

Adaptive wavelet-based estimator of the memory parameter for stationary Gaussian processes

JEAN-MARC BARDET^{1,*}, HATEM BIBI^{1,**} and ABDELLATIF JOUINI²

¹*Samos-Matisse-CES, Université Paris 1, CNRS UMR 8174, 90 rue de Tolbiac, 75013 Paris, France.*
*E-mail: *bardet@univ-paris1.fr; **hatem.bibi@malix.univ-paris1.fr*

²*Département de Mathématiques, Faculté des Sciences de Tunis, 1060 Tunis, Tunisia.*
E-mail: Abdellatif.jouini@fst.mu.tn

This work is intended as a contribution to the theory of a wavelet-based adaptive estimator of the memory parameter in the classical semi-parametric framework for Gaussian stationary processes. In particular, we introduce and develop the choice of a data-driven optimal bandwidth. Moreover, we establish a central limit theorem for the estimator of the memory parameter with the minimax rate of convergence (up to a logarithm factor). The quality of the estimators is demonstrated via simulations.

Keywords: wavelet analysis; long range dependence; memory parameter; semi-parametric estimation; adaptive estimation

1. Introduction

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a second order zero-mean stationary process and its covariogram be defined by

$$r(t) = \mathbb{E}(X_0 \cdot X_t) \quad \text{for } t \in \mathbb{Z}.$$

Assume the spectral density f of X , with

$$f(\lambda) = \frac{1}{2\pi} \cdot \sum_{k \in \mathbb{Z}} r(k) \cdot e^{-ik},$$

exists and represents a continuous function on $[-\pi, 0) \cup (0, \pi]$. Consequently, the spectral density of X should satisfy the asymptotic property

$$f(\lambda) \sim C \cdot \frac{1}{\lambda^D} \quad \text{when } \lambda \rightarrow 0,$$

with $D < 1$ called the ‘memory parameter’ and $C > 0$. If $D \in (0, 1)$, then the process X is a so-called long-memory process; if not, X is called a short-memory process (see Doukhan *et al.* (2003), for more details). This paper deals with two semi-parametric frameworks, based, respectively, on the following assumptions.

Assumption A1. X is a zero-mean stationary Gaussian process with spectral density satisfying

$$f(\lambda) = |\lambda|^{-D} \cdot f^*(\lambda) \quad \text{for all } \lambda \in [-\pi, 0) \cup (0, \pi],$$

with $f^*(0) > 0$ and $f^* \in \mathcal{H}(D', C_{D'})$, where $0 < D', 0 < C_{D'}$ and

$$\mathcal{H}(D', C_{D'}) = \{g : [-\pi, \pi] \rightarrow \mathbb{R}^+ \text{ such that } |g(\lambda) - g(0)| \leq C_{D'} \cdot |\lambda|^{D'} \text{ for all } \lambda \in [-\pi, \pi]\}.$$

Assumption A1'. X is a zero-mean stationary Gaussian process with spectral density satisfying

$$f(\lambda) = |\lambda|^{-D} \cdot f^*(\lambda) \quad \text{for all } \lambda \in [-\pi, 0) \cup (0, \pi],$$

with $f^*(0) > 0$ and $f^* \in \mathcal{H}'(D', C_{D'})$, where $0 < D', C_{D'} > 0$ and

$$\mathcal{H}'(D', C_{D'}) = \{g : [-\pi, \pi] \rightarrow \mathbb{R}^+ \text{ such that } g(\lambda) = g(0) + C_{D'}|\lambda|^{D'} + o(|\lambda|^{D'}) \text{ when } \lambda \rightarrow 0\}.$$

Remark 1. A great number of earlier works concerning the estimation of the long-range parameter in a semi-parametric framework (see, e.g., Giraitis *et al.* (1997, 2000)) are based on Assumption A1 or an equivalent assumption on f . Another expression (see Robinson (1995), Moulines and Soulier (2003) or Moulines *et al.* (2007)) is $f(\lambda) = |1 - e^{i\lambda}|^{-2d} \cdot f^*(\lambda)$, with f^* a function such that $|f^*(\lambda) - f^*(0)| \leq f^*(0) \cdot \lambda^\beta$ and $0 < \beta$. It is obvious that for $\beta \leq 2$, such an assumption corresponds to Assumption A1 with $D' = \beta$. Moreover, following arguments developed in Giraitis *et al.* (1997, 2000), if $f^* \in \mathcal{H}(D', C_{D'})$ with $D' > 2$ is such that f^* is $s \in \mathbb{N}^*$ times differentiable around $\lambda = 0$ with $f^{*(s)}$ satisfying a Lipschitzian condition of degree $0 < \ell < 1$ around 0, then $D' \leq s + \ell$. So, for our purposes, D' is a more pertinent parameter than $s + \ell$ (which is often used in nonparametric literature). Finally, Assumption A1' is a necessary condition to study the following adaptive estimator of D .

We have $\mathcal{H}'(D', C_{D'}) \subset \mathcal{H}(D', C_{D'})$. Fractional Gaussian noises (with $D' = 2$) and FARIMA[p, d, q] processes (also with $D' = 2$) represent the first, and well-known, examples of processes satisfying Assumption A1' (and, therefore, Assumption A1).

Remark 2. In Andrews and Sun (2004), an adaptive procedure covers a more general class of functions than $\mathcal{H}(D', C_{D'})$, that is, $\mathcal{H}_{AS}(D', C_{D'})$, defined by

$$\mathcal{H}_{AS}(D', C_{D'}) = \left\{ \begin{array}{l} g : [-\pi, \pi] \rightarrow \mathbb{R}^+ \text{ such that, as } \lambda \rightarrow 0, \\ g(\lambda) = g(0) + \sum_{i=0}^k C'_i \lambda^{2i} + C_{D'} |\lambda|^{D'} + o(|\lambda|^{D'}) \text{ with } 2k < D' \leq 2k + 2 \end{array} \right\}.$$

Unfortunately, the adaptive wavelet-based estimator defined below, as a local or global log-periodogram estimator, cannot be adapted to such a class (and therefore, when $D' > 2$, its convergence rate will be the same as if the spectral density is included in $\mathcal{H}_{AS}(2, C_2)$, contrary to the estimator of Andrews and Sun).

This work is intended to provide a wavelet-based semi-parametric estimation of the parameter D . This method was introduced by Flandrin (1992) and numerically developed by Abry *et al.* (1998) and Veitch *et al.* (2003). Asymptotic results are reported in Bardet *et al.* (2000) and, more recently, in Moulines *et al.* (2007). Taking these papers into account, two points of our work can be highlighted: first, a central limit theorem based on conditions which are weaker than those in Bardet *et al.* (2000); second, we define an auto-driven estimator \tilde{D}_n of D (its definition being different from that in Veitch *et al.* (2003)). This results in a central limit theorem followed by \tilde{D}_n and this estimator is proved rate-optimal up to a logarithm factor (see below). Below, we shall develop this point.

Define the usual Sobolev space $\tilde{W}(\beta, L)$ for $\beta > 0$ and $L > 0$,

$$\tilde{W}(\beta, L) = \left\{ g(\lambda) = \sum_{\ell \in \mathbb{Z}} g_\ell e^{2\pi i \ell \lambda} \in \mathbb{L}^2([0, 1]) \mid \sum_{\ell \in \mathbb{Z}} (1 + |\ell|)^\beta |g_\ell| < \infty \text{ and } \sum_{\ell \in \mathbb{Z}} |g_\ell|^2 \leq L \right\}.$$

Let ψ be a ‘mother’ wavelet satisfying the following assumption.

Assumption $W(\infty)$. $\psi : \mathbb{R} \mapsto \mathbb{R}$ has support $[0, 1]$ and is such that:

1. ψ is included in the Sobolev class $\tilde{W}(\infty, L)$ with $L > 0$;
2. $\int_0^1 \psi(t) dt = 0$ and $\psi(0) = \psi(1) = 0$.

A consequence of the first point of this assumption is that for all $p > 0$, $\sup_{\lambda \in \mathbb{R}} |\hat{\psi}(\lambda)|(1 + |\lambda|)^p < \infty$, where $\hat{\psi}(u) = \int_0^1 \psi(t)e^{-iut} dt$ is the Fourier transform of ψ . A useful consequence of the second point is that $\hat{\psi}(u) \sim Cu$ for $u \rightarrow 0$, with $|C| < \infty$ a real number not depending on u .

ψ is a smooth, compactly supported function (the interval $[0, 1]$ is adopted for better readability, but the following results can be extended to another interval) with its first m moments vanishing. If $D' \leq 2$ and $0 < D < 1$ in Assumption A1, then Assumption $W(\infty)$ can be replaced by the following, weaker, assumption.

Assumption $W(5/2)$. $\psi : \mathbb{R} \mapsto \mathbb{R}$ has support $[0, 1]$ and is such that:

1. ψ is included in the Sobolev class $\tilde{W}(5/2, L)$ with $L > 0$;
2. $\int_0^1 \psi(t) dt = 0$ and $\psi(0) = \psi(1) = 0$.

Remark 3. The choice of a wavelet satisfying Assumption $W(\infty)$ is quite restricted because of the required smoothness of ψ . For instance, the function $\psi(t) = (t^2 - t + a) \exp(-1/t(1 - t))$ with $a \simeq 0.23087577$ satisfies Assumption $W(\infty)$. The class of ‘wavelets’ satisfying Assumption $W(5/2)$ is larger. For instance, ψ can be a dilated Daubechies ‘mother’ wavelet of order d with $d \geq 6$ to ensure the smoothness of the function ψ . It is also possible to apply the following theory to an ‘essentially’ compactly supported ‘mother’ wavelet like the Lemarié–Meyer wavelet. Note that it is not necessary to choose ψ to be a ‘mother’ wavelet associated with a multiresolution analysis of $\mathbb{L}^2(\mathbb{R})$ as in the recent paper of Moulines *et al.* (2007). The whole theory can be developed without this assumption, in which case the choice of ψ is larger.

If $Y = (Y_t)_{t \in \mathbb{R}}$ is a continuous-time process for $(a, b) \in \mathbb{R}_+^* \times \mathbb{R}$, the ‘classical’ wavelet coefficient $d(a, b)$ of the process Y for the scale a and the shift b is

$$d(a, b) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi\left(\frac{t}{a} - b\right) Y_t dt. \tag{1}$$

However, this formula for a wavelet coefficient cannot be computed from a time series. The support of ψ being $[0, 1]$, let us take the following approximation of formula (1) and define the wavelet coefficients of $X = (X_t)_{t \in \mathbb{Z}}$ by

$$e(a, b) = \frac{1}{\sqrt{a}} \sum_{k=1}^a \psi\left(\frac{k}{a}\right) X_{k+ab} \tag{2}$$

for $(a, b) \in \mathbb{N}_+^* \times \mathbb{Z}$. Note that this approximation is the same as the wavelet coefficient computed from the Mallat algorithm for an orthogonal discrete wavelet basis (e.g., with Daubechies mother wavelet).

Remark 4. Here, a continuous wavelet transform is considered. The discrete wavelet transform where $a = 2^j$, numerically very interesting (using the Mallat cascade algorithm), is just a particular case. The main point in studying a continuous transform is to offer a larger number of ‘scales’ for computing the data-driven optimal bandwidth (see below).

