \infty. \quad (21)$$

- There exists a constant C such that, for all $j \geq 0$ and $N \geq 1$,

$$\text{Var}_f \left(\sum_{k=1}^N W_{j,k}^2 \right) \leq C N \sigma_j^4(f). \quad (22)$$

Condition (21) states that, up to the multiplicative constant σ^2 , the variance $\sigma_j^2(f)$ is approximated by $2^{2d_0 j}$ with a rate given by the exponent β . Condition (22) imposes a bound on the variance of the normalized partial sum of the stationary centered sequence $\{W_{j,k}^2/\sigma_j^2(f)\}$ which is similar to what happens when these variables are independent. We stress that the wavelet coefficients $W_{j,k}$, however, are not independent since their spectral density is not constant, see Moulines et al. (2005).

Theorem 1. *Suppose that (19) holds with $f^* \in \mathcal{H}(\beta, L)$, $\beta \in (0, 2]$ and $f^*(0) > 0$. Assume (W-1)-(W-4) with $M \geq d_0$ and $\alpha > (1 + \beta)/2 - d_0$. Then,*

- (i) *Condition (21) holds with $\sigma^2 = f^*(0)K(d_0, \psi)$, where*

$$K(d, \psi) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} |\xi|^{-2d} |\widehat{\psi}(\xi)|^2 d\xi. \quad (23)$$

- (ii) *Condition (22) holds.*

Proof. See Section 6. □

The estimator $\widehat{d}_{\mathcal{I}}$ was defined in Section 1.3. The following theorem states that it is consistent under Conditions (21) and (22). We consider the two different cases (C-1) and (C-2) discussed above.

Theorem 2 (Weak Consistency). *Assume Conditions (21) and (22). Then,*

(a) *For any fixed positive integer ℓ , if $m/n + m^{-1} \rightarrow 0$ with $m = \sum_{j=J_0}^{J_0+\ell} n_j$, then*

$$\widehat{d}_{\mathcal{I}_n(J_0, J_0+\ell)} = d_0 + O_{\mathbb{P}_f} \left\{ \left(\frac{m}{n} \right)^{\beta/2} + m^{-1/4} \right\}. \quad (24)$$

(b) *If $m/n + m^{-1} \rightarrow 0$ with $m = \sum_{j=J_0}^J n_j$, then there exists $\gamma > 0$ such that*

$$\widehat{d}_{\mathcal{I}_n(J_0, J)} = d_0 + O_{\mathbb{P}_f} \left\{ \left(\frac{m}{n} \right)^{\beta/2} + m^{-\gamma} \right\}. \quad (25)$$

Proof. See Section 4. □

Remark 1. Eqs. (24) and (25) imply that $\widehat{d}_{\mathcal{I}_n(J_0, J_0+\ell)} \rightarrow_{\mathbb{P}_f} d_0$ and $\widehat{d}_{\mathcal{I}_n(J_0, J)} \rightarrow_{\mathbb{P}_f} d_0$, respectively. While the rates in (24) and (25) are not optimal, they will be used to derive the optimal rates of convergence (Theorem 3) and the asymptotic normality of the estimators (Theorem 4).

Theorem 3 (Rate of Convergence). *Assume Conditions (21) and (22). In addition, assume either that*

(a) $\mathcal{I}_n = \mathcal{I}_n(J_0, J_0 + \ell)$ for some fixed positive integer ℓ and $m = \sum_{j=J_0}^{J_0+\ell} n_j \asymp n^\gamma$ for some $\gamma \in (0, 1)$,

(b) or $\mathcal{I}_n = \mathcal{I}_n(J_0, J)$ and $m = \sum_{j=J_0}^J n_j \asymp n^\gamma$ for some $\gamma \in (0, 1)$.

Then,

$$\widehat{d}_{\mathcal{I}_n} = d_0 + O_{\mathbb{P}_f} \left\{ m^{-1/2} + \left(\frac{m}{n} \right)^\beta \right\}. \quad (26)$$

Proof. See Section 5. □

Remark 2. The O -term appearing in (26) can be interpreted as a fluctuation + bias decomposition which is similar to the one obtained for the local Whittle estimator (see Robinson (1995a)).

Remark 3. By (18), the two terms in (26) are of the same order of magnitude if we set $J_0 = \lceil \log_2(n)/(1 + 2\beta) \rceil$; in that case,

$$\widehat{d}_{\mathcal{I}_n(J_0, J)} - d_0 = O_{\mathbb{P}_f} \left(n^{\beta/(1+2\beta)} \right).$$

The rate of convergence $n^{\beta/(1+2\beta)}$ is thus identical to the one for the local Whittle estimator derived by Robinson (1995a) in a similar semiparametric context.

We also obtain a central limit theorem for the estimator $\widehat{d}_{\mathcal{I}}$ of d_0 . For all $d > 1/2 - \alpha$ and $u \in \mathbb{N}$, define

$$I_u(d, \psi) \stackrel{\text{def}}{=} \sum_{v=0}^{2^u-1} \int_{-\pi}^{\pi} \left| \sum_{l \in \mathbb{Z}} \frac{e^{-i v 2^{-u}(\lambda + 2l\pi)}}{|\lambda + 2l\pi|^{2d}} \overline{\widehat{\psi}(\lambda + 2l\pi)} \widehat{\psi}(2^{-u}(\lambda + 2l\pi)) \right|^2 d\lambda \quad (27)$$

and, for all $\ell \geq 1$,

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

Theorem 4 (CLT). *Set $\beta \in (0, 2]$ and $L < \infty$. Let X be a Gaussian process with generalized spectral density $f(\lambda) = |1 - e^{i\lambda}|^{-2d_0} f^*(\lambda)$ where $f^* \in \mathcal{H}(\beta, L)$ and satisfies $f^*(0) > 0$. Assume (W-1)-(W-4) with $M \geq d_0$ and $\alpha > (1 + \beta)/2 - d_0$. Then,*

- (a) *For any positive integer ℓ , if $m = \sum_{j=J_0}^{J_0+\ell} n_j \asymp n^\gamma$ for some $\gamma \in (2\beta/\{1 + 2\beta\}, 1)$, then*

$$\sqrt{m} (\widehat{d}_{\mathcal{I}_n(J_0, J_0+\ell)} - d_0) \xrightarrow{\mathcal{L}} \mathcal{N}[0, V(d_0, \ell, \psi)],$$

where

$$V(d_0, \ell, \psi) \stackrel{\text{def}}{=} \frac{\pi}{\kappa_\ell (\log(2)K(d_0, \psi))^2} \times \left\{ I_0(d_0, \psi) + \frac{2}{\kappa_\ell} \sum_{u=1}^{\ell} I_u(d_0, \psi) 2^{(2d_0-1)u} \sum_{i=0}^{\ell-u} \frac{2^{-i}}{2 - 2^{-\ell}} (i - \eta_\ell)(i + u - \eta_\ell) \right\}. \quad (29)$$

- (b) *If $m = \sum_{j=J_0}^J n_j \asymp n^\gamma$ for some $\gamma \in (2\beta/\{1 + 2\beta\}, 1)$, then*

$$\sqrt{m} (\widehat{d}_{\mathcal{I}_n(J_0, J)} - d_0) \xrightarrow{\mathcal{L}} \mathcal{N}[0, V(d_0, \psi)],$$

where

$$V(d_0, \psi) \stackrel{\text{def}}{=} \frac{\pi}{2(\log(2)K(d_0, \psi))^2} \left\{ I_0(d_0, \psi) + 2 \sum_{u=1}^{\infty} I_u(d_0, \psi) 2^{(2d_0-1)u} \right\}. \quad (30)$$

Proof. See Section 7. □

Remark 4. It will follow from the proofs that $V(d_0, \ell, \psi) \rightarrow V(d_0, \psi)$ as $\ell \rightarrow \infty$. In practice, the number of observation n is finite. Since the estimator $\widehat{d}_{\mathcal{I}_n(J_0, J)}$ may always be interpreted as an estimator $\widehat{d}_{\mathcal{I}_n(J_0, J_0 + \ell)}$, with $\ell = J - J_0$, we have two approximations at hand for the distribution of $(n2^{-J_0})^{1/2} (\widehat{d}_{\mathcal{I}_n(J_0, J)} - d_0)$, namely, a $\mathcal{N}(0, V(d_0, \ell, \psi))$ and a $\mathcal{N}(0, V(d_0, \psi))$. The first one is obtained by letting $J \rightarrow \infty$ with ℓ fixed and the second by letting both J and ℓ tend to ∞ . Since the first one involves only a single limit, it is likely to provide a better approximation at finite n .

Both the local Whittle method –which is Fourier based – and the Wavelet Whittle method have advantages and disadvantages. The main advantage of the local Whittle method is that the asymptotic variance does not depend on the unknown value of d_0 (see (Robinson, 1995a, Theorem 2)). The advantage of the Wavelet Whittle method is that it is robust to the presence of additive polynomial trends and continues to apply when d takes value outside the interval $(-1/2, 1/2)$ where X is stationary and invertible. One has only to choose appropriate wavelets, for example the Daubechies wavelets with a sufficiently large number of vanishing moments M (see Section 1.2). We will compare systematically the wavelet Whittle estimator to other estimators in a subsequent work.

3. ASYMPTOTIC BEHAVIOR OF THE CONTRAST PROCESS

We shall decompose the contrast (13) into a sum of deterministic and empirical terms and provide uniform approximations of the deterministic component and uniform bounds of the empirical component. More precisely, it is easily verified from (13) that there exists a constant C such that for all d

$$\widetilde{L}_{\mathcal{I}}(d) \stackrel{\text{def}}{=} L_{\mathcal{I}}(d) + E_{\mathcal{I}}(d) + C, \quad (31)$$

where

$$L_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log \left(\frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} 2^{2(d_0-d)j} \right) - \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} \log(2^{2(d_0-d)j}), \quad (32)$$

$$E_{\mathcal{I}}(d) \stackrel{\text{def}}{=} \log \left[1 + \sum_{(j,k) \in \mathcal{I}} \frac{2^{2(d_0-d)j}}{\sum_{\mathcal{I}} 2^{2(d_0-d)j}} \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}} - 1 \right) \right] \quad (33)$$

and σ^2 is the constant appearing in Condition (21). Let us comment briefly on the terms appearing in (31).

- (1) $d \mapsto L_{\mathcal{I}}(d)$ is a deterministic function. The Jensen inequality implies that this function has a single minimum at $d = d_0$, where it vanishes. Asymptotic approximations for $L_{\mathcal{I}}(d)$ as the number of scales tend to infinity will be given in Section 3.1.
- (2) $d \mapsto E_{\mathcal{I}}(d)$ is a random process. Consistency is obtained by showing that it is asymptotically negligible uniformly in d , see Section 3.2.
- (3) The constant C may be discarded, as it does not depend on d .

3.1. Deterministic component of the contrast.

Proposition 5. *For any finite and non-empty set $\mathcal{I} \subset \mathbb{N} \times \mathbb{Z}$, the function $d \mapsto L_{\mathcal{I}}(d)$ is non-negative, convex and vanishes at $d = d_0$.*

Assume that $n2^{-J_0} \rightarrow \infty$ and that (17) holds for n large enough. Then, for any constants d_ and d^* satisfying $d_0 - 1/2 < d_* \leq d^*$, one has*

$$\liminf_{n \rightarrow \infty} \max_{d \in [d_*, d^*]} \ddot{L}_{\mathcal{I}_n(J_0, J_1)}(d) > 0. \quad (34)$$

where $\ddot{L}_{\mathcal{I}}$ denotes the second derivative of the function $L_{\mathcal{I}}$.

Proof. By concavity of the log function, $L_{\mathcal{I}}(d) \geq 0$ and is zero if $d = d_0$. The second derivative of the function $d \mapsto L_{\mathcal{I}}(d)$ is given by

$$\ddot{L}_{\mathcal{I}}(d) = (2 \log(2))^2 \left\{ \sum_{(j,k) \in \mathcal{I}} j^2 \frac{2^{2(d_0-d)j}}{\sum_{(j',k') \in \mathcal{I}} 2^{2(d_0-d)j'}} - \left(\sum_{(j,k) \in \mathcal{I}} j \frac{2^{2(d_0-d)j}}{\sum_{(j',k') \in \mathcal{I}} 2^{2(d_0-d)j'}} \right)^2 \right\}.$$

It is easily seen that $\ddot{L}_{\mathcal{I}}(d) \geq 0$ for any $d \in \mathbb{R}$ showing that $d \mapsto L_{\mathcal{I}}(d)$ is convex.

Observe that, in the above expression of $\ddot{L}_{\mathcal{I}}(d)$, the term between the curly brackets is $\text{Var}(N)$, where N is an integer valued random variable with distribution

$$\mathbb{P}(N = j) = \frac{2^{2(d_0-d)j} \#\{k : (j, k) \in \mathcal{I}\}}{\sum_{(j',k) \in \mathcal{I}} 2^{2(d_0-d)j'}}, \quad j \geq 0.$$

Let now $d \geq d_\star > d_0 - 1/2$. For $\mathcal{I} = \mathcal{I}_n(J_0, J_1)$ such that (17) holds, one has

$$\begin{aligned} \mathbb{P}(N = J_0) &= \frac{2^{2(d_0-d)J_0} n_{J_0}}{\sum_{j=J_0}^{J_1} 2^{2(d_0-d)j} n_j} \\ &\geq \frac{2^{2(d_0-d)J_0} \{2^{-J_0}(n - \mathbb{T} + 1) - \mathbb{T}\}}{(n - \mathbb{T} + 1) \sum_{j=J_0}^{\infty} 2^{\{2(d_0-d)-1\}j}} \\ &\geq (1 - 2^{2(d_0-d_\star)-1}) \frac{2^{2(d_0-d)J_0} \{2^{-J_0}(n - \mathbb{T} + 1) - \mathbb{T}\}}{(n - \mathbb{T} + 1) 2^{\{2(d_0-d)-1\}J_0}} \\ &= (1 - 2^{2(d_0-d_\star)-1}) \{1 - \mathbb{T} 2^{J_0} (n - \mathbb{T} + 1)^{-1}\} , \end{aligned}$$

Since $n2^{-J_0} \rightarrow \infty$, the term between brackets tends to 1 as $n \rightarrow \infty$. Hence, for n large enough, we have $\inf_{d \geq d_\star} \mathbb{P}(N = J_0) \geq (1 - 2^{2(d_0-d_\star)-1})/2$. Similarly, one finds, for n large enough, $\inf_{d \in [d_\star, d^\star]} \mathbb{P}(N = J_0 + 1) \geq (1 - 2^{2(d_0-d_\star)-1})2^{2(d_0-d^\star)-1}/2$. Hence,

$$\begin{aligned} \inf_{d \in [d_\star, d^\star]} \text{Var}(N) &\geq \{J_0 - \mathbb{E}(N)\}^2 \mathbb{P}(N = J_0) + \{J_0 + 1 - \mathbb{E}(N)\}^2 \mathbb{P}(N = J_0 + 1) \\ &\geq (1 - 2^{2(d_0-d_\star)-1}) 2^{2(d_0-d^\star)-2} (\{J_0 - \mathbb{E}(N)\}^2 + \{J_0 + 1 - \mathbb{E}(N)\}^2) \\ &\geq (1 - 2^{2(d_0-d_\star)-1}) 2^{2(d_0-d^\star)-4} , \end{aligned}$$

where the last inequality is obtained by observing that either $\mathbb{E}(N) - J_0 \geq 1/2$ or $J_0 + 1 - \mathbb{E}(N) < 1/2$. This achieves the proof of (34). \square

3.2. Random component of the contrast. We will show that, under appropriate assumptions, the random term (33) tends to zero uniformly in d . For all $C > 0$, $q \geq 0$ and $\delta \geq 0$, define the set of real-valued sequences

$$\mathcal{B}(C, q, \delta) \stackrel{\text{def}}{=} \left\{ (\mu_j)_{j \geq 0} : |\mu_j| \leq C (1 + j^q) 2^{j\delta} \text{ for all } j \geq 0 \right\}. \quad (35)$$

Let \mathcal{I} be a finite subset of $\mathbb{N} \times \mathbb{Z}$. Define, for any sequence $\boldsymbol{\mu} \stackrel{\text{def}}{=} \{\mu_j\}_{j \geq 0}$,

$$\tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) \stackrel{\text{def}}{=} \sum_{j=J_0}^{J_1} \mu_{j-J_0} \sum_{k=0}^{n_j-1} \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1 \right). \quad (36)$$

Observe that, in (36), the summation in j starts at $j = J_0$ but that the corresponding indices for $\boldsymbol{\mu}$ are $j - J_0 = 0, 1, \dots, J_1 - J_0$.

We first provide some general uniform bounds for $\tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu})$ and then, in a corollary, we focus on the random component (33) appearing in the contrast (31).

Proposition 6. *Let $\delta < 1$. Conditions (21) and (22) imply the following asymptotic results for any $q \geq 0$ and $C > 0$:*

(a) *if $J_0^{-1} + (n2^{-J_1})^{-1} \rightarrow 0$ as $n \rightarrow \infty$, then,*

$$\begin{aligned} \sup_{\boldsymbol{\mu} \in \mathcal{B}(C, q, \delta)} \left| \tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) \right| \\ = n2^{-J_0} \{1 + (J_1 - J_0)^q\} \{O_{\mathbb{P}_f}((n2^{-J_1})^{-1/2}) + O(2^{-\beta J_0})\} ; \end{aligned}$$

(b) *if $J_0^{-1} + n2^{-J_0} \rightarrow 0$ as $n \rightarrow \infty$, then,*

$$\begin{aligned} \sup_{\boldsymbol{\mu} \in \mathcal{B}(C, q, \delta)} \left| \tilde{\mathbf{S}}_{n, J_0, J}(\boldsymbol{\mu}) \right| \\ = n2^{-J_0} \log^q(n2^{-J_0}) \{O_{\mathbb{P}_f}(\log(n2^{-J_0}) (n2^{-J_0})^{-(1-\delta)/(3-2\delta)}) + O(2^{-\beta J_0})\} . \end{aligned}$$

To prove Part (b), we will use Part (a) for some J_1 chosen in such a way that both $J_1 - J_0$ and $n2^{-J_1}$ tend to infinity.