Under Assumption A1, for all $b \in \mathbb{Z}$, the asymptotic behavior of the variance of $e(a, b)$ is a power law in scale a (when $a \rightarrow \infty$). Indeed, for all $a \in \mathbb{N}^*$, $(e(a, b))_{b \in \mathbb{Z}}$ is a Gaussian stationary process and (see Section 2 for more details).

$$\mathbb{E}(e^2(a, 0)) \sim K_{(\psi, D)} \cdot a^D \quad \text{when } a \rightarrow \infty, \tag{3}$$

with a constant $K_{(\psi, D)}$ such that

$$K_{(\psi, \alpha)} = \int_{-\infty}^{\infty} |\widehat{\psi}(u)|^2 \cdot |u|^{-\alpha} du > 0 \quad \text{for all } \alpha < 1, \tag{4}$$

where $\widehat{\psi}$ is the Fourier transform of ψ (the existence of $K_{(\psi, \alpha)}$ is established in Section 5). Note that (3) is also satisfied without the Gaussian hypothesis in Assumption A1 (the existence of the second order moment of X is sufficient).

The principle of the wavelet-based estimation of D is linked to this power law a^D . Indeed, let (X_1, \dots, X_N) be a sample path of X and define $\widehat{T}_N(a)$ to be a sample variance of $e(a, \cdot)$ obtained from an appropriate choice of shifts b , that is,

$$\widehat{T}_N(a) = \frac{1}{[N/a]} \sum_{k=1}^{[N/a]} e^2(a, k-1). \tag{5}$$

Then, when $a = a_N \rightarrow \infty$ satisfies $\lim_{N \rightarrow \infty} a_N \cdot N^{-1/(2D'+1)} = \infty$, a central limit theorem for $\log(\widehat{T}_N(a_N))$ can be proven. More precisely, we get

$$\log(\widehat{T}_N(a_N)) = D \log(a_N) + \log(f^*(0)K_{(\psi, D)}) + \sqrt{\frac{a_N}{N}} \cdot \varepsilon_N,$$

with $\varepsilon_N \xrightarrow{\mathcal{L}}_{N \rightarrow \infty} \mathcal{N}(0, \sigma_{(\psi, D)}^2)$ and $\sigma_{(\psi, D)}^2 > 0$. As a consequence, using different scales $(r_1 a_N, \dots, r_\ell a_N)$, where $(r_1, \dots, r_\ell) \in (\mathbb{N}^*)^\ell$ with a_N a ‘large enough’ scale, a linear regression of $(\log(\widehat{T}_N(r_i a_N)))_i$ by $(\log(r_i a_N))_i$ provides an estimator $\widehat{D}(a_N)$ which simultaneously satisfies a central limit theorem with a convergence rate $\sqrt{\frac{N}{a_N}}$.

But the main problem is how to select a large enough scale a_N , considering that the smaller the a_N , the faster the convergence rate of $\widehat{D}(a_N)$. An optimal solution would be to chose a_N larger, but closer to $N^{1/(2D'+1)}$, but the parameter D' is supposed to be unknown. In Veitch *et al.* (2003), an automatic selection procedure is proposed, using a chi-squared goodness-of-fit statistic. This procedure is applied successfully on a large number of numerical examples, but without any theoretical proofs. Our present method is close to the latter. Roughly speaking, the ‘optimal’ choice of scale (a_N) is based on the ‘best’ linear regression among all possible linear regressions of ℓ consecutive points $(a, \log(\widehat{T}_N(a)))$, where ℓ is a fixed integer. Formally speaking, a squared distance is minimized and the chosen scale \tilde{a}_N satisfies

$$\frac{\log(\tilde{a}_N)}{\log N} \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \frac{1}{2D' + 1}.$$

Thus, the adaptive estimator \tilde{D}_N of D for this scale \tilde{a}_N is such that

$$\sqrt{\frac{N}{\tilde{a}_N}} (\tilde{D}_N - D) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_D^2),$$

with $\sigma_D^2 > 0$. Consequently, the minimax rate of convergence $N^{D'/(1+2D')}$, up to a logarithm factor, for the estimation of the long memory parameter D in this semi-parametric setting (see Giraitis *et al.* (1997)) is given by \tilde{D}_N .

Such a rate of convergence can also be obtained by other adaptive estimators (for more details see below). However, \tilde{D}_N has several ‘theoretical’ advantages. First, it can be applied to all $D < -1$ and $D' > 0$ (which are very general conditions covering long and short memory – in fact, more general conditions than those usually required for adaptive log-periodogram or local Whittle estimators) with a nearly optimal convergence rate. Second, \tilde{D}_N satisfies a central limit theorem and sharp confidence intervals for D can be computed (in such a case, the asymptotic σ_D^2 is replaced by $\sigma_{\tilde{D}_N}^2$ – for more details, see below). Finally, under additive assumptions on ψ (ψ is supposed to have its first m moments vanish), \tilde{D}_N can also be applied to a process with a polynomial trend of degree $\leq m - 1$.

We then provide several simulations in order to demonstrate empirical properties of the adaptive estimator \tilde{D}_N . First, using a benchmark composed of five different ‘test’ processes satisfying Assumption A1' (see below), the central limit theorem satisfied by \tilde{D}_N is empirically checked.

The empirical choice of the parameter ℓ is also studied. Moreover, the robustness of \tilde{D}_N is successfully tested. Finally, the adaptive wavelet-based estimator is compared with several existing adaptive estimators of the memory parameter from generated paths of the five different ‘test’ processes (Giraitis–Robinson–Samarov adaptive local log-periodogram, Moulines–Soulier adaptive global log-periodogram, Robinson local Whittle, Abry–Taqqu–Veitch data-driven wavelet-based and Bhansali–Giraitis–Kokoszka FAR estimators). The simulation results of \tilde{D}_N are convincing. The convergence rate of \tilde{D}_N is often among the best of the five test processes (however, the Robinson local Whittle estimator \hat{D}_R provides more uniformly accurate estimations of D). Three other numerical advantages are offered by the adaptive wavelet-based estimator (and not by \hat{D}_R). First, it is not a very time-consuming estimator. Second, it is a very robust estimator: it is not sensitive to possible polynomial trends and seems to be consistent in non-Gaussian cases. Finally, the graph of the log-log regression of the sample variance of wavelet coefficients is meaningful and may lead us to model data with more general processes like locally fractional Gaussian noise (see Bardet and Bertrand (2007)).

The central limit theorem for sample variance of wavelet coefficients is the subject of Section 2. Section 3 is concerned with the automatic selection of the scale, as well as the asymptotic behavior of \tilde{D}_N . Finally, simulations are given in Section 4 and proofs in Section 5.

2. A central limit theorem for the sample variance of wavelet coefficients

The following asymptotic behavior of the variance of wavelet coefficients is the basis of all further developments. The first point explains all that follows.

Property 1. *Under Assumption A1 and Assumption $W(\infty)$, for $a \in \mathbb{N}^*$, $(e(a, b))_{b \in \mathbb{Z}}$ is a zero-mean Gaussian stationary process and there exists some $M > 0$, not depending on a , such that for all $a \in \mathbb{N}^*$,*

$$|\mathbb{E}(e^2(a, 0)) - f^*(0)K_{(\psi, D)} \cdot a^D| \leq M \cdot a^{D-D'}. \quad (6)$$

See Section 5 for the proofs. The paper of Moulines *et al.* (2007) gives similar results for multiresolution wavelet analysis. The special case of long-memory processes can also be studied with the weaker Assumption $W(5/2)$.

Property 2. *Under Assumption $W(5/2)$ and Assumption A1 with $0 < D < 1$ and $0 < D' \leq 2$, for $a \in \mathbb{N}^*$, $(e(a, b))_{b \in \mathbb{Z}}$ is a zero-mean Gaussian stationary process and (6) holds.*

Two corollaries can be added to these properties. First, under Assumption A1', a more precise result can be established.

Corollary 1. *Under*

- *Assumption A1' and Assumption $W(\infty)$,*

• or Assumption A1' with $0 < D < 1, 0 < D' \leq 2$ and Assumption W(5/2), $(e(a, b))_{b \in \mathbb{Z}}$ is a zero-mean Gaussian stationary process and

$$\begin{aligned} \mathbb{E}(e^2(a, 0)) &= f^*(0) (K_{(\psi, D)} \cdot a^D + C_{D'} K_{(\psi, D-D')} \cdot a^{D-D'}) \\ &\quad + o(a^{D-D'}) \quad \text{when } a \rightarrow \infty. \end{aligned} \tag{7}$$

This corollary is a key point for the estimation of an appropriated sequence of scales $a = (a_N)$. Indeed, if $f^* \in \mathcal{H}'(D', C_{D'})$, then $f^* \in \mathcal{H}(D'', C_{D''})$ for all D'' satisfying $0 < D'' \leq D'$. Therefore, Assumption A1' is required to obtain the optimal choice of a_N , that is, $a_N \simeq N^{1/(2D'+1)}$ (see below for more details). The following corollary generalizes the above Properties 1 and 2.

Corollary 2. *Properties 1 and 2 are also satisfied when the Gaussian hypothesis of X is replaced by $\mathbb{E}X_k^2 < \infty$ for all $k \in \mathbb{Z}$.*

Remark 5. In this paper, the Gaussian hypothesis has been taken into account merely to ensure the convergence of the sample variance (5) of wavelet coefficients following a central limit theorem (see below). Such a convergence can also be obtained for more general processes using a different proof of the central limit theorem, for instance, for linear processes (see a forthcoming work).

As mentioned in the Introduction, this property allows an estimation of D from a log-log regression, as soon as a consistent estimator of $\mathbb{E}(e^2(a, 0))$ is provided from a sample (X_1, \dots, X_N) of the time series X . We then define the normalized wavelet coefficient such that

$$\tilde{e}(a, b) = \frac{e(a, b)}{(f^*(0)K_{(\psi, D)} \cdot a^D)^{1/2}} \quad \text{for } a \in \mathbb{N}^* \text{ and } b \in \mathbb{Z}. \tag{8}$$

From Property 1, it is obvious that under Assumption A1, there exists some $M' > 0$ satisfying, for all $a \in \mathbb{N}^*$,

$$|\mathbb{E}(\tilde{e}^2(a, 0)) - 1| \leq M' \cdot \frac{1}{a^{D'}}.$$

To use this formula to estimate D by a log-log regression, an estimator of the variance of $e(a, 0)$ should be considered (recall that a sample (X_1, \dots, X_N) is supposed to be known, but parameters D, D' and $C_{D'}$ are unknown). Consider the sample variance and the normalized sample variance of the wavelet coefficients for $1 \leq a < N$,

$$\hat{T}_N(a) = \frac{1}{[N/a]} \sum_{k=1}^{[N/a]} e^2(a, k-1) \quad \text{and} \quad \tilde{T}_N(a) = \frac{1}{[N/a]} \sum_{k=1}^{[N/a]} \tilde{e}^2(a, k-1). \tag{9}$$

The following proposition specifies a central limit theorem satisfied by $\log \tilde{T}_N(a)$, which provides the first step for obtaining the asymptotic properties of the estimator by log-log regression. More generally, the following multidimensional central limit theorem for a vector $(\log \tilde{T}_N(a_i))_i$ can be established.

Proposition 1. Define $\ell \in \mathbb{N} \setminus \{0, 1\}$ and $(r_1, \dots, r_\ell) \in (\mathbb{N}^*)^\ell$. Let $(a_n)_{n \in \mathbb{N}}$ be such that $N/a_N \xrightarrow{N \rightarrow \infty} \infty$ and $a_N \cdot N^{-1/(1+2D')} \xrightarrow{N \rightarrow \infty} \infty$. Under Assumption A1 and Assumption $W(\infty)$,

$$\sqrt{\frac{N}{a_N}} (\log \widehat{T}_N(r_i a_N))_{1 \leq i \leq \ell} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}_\ell(0; \Gamma(r_1, \dots, r_\ell, \psi, D)), \tag{10}$$

with $\Gamma(r_1, \dots, r_\ell, \psi, D) = (\gamma_{ij})_{1 \leq i, j \leq \ell}$ the covariance matrix such that

$$\gamma_{ij} = \frac{8(r_i r_j)^{2-D}}{K_{(\psi, D)}^2 d_{ij}} \sum_{m=-\infty}^{\infty} \left(\int_0^\infty \frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)}{u^D} \cos(u d_{ij} m) du \right)^2. \tag{11}$$

The same result under weaker assumptions on ψ can be also established when X is a long-memory process.

Proposition 2. Define $\ell \in \mathbb{N} \setminus \{0, 1\}$ and $(r_1, \dots, r_\ell) \in (\mathbb{N}^*)^\ell$. Let $(a_n)_{n \in \mathbb{N}}$ be such that $N/a_N \xrightarrow{N \rightarrow \infty} \infty$ and $a_N \cdot N^{-1/(1+2D')} \xrightarrow{N \rightarrow \infty} \infty$. Under Assumption $W(5/2)$ and Assumption A1 with $D \in (0, 1)$ and $D' \in (0, 2)$, the CLT (10) holds.

These results can be easily generalized for processes with polynomial trends if ψ is considered assumed to have its first m moments vanishing. That is, we have the following.

Corollary 3. Given the same hypothesis as in Proposition 1 or 2 and if ψ is such that, for $m \in \mathbb{N} \setminus \{0, 1\}$, $\int t^p \psi(t) dt = 0$ for all $p \in \{0, 1, \dots, m-1\}$, then the CLT (10) also holds for any process $X' = (X'_t)_{t \in \mathbb{Z}}$ such that for all $t \in \mathbb{Z}$, $\mathbb{E}X'_t = P_m(t)$, where $P_m(t) = a_0 + a_1 t + \dots + a_{m-1} t^{m-1}$ is a polynomial function and $(a_i)_{0 \leq i \leq m-1}$ are real numbers.

3. Adaptive estimator of memory parameter using data driven optimal scales

The CLT (10) implies the following CLT for the vector $(\log \widehat{T}_N(r_i a_N))_i$:

$$\sqrt{\frac{N}{a_N}} (\log \widehat{T}_N(r_i a_N) - D \log(r_i a_N) - \log(f^*(0)K_{(\psi, D)}))_{1 \leq i \leq \ell} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}_\ell(0; \Gamma(r_1, \dots, r_\ell, \psi, D)).$$

Therefore,

$$(\log \widehat{T}_N(r_i a_N))_{1 \leq i \leq \ell} = A_N \cdot \left(\frac{D}{K} \right) + \frac{1}{\sqrt{N/a_N}} (\varepsilon_i)_{1 \leq i \leq \ell},$$

with

$$A_N = \begin{pmatrix} \log(r_1 a_N) & 1 \\ \vdots & \vdots \\ \log(r_\ell a_N) & 1 \end{pmatrix}, \quad K = -\log(f^*(0) \cdot K(\psi, D))$$

and

$$(\varepsilon_i)_{1 \leq i \leq \ell} \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}_\ell(0; \Gamma(r_1, \dots, r_\ell, \psi, D)).$$

Therefore, a log-log regression of $(\widehat{T}_N(r_i a_N))_{1 \leq i \leq \ell}$ on scales $(r_i a_N)_{1 \leq i \leq \ell}$ provides an estimator $\begin{pmatrix} \widehat{D}(a_N) \\ \widehat{K}(a_N) \end{pmatrix}$ of $\begin{pmatrix} D \\ K \end{pmatrix}$ such that

$$\begin{pmatrix} \widehat{D}(a_N) \\ \widehat{K}(a_N) \end{pmatrix} = (A'_N \cdot A_N)^{-1} \cdot A'_N \cdot Y_{a_N}^{(r_1, \dots, r_\ell)} \quad \text{with } Y_{a_N}^{(r_1, \dots, r_\ell)} = (\log \widehat{T}_N(r_i a_N))_{1 \leq i \leq \ell}, \quad (12)$$

which satisfies the following CLT.

Proposition 3. *Under the Assumptions of Proposition 1,*

$$\begin{aligned} & \sqrt{\frac{N}{a_N}} \left(\begin{pmatrix} \widehat{D}(a_N) \\ \widehat{K}(a_N) \end{pmatrix} - \begin{pmatrix} D \\ K \end{pmatrix} \right) \\ & \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}_2(0; (A' \cdot A)^{-1} \cdot A' \cdot \Gamma(r_1, \dots, r_\ell, \psi, D) \cdot A \cdot (A' \cdot A)^{-1}), \end{aligned} \quad (13)$$

with

$$A = \begin{pmatrix} \log(r_1) & 1 \\ \vdots & \vdots \\ \log(r_\ell) & 1 \end{pmatrix} \quad \text{and } \Gamma(r_1, \dots, r_\ell, \psi, D) \text{ given by (11).}$$

Moreover, under Assumption $A1'$ and if $D \in (-1, 1)$, $\widehat{D}(a_N)$ is a semi-parametric estimator of D and its asymptotic mean square error can be minimized with an appropriate scales sequence (a_N) reaching the well-known minimax rate of convergence for memory parameter D in this semi-parametric setting (see, e.g., Giraitis *et al.* (1997, 2000)). Indeed, we have the following

Proposition 4. *Let X satisfy Assumption $A1'$ with $D \in (-1, 1)$ and let ψ satisfy Assumption $W(\infty)$. Let (a_N) be a sequence such that $a_N = [N^{1/(1+2D')}]$. Then the estimator $\widehat{D}(a_N)$ is rate-optimal in the minimax sense, that is,*

$$\limsup_{N \rightarrow \infty} \sup_{D \in (-1, 1)} \sup_{f^* \in \mathcal{H}(D', C_{D'})} N^{2D'/(1+2D')} \cdot \mathbb{E}[\widehat{D}(a_N) - D]^2 < +\infty.$$

Remark 6. As far as we know, there are no theoretical results concerning optimality in the case $D \leq -1$, but, according to the usual following nonparametric theory, such minimax results can also be obtained. Moreover, in the case of long-memory processes (if $D \in (0, 1)$), under Assumption $A1'$ for X and Assumption $W(5/2)$ for ψ , the estimator $\widehat{D}(a_N)$ is also rate-optimal in the minimax sense.

In the previous Propositions 1 and 3, the rate of convergence of scale a_N obeys the following condition:

$$\frac{N}{a_N} \xrightarrow{N \rightarrow \infty} \infty \quad \text{and} \quad \frac{a_N}{N^{1/(1+2D')}} \xrightarrow{N \rightarrow \infty} \infty \quad \text{with } D' \in (0, \infty).$$

Now, for improved readability, we take $a_N = N^\alpha$. The above condition then becomes

$$a_N = N^\alpha \quad \text{with } \alpha^* < \alpha < 1 \quad \text{and} \quad \alpha^* = \frac{1}{1 + 2D'}. \tag{14}$$

Thus, an optimal choice (leading to a faster convergence rate of the estimator) is obtained for $\alpha = \alpha^* + \varepsilon$ with $\varepsilon \rightarrow 0+$. But α^* depends on D' , which is unknown. To solve this problem, Veitch *et al.* (2003) suggest a chi-square-based test (constructed from the distance between the regression line and the different points $(\log \widehat{T}_N(r_i a_N), \log(r_i a_N))$). It seems to be an efficient and interesting numerical way to estimate D , but without theoretical proofs (contrary to global or local log-periodogram procedures which are proved to reach the minimax convergence rate – see, e.g., Moulines and Soulier (2003)).

We suggest a new procedure for the data-driven selection of optimal scales, that is, optimal α . Let us consider an important parameter – the number of considered scales $\ell \in \mathbb{N} \setminus \{0, 1, 2\}$ – and set $(r_1, \dots, r_\ell) = (1, \dots, \ell)$. For $\alpha \in (0, 1)$, we also define:

- the vector $Y_N(\alpha) = (\log \widehat{T}_N(i \cdot N^\alpha))_{1 \leq i \leq \ell}$;
- the matrix $A_N(\alpha) = \begin{pmatrix} \log(N^\alpha) & 1 \\ \vdots & \vdots \\ \log(\ell \cdot N^\alpha) & 1 \end{pmatrix}$;
- the non-negative number, $Q_N(\alpha, D, K) = (Y_N(\alpha) - A_N(\alpha) \cdot \begin{pmatrix} D \\ K \end{pmatrix})' \cdot (Y_N(\alpha) - A_N(\alpha) \cdot \begin{pmatrix} D \\ K \end{pmatrix})$.

$Q_N(\alpha, D, K)$ corresponds to a squared distance between the ℓ points $(\log(i \cdot N^\alpha), \log T_N(i \cdot N^\alpha))_i$ and a line. The point is to minimize this non-negative number for these three parameters. It is obvious that for a fixed $\alpha \in (0, 1)$, Q is minimized from the previous least-square regression and, therefore,

$$Q_N(\widehat{\alpha}_N, \widehat{D}(a_N), \widehat{K}(a_N)) = \min_{\alpha \in (0,1), D < 1, K \in \mathbb{R}} Q_N(\alpha, D, K),$$

with $(\widehat{D}(a_N), \widehat{K}(a_N))$ obtained as in relation (12). However, since $\widehat{\alpha}_N$ has to be obtained from numerical computations, the interval $(0, 1)$ can be discretized as follows:

$$\widehat{\alpha}_N \in \mathcal{A}_N = \left\{ \frac{2}{\log N}, \frac{3}{\log N}, \dots, \frac{\log[N/\ell]}{\log N} \right\}.$$

Hence, if $\alpha \in \mathcal{A}_N$, then there exists some $k \in \{2, 3, \dots, \log[N/\ell]\}$ such that $k = \alpha \cdot \log N$.

Remark 7. This choice of discretization is implied by the following proof of the consistency of $\widehat{\alpha}_N$. If the interval $(0, 1)$ is stepped in N^β points, with $\beta > 0$, the proof used cannot attest this consistency. Finally, it is the same framework as the usual discrete wavelet transform (see, e.g., Veitch *et al.* (2003)), but less restricted since $\log N$ may be replaced in the previous expression for \mathcal{A}_N by any function of N which is negligible compared to functions N^β with $\beta > 0$ (e.g., $(\log N)^d$ or $d \log N$ can be used).