Proof. We set $C = 1$ without loss of generality. We may write

$$\tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) = \sum_{j=J_0}^{J_1} \nu_j \mu_{j-J_0} \sum_{k=0}^{n_j} \left(\frac{W_{j,k}^2}{\sigma_j^2(f)} - 1 \right) + \sum_{j=J_0}^{J_1} n_j \mu_{j-J_0} (\nu_j - 1), \quad (37)$$

where $\nu_j \stackrel{\text{def}}{=} \frac{\sigma_j^2(f)}{\sigma^2 2^{2d_0 j}}$ for all j . By convexity of the square function, we have

$$\begin{aligned} \left| \sum_{j=J_0}^{J_1} \nu_j \mu_{j-J_0} \sum_{k=0}^{n_j-1} \left(\frac{W_{j,k}^2}{\sigma_j^2(f)} - 1 \right) \right|^2 \\ \leq \left[\sum_{j=J_0}^{J_1} n_j |\nu_j \mu_{j-J_0}| \right]^2 \sum_{j=J_0}^{J_1} \frac{1}{n_j} \left| \sum_{k=0}^{n_j-1} \left(\frac{W_{j,k}^2}{\sigma_j^2(f)} - 1 \right) \right|^2 . \quad (38) \end{aligned}$$

By (21), $\sup_{j \geq 0} 2^{\beta j} |\nu_j - 1| < \infty$. In particular, $\|\boldsymbol{\nu}\|_\infty \stackrel{\text{def}}{=} \sup_{j \geq 0} |\nu_j| < \infty$ and, since $n_j \leq n2^{-j}$ and $\delta < 1$, for all $\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)$,

$$\sum_{j=J_0}^{J_1} n_j |\nu_j \mu_{j-J_0}| \leq \|\boldsymbol{\nu}\|_\infty \{1 + (J_1 - J_0)^q\} \frac{n 2^{-J_0}}{1 - 2^{\delta-1}} .$$

Inserting this bound in (38) and using (22), we get that there exists a constant C' such that

$$\begin{aligned} \mathbb{E}_f \left\{ \sup_{\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)} \left| \sum_{j=J_0}^{J_1} \nu_j \mu_{j-J_0} \sum_{k=0}^{n_j-1} \left(\frac{W_{j,k}^2}{\sigma_j^2(f)} - 1 \right) \right|^2 \right\} \\ \leq C' \{1 + (J_1 - J_0)^q\}^2 (n 2^{-J_0})^2 \sum_{j=J_0}^{J_1} n_j^{-1}. \end{aligned}$$

Observe that, for $n 2^{-J_1}$ large enough,

$$\sum_{j=J_0}^{J_1} n_j^{-1} \leq \frac{\sum_{j=J_0}^{J_1} 2^j}{n - (T-1) - 2^{J_1}(T-1)} \leq \frac{2(n 2^{-J_1})^{-1}}{1 - 2(T-1)(n 2^{-J_1})^{-1}}.$$

Applying the Markov inequality, the two last bounds give, as $n 2^{-J_1} \rightarrow \infty$,

$$\sup_{\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)} \left| \sum_{j=J_0}^{J_1} \nu_j \mu_{j-J_0} \sum_{k=0}^{n_j} \left(\frac{W_{j,k}^2}{\sigma_j^2(f)} - 1 \right) \right| = O_{\mathbb{P}_f} \left(\frac{n 2^{-J_0} \{1 + (J_1 - J_0)^q\}}{(n 2^{-J_1})^{1/2}} \right).$$

We now consider the second term in the RHS of (37). Using again the bound $\sup_{j \geq 0} 2^{\beta j} |\nu_j - 1| \leq C$, we get, for all $\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)$,

$$\left| \sum_{j=J_0}^{J_1} n_j \mu_{j-J_0} (\nu_j - 1) \right| \leq C \{1 + (J_1 - J_0)^q\} \frac{n 2^{-(\beta+1)J_0}}{1 - 2^{\delta-\beta-1}}.$$

Inserting the two last equations in (37) concludes the proof of Part (a).

We now turn to Part (b). Let $J_1 \geq J_0$ depending on n such that $n 2^{-J_1} \rightarrow \infty$ and write

$$\tilde{\mathbf{S}}_{n, J_0, J}(\boldsymbol{\mu}) = \tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) + \tilde{\mathbf{S}}_{n, J_1+1, J}(\bar{\boldsymbol{\mu}}), \quad (39)$$

where $\bar{\boldsymbol{\mu}} \stackrel{\text{def}}{=} \{\mu_{j+J_1+1-J_0}\}_{j \geq 0}$. Using Part (a), we have

$$\begin{aligned} \sup_{\boldsymbol{\mu} \in \mathcal{B}(C, q, \delta)} \left| \tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) \right| \\ = n 2^{-J_0} \{1 + (J_1 - J_0)^q\} \left\{ O_{\mathbb{P}_f} \left((n 2^{-J_1})^{-1/2} \right) + O \left(2^{-\beta J_0} \right) \right\}. \quad (40) \end{aligned}$$

Let us consider the last term in (39). We have

$$\left| \tilde{\mathbf{S}}_{n, J_1+1, J}(\bar{\boldsymbol{\mu}}) \right| \leq \left[\sup_{j > J_1} |\mu_{j-J_0}| 2^{-j} \right] \sum_{j=J_1+1}^J 2^j \sum_{k=0}^{n_j-1} \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} + 1 \right).$$

Observe that, since $\delta < 1$, there exists $C > 0$ only depending on q and δ such that for all $\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)$,

$$\sup_{j > J_1} |\mu_{j-J_0}| 2^{-j} \leq C 2^{(\delta-1)(J_1-J_0)} \{1 + (J_1 - J_0)^q\} 2^{-J_0} .$$

From (21), $\mathbb{E}_f W_{j,k}^2 / \sigma^2 2^{2d_0j}$ is bounded by a constant independent of $j \geq 0$, hence

$$\mathbb{E}_f \sup_{\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)} \left| \tilde{\mathbf{S}}_{n, J_1+1, J}(\bar{\boldsymbol{\mu}}) \right| = O \left(2^{(\delta-1)(J_1-J_0)} \{1 + (J_1 - J_0)^q\} 2^{-J_0} \sum_{j=J_1+1}^J 2^j n_j \right) .$$

By the Markov inequality and $\sum_{j=J_1+1}^J 2^j n_j \leq n(J - J_1) = O(n \log(n 2^{-J_1}))$, we get

$$\begin{aligned} \sup_{\boldsymbol{\mu} \in \mathcal{B}(1, q, \delta)} \left| \tilde{\mathbf{S}}_{n, J_1+1, J}(\bar{\boldsymbol{\mu}}) \right| \\ = O_{\mathbb{P}_f} \left(2^{(\delta-1)(J_1-J_0)} \{1 + (J_1 - J_0)^q\} n 2^{-J_0} \log(n 2^{-J_1}) \right) . \end{aligned} \quad (41)$$

We may set $J_1 = J_0 + \lceil \log_2(n 2^{-J_0}) / (3 - 2\delta) \rceil$ as under the assumptions of (b), this implies $n 2^{-J_1} \rightarrow \infty$. For this choice of J_1 , the conclusion in (b) then follows from (39), (40), (41) and by using that $\{1 + (J_1 - J_0)^q\} = O\{\log^q(n 2^{-J_0})\}$ and $\log(n 2^{-J_1}) \leq \log(n 2^{-J_0})$. \square

Corollary 7. *Conditions (21) and (22) imply the two following bounds.*

(a) *For any $\ell \geq 0$, if $J_0^{-1} + (n 2^{-J_0})^{-1} \rightarrow 0$, then*

$$\sup_{d \in \mathbb{R}} \left| \mathbb{E}_{\mathcal{I}_n(J_0, J_0+\ell)}(d) \right| = O_{\mathbb{P}_f} \left(2^{-\beta J_0} + (n 2^{-J_0})^{-1/2} \right) .$$

(b) *For all $d_* > d_0 - 1/2$, if $J_0^{-1} + (n 2^{-J_0})^{-1} \rightarrow 0$,*

$$\sup_{d \geq d_*} \left| \mathbb{E}_{\mathcal{I}_n(J_0, J)}(d) \right| = O_{\mathbb{P}_f} \left(2^{-\beta J_0} + \log(n 2^{-J_0}) (n 2^{-J_0})^{-\{1-2(d_0-d_*)\}/\{3-4(d_0-d_*)\}} \right) .$$

Proof. In view of (33) and (36), we have

$$\mathbb{E}_{\mathcal{I}_n(J_0, J_1)}(d) = \log \left[1 + (n 2^{-J_0})^{-1} \tilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}) \right]$$

with $\boldsymbol{\mu} = \{\mu_j\}_{j \geq 0}$ defined by

$$\mu_j \stackrel{\text{def}}{=} n 2^{-J_0} \frac{2^{2(d_0-d)(j+J_0)}}{\sum_{j'=J_0}^{J_1} 2^{2(d_0-d)j'} n_{j'}} \mathbb{1}(j \leq J_1 - J_0) .$$

Part (a) We set $J_1 = J_0 + \ell$ and apply Proposition 6 (a) with $\delta = 0$. Indeed, observe that $\mu_j \leq n2^{-J_0} \sup_{j=0, \dots, \ell} n_{j+J_0}^{-1} = n2^{-J_0}/n_{J_0} \rightarrow 1$ as $n2^{-J_0} \rightarrow \infty$. Then $\boldsymbol{\mu} \in \mathcal{B}(2, 0, 0)$ for all $d \in \mathbb{R}$, provided that n is large enough.