Consequently, taking

$$\widehat{Q}_N(\alpha) = Q_N(\alpha, \widehat{D}(a_N), \widehat{K}(a_N))$$

and then minimizing Q_N for the variables α, D and K is equivalent to minimizing \widehat{Q}_N for the variable $\alpha \in \mathcal{A}_N$, that is,

$$\widehat{Q}_N(\widehat{\alpha}_N) = \min_{\alpha \in \mathcal{A}_N} \widehat{Q}_N(\alpha).$$

From this central limit theorem, we derive the following.

Proposition 5. *Let X satisfy Assumption A1' and let ψ satisfy Assumption $W(\infty)$ (or Assumption $W(5/2)$, if $0 < D < 1$ and $0 < D' \leq 2$). Then*

$$\widehat{\alpha}_N = \frac{\log \widehat{a}_N}{\log N} \xrightarrow[N \rightarrow \infty]{\mathcal{P}} \alpha^* = \frac{1}{1 + 2D'}.$$

This also proves the consistency of an estimator \widehat{D}'_N of the parameter D' , as follows.

Corollary 4. *Taking the hypothesis of Proposition 5, we have*

$$\widehat{D}'_N = \frac{1 - \widehat{\alpha}_N}{2\widehat{\alpha}_N} \xrightarrow[N \rightarrow \infty]{\mathcal{P}} D'.$$

The estimator $\widehat{\alpha}_N$ defines the selected scale \widehat{a}_N such that $\widehat{a}_N = N^{\widehat{\alpha}_N}$. From a straightforward application of the proof of Proposition 5 (see the details in the proof of Theorem 1), the asymptotic behavior of \widehat{a}_N can be specified, that is,

$$\Pr\left(\frac{N^{\alpha^*}}{(\log N)^\lambda} \leq N^{\widehat{\alpha}_N} \leq N^{\alpha^*} \cdot (\log N)^\mu\right) \xrightarrow[N \rightarrow \infty]{\mathcal{P}} 1 \tag{15}$$

for all positive real numbers λ and μ such that $\lambda > \frac{2}{(\ell-2)D'}$ and $\mu > \frac{12}{\ell-2}$. Consequently, the selected scale is asymptotically equal to N^{α^*} , up to a logarithm factor.

Finally, Proposition 5 can be used to define an adaptive estimator of D . First, define the straightforward estimator

$$\widehat{\widehat{D}}_N = \widehat{D}(\widehat{a}_N),$$

which should minimize the mean square error using \widehat{a}_N . However, the estimator $\widehat{\widehat{D}}_N$ does not satisfy a CLT since $\Pr(\widehat{\alpha}_N \leq \alpha^*) > 0$ and therefore it cannot be asserted that $\mathbb{E}(\sqrt{N/\widehat{a}_N}(\widehat{\widehat{D}}_N - D)) = 0$. To establish a CLT satisfied by an adaptive estimator \widetilde{D}_N of D , an adaptive scale sequence $(\widetilde{a}_N) = (N^{\widetilde{\alpha}_N})$ has to be defined to ensure $\Pr(\widetilde{\alpha}_N \leq \alpha^*) \rightarrow_{N \rightarrow \infty} 0$. The following theorem provides the asymptotic behavior of such an estimator.

Theorem 1. Let X satisfy Assumption $A1'$ and let ψ satisfy Assumption $W(\infty)$ (or Assumption $W(5/2)$, if $0 < D < 1$ and $0 < D' \leq 2$). Define

$$\tilde{\alpha}_N = \hat{\alpha}_N + \frac{3}{(\ell - 2)\hat{D}'_N} \cdot \frac{\log \log N}{\log N}, \quad \tilde{a}_N = N^{\tilde{\alpha}_N} = N^{\hat{\alpha}_N} \cdot (\log N)^{3/((\ell-2)\hat{D}'_N)}$$

and

$$\tilde{D}_N = \hat{D}(\tilde{a}_N).$$

Then, with $\sigma_D^2 = (1 \ 0) \cdot (A' \cdot A)^{-1} \cdot A' \cdot \Gamma(1, \dots, \ell, \psi, D) \cdot A \cdot (A' \cdot A)^{-1} \cdot (1 \ 0)'$,

$$\begin{aligned} & \sqrt{\frac{N}{N^{\tilde{\alpha}_N}}} (\tilde{D}_N - D) \xrightarrow[N \rightarrow \infty]{D} \mathcal{N}(0; \sigma_D^2) \quad \text{and} \\ & \forall \rho > \frac{2(1 + 3D')}{(\ell - 2)D'}, \quad \frac{N^{D'/(1+2D')}}{(\log N)^\rho} \cdot |\tilde{D}_N - D| \xrightarrow[N \rightarrow \infty]{\mathcal{P}} 0. \end{aligned} \tag{16}$$

Remark 8. Both the adaptive estimators \hat{D}_N and \tilde{D}_N converge to D with a rate of convergence equal to the minimax rate of convergence $N^{D'/(1+2D')}$, up to a logarithm factor (this result being classical within this semi-parametric framework). Unfortunately, our method cannot prove that the mean square error of both of these estimators reaches the optimal rate and therefore that they are oracles.

To conclude this theoretical approach, the main properties satisfied by the estimators \hat{D}_N and \tilde{D}_N can be summarized as follows:

1. Both the adaptive estimators \hat{D}_N and \tilde{D}_N converge to D with a rate of convergence rate equal to the minimax rate of convergence $N^{D'/(1+2D')}$, up to a logarithm factor for all $D < -1$ and $D' > 0$ (these being very general conditions covering long and short memory, even more general than the usual conditions required for adaptive log-periodograms or local Whittle estimators) with X considered a Gaussian process.
2. The estimator \tilde{D}_N satisfies the CLT (16) and therefore sharp confidence intervals for D can be computed (in which case the asymptotic matrix $\Gamma(1, \dots, \ell, \psi, D)$ is replaced by $\Gamma(1, \dots, \ell, \psi, \hat{D}_N)$). This is not applicable to an adaptive log-periodogram or to local Whittle estimators.
3. Property 1 is also satisfied without the Gaussian hypothesis. Therefore, adaptive estimators \hat{D}_N and \tilde{D}_N can also be interesting estimators of D for non-Gaussian processes like linear, or more general, processes (but a CLT similar to Theorem 1 must be established).
4. Under additive assumptions on ψ (ψ is supposed to have its first m moments vanishing), both estimators \hat{D}_N and \tilde{D}_N can also be used for a process X with a polynomial trend of degree $\leq m - 1$, which, again, cannot be yielded by an adaptive log-periodogram or by local Whittle estimators.

4. Simulations

The adaptive wavelet-based estimators \widehat{D}_N and \widetilde{D}_N are new estimators of the memory parameter D in the semi-parametric setting. In other research, different estimators of this kind are also reported to have proven optimal. In this paper, some theoretical advantages of adaptive wavelet-based estimators have been highlighted. But what about concrete procedures and the results of such estimators applied to an observed sample? The following simulations will help to answer this question.

First, the properties (consistency, robustness, choice of the parameter ℓ and mother wavelet function ψ) of \widehat{D}_N and \widetilde{D}_N are investigated. Second, in the cases of Gaussian long-memory processes (with $D \in (0, 1)$ and $D' \leq 2$), the simulation results of the estimator \widehat{D}_N are compared to those obtained with the best known semi-parametric long-memory estimators.

To begin, the simulation conditions must be specified. The results are obtained from 100 generated independent samples of each process belonging to the following ‘benchmark’. The concrete procedures of generation for these processes are obtained from the circulant matrix method, as detailed in Doukhan *et al.* (2003). The simulations are realized for different values of D , N and processes which satisfy Assumption A1' and therefore Assumption A1 (the article of Moulines *et al.* (2007), provides much detail on this point):

1. the fractional Gaussian noise (fGn) with parameter $H = (D + 1)/2$ (for $-1 < D < 1$) and $\sigma^2 = 1$ the spectral density f_{fGn} of a fGn being such that f_{fGn}^* is included in $\mathcal{H}(2, C_2)$ (thus, $D' = 2$);
2. the FARIMA[p, d, q] process with parameter d such that $d = D/2 \in (-0.5, 0.5)$ (therefore, $-1 < D < 1$), the innovation variance σ^2 satisfying $\sigma^2 = 1$ and $p, q \in \mathbb{N}$, and the spectral density f_{FARIMA} of such a process being such that f_{FARIMA}^* is included in the set $\mathcal{H}(2, C_2)$ (thus, $D' = 2$);
3. the Gaussian stationary process $X^{(D, D')}$, such that its spectral density is

$$f_3(\lambda) = \frac{1}{\lambda^D} (1 + \lambda^{D'}) \quad \text{for } \lambda \in [-\pi, \pi], \tag{17}$$

with $D \in (-\infty, 1)$ and $D' \in (0, \infty)$, therefore $f_3^* = 1 + \lambda^{D'} \in \mathcal{H}(D', 1)$ with $D' \in (0, \infty)$.

In the long-memory setting, a ‘benchmark’ of processes is considered for $D = 0.1, 0.3, 0.5, 0.7, 0.9$:

- fGn processes with parameters $H = (D + 1)/2$ and $\sigma^2 = 1$;
- FARIMA[0, d , 0] processes with $d = D/2$ and standard Gaussian innovations;
- FARIMA[1, d , 0] processes with $d = D/2$, standard Gaussian innovations and AR coefficient $\phi = 0.95$;
- FARIMA[1, d , 1] processes with $d = D/2$, standard Gaussian innovations, AR coefficient $\phi = -0.3$ and MA coefficient $\phi = 0.7$;
- $X^{(D, D')}$ Gaussian processes with $D' = 1$.

4.1. Properties of adaptive wavelet-based estimators from simulations

Below, we give the different properties of the adaptive wavelet-based method.

Choice of the mother wavelet ψ

For short-memory processes ($D \leq 0$), let the wavelet ψ_{SM} be such that $\psi_{SM}(t) = (t^2 - t + a) \exp(-1/t(1 - t))$ with $a \simeq 0.23087577$. This satisfies Assumption $W(\infty)$. Lemarié–Meyer wavelets can be also investigated but this will lead to quite different theoretical studies since its support is not bounded (but is ‘essentially’ compact).

For long-memory processes ($0 < D < 1$), let the mother wavelet ψ_{LM} be such that $\psi_{LM}(t) = 100 \cdot t^2(t - 1)^2(t^2 - t + 3/14)\mathbb{I}_{0 \leq t \leq 1}$, which satisfies Assumption $W(5/2)$. Note that the Daubechies mother wavelet or ψ_{SM} leads to ‘similar’ (but not as good) results.

Choice of the parameter ℓ

This parameter is very important to estimate the ‘beginning’ of the linear part of the graph drawn by the points $(\log(a_i), \log \widehat{T}(a_i))_i$. On the one hand, if ℓ is a too small (e.g., $\ell = 3$), another small linear part of this graph (even before the ‘true’ beginning N^{α^*}) may be chosen; consequently, the \sqrt{MSE} (square root of the mean square error) of $\widehat{\alpha}_N$, and therefore of \widehat{D}_N or \check{D}_N , will be too large. On the other hand, if ℓ is a too large (e.g., $\ell = 50$ for $N = 1000$), the estimator $\widehat{\alpha}_N$ will certainly satisfy $\widehat{\alpha}_N < \alpha^*$ since it will not be possible to consider ℓ different scales larger than N^{α^*} (if $D' = 1$ and therefore $\alpha' = 1/3$, then a_N has to satisfy the property that $N/(50a_N) = 20/a_N$ is a large number and $a_N > N^{1/3} = 10$ this is not really possible). Moreover, it is possible that a ‘good’ choice of ℓ depends on the ‘flatness’ of the spectral density f , that is, on D' . We have carried out simulations for each different value of ℓ (and N and D). Only the \sqrt{MSE} of the estimators are presented. The results are displayed in Table 1.