Part (b) Set now $J_1 = J$ and apply Proposition 6 (b) with $\delta \stackrel{\text{def}}{=} 2(d_0 - d_*) < 1$. Then, since, for all $d \geq d_*$, $2^{2(d_0-d)(j+J_0)} \leq 2^{\delta(j+J_0)}$ and $\sum_{j'=J_0}^J 2^{2(d_0-d)j'} n_{j'} \geq 2^{\delta J_0} n_{J_0} \sim n2^{(\delta-1)J_0}$ as $n2^{-J_0} \rightarrow \infty$, we get $\boldsymbol{\mu} \in \mathcal{B}(2, 0, \delta)$ for all $d \geq d_*$, provided that n is large enough. \square

Notation convention. The main differences between cases (C-1) where $J_1 - J_0$ is fixed and (C-2) where $J_1 = J$ lie in Proposition 6 and in the computations of the asymptotic variances in Theorem 4 (CLT). Therefore, in the following, we shall often write, for convenience, $\mathcal{I}, \eta, \widehat{d}, \widehat{\mathbf{S}}(d), \widetilde{\mathbf{S}}(\boldsymbol{\mu}), \widetilde{\mathbf{L}}, \mathbf{E}$ and \mathbf{L} for $\mathcal{I}_n(J_0, J_1), \eta(\mathcal{I}_n(J_0, J_1)), \widehat{d}_{\mathcal{I}_n(J_0, J_1)}, \widehat{\mathbf{S}}_{\mathcal{I}_n(J_0, J_1)}(d), \widetilde{\mathbf{S}}_{n, J_0, J_1}(\boldsymbol{\mu}), \widetilde{\mathbf{L}}_{\mathcal{I}_n(J_0, J_1)}, \mathbf{E}_{\mathcal{I}_n(J_0, J_1)}$ and $\mathbf{L}_{\mathcal{I}_n(J_0, J_1)}$ respectively, with either $J_1 = J_0 + \ell$ or $J_1 = J$. We will use the complete notation when the distinction between these two cases is necessary.

4. PROOF OF THEOREM 2

In view of (18), the assumptions of the theorem are equivalent to $J_0^{-1} + (n2^{-J_0})^{-1} \rightarrow 0$ and 2^{J_0} and $n2^{-J_0}$ are asymptotically equivalent to n/m and m , respectively. We will develop the proof using J_0 . By definition of \widehat{d} and (31), we have

$$0 \geq \widetilde{\mathbf{L}}(\widehat{d}) - \widetilde{\mathbf{L}}(d_0) = \mathbf{L}(\widehat{d}) + \mathbf{E}(\widehat{d}) - \mathbf{E}(d_0). \quad (42)$$

The basic idea for proving consistency is to show that 1) the function $d \mapsto \widetilde{\mathbf{L}}(d)$ behaves as $(d - d_0)^2$ up to a multiplicative positive constant and 2) the function $d \mapsto \mathbf{E}(d)$ tends to zero in probability uniformly in d . Proposition 5 will provide all the results needed for 1) and Corollary 7 (a) and (b) will respectively provide those needed for 2) when either $J_1 = J_0 + \ell$ or $J_1 = J$. The case $J_1 = J$ will be more involved because $d \mapsto \mathbf{E}_{\mathcal{I}_n(J_0, J)}(d)$ tends to zero in probability uniformly only over $d \geq d_{\min}$ with $d_{\min} > d_0 - 1/2$ (see Corollary 7(b)).

Proof of Part (a) Here $J_1 = J_0 + \ell$ with a fixed $\ell \geq 1$. Using classical argument for contrast estimation (similar to those detailed in Step 2 and Step 3 in the proof of Part (b) below), (42), the properties of $\mathbf{L}(d)$ established in Proposition 5 and Corollary 7(a) give the result.

Proof of Part (b) Here $J_1 = J$. The proof proceeds in three steps, namely

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

$$\mathbb{P}_f\{\widehat{d} \leq d_{\min}\} \rightarrow 0. \quad (43)$$

Step 2 For any $d_{\max} > d_0$,

$$\mathbb{P}_f\{\widehat{d} \geq d_{\max}\} \rightarrow 0. \quad (44)$$

Step 3 For all $d_{\min} \in (d_0 - 1/2, d_0)$ and $d_{\max} > d_0$,

$$\begin{aligned} \mathbb{1}(d_{\min} \leq \widehat{d} \leq d_{\max}) (\widehat{d} - d_0)^2 = \\ O_{\mathbb{P}_f}(2^{-\beta J_0}) + O_{\mathbb{P}_f}\left(\log(n2^{-J_0}) (n2^{-J_0})^{-\{1-2(d_0-d_{\min})\}/\{3-4(d_0-d_{\min})\}}\right). \end{aligned} \quad (45)$$

Proof of Step 1. We have, for all $d \in \mathbb{R}$,

$$\begin{aligned} \widetilde{\mathbb{L}}(d) - \widetilde{\mathbb{L}}(d_0) = \\ \log\left(\frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} 2^{2(d-d_0)(\eta-j)} \frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}}\right) - \log\left(\frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} \frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}}\right). \end{aligned} \quad (46)$$

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

$$w_j(d) \stackrel{\text{def}}{=} \begin{cases} 2^{2(j-\eta)(d_0-d)} & \text{for } j \leq \eta \\ 2^{2(j-\eta)(d_0-d_{\min})} & \text{otherwise.} \end{cases} \quad (47)$$

so that, for all j and $d \leq d_{\min}$, $w_j(d) \leq 2^{2(j-\eta)(d_0-d)}$. We further obtain

$$\begin{aligned} \widetilde{\mathbb{L}}(d) - \widetilde{\mathbb{L}}(d_0) \\ \geq \log\left(\frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} w_j(d) \frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}}\right) - \log\left(\frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} \frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}}\right) \\ = \log \frac{\Sigma(d) + \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} w_j(d) \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}} - 1\right)}{1 + \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}} - 1\right)}, \end{aligned} \quad (48)$$

where

$$\Sigma(d) \stackrel{\text{def}}{=} \frac{1}{\#\mathcal{I}} \sum_{(j,k) \in \mathcal{I}} w_j(d).$$

The proof consists of showing that one may choose $d_{\min} \in (d_0 - 1/2, d_0)$ so that $\inf_{d \leq d_{\min}} \Sigma(d)$ is strictly larger than 1 for n large enough and that the two

(stochastic) terms involving summations converge to zero in probability uniformly for $d \leq d_{\min}$ (the second sum does not depend on d). Since by construction $\tilde{L}(\hat{d}) \leq \tilde{L}(d_0)$, this implies that $\mathbb{P}_f\{\hat{d} \leq d_{\min}\} \leq \mathbb{P}_f\{\inf_{d \leq d_{\min}} \tilde{L}(d) \leq \tilde{L}(d_0)\} \rightarrow 0$, which establishes (43).

By Lemma 11, we have, for n large enough,

$$J_0 \leq \eta < J_0 + 2. \quad (49)$$

We shall now use (47) with $j \geq J_0 + 2 > \eta$. Since $d_0 > d_{\min}$, we get, for n large enough,

$$\begin{aligned} \inf_{d \leq d_{\min}} \Sigma(d) &\geq \frac{1}{\#\mathcal{I}} \sum_{j=J_0+2}^J 2^{2(j-\eta)(d_0-d_{\min})} n_j \\ &\geq \frac{2^{-2\eta(d_0-d_{\min})}}{\#\mathcal{I}} \sum_{j=J_0+2}^J 2^{2j(d_0-d_{\min})} \{n2^{-j} - 2(\mathbb{T} - 1)\} \end{aligned}$$

Using that $n2^{-J_0} \rightarrow \infty$, $n \asymp 2^J$ and that $2(d_0 - d_{\min}) - 1 < 0$, straightforward computations give that the term involving a summation in the previous display is asymptotically equivalent to $n2^{\{2(d_0-d_{\min})-1\}J_0} (4\{1 - 2^{2(d_0-d_{\min})}\})^{-1}$. By Lemma 11, we have $\#\mathcal{I} \sim 2n2^{-J_0}$ and $\eta - J_0 \rightarrow 1$. Hence,

$$\liminf_{n \rightarrow \infty} \inf_{d \leq d_{\min}} \Sigma(d) \geq \frac{2^{-2(d_0-d_{\min})}}{8\{1 - 2^{2(d_0-d_{\min})}\}}.$$

There is a value of $d_{\min} \in (d_0 - 1/2, d_0)$ such that the RHS in the previous display is strictly larger than 1. For such choice,

$$\inf_{d \leq d_{\min}} \Sigma(d) > 1. \quad (50)$$

for n large enough.

Let us now check that the two terms in (48) involving summations converge to zero uniformly in $d \leq d_{\min}$. The one in the denominator equals $\tilde{\mathbf{S}}(\mathbf{1})/(\#\mathcal{I})$, and thus converges to zero by Proposition 6 (b), since $\#\mathcal{I} \sim 2n2^{-J_0}$ and $J_0^{-1} + (n2^{-J_0})^{-1} \rightarrow \infty$.

The one in the numerator equals $\tilde{\mathbf{S}}(\{w_{J_0+j}(d)\}_{j \geq 0})/(\#\mathcal{I})$. Using again Proposition 6 (b), $\#\mathcal{I} \sim 2n2^{-J_0}$ and $J_0^{-1} + (n2^{-J_0})^{-1} \rightarrow \infty$ yield

$$\sup_{d \leq d_{\min}} \frac{2^{-2(d_0-d)J_0}}{\#\mathcal{I}} \tilde{\mathbf{S}}(\{w_{J_0+j}(d)\}_{j \geq 0}) = o_{\mathbb{P}_f}(1),$$

provided that there exist $C > 0$, $q \geq 0$ and $\delta < 1$ such that $\{w_{J_0+j}(d)\}_{j \geq 0} \in \mathcal{B}(C, q, \delta)$ for all $d \leq d_{\min}$ and all n . It now only remains to check this latter fact. Let $d \leq d_{\min}$. Applying the definition (47) and using that $d_0 - d \geq d_0 - d_{\min}$ in the case $J_0 + j \leq \eta$, we have, for all $j \geq 0$,

$$w_{J_0+j}(d) \leq 2^{2(J_0+j-\eta)(d_0-d_{\min})} \leq 2^{2j(d_0-d_{\min})} ,$$

where the last inequality follows by observing that $J_0 \leq \eta$. Hence $\{w_{J_0+j}(d)\}_{j \geq 0} \in \mathcal{B}(1, 0, 2(d_0 - d_{\min}))$ for all $d \leq d_{\min}$. Since $2(d_0 - d_{\min}) < 1$, this achieves the proof of Step 1. \square

Proof of Step 2 : By (42), $L(\widehat{d}) \leq E(d_0) - E(\widehat{d})$ so that, for any $d_{\max} \geq d_0$, one has $\inf_{d \geq d_{\max}} L(d) \leq 2 \sup_{d \geq d_{\max}} |E(d)|$ on the event $\{\widehat{d} \geq d_{\max}\}$. Since the function $d \mapsto L(d)$ vanishes at $d = d_0 < d_{\max}$ and is convex (see Proposition 5), we have

$$\inf_{d \geq d_{\max}} L(d) = L(d_{\max}) \geq \frac{1}{2}(d_{\max} - d_0)^2 \inf_{d \in [d_0, d_{\max}]} \ddot{L}(d) .$$

Using Proposition 5, there exists $c > 0$ such that, for all n sufficiently large, $\inf_{d \in [d_0, d_{\max}]} \ddot{L}(d) \geq c$. Thus, $\mathbb{P}_f\{\widehat{d} \geq d_{\max}\} \leq \mathbb{P}_f\{c(d_0 - d_{\max})^2 \leq E(d)\}$, which converges to zero by corollary 7-(b). \square

Proof of Step 3 : Let $d_{\min} \in (d_0 - 1/2, d_0)$ and $d_{\max} > d_0$. Relation (42) implies

$$\mathbb{1}\{d_{\min} \leq \widehat{d} \leq d_{\max}\}L(\widehat{d}) \leq 2 \sup_{d \geq d_{\min}} |E(d)| \quad (51)$$

When $d_{\min} \leq \widehat{d} \leq d_{\max}$, we have $L(\widehat{d}) \geq \frac{1}{2}(\widehat{d} - d_0)^2 \inf_{d \in [d_{\min}, d_{\max}]} \ddot{L}(d)$; hence, by Proposition 5, there exists $c > 0$ such that, for n large enough,

$$\mathbb{1}\{d_{\min} \leq \widehat{d} \leq d_{\max}\}(\widehat{d} - d_0)^2 \leq c \sup_{d \geq d_{\min}} |E(d)| .$$

Corollary 7 (b) then gives (45). \square

5. PROOF OF THEOREM 3

We prove here Theorem 3 which concerns rates of convergence. By (18), the assumptions are equivalent to $n2^{-J_0} \asymp n^\gamma$ for $\gamma \in (0, 1)$ and 2^{J_0} and $n2^{-J_0}$ are asymptotically equivalent to n/m and m , respectively. We thus again develop

the proof using J_0 . We established in Section 1.3 that the estimator \widehat{d} satisfies $\widehat{\mathbf{S}}(\widehat{d}) = 0$. Performing a Taylor expansion of $\widehat{\mathbf{S}}(d)$ around $d = \widehat{d}$ gives

$$\widehat{\mathbf{S}}(d_0) = 2 \log(2) (\widehat{d} - d_0) \sum_{(j,k) \in \mathcal{I}} [j - \eta] j \frac{W_{j,k}^2}{2^{2\widehat{d}j}}. \quad (52)$$

where \widetilde{d} lies between d_0 and \widehat{d} . The proof of Theorem 3 now consists in bounding $\widehat{\mathbf{S}}(d_0)$ from above (Step 1) and in showing that $\sum_{\mathcal{I}} [j - \eta] j W_{j,k}^2 / 2^{2\widetilde{d}j}$ has a positive limit (Step 2).

Proof of Step 1. We define the empirical variance of the wavelet coefficients at scale index j as

$$\widehat{\sigma}_j^2 \stackrel{\text{def}}{=} \frac{1}{n_j} \sum_{k=0}^{n_j-1} W_{j,k}^2. \quad (53)$$

Using that $\sum_{j=J_0}^{J_1} (j - \eta) n_j = 0$ we may then write (16) as

$$\widehat{\mathbf{S}}(d_0) = \sum_{j=J_0}^{J_1} (j - \eta) n_j \{2^{-2jd_0} \widehat{\sigma}_j^2 - \sigma^2\},$$

where σ^2 is the constant appearing in (21). Applying the Minkowski inequality in (16), and $\mathbb{E}_f[\widehat{\sigma}_j^2] = \sigma_j^2(f)$, we get

$$\left(\mathbb{E}_f[|\widehat{\mathbf{S}}(d_0)|^2] \right)^{1/2} \leq \sum_{j=J_0}^{J_1} |j - \eta| \left\{ \text{Var}_f^{1/2} \left(\frac{n_j}{2^{2jd_0}} \widehat{\sigma}_j^2 \right) + n_j \left| \frac{\sigma_j^2(f)}{2^{2jd_0}} - \sigma^2 \right| \right\}.$$

Condition (21) and Condition (22) imply that the term between curly brackets is $O(n_j^{1/2} + n_j 2^{-\beta j})$. Hence, since $n_j \leq n 2^{-j}$ and $\eta - J_0$ is bounded independently of n (see Lemma 11), we finally obtain

$$\widehat{\mathbf{S}}(d_0) = O_{\mathbb{P}_f} \left((n 2^{-J_0})^{1/2} + n 2^{-(1+\beta)J_0} \right). \quad (54)$$

□

Proof of Step 2. We will establish the following asymptotic relations:

$$(n 2^{-J_0})^{-1} \sum_{(j,k) \in \mathcal{I}_n(J_0, J_0 + \ell)} [j - \eta(\mathcal{I}_n(J_0, J_0 + \ell))] j \frac{W_{j,k}^2}{2^{2\widetilde{d}j}} \xrightarrow{\mathbb{P}_f} (2 - 2^{-\ell}) \kappa_\ell \sigma^2, \quad (55)$$

where \tilde{d} lies between d_0 and $\widehat{d}_{\mathcal{I}_n(J_0, J_0+\ell)}$ and κ_ℓ is defined in (28), and

$$(n2^{-J_0})^{-1} \sum_{(j,k) \in \mathcal{I}_n(J_0, J)} [j - \eta(\mathcal{I}_n(J_0, J))] j \frac{W_{j,k}^2}{2^{2\tilde{d}j}} \xrightarrow{\mathbb{P}_f} 4\sigma^2, \quad (56)$$

where \tilde{d} lies between d_0 and $\widehat{d}_{\mathcal{I}_n(J_0, J)}$. To prove (55) and (56), write

$$\begin{aligned} & \left| \sum_{(j,k) \in \mathcal{I}} (j - \eta) j \frac{W_{j,k}^2}{2^{2\tilde{d}j}} - \sum_{(j,k) \in \mathcal{I}} (j - \eta) j \frac{W_{j,k}^2}{2^{2d_0j}} \right| \\ & \leq \left[\sup_{j=J_0, \dots, J_1} |2^{2(d_0 - \tilde{d})j} - 1| \right] \sum_{(j,k) \in \mathcal{I}} |j - \eta| j \frac{W_{j,k}^2}{2^{2d_0j}}. \quad (57) \end{aligned}$$

Applying Theorem 2 with $J_1 = J_0 + \ell$ or $J_1 = J$, there exists $\gamma > 0$ such that

$$|\tilde{d} - d_0| \leq |\widehat{d} - d_0| = O_{\mathbb{P}_f} \left(2^{-\beta J_0/2} + (n2^{-J_0})^{-\gamma} \right).$$