In Table 1, two phenomena can be distinguished: the detection of α^* and the estimation of D .

- To estimate α^* , ℓ has to be small enough, especially because of ‘ D' close to 0’ and so ‘ α' close to 1’ is possible. However, our simulations indicate that ℓ must not be too small (e.g., $\ell = 5$ leads to an important MSE for $\widehat{\alpha}_N$ implying an important MSE for \widehat{D}_N) and seems to be independent of N (cases $N = 1000$ and $N = 10000$ are quite similar). Hence, our choice is $\ell_1 = 15$ **to estimate α^* for any N** .
- To estimate D , once α^* is estimated, a second value ℓ_2 of ℓ can be chosen. We use an adaptive procedure which, roughly speaking, consists of determining the ‘end’ of the acceptable linear zone. First, we again use the same procedure as was used for estimating $\widehat{\alpha}_N$, but with scales $(a_N/i)_{1 \leq i \leq \ell_1}$ and $\ell_1 = 15$. It provides an estimator \widehat{b}_N corresponding to the maximum of acceptable (for a linear regression) scales. Second, **the adaptive number of scales ℓ_2 is computed from the formula $\ell_2 = \widehat{\ell} = \lceil \widehat{b}_N / \widehat{\alpha}_N \rceil$** . The simulations carried out with such values of ℓ_1 and ℓ_2 are detailed in Table 1.

As may be seen in Table 1, the choice of parameters ($\ell_1 = 15, \ell_2 = \widehat{\ell}$) provides the best results for estimating D , almost uniformly for all processes.

Table 1. Consistency of estimators $\widehat{D}_N, \widetilde{D}_N, \widehat{\alpha}_N, \widetilde{\alpha}_N$ following ℓ from simulations of the different long-memory processes of the benchmark – for each value of N ($10^3, 10^4$ and of 10^5), of D (0.1, 0.3, 0.5, 0.7 and 0.9) and ℓ (5, 10, 15, 20, 25 and $(15, \widehat{\ell})$), 100 independent samples of each process are generated; the \sqrt{MSE} of each estimator is obtained from a mean of \sqrt{MSE} obtained for the different values of D

	\sqrt{MSE}	$\ell = 5$	$\ell = 10$	$\ell = 15$	$\ell = 20$	$\ell = 25$	$\left. \begin{matrix} \ell_1 = 15 \\ \ell_2 = \widehat{\ell} \end{matrix} \right\}$
$N = 10^3$							
fGn ($H = \frac{D+1}{2}$)	$\widehat{D}_N, \widetilde{D}_N$	0.16, 0.75	0.14, 0.19	0.13, 0.17	0.14, 0.15	0.14, 0.15	0.15, 0.18
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.12, 0.32	0.07, 0.13	0.05, 0.08	0.04, 0.05	0.04, 0.04	0.05, 0.08
FARIMA(0, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.21, 0.81	0.15, 0.20	0.14, 0.17	0.15, 0.15	0.15, 0.15	0.15, 0.19
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.14, 0.34	0.07, 0.13	0.05, 0.09	0.05, 0.06	0.04, 0.04	0.05, 0.09
FARIMA(1, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.30, 0.96	0.28, 0.35	0.27, 0.29	0.29, 0.27	0.30, 0.30	0.31, 0.35
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.19, 0.44	0.15, 0.24	0.12, 0.17	0.11, 0.15	0.11, 0.12	0.12, 0.17
FARIMA(1, $\frac{D}{2}$, 1)	$\widehat{D}_N, \widetilde{D}_N$	0.60, 0.92	0.43, 0.41	0.39, 0.35	0.36, 0.35	0.32, 0.33	0.21, 0.20
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.17, 0.38	0.11, 0.18	0.09, 0.12	0.07, 0.09	0.06, 0.07	0.09, 0.12
$X^{(D, D')}, D' = 1$	$\widehat{D}_N, \widetilde{D}_N$	0.33, 0.68	0.29, 0.28	0.27, 0.26	0.26, 0.27	0.25, 0.25	0.29, 0.30
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.10, 0.22	0.10, 0.07	0.11, 0.07	0.12, 0.12	0.13, 0.13	0.11, 0.07
$N = 10^4$							
fGn ($H = \frac{D+1}{2}$)	$\widehat{D}_N, \widetilde{D}_N$	0.08, 0.26	0.05, 0.05	0.05, 0.05	0.04, 0.04	0.04, 0.04	0.04, 0.04
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.08, 0.22	0.05, 0.06	0.04, 0.05	0.04, 0.05	0.05, 0.05	0.04, 0.05
FARIMA(0, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.08, 0.31	0.06, 0.06	0.05, 0.05	0.05, 0.05	0.05, 0.05	0.05, 0.05
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.09, 0.24	0.05, 0.07	0.04, 0.05	0.04, 0.05	0.05, 0.05	0.04, 0.05
FARIMA(1, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.13, 0.57	0.10, 0.10	0.09, 0.08	0.09, 0.08	0.09, 0.09	0.09, 0.08
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.15, 0.36	0.09, 0.16	0.08, 0.11	0.07, 0.09	0.06, 0.08	0.08, 0.11
FARIMA(1, $\frac{D}{2}$, 1)	$\widehat{D}_N, \widetilde{D}_N$	0.22, 0.63	0.17, 0.15	0.16, 0.13	0.15, 0.14	0.15, 0.14	0.09, 0.09
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.16, 0.38	0.11, 0.17	0.08, 0.11	0.07, 0.09	0.06, 0.07	0.08, 0.11
$X^{(D, D')}, D' = 1$	$\widehat{D}_N, \widetilde{D}_N$	0.23, 0.36	0.19, 0.15	0.18, 0.17	0.17, 0.17	0.15, 0.14	0.15, 0.14
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.10, 0.18	0.12, 0.08	0.13, 0.12	0.14, 0.14	0.15, 0.15	0.13, 0.12
$N = 10^5$							
fGn ($H = \frac{D+1}{2}$)	$\widehat{D}_N, \widetilde{D}_N$	0.04, 0.09	0.03, 0.03	0.02, 0.03	0.02, 0.02	0.02, 0.02	0.02, 0.02
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.07, 0.16	0.06, 0.04	0.06, 0.06	0.07, 0.07	0.07, 0.07	0.06, 0.06
FARIMA(0, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.03, 0.13	0.02, 0.02	0.02, 0.02	0.02, 0.02	0.02, 0.02	0.02, 0.02
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.07, 0.18	0.04, 0.05	0.04, 0.03	0.04, 0.04	0.05, 0.05	0.04, 0.03
FARIMA(1, $\frac{D}{2}$, 0)	$\widehat{D}_N, \widetilde{D}_N$	0.05, 0.25	0.05, 0.04	0.04, 0.03	0.04, 0.03	0.04, 0.04	0.03, 0.02
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.12, 0.30	0.07, 0.12	0.05, 0.07	0.04, 0.06	0.04, 0.05	0.05, 0.07
FARIMA(1, $\frac{D}{2}$, 1)	$\widehat{D}_N, \widetilde{D}_N$	0.08, 0.30	0.06, 0.04	0.05, 0.04	0.05, 0.04	0.05, 0.05	0.04, 0.03
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.13, 0.33	0.09, 0.15	0.08, 0.11	0.07, 0.09	0.06, 0.08	0.08, 0.11
$X^{(D, D')}, D' = 1$	$\widehat{D}_N, \widetilde{D}_N$	0.13, 0.19	0.11, 0.08	0.10, 0.08	0.09, 0.09	0.09, 0.09	0.08, 0.07
	$\widehat{\alpha}_N, \widetilde{\alpha}_N$	0.09, 0.15	0.10, 0.07	0.11, 0.09	0.12, 0.11	0.13, 0.13	0.11, 0.09

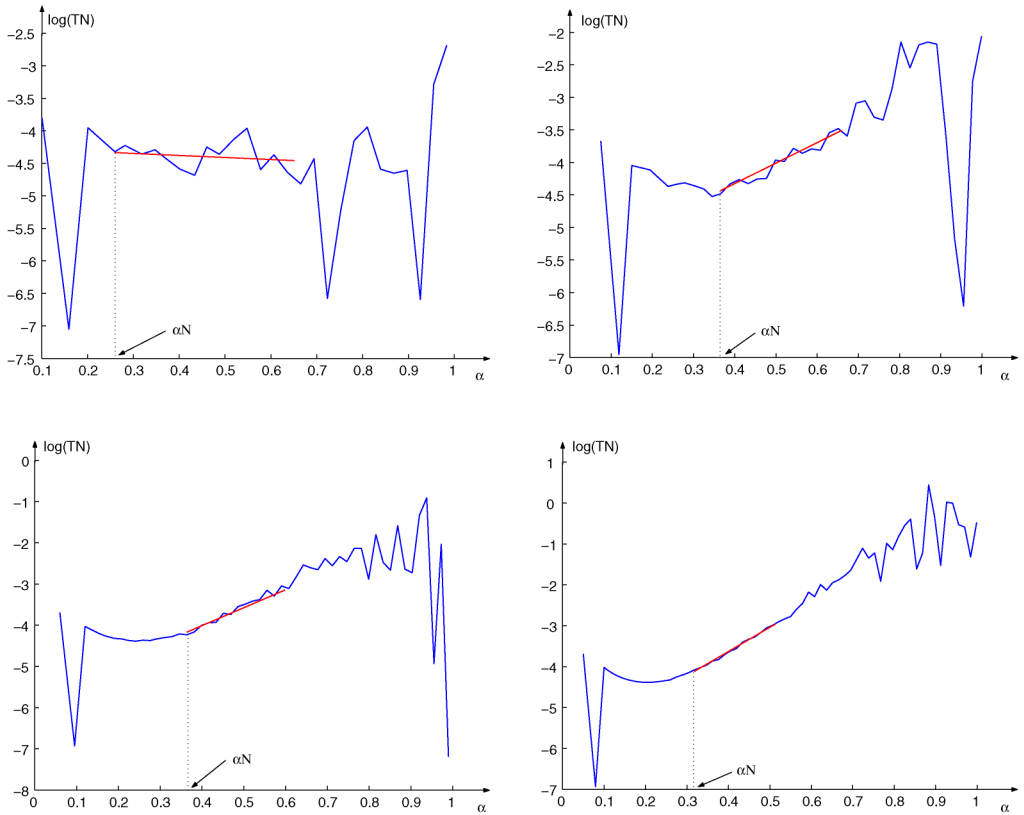


Figure 1. Log-log graphs for different samples of $X^{(D, D')}$ with $D = 0.5$ and $D' = 1$ when $N = 10^3$ (upper-left, $\widehat{D}_N \simeq 1.04$), $N = 10^4$ (upper-right, $\widehat{D}_N \simeq 0.66$), $N = 10^5$ (lower-left, $\widehat{D}_N \simeq 0.62$) and $N = 10^6$ (lower-right, $\widehat{D}_N \simeq 0.54$).