Notice that $\sup_{j \leq J_1} |2^{2jx} - 1| \leq 2^{2J_1|x|} - 1 = O(J_1|x|)$ as $J_1x \rightarrow 0$. Since $n2^{-J_0} \asymp n^\gamma$ with $\gamma \in (0, 1)$, we obtain $J_0 \sup_{j=J_0, \dots, J_1} |2^{2(d_0 - \tilde{d})j} - 1| = o_{\mathbb{P}_f}(1)$. By Condition (21), $\{2^{-2d_0j} \sigma_j^2(f)\}_{j \geq 0}$ is a bounded sequence; hence there exists a constant $C > 0$ such that

$$\begin{aligned} \mathbb{E}_f \sum_{(j,k) \in \mathcal{I}} |j - \eta| j \frac{W_{j,k}^2}{2^{2d_0j}} & \leq C \sum_{(j,k) \in \mathcal{I}} j |j - \eta| \\ & \leq C \left\{ \sum_{j=J_0}^{J_1} (j - \eta)^2 n_j + \eta \sum_{j=J_0}^{J_1} |j - \eta| n_j \right\} = O(J_0 n2^{-J_0}), \end{aligned}$$

where the last bound follows from Lemma 11. Inserting the last two displays in (57), we obtain

$$\begin{aligned} \sum_{(j,k) \in \mathcal{I}} (j - \eta) j \frac{W_{j,k}^2}{2^{2\tilde{d}j}} & = o_{\mathbb{P}_f}(n2^{-J_0}) + \\ & \sigma^2 \left(\sum_{(j,k) \in \mathcal{I}} (j - \eta) j + \sum_{(j,k) \in \mathcal{I}} (j - \eta) j \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0j}} - 1 \right) \right). \quad (58) \end{aligned}$$

From Lemma 11, we get that $(n2^{-J_0})^{-1} \sum_{(j,k) \in \mathcal{I}} (j - \eta) j$ converges to $(2 - 2^{-\ell})\kappa_\ell$ when $J_1 = J_0 + \ell$ and to 4 when $J_1 = J$. This establishes (55) and (56).

We finally show that the random term in (58) is negligible. By Proposition 6 (a) for $J_1 = J_0 + \ell$ or Proposition 6 (b) for $J_1 = J$ and by using (49) and $n2^{-J_0} \asymp n^\gamma$

with $\gamma \in (0, 1)$, we have, for any $q, q' \geq 0$, $\eta^{q'} \sum_{(j,k) \in \mathcal{I}} (j-\eta)^q (W_{j,k}^2 / \sigma^2 2^{2d_0 j} - 1) = o_{\mathbb{P}_f}(n2^{-J_0})$. Thus writing $(j-\eta)j = (j-\eta)^2 + \eta(j-\eta)$, we conclude

$$(n2^{-J_0})^{-1} \sum_{(j,k) \in \mathcal{I}} (j-\eta)j \left(\frac{W_{j,k}^2}{\sigma^2 2^{2d_0 j}} - 1 \right) = o_{\mathbb{P}_f}(1).$$

□

6. THE SPECTRAL DENSITY OF WAVELET COEFFICIENTS: DEFINITIONS AND APPROXIMATIONS

We provide some definitions and approximation results introduced and obtained in Moulines et al. (2005) and then prove Theorem 1.

A convenient way to define a joint spectral density for wavelet coefficients is to consider the *between-scale* process $\{[W_{j,k}^X \mathbf{W}_{j,k}^X(j-j')^T]^T\}_{k \in \mathbb{Z}}$, where for any $j \geq u \geq 0$,

$$\mathbf{W}_{j,k}^X(u) \stackrel{\text{def}}{=} [W_{j-u, 2^u k}^X, \dots, W_{j-u, 2^u k + 2^u - 1}^X]^T, \quad (59)$$

and the *within-scale* process $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$, which is the between-scale process with $u = 0$. For all $j \geq 0$, both the within-scale process and the between-scale process are covariance stationary in k . We denote by $\mathbf{D}_{j,0}(\cdot; f, \phi, \psi)$ the spectral density of the within-scale process at scale index j and by $\mathbf{D}_{j,u}(\cdot; f, \phi, \psi)$ respectively the cross spectral density between $\{W_{j,k}^X\}_{k \in \mathbb{Z}}$ and $\{\mathbf{W}_{j,k}^X(j-j')^T\}_{k \in \mathbb{Z}}$. Applying (Moulines et al., 2005, Theorem 3), in the setting of Section 2, under (W-1)-(W-4), we obtain the following approximations for these quantities.

- (a) If $d_0 \in ((1+\beta)/2 - \alpha, M + 1/2)$, then, there exists a constant $C > 0$ such that, for all $j \geq 0$,

$$|\sigma_j^2(f) - f^*(0) \mathbf{K}(d_0, \psi) 2^{2jd_0}| \leq C f^*(0) L 2^{(2d_0 - \beta)j} \quad (60)$$

where $\sigma_j^2(f)$ is defined in (20) and $\mathbf{K}(d, \psi)$ in (23);

- (b) If $d_0 \in ((1+\beta)/2 - \alpha, M]$, then, for all $u \geq 0$, there exists $C > 0$ such that, for all $\lambda \in (-\pi, \pi)$ and $j \geq 0$,

$$|\mathbf{D}_{j,u}(\lambda; f, \phi, \psi) - f^*(0) \mathbf{D}_{\infty,u}(\lambda; d_0, \psi) 2^{2jd_0}| \leq C f^*(0) L 2^{(2d_0 - \beta)j} \quad (61)$$

where, for all $u \geq 0$ and $\lambda \in (-\pi, \pi)$,

$$\mathbf{D}_{\infty,u}(\lambda; d, \psi) \stackrel{\text{def}}{=} \sum_{l \in \mathbb{Z}} |\lambda + 2l\pi|^{-2d} \mathbf{e}_u(\lambda + 2l\pi) \overline{\widehat{\psi}(\lambda + 2l\pi)} \widehat{\psi}(2^{-u}(\lambda + 2l\pi)). \quad (62)$$

and where $\mathbf{e}_u(\xi) \stackrel{\text{def}}{=} [1, e^{-i2^{-u}\xi}, \dots, e^{-i(2^u-1)2^{-u}\xi}]^T$. The spectral density defined in (62) is related to the quantity appearing in the expressions of the asymptotic variance of the CLT as follows. From (27), we have

$$I_u(d, \psi) = \int_{-\pi}^{\pi} |\mathbf{D}_{\infty, u}(\lambda; d, \psi)|^2 d\lambda.$$

Proof of Theorem 1. Condition (21) follows from (60) and by observing that $f^*(0)$ and $K(d_0, \psi)$ are positive constants.

From (60) and (61), we have that, for all $j \geq 0$, $\{W_{j,k}/\sqrt{\sigma_j^2(f)}, k \in \mathbb{Z}\}$ is a stationary unit variance sequence whose spectral density is bounded above by a constant independent of j . Now, Condition (22) is a simple consequence of Lemma 8. \square

7. PROOF OF THEOREM 4

We prove here Theorem 4 (CLT). As the estimator is invariant under multiplication of the observations by a positive constant, we may set, without loss of generality, $f^*(0) = 1$.

By Theorem 1, Conditions (21) and (22) with $\sigma^2 = K(d_0, \psi)$ under the stated assumptions. Thus using (18), (52) with (55) and (56), respectively, we obtain

$$m_\ell^{1/2} (\widehat{d}_{\mathcal{I}_n(J_0, J_0+\ell)} - d_0) = \frac{m_\ell^{-1/2} \widehat{\mathbf{S}}_{\mathcal{I}_n(J_0, J_0+\ell)}(d_0)}{2 \kappa_\ell \sigma^2 \log(2)} (1 + o_{\mathbb{P}_f}(1)); \quad (63)$$

$$m^{1/2} (\widehat{d}_{\mathcal{I}_n(J_0, J)} - d_0) = \frac{m^{-1/2} \widehat{\mathbf{S}}_{\mathcal{I}_n(J_0, J)}(d_0)}{4 \sigma^2 \log(2)} (1 + o_{\mathbb{P}_f}(1)) \quad (64)$$

Let us write

$$\widehat{\mathbf{S}}(d_0) = \mathbb{E}_f [\widehat{\mathbf{S}}(d_0)] + \left(\widehat{\mathbf{S}}(d_0) - \mathbb{E}_f [\widehat{\mathbf{S}}(d_0)] \right). \quad (65)$$

Since $\sum_{(j,k) \in \mathcal{I}} (j - \eta)$ vanishes and $\mathbb{E}_f W_{j,k}^2 = \sigma_j^2(f)$, we have

$$\mathbb{E}_f \widehat{\mathbf{S}}(d_0) = \sum_{(j,k) \in \mathcal{I}} (j - \eta) (2^{-2d_0 j} \sigma_j^2(f) - \sigma^2).$$

Applying (18), (21), (49) and $n_j \leq n2^{-j}$, we get

$$\mathbb{E}_f \widehat{\mathbf{S}}(d_0) = O\left(\left(\frac{m}{n}\right)^\beta\right). \quad (66)$$

The following assertions proved below then provide the needed convergence results on $\widehat{\mathbf{S}}(d_0)$.

$$m_\ell^{-1/2} \frac{\widehat{\mathbf{S}}_{\mathcal{I}(J_0, J_0+\ell, J)}(d_0) - \mathbb{E}_f \widehat{\mathbf{S}}_{\mathcal{I}(J_0, J_0+\ell, J)}(d_0)}{2 \kappa_\ell \sigma^2 \log(2)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}(d_0, \ell, \psi)); \quad (67)$$

$$m^{-1/2} \frac{\widehat{\mathbf{S}}_{\mathcal{I}(J_0, J)}(d_0) - \mathbb{E}_f \widehat{\mathbf{S}}_{\mathcal{I}(J_0, J)}(d_0)}{4 \sigma^2 \log(2)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}(d_0, \psi)). \quad (68)$$

Eq. (66) provides a bound for the bias term in the decomposition (65), while (67) and (68) provide the Gaussian limit of the fluctuation term in the two cases $J_1 - J_0 = \ell$ fixed and $J_1 = J$. The condition on γ moreover implies that (66) tends to zero so that the result follows from Slutsky's lemma.

Proof of (67) and (68). We first establish the following limits. Let $\mathbf{I}_u(d, \psi)$, $u \in \mathbb{N}$ be defined as in (27). Then,

$$m_\ell^{-1} \text{Var}_f \left(\widehat{\mathbf{S}}_{\mathcal{I}_n(J_0, J_0+\ell)}(d_0) \right) \rightarrow 4 \pi \sum_{j_0, j_1=0}^{\ell} \frac{(j_0 - \eta_\ell)(j_1 - \eta_\ell) \mathbf{I}_{|j_1-j_0|}(d_0, \psi)}{(2 - 2^{-\ell}) 2^{2d_0|j_1-j_0|+(j_1 \vee j_0)}}, \quad (69)$$

$$m^{-1} \text{Var}_f \left(\widehat{\mathbf{S}}_{\mathcal{I}_n(J_0, J)}(d_0) \right) \rightarrow 2 \pi \sum_{j_0, j_1=0}^{\ell} \frac{(j_0 - 1)(j_1 - 1) \mathbf{I}_{|j_1-j_0|}(d_0, \psi)}{2^{2d_0|j_1-j_0|+(j_1 \vee j_0)}} \quad (70)$$

Expressing the RHS of (16) as a weighted sum of the empirical variances $\widehat{\sigma}_j^2$, $j = J_0, \dots, J_1$ defined in (53), we get

$$(n2^{-J_0})^{-1} \text{Var}_f \left(\widehat{\mathbf{S}}(d_0) \right) = \sum_{j_0, j_1=0}^{J_1-J_0} u_{j_0, j_1}, \quad (71)$$

where

$$u_{j_0, j_1} \stackrel{\text{def}}{=} \frac{n2^{-J_0}}{2^{4d_0 J_0}} \left[\prod_{j=j_0, j_1} \frac{j - (\eta - J_0)}{2^{(1+2d_0)j}} \right] \text{Cov}_f \left(\widehat{\sigma}_{J_0+j_0}^2, \widehat{\sigma}_{J_0+j_1}^2 \right). \quad (72)$$

By (49), we have, for n large enough and all $j \geq 0$, $|j - (\eta - J_0)| \leq j + 2$. Conditions (21) and (22) imply that, for some constant C , for all $j \geq 0$,

$$\text{Var}_f \left(\widehat{\sigma}_{J_0+j}^2 \right) \leq C (n2^{-J_0})^{-1} 2^{j+4d_0(J_0+j)}.$$

The two last bounds and the Cauchy-Schwarz inequality give that, for n large enough,

$$|u_{j_0, j_1}| \leq C (j_0 + 2)(j_1 + 2) 2^{-(j_0+j_1)/2} \quad (73)$$

which is summable in $j_0, j_1 \geq 0$ and does not depend on n . Hence, by dominated convergence, to show (70), we may compute the limit of u_{j_0, j_1} for any given fixed indices j_0 and j_1 and then compute the sum of the limits over j_0 and j_1 . As shown in (Moulines et al., 2005, Theorem 5), we have, for any fixed $j_0 \leq j_1$,

$$n2^{-J_0-4d_0J_0} \text{Cov}_f(\widehat{\sigma}_{J_0+j_0}^2, \widehat{\sigma}_{J_0+j_1}^2) \rightarrow 4\pi 2^{j_0} 2^{4d_0j_1} \mathbf{I}_{j_1-j_0}(d_0, \psi), \quad (74)$$

where $\mathbf{I}_{j_1-j_0}(d_0, \psi)$ is defined in (27). The limits of $\eta - J_0$ are computed in Lemma 11 for $J_1 = J_0 + \ell$ and $J_1 = J$ so that (72) and (74) yield

$$u_{j_0, j_1} \rightarrow \frac{4\pi (j_0 - \eta_\ell) (j_1 - \eta_\ell)}{2^{2d_0(j_0-j_1)+j_1}} \mathbf{I}_{j_1-j_0}(d_0, \psi) \quad (75)$$

if $J_1 = J_0 + \ell$ and

$$u_{j_0, j_1} \rightarrow \frac{4\pi (j_0 - 1) (j_1 - 1)}{2^{2d_0(j_0-j_1)+j_1}} \mathbf{I}_{j_1-j_0}(d_0, \psi) \quad (76)$$

if $J_1 = J$. The case $j_0 > j_1$ is obtained by symmetry. Up to the multiplicative constants appearing in (18), summing the RHS of (75) over $j_0, j_1 = 0, 1, \dots, \ell$ gives (69), while summing the RHS of (76) over $j_0, j_1 = 0, 1, \dots$ gives (70).

The normal limits (67) and (68) now follow from Lemma 10. By proving (69) and (70), we already computed the limits of the variances. It is indeed a simple verification to show that the RHS of (69) (resp. (70)) divided by $(2\kappa_\ell \sigma^2 \log(2))^2$ (resp. $(4\sigma^2 \log(2))^2$) gives $\mathbf{V}(d_0, \ell, \psi)$ (resp. $\mathbf{V}(d_0, \psi)$). It remains to check (79). Let A_n and Γ_n be the square matrices indexed on $\mathcal{I} \times \mathcal{I}$ and defined as follows.

(1) A_n is the diagonal matrix such that $[A_n]_{(j,k), (j,k)} = (n2^{-J_0})^{-1/2} \text{sign}(j - \eta)$ for all $(j, k) \in \mathcal{I}$,

(2) Γ_n is the covariance matrix of the vector $[|j - \eta|^{1/2} 2^{-d_0j} W_{j,k}]_{(j,k) \in \mathcal{I}}$.

Of course $\rho[A_n] = (n2^{-J_0})^{-1/2}$. Moreover, $\rho[\Gamma_n] \leq \sum_{j=J_0}^{J_1} \rho[\Gamma_{n,j}]$, where $\Gamma_{n,j}$ is the covariance matrix of the vector $[|j - \eta|^{1/2} 2^{-d_0j} W_{j,k}]_{k=0, \dots, n_j-1}$. Since $\{W_{j,k}\}_{k \in \mathbb{Z}}$ is a stationary time series, by Lemma 9,

$$\rho[\Gamma_{n,j}] \leq |j - \eta| 2^{-2d_0j} 2\pi \sup_{\lambda \in (-\pi, \pi)} \mathbf{D}_{j,0}(\lambda; f).$$

From (61), noticing that $\mathbf{D}_{\infty,0}(\cdot; d_0)$ is bounded on $(-\pi, \pi)$, we then get, for a constant C not depending on n ,

$$\rho[\Gamma_n] \leq C \sum_{j=J_0}^{J_1} |j - \eta|.$$

By (49), the latter sum is $O((1 + J_1 - J_0)^2)$. Hence

$$\rho[A_n]\rho[\Gamma_n] = O((n2^{-J_0})^{-1/2}(J_1 - J_0)^2)$$

so that the conditions of Lemma 10 are met for both cases $J_1 = J_0 + \ell$ and $J_1 = J$, which completes the proof.

APPENDIX A. AUXILIARY RESULTS ON GAUSSIAN VECTORS

Lemma 8. *Let $\xi = \{\xi_\ell, \ell \in \mathbb{Z}\}$ be a stationary centered Gaussian process with spectral density g . Then, for all $n \geq 1$,*

$$\text{Var} \left(\sum_{\ell=1}^n \xi_\ell^2 \right) \leq 4\pi n \int_{-\pi}^{\pi} g^2(\lambda) d\lambda, \quad (77)$$

$$\text{Var} \left(\sum_{\ell=1}^n \xi_\ell^2 \right) \leq 4\pi n \text{Var}(\xi_1) \|g\|_\infty. \quad (78)$$

Inequality (77) becomes an asymptotic equivalence as $n \rightarrow \infty$ if $\int_{-\pi}^{\pi} g^2(\lambda) d\lambda < \infty$.