Consistency of the estimators $\widehat{\alpha}_N$ and $\widetilde{\alpha}_N$

The previous numerical results (here, we consider $\ell_1 = 15$) show that $\widehat{\alpha}_N$ and $\widetilde{\alpha}_N$ converge (very slowly) to the optimal rate α^* , that is, 0.2 for the first four processes and 1/3 for the fifth. Figure 1 illustrates the evolution with N of the log-log plotting and the choice of $\widehat{\alpha}_N$.

Figure 1 shows that $\log T_N(i \cdot N^\alpha)$ is not a linear function of the logarithm of the scales $\log(i \cdot N^\alpha)$ when N increases and $\alpha < \alpha^*$ (a consequence of Property 1 – it means there is a bias). Moreover, if $\alpha > \alpha^*$ and α increases, a linear model appears with an increasing error variance.

Consistency and distribution of the estimators \widehat{D}_N and \widetilde{D}_N

The results of Table 1 show the consistency with N of \widehat{D}_N and \widetilde{D}_N by only using $\ell_1 = 15$. Figure 2 provides the histograms of \widehat{D}_N and \widetilde{D}_N for 100 independent samples of FARIMA(1, d , 1)

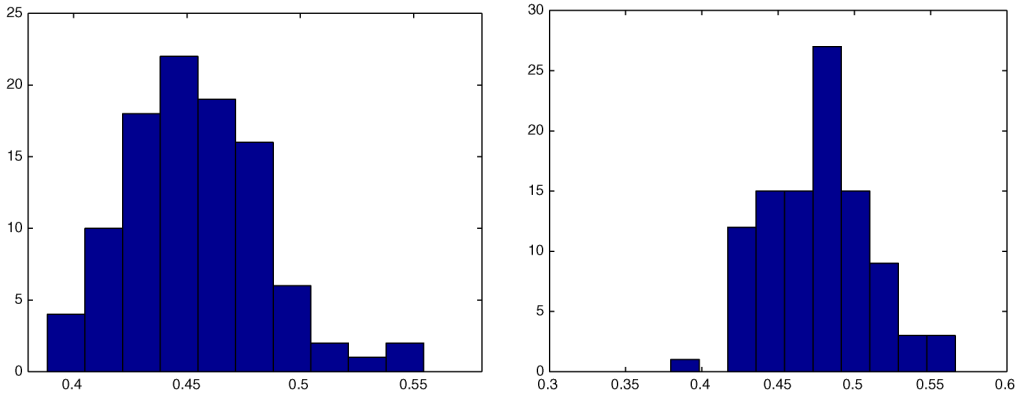


Figure 2. Histograms of \widehat{D}_N and \tilde{D}_N for 100 samples of FARIMA(1, d , 1) with $D = 0.5$ for $N = 10^5$.

processes with $D = 0.5$ and $N = 10^5$. Both histograms in Figure 2 are similar to Gaussian distribution histograms. It is not surprising for \tilde{D}_N since Theorem 1 shows that the asymptotic distribution of \tilde{D}_N is a Gaussian distribution with mean equal to D . The asymptotic distribution of \widehat{D}_N and the Gaussian distribution also seem to be similar. A Cramér–von Mises test of normality indicates that both distributions of \widehat{D}_N and \tilde{D}_N can be considered Gaussian (resp. $W \simeq 0.07$, p -value $\simeq 0.24$ and $W \simeq 0.05$, p -value $\simeq 0.54$).

Consistency in case of short memory

Table 2 provides the behavior of \widehat{D}_N and \tilde{D}_N if $D \leq 0$ and $D' > 0$. Two processes are considered in such a setting: a FARIMA(0, d , 0) process with $-0.5 < d < 0$ and therefore $-1 < D \leq 0$ (always with $D' = 2$) and a process $X^{(D,D')}$ with $D < 0$ and $D' > 0$. The results are displayed in Table 2 (here $N = 1000$, $N = 10000$ and $N = 100000$, $\ell_1 = 15$ and $\ell_2 = \lceil 5N^{0.1} \rceil$) for different choices of D and D' . Thus, it appears that \widehat{D}_N and \tilde{D}_N can be successively applied to short memory processes as well. Moreover, the larger the D' , the faster their convergence rates.

Table 2. Estimation of the memory parameter from 100 independent samples in case of short memory ($D \leq 0$)

		FARIMA(0, -0.25, 0)	$X^{(-1,1)}$	$X^{(-1,3)}$	$X^{(-3,1)}$	$X^{(-3,3)}$
$N = 10^3$	$\sqrt{MSE} \widehat{D}_N, \tilde{D}_N$	0.15, 0.20	0.30, 0.30	0.38, 0.37	0.36, 0.37	0.39, 0.38
$N = 10^4$	$\sqrt{MSE} \widehat{D}_N, \tilde{D}_N$	0.04, 0.04	0.15, 0.14	0.08, 0.08	0.13, 0.14	0.13, 0.13
$N = 10^5$	$\sqrt{MSE} \widehat{D}_N, \tilde{D}_N$	0.03, 0.03	0.06, 0.05	0.04, 0.03	0.04, 0.04	0.03, 0.03

Table 3. Estimation of the long-memory parameter from 100 independent samples in case of processes P1–4 defined above

		P1	P2	P3	P4
$N = 10^3$	$\sqrt{MSE} \widehat{\widehat{D}}_N, \widetilde{D}_N$	0.22, 0.23	0.32, 0.41	0.47, 0.76	0.40, 0.41
$N = 10^4$	$\sqrt{MSE} \widehat{\widehat{D}}_N, \widetilde{D}_N$	0.06, 0.06	0.18, 0.28	0.24, 0.65	0.13, 0.13
$N = 10^5$	$\sqrt{MSE} \widehat{\widehat{D}}_N, \widetilde{D}_N$	0.02, 0.02	0.02, 0.02	0.14, 0.47	0.03, 0.04

Robustness of $\widehat{\widehat{D}}_N, \widetilde{D}_N$

To conclude our examination of the numerical properties of the estimators, four different processes not satisfying Assumption A1' are considered:

- a FARIMA(0, d , 0) process (denoted P1) with innovations satisfying a uniform law (and $\mathbb{E}X_i^2 < \infty$);
- a FARIMA(0, d , 0) process (denoted P2) with innovations satisfying a distribution with density with respect to Lebesgue measure $f(x) = 3/4 * (1 + |x|)^{-5/2}$ for $x \in \mathbb{R}$ (and therefore $\mathbb{E}|X_i|^2 = \infty$, but $\mathbb{E}|X_i| < \infty$);
- a FARIMA(0, d , 0) process (denoted P3) with innovations satisfying a Cauchy distribution (and $\mathbb{E}|X_i| = \infty$);
- a Gaussian stationary process (denoted P4) with a spectral density $f(\lambda) = (|\lambda| - \pi/2)^{-1/2}$ for all $\lambda \in [-\pi, \pi] \setminus \{-\pi/2, \pi/2\}$, the local behavior of f in 0 being $f(|\lambda|) \sim \sqrt{\pi/2}|\lambda|^D$ with $D = 0$, but the smoothness condition for f in Assumption A1 not being satisfied.

For the first three processes, D varies in {0.1, 0.3, 0.5, 0.7, 0.9} and 100 independent replications are taken into account. The results of these simulations are given in Table 3. As outlined in the theoretical part of this paper, the estimators $\widehat{\widehat{D}}_N$ and \widetilde{D}_N also seem to be accurate for \mathbb{L}^2 -linear processes. For \mathbb{L}^α -linear processes with $1 \leq \alpha < 2$, they are also convergent, with a slower rate of convergence. Despite the fact that the spectral density of process P4 does not satisfies the smoothness hypothesis required in Assumptions A1 or A1', the convergence rates of $\widehat{\widehat{D}}_N$ and \widetilde{D}_N are still convincing. These results confirm the robustness of wavelet-based estimators.

4.2. Comparisons with other semi-parametric long-memory parameter estimators from simulations

Here, we consider only long-memory Gaussian processes ($D \in (0, 1)$) based on the usual hypothesis $0 < D' \leq 2$. More precisely, the ‘benchmark’ is 100 generated independent samples of each process with length $N = 10^3$ and $N = 10^4$ and different values of D , that is, $D = 0.1, 0.3, 0.5, 0.7, 0.9$. Several different semi-parametric estimators of D are considered:

- \widehat{D}_{BGK} , an ‘optimal’ parametric Whittle estimator obtained from a BIC criterium model selection of fractionally differenced autoregressive models (introduced by Bhansali *et al.*

(2006)), the required confidence interval of the estimation \widehat{D}_{BGK} being $[\widehat{D}_R - 2/N^{1/4}, \widehat{D}_R + 2/N^{1/4}]$;

- \widehat{D}_{GRS} , is an adaptive local periodogram estimator introduced by Giraitis *et al.* (2000) which requires two parameters – a bandwidth parameter m with a procedure of determination provided in this article, and a number of low trimmed frequencies l (satisfying different conditions, but without being fixed in this paper – after a number of simulations, $l = \max(m^{1/3}, 10)$ is chosen);
- \widehat{D}_{MS} , is an adaptive global periodogram estimator introduced by Moulines and Soulier (1998, 2003), also called FEXP estimator, with bias-variance balance parameter $\kappa = 2$;
- \widehat{D}_R , is a local Whittle estimator introduced by Robinson (1995), the trimming parameter being $m = N/30$;
- \widehat{D}_{ATV} , an adaptive wavelet-based estimator introduced by Veitch *et al.* (2003) using a Db4 wavelet (and described above);
- \widehat{D}_N defined previously with $\ell_1 = 15$ and $\ell_2 = N^{1-\hat{\alpha}_N}/10$ and a mother wavelet $\psi(t) = 100 \cdot t^2(t-1)^2(t^2-t+3/14)\mathbb{I}_{0 \leq t \leq 1}$ satisfying Assumption $W(5/2)$.

Software (using the MATLAB language) for computing some of these estimators is available on Internet (see the website of D. Veitch, <http://www.cubinlab.ee.unimelb.edu.au/~darryl/>, for \widehat{D}_{ATV} and the homepage of E. Moulines, <http://www.tsi.enst.fr/~moulines/>, for \widehat{D}_{MS} and \widehat{D}_R). Other software is available at <http://samos.univ-paris1.fr/spip/-Jean-Marc-Bardet>. Simulation results are reported in Table 4.

Comments on the results of Table 4

These simulations allow four ‘clusters’ of estimators to be distinguished.

- \widehat{D}_{BGK} is obtained from a BIC-criterion hierarchical model selection (from 2 to 11 parameters, corresponding to the length of the approximation of the Fourier expansion of the spectral density) using Whittle estimation. For these simulations, the BIC criterion is generally minimal for 5 to 7 parameters to be estimated. Simulation results are not very satisfactory, except for $D = 0.1$ (close to the short memory). Moreover, this procedure is rather time-consuming.
- \widehat{D}_{GRS} offers good results for fGn and FARIMA(0, d , 0). However, this estimator does not converge fast enough for the other processes.
- Estimators \widehat{D}_{MS} and \widehat{D}_R have similar properties. They (especially \widehat{D}_R) are very interesting because they offer the same fairly good rates of convergence for any time series of the benchmark.
- Being built on similar principles, estimators \widehat{D}_{ATV} and \widehat{D}_N also have similar behavior. Their convergence rates are the fastest for fGn and FARIMA(0, d , 0) and are fairly close to fast ones for the other processes. Their computation times especially for \widehat{D}_{ATV} , for which the computations of wavelet coefficients were carried out with the Mallat algorithm, are the shortest.