Proof. Denote by $\gamma(\cdot)$ the auto-covariance function of ξ . Because ξ is Gaussian, $\{\xi_\ell^2\}$ is also stationary and its autocovariance function is given by $2\gamma^2(\cdot)$. Thus,

$$\begin{aligned} \text{Var} \left(\sum_{\ell=1}^n \xi_\ell^2 \right) &= 2n \sum_{\tau=-n+1}^{n-1} \left(1 - \frac{|\tau|}{n} \right) \gamma^2(\tau) \\ &\leq 2n \sum_{\tau=-\infty}^{\infty} \gamma^2(\tau) = 4\pi n \int_{-\pi}^{\pi} g^2(\lambda) d\lambda \end{aligned}$$

by the Parseval Formula. This shows (77). The asymptotic equivalence in (77) as $n \rightarrow \infty$ follows by dominated convergence. Since $\int_{-\pi}^{\pi} g^2(\lambda) d\lambda \leq \int_{-\pi}^{\pi} g(\lambda) d\lambda \|g\|_\infty$, one also gets (78). \square

Let $\rho(A)$ denote the spectral radius of the square matrix A , that is, the maximum of the absolute value of its eigenvalues.

Lemma 9. *Let $\{\xi_\ell, \ell \in \mathbb{Z}\}$ be a stationary process with spectral density g and let Γ_n be the covariance matrix of $[\xi_1, \dots, \xi_n]$. Then, $\rho(\Gamma_n) \leq 2\pi \|g\|_\infty$.*

Proof. Since Γ_n is a non-negative definite matrix, $\rho(\Gamma_n) = \sup_{\mathbf{x} \in \mathbb{R}^n, |\mathbf{x}| \leq 1} \mathbf{x}^T \Gamma_n \mathbf{x}$, where $|\mathbf{x}|$ is the Euclidean norm of \mathbf{x} . For all $\mathbf{x} \in \mathbb{R}^n$, we may write

$$\begin{aligned} \mathbf{x}^T \Gamma_n \mathbf{x} &= \int_{-\pi}^{\pi} g(\lambda) \left| \sum_{\ell=1}^n \mathbf{x}_\ell e^{-i\ell\lambda} \right|^2 d\lambda, \\ &\leq \|g\|_\infty \int_{-\pi}^{\pi} \left| \sum_{\ell=1}^n \mathbf{x}_\ell e^{-i\ell\lambda} \right|^2 d\lambda = 2\pi \|g\|_\infty |\mathbf{x}|^2. \end{aligned}$$

□

Lemma 10. *Let $\{\xi_n, n \geq 1\}$ be a sequence of centered Gaussian vectors and let Γ_n be the covariance matrix of ξ_n . Let $(A_n)_{n \geq 1}$ be a sequence of deterministic symmetric matrices with adapted dimensions such that*

$$\lim_{n \rightarrow \infty} \text{Var}(\xi_n^T A_n \xi_n) = \sigma^2 \in [0, \infty).$$

Assume that

$$\lim_{n \rightarrow \infty} [\rho(A_n)\rho(\Gamma_n)] = 0. \quad (79)$$

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 let Q_n denote a $n \times k_n$ 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 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

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

and

$$\rho(\Lambda_n) = \rho(Q_n^T A_n Q_n) \leq \rho(A_n) \sup_{\|x\|=1} \|Q_n x\| = \rho(A_n) \rho(\Gamma_n) \rightarrow 0 .$$

Since $\rho(\Lambda_n) = \max_{1 \leq k \leq k_n} |\lambda_{k,n}|$, for all $\epsilon > 0$,

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

Hence the Lindeberg conditions hold as soon as $\sigma > 0$. \square

APPENDIX B. TECHNICAL RESULTS

Lemma 11. *Let $0 \leq J_0 < J_1 \leq J$ and denote $\mathcal{I}_n = \mathcal{I}_n(J_0, J_1)$. Let $\ell \geq 0$, η_ℓ and κ_ℓ be defined as in (28), $\eta(\mathcal{I})$ as in (15) and*

$$\kappa(\mathcal{I}) \stackrel{\text{def}}{=} (\#\mathcal{I})^{-1} \sum_{(j,k) \in \mathcal{I}} (j - \eta(\mathcal{I}))^2 = (\#\mathcal{I})^{-1} \sum_{(j,k) \in \mathcal{I}} j(j - \eta(\mathcal{I})) .$$

Assume that $n2^{-J_0} \rightarrow \infty$. Then for all $p = 0, 1, \dots$,

$$\eta_\ell = \frac{1 - 2^{-\ell}(1 + \ell/2)}{1 - 2^{-(\ell+1)}} ; \quad (80)$$

$$\lim_{\ell \rightarrow \infty} \kappa_\ell = 2 , \quad (81)$$

$$\left| \sum_{j=J_0}^{J_1} (j - J_0)^p n_j - n2^{-J_0} \sum_{i=0}^{J_1-J_0} i^p 2^{-i} \right| \leq 2(\mathbb{T} - 1) (J_1 - J_0)^{p+1} ; \quad (82)$$

$$\eta(\mathcal{I}_n) - J_0 = \eta_{J_1-J_0} + O(\{J_1 - J_0\}^2 n2^{-J_0}) ; \quad (83)$$

$$(n2^{-J_0})^{-1} \kappa(\mathcal{I}_n) = \kappa_{J_1-J_0} + O(\{J_1 - J_0\}^3 n2^{-J_0}) , \quad (84)$$

Note that, since $J_1 \leq J \leq \log_2(n)$, the O -terms of (83) and (84) go to zero as $n2^{-J_0} \rightarrow \infty$.

Proof (sketch). Eq (80) follows from standard computations. This gives $\eta_\ell \rightarrow 1$ as $\ell \rightarrow \infty$, and by dominated convergence $\kappa_\ell \rightarrow \sum_{i \geq 0} (i-1)^2 2^{-i} / 2$, which is equal to 2, hence (81). To show (82), use that

$$n2^{-j} - 2(\mathbb{T} - 1)n_j \leq n2^{-j} .$$

To show (83), use (82) with $p = 0, 1$ and

$$\eta(\mathcal{I}_n) = J_0 + \frac{\sum_{j=J_0}^{J_1} (j - J_0)n_j}{\sum_{j=J_0}^{J_1} n_j}$$

To show (84), use (82), (83) and

$$\kappa(\mathcal{I}_n) = \frac{\sum_{j=J_0}^{J_1} (j - J_0)^2 n_j}{\sum_{j=J_0}^{J_1} n_j} - (\eta(\mathcal{I}_n) - J_0)^2$$

to be compared with

$$\kappa_{J_1-J_0} = \frac{\sum_{i=0}^{J_1-J_0} i^2 2^{-i}}{\sum_{i=0}^{J_1-J_0} 2^{-i}} - \eta_{J_1-J_0}^2.$$

□

REFERENCES

- ABRY, P. and VEITCH, D. (1998). Wavelet analysis of long-range-dependent traffic. *IEEE Trans. Inform. Theory* **44** 2–15.
- BAILLIE, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics* **73** 5–59.
- BARDET, J. M., LANG, G., MOULINES, E. and SOULIER, P. (2000). Wavelet estimator of long-range dependent processes. *Stat. Inference Stoch. Process.* **3** 85–99. 19th “Rencontres Franco-Belges de Statisticiens” (Marseille, 1998).
- COHEN, A. (2003). *Numerical analysis of wavelet methods*, vol. 32 of *Studies in Mathematics and its Applications*. North-Holland Publishing Co., Amsterdam.
- CRAIGMILE, P. and PERCIVAL, D. (2005). Asymptotic decorrelation of between-scale wavelet coefficients. *IEEE Trans. Inform. Theory* **51** 1039–1048.
- DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S. (eds.) (2003). *Theory and Applications of Long-range Dependence*. Birkhäuser, Boston.
- GEWEKE, J. and PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.* **4** 221–238.
- KAPLAN, L. and KUO, C.-C. (1993). Fractal estimation from noisy data via discrete fractional Gaussian noise (DFGN) and the Haar basis. *IEEE Trans. Signal Process.* **41** 3554–3562.
- KÜNSCH, H. R. (1987). Statistical aspects of self-similar processes. In *Probability Theory and Applications, Proc. World Congr. Bernoulli Soc.*

- MCCOY, E. J. and WALDEN, A. T. (1996). Wavelet analysis and synthesis of stationary long-memory processes. *J. Comput. Graph. Statist.* **5** 26–56.
- MOULINES, E., ROUEFF, F. and TAQQU, M. (2005). On the spectral density of the wavelet coefficients of long memory time series with application to the log-regression estimation of the memory parameter. Arxiv preprint: <http://arxiv.org/abs/math.ST/0512635>.
- ROBINSON, P. (1995a). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661.
- ROBINSON, P. M. (1995b). Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* **23** 1048–1072.
- VELASCO, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.* **20** 87–127.
- WORNELL, G. and OPPENHEIM, A. (1992). Estimation of fractal signals from noisy measurements using wavelets. *IEEE Trans. Signal Process.* **40** 611 – 623.

GET/TÉLÉCOM PARIS, CNRS LTCI, 46, RUE BARRAULT, 75634 PARIS CÉDEX 13,
FRANCE

E-mail address: `moulines@tsi.enst.fr`

E-mail address: `roueff@tsi.enst.fr`

DEPARTMENT OF MATHEMATICS AND STATISTICS, BOSTON UNIVERSITY, BOSTON MA
02215, USA

E-mail address: `murad@math.bu.edu`