Table 4. Comparison of the different log-memory parameter estimators for processes of the benchmark – for each process and value of D and N , \sqrt{MSE} are computed from 100 independent generated samples

		$D = 0.1$	$D = 0.3$	$D = 0.5$	$D = 0.7$	$D = 0.9$
		$N = 10^3$				
fGn ($H = (D + 1)/2$)	\widehat{D}_{BGK}	0.089	0.171	0.259	0.341	0.369
	\widehat{D}_{GRS}	0.114	0.132	0.147	0.155	0.175
	\widehat{D}_{MS}	0.163	0.169	0.181	0.195	0.191
	\widehat{D}_R	0.211	0.220	0.215	0.218	0.128
	\widehat{D}_{ATV}	0.176	0.153	0.156	0.164	0.162
	$\widehat{\widehat{D}}_N$	0.139	0.147	0.133	0.140	0.150
FARIMA($0, \frac{D}{2}, 0$)	\widehat{D}_{BGK}	0.094	0.138	0.239	0.326	0.413
	\widehat{D}_{GRS}	0.131	0.139	0.150	0.150	0.162
	\widehat{D}_{MS}	0.172	0.167	0.174	0.197	0.188
	\widehat{D}_R	0.246	0.189	0.223	0.234	0.181
	\widehat{D}_{ATV}	0.128	0.107	0.081	0.074	0.065
	$\widehat{\widehat{D}}_N$	0.161	0.146	0.149	0.149	0.161
FARIMA($1, \frac{D}{2}, 0$)	\widehat{D}_{BGK}	0.146	0.203	0.239	0.236	0.212
	\widehat{D}_{GRS}	0.519	0.545	0.588	0.585	0.830
	\widehat{D}_{MS}	0.235	0.258	0.256	0.252	0.249
	\widehat{D}_R	0.242	0.241	0.234	0.202	0.144
	\widehat{D}_{ATV}	0.248	0.267	0.280	0.268	0.375
	$\widehat{\widehat{D}}_N$	0.340	0.319	0.314	0.315	0.334
FARIMA($1, \frac{D}{2}, 1$)	\widehat{D}_{BGK}	0.204	0.253	0.342	0.363	0.384
	\widehat{D}_{GRS}	0.901	0.894	0.866	0.870	0.893
	\widehat{D}_{MS}	0.181	0.175	0.180	0.185	0.181
	\widehat{D}_R	0.204	0.200	0.200	0.191	0.130
	\widehat{D}_{ATV}	0.392	0.380	0.371	0.343	0.355
	$\widehat{\widehat{D}}_N$	0.170	0.218	0.225	0.226	0.213
$X^{(D, D')}, D' = 1$	\widehat{D}_{BGK}	0.090	0.139	0.261	0.328	0.388
	\widehat{D}_{GRS}	0.342	0.339	0.331	0.300	0.315
	\widehat{D}_{MS}	0.176	0.178	0.182	0.166	0.177
	\widehat{D}_R	0.219	0.232	0.231	0.173	0.167
	\widehat{D}_{ATV}	0.153	0.161	0.168	0.176	0.176
	$\widehat{\widehat{D}}_N$	0.284	0.294	0.293	0.292	0.288

Conclusion

Which estimator, of those studied above, should be chosen in a practical setting, that is, an observed time series? We propose the following procedure for estimating an eventual long-memory parameter:

1. First, since this procedure is not very time-consuming and is applicable to processes with smooth trends, draw the plot of the log-log regression of wavelet coefficients' f variances

Table 4. (Continued)

		$D = 0.1$	$D = 0.3$	$D = 0.5$	$D = 0.7$	$D = 0.9$
		$N = 10^4$				
fGn ($H = (D + 1)/2$)	\widehat{D}_{BGK}	0.062	0.143	0.182	0.171	0.182
	\widehat{D}_{GRS}	0.040	0.047	0.054	0.068	0.066
	\widehat{D}_{MS}	0.069	0.064	0.061	0.071	0.063
	\widehat{D}_R	0.063	0.055	0.058	0.063	0.052
	\widehat{D}_{ATV}	0.036	0.042	0.041	0.047	0.045
	\widehat{D}_N	0.050	0.040	0.041	0.039	0.040
FARIMA($0, \frac{D}{2}, 0$)	\widehat{D}_{BGK}	0.059	0.141	0.195	0.187	0.178
	\widehat{D}_{GRS}	0.042	0.048	0.050	0.046	0.057
	\widehat{D}_{MS}	0.072	0.055	0.066	0.059	0.065
	\widehat{D}_R	0.073	0.053	0.064	0.057	0.059
	\widehat{D}_{ATV}	0.026	0.038	0.039	0.032	0.022
	\widehat{D}_N	0.053	0.050	0.056	0.055	0.044
FARIMA($1, \frac{D}{2}, 0$)	\widehat{D}_{BGK}	0.085	0.148	0.146	0.164	0.120
	\widehat{D}_{GRS}	0.179	0.175	0.182	0.192	0.190
	\widehat{D}_{MS}	0.109	0.105	0.099	0.100	0.094
	\widehat{D}_R	0.063	0.059	0.057	0.054	0.054
	\widehat{D}_{ATV}	0.118	0.101	0.088	0.120	0.081
	\widehat{D}_N	0.095	0.085	0.093	0.081	0.097
FARIMA($1, \frac{D}{2}, 1$)	\widehat{D}_{BGK}	0.111	0.201	0.189	0.202	0.181
	\widehat{D}_{GRS}	0.308	0.321	0.306	0.314	0.311
	\widehat{D}_{MS}	0.070	0.064	0.065	0.064	0.069
	\widehat{D}_R	0.063	0.057	0.060	0.064	0.052
	\widehat{D}_{ATV}	0.114	0.118	0.103	0.102	0.093
	\widehat{D}_N	0.095	0.099	0.087	0.101	0.090
$X^{(D, D')}, D' = 1$	\widehat{D}_{BGK}	0.069	0.110	0.204	0.190	0.197
	\widehat{D}_{GRS}	0.192	0.185	0.172	0.177	0.190
	\widehat{D}_{MS}	0.083	0.059	0.071	0.066	0.068
	\widehat{D}_R	0.066	0.057	0.068	0.054	0.064
	\widehat{D}_{ATV}	0.124	0.131	0.139	0.147	0.153
	\widehat{D}_N	0.158	0.143	0.152	0.158	0.155

onto scales. If a linear zone appears in this graph, consider the estimator \widehat{D}_N (or \widehat{D}_{ATV}) of D .

2. If a linear zone appears in the previous graph and if the observed time series seems to be without a trend, compute \widehat{D}_R .
3. Compare both the estimated values of D from confidence intervals (available for \widehat{D}_N or \widehat{D}_{ATV} and \widehat{D}_R).

5. Proofs

Proof of Property 1. The arguments of this proof are similar to those of Abry *et al.* (1998) or Moulines *et al.* (2007). First, for $a \in \mathbb{N}^*$,

$$\begin{aligned}
 & \mathbb{E}(e^2(a, 0)) \\
 &= \frac{1}{a} \sum_{k=1}^a \sum_{k'=1}^a \psi(k/a) \psi(k'/a) \mathbb{E}(X_k X_{k'}) \\
 &= \frac{1}{a} \sum_{k=1}^a \sum_{k'=1}^a \psi(k/a) \psi(k'/a) r(k - k') \\
 &= \frac{1}{a} \sum_{k=1}^a \sum_{k'=1}^a \psi(k/a) \psi(k'/a) \int_{-\pi}^{\pi} f(\lambda) e^{i\lambda(k-k')} d\lambda \\
 &= \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times \frac{1}{a^2} \sum_{k=1}^a \sum_{k'=1}^a \psi\left(\frac{k}{a}\right) \psi\left(\frac{k'}{a}\right) e^{iu(k/a-k'/a)} du \\
 &= \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times \left\{ \left(\frac{1}{a} \sum_{k=1}^a \psi\left(\frac{k}{a}\right) \cos\left(\frac{k}{a}u\right) \right)^2 + \left(\frac{1}{a} \sum_{k=1}^a \psi\left(\frac{k}{a}\right) \sin\left(\frac{k}{a}u\right) \right)^2 \right\} du.
 \end{aligned} \tag{18}$$

Now, it is well known that if $\psi \in \tilde{W}(\beta, L)$, the Sobolev space with parameters $\beta > 1/2$ and $L > 0$, then

$$\sup_{|u| \leq a\pi} \Delta_a(u) \leq C_{\beta,L} \frac{1}{a^{\beta-1/2}} \quad \text{with } \Delta_a(u) := \left| \frac{1}{a} \sum_{k=1}^a \psi\left(\frac{k}{a}\right) e^{-iuk/a} - \int_0^1 \psi(t) e^{-iut} dt \right|, \tag{19}$$

with $C_{\beta,L} > 0$ depending only on β and L (see, e.g., Devore and Lorentz (1993)). Therefore, if ψ satisfies Assumption $W(\infty)$ and X satisfies Assumption **A1**, then for all $\beta > 1/2$, since $\sup_{u \in \mathbb{R}} |\widehat{\psi}(u)| < \infty$,

$$\begin{aligned}
 & \left| \mathbb{E}(e^2(a, 0)) - \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \\
 & \leq 2C_{\beta,L} \frac{2}{a^{\beta-3/2}} \int_0^{a\pi} f\left(\frac{u}{a}\right) |\widehat{\psi}(u)| du + C_{\beta,L}^2 \frac{2}{a^{2\beta-2}} \int_0^{a\pi} f\left(\frac{u}{a}\right) du \\
 & \leq 2 \cdot C_{\beta,L}^2 \frac{2}{a^{2\beta-3}} \int_0^{\pi} f(v) dv
 \end{aligned} \tag{20}$$

since $\sup_{u \in \mathbb{R}} (1 + u^n) |\widehat{\psi}(u)| < \infty$ for all $n \in \mathbb{N}$. Consequently, if ψ satisfies Assumption $W(\infty)$, then for all $n > 0$ and all $a \in \mathbb{N}^*$, there exists $C(n) > 0$, not depending on a , such that

$$\left| \mathbb{E}(e^2(a, 0)) - \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \leq C(n) \frac{1}{a^n}. \quad (21)$$

But, from Assumption $W(\infty)$, for all $c < 1$,

$$K_{(\psi, c)} = \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(u)|^2}{|u|^c} du < \infty$$

because Assumption $W(\infty)$ implies that $|\widehat{\psi}(u)| = O(|u|)$ when $u \rightarrow 0$ and there exists $p > 1 - c$ such that $\sup_{u \in \mathbb{R}} |\widehat{\psi}(u)|^2 (1 + |u|)^p < \infty$. Moreover, for all $p > 1 - c$,

$$\begin{aligned} \left| \int_{-a\pi}^{a\pi} \frac{|\widehat{\psi}(u)|^2}{|u|^c} du - K_{(\psi, c)} \right| &= 2 \int_{a\pi}^{\infty} \frac{|\widehat{\psi}(u)|^2}{u^c} du \\ &\leq C \cdot \int_{a\pi}^{\infty} \frac{1}{u^{p+c}} du \\ &\leq C' \cdot \frac{1}{a^{p+c-1}} \end{aligned}$$

with $C > 0$ and $C' > 0$ not depending on a . As a consequence, under Assumption **A1**, for all $p > 1 - D$, all $n \in \mathbb{N}$ and all $a \in \mathbb{N}^*$,

$$\begin{aligned} &\left| \mathbb{E}(e^2(a, 0)) - f^*(0) \cdot \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(u)|^2}{|u/a|^D} du \right| \\ &\leq 2f^*(0)a^D \int_{a\pi}^{\infty} \frac{|\widehat{\psi}(u)|^2}{u^D} du + C_{D'} a^{D-D'} \int_{-a\pi}^{a\pi} \frac{|\widehat{\psi}(u)|^2}{|u|^{D-D'}} du + C(n) \frac{1}{a^n} \\ \implies &|\mathbb{E}(e^2(a, 0)) - f^*(0)K_{(\psi, D)} \cdot a^D| \leq C' f^*(0) \cdot a^{1-p} + C_{D'} K_{(\psi, D-D')} \cdot a^{D-D'}. \end{aligned}$$

Now, by choosing p such that $1 - p < D - D'$, the inequality (6) is obtained. □

Proof of Property 2. Using the proof of Property 1, with Assumption $W(5/2)$, ψ is included in a Sobolev space $\tilde{W}(5/2, L)$, inequality (19) is satisfied with $\beta = 5/2$ and (20) is replaced by

$$\left| \mathbb{E}(e^2(a, 0)) - a \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \leq 2 \cdot C_{5/2, L}^2 \frac{2}{a^2} \int_0^\pi f(v) dv \quad (22)$$

since $\sup_{u \in \mathbb{R}} (1 + u^{3/2}) |\widehat{\psi}(u)| < \infty$. Therefore, inequality (21) is replaced by

$$\left| \mathbb{E}(e^2(a, 0)) - a \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \leq C(2) \frac{1}{a^2}.$$

The end of the proof is similar to the end of the previous proof, but now $K_{(\psi,c)}$ exists for $-2 < c < 1$ and

$$\left| \int_{-a\pi}^{a\pi} \frac{|\widehat{\psi}(u)|^2}{|u|^c} du - K_{(\psi,c)} \right| \leq C' \cdot \frac{1}{a^{2+c}}.$$

Finally, under Assumption A1', for all $a \in \mathbb{N}^*$, since $-2 < D - D' < 1$,

$$|\mathbb{E}(e^2(a, 0)) - f^*(0)K_{(\psi,D)} \cdot a^D| \leq C_{D'}K_{(\psi,D-D')} \cdot a^{D-D'} + C' \frac{1}{a^2},$$

which completes the proof. □

Proof of Corollary 1. Both of these proofs provide the main arguments needed to establish (7). For improved readability, we will consider only Assumption A1' and Assumption $W(\infty)$ (the long-memory process being similar). The main difference consists of specifying the asymptotic behavior of $\int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du$. But,

$$\int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du = \int_{-\sqrt{a}}^{\sqrt{a}} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du + 2 \int_{\sqrt{a}}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du. \tag{23}$$

The asymptotic behavior of $\widehat{\psi}(u)$ when $u \rightarrow \infty$ (ψ is supposed to satisfy Assumption $W(\infty)$) implies

$$\int_{\sqrt{a}}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \leq Ca^D \int_{\sqrt{a}}^{\infty} u^{-D} \times |\widehat{\psi}(u)|^2 du \leq \frac{C(n)}{a^n}, \tag{24}$$

for all $n \in \mathbb{N}$. Moreover,

$$\begin{aligned} & \int_{-\sqrt{a}}^{\sqrt{a}} f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du \\ &= f^*(0) \int_{-\sqrt{a}}^{\sqrt{a}} \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) |\widehat{\psi}(u)|^2 du \\ &+ \int_{-\sqrt{a}}^{\sqrt{a}} \left(f\left(\frac{u}{a}\right) - f^*(0) \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) \right) |\widehat{\psi}(u)|^2 du. \end{aligned} \tag{25}$$

From computations in previous proofs,

$$\begin{aligned} & \int_{-\sqrt{a}}^{\sqrt{a}} \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) |\widehat{\psi}(u)|^2 du \\ &= K_{(\psi,D)} \cdot a^D + C_{D'}K_{(\psi,D-D')} \cdot a^{D-D'} + \Lambda(a) \end{aligned} \tag{26}$$

and $|\Lambda(a)| \leq \frac{C(n)}{a^n}$. Finally, using $f(\lambda) = f^*(0)(|\lambda|^{-D} + C_{D'}|\lambda|^{D'-D}) + o(|\lambda|^{D'-D})$ when $\lambda \rightarrow 0$, we obtain

$$\begin{aligned} & \int_{-\sqrt{a}}^{\sqrt{a}} \left(f\left(\frac{u}{a}\right) - f^*(0) \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) \right) |\widehat{\psi}(u)|^2 du \\ &= \int_{-\sqrt{a}}^{\sqrt{a}} \left| \frac{u}{a} \right|^{D-D'} \left(f\left(\frac{u}{a}\right) - f^*(0) \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) \right) |\widehat{\psi}(u)|^2 \left| \frac{u}{a} \right|^{D'-D} du \\ &= a^{D-D'} \int_{-\sqrt{a}}^{\sqrt{a}} g(u, a) |\widehat{\psi}(u)|^2 |u|^{D'-D} du \end{aligned}$$

for all $u \in [-\sqrt{a}, \sqrt{a}]$, $g(u, a) \rightarrow 0$ when $a \rightarrow \infty$. Therefore, from Lebesgue's theorem (satisfied by the asymptotic behavior of $\widehat{\psi}$),

$$\lim_{a \rightarrow \infty} a^{D-D'} \int_{-\sqrt{a}}^{\sqrt{a}} \left(f\left(\frac{u}{a}\right) - f^*(0) \left(\left| \frac{u}{a} \right|^{-D} + C_{D'} \left| \frac{u}{a} \right|^{D'-D} \right) \right) |\widehat{\psi}(u)|^2 du = 0. \tag{27}$$

As a consequence, by (23), (24), (25), (26) and (27), the corollary is proved. □

Proof of Proposition 1. This proof can be decomposed into three steps: **Step 1**, **Step 2** and **Step 3**.

Step 1. In this step, $\frac{N}{a_N} \cdot \text{Cov}(\tilde{T}_N(r_i a_N), \tilde{T}_N(r_j a_N))_{1 \leq i, j \leq \ell}$ is proved to converge to an asymptotic covariance matrix Γ . First, for all $(i, j) \in \{1, \dots, \ell\}^2$,

$$\begin{aligned} & \text{Cov}(\tilde{T}_N(r_i a_N), \tilde{T}_N(r_j a_N)) \\ &= 2 \frac{1}{[N/r_i a_N]} \frac{1}{[N/r_j a_N]} \sum_{p=1}^{[N/r_i a_N]} \sum_{q=1}^{[N/r_j a_N]} (\text{Cov}(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q)))^2 \end{aligned} \tag{28}$$

because X is a Gaussian process. Therefore, by considering only $i = j$ and $p = q$, for N and a_N large enough,

$$\text{Cov}(\tilde{T}_N(r_i a_N), \tilde{T}_N(r_i a_N)) \geq \frac{1}{r_i} \frac{N}{a_N}. \tag{29}$$

Now, for $(p, q) \in \{1, \dots, [N/r_i a_N]\} \times \{1, \dots, [N/r_i a_N]\}$,

$$\begin{aligned} & \text{Cov}(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q)) \\ &= \frac{a_N^{1-D} (r_i r_j)^{(1-D)/2}}{f^*(0) K(\psi, D)} \frac{1}{r_i a_N} \frac{1}{r_j a_N} \sum_{k=1}^{r_i a_N} \sum_{k'=1}^{r_j a_N} \psi\left(\frac{k}{r_i a_N}\right) \psi\left(\frac{k'}{r_j a_N}\right) r(k - k' + a_N(r_i p - r_j q)) \\ &= \frac{a_N^{1-D} (r_i r_j)^{(1-D)/2}}{f^*(0) K(\psi, D)} \frac{1}{r_i a_N} \frac{1}{r_j a_N} \end{aligned}$$

$$\begin{aligned} & \times \sum_{k=1}^{r_i a_N} \sum_{k'=1}^{r_j a_N} \psi\left(\frac{k}{r_i a_N}\right) \psi\left(\frac{k'}{r_j a_N}\right) \int_{-\pi}^{\pi} d\lambda f(\lambda) e^{-i\lambda(k-k'+a_N(r_i p-r_j q))} \\ &= \frac{(r_i r_j)^{(1-D)/2}}{a_N^D f^*(0) K_{(\psi, D)}} \frac{1}{r_i a_N} \frac{1}{r_j a_N} \\ & \times \sum_{k=1}^{r_i a_N} \sum_{k'=1}^{r_j a_N} \psi\left(\frac{k}{r_i a_N}\right) \psi\left(\frac{k'}{r_j a_N}\right) \int_{-\pi a_N}^{\pi a_N} du f\left(\frac{u}{a_N}\right) e^{-iu\left(\frac{k}{a_N}-\frac{k'}{a_N}+r_i p-r_j q\right)}. \end{aligned}$$

Using the same expansion as in (21), under Assumption $W(\infty)$ the previous equality becomes, for all $n \in \mathbb{N}^*$,

$$\begin{aligned} & \left| \text{Cov}(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q)) \right. \\ & \left. - \frac{(r_i r_j)^{(1-D)/2}}{a_N^D f^*(0) K_{(\psi, D)}} \int_{-\pi a_N}^{\pi a_N} du \widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j) f\left(\frac{u}{a_N}\right) e^{-iu(r_i p-r_j q)} \right| \\ & \leq \frac{C(n)}{a_N^{n+D}} \int_{-\pi a_N}^{\pi a_N} du \left| \widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j) f\left(\frac{u}{a_N}\right) \right| \tag{30} \\ & \leq \frac{C'(n)}{a_N^n} \int_{-\infty}^{\infty} du |u|^{-D} |\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)| \\ & \leq \frac{C''(n)}{a_N^n}, \end{aligned}$$

with $C(n), C'(n), C''(n) > 0$ not depending on a_N , due to the asymptotic behaviors of $\widehat{\psi}(u)$ when $u \rightarrow 0$ and $u \rightarrow \infty$. Now, under Assumption A1,

$$\begin{aligned} & \left| \int_{-\pi a_N}^{\pi a_N} du \widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j) f\left(\frac{u}{a_N}\right) e^{-iu(r_i p-r_j q)} \right. \\ & \left. - a_N f^*(0) \int_{-\pi}^{\pi} du \frac{\widehat{\psi}(ur_i a_N) \overline{\widehat{\psi}}(ur_j a_N)}{|u|^D} e^{-iu a_N(r_i p-r_j q)} \right| \tag{31} \\ & \leq a_N^{D-D'} f^*(0) C_{D'} \int_{-\pi a_N}^{\pi a_N} du \frac{|\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)|}{|u|^{D-D'}} \\ & \leq a_N^{D-D'} f^*(0) C_{D'} \int_{-\infty}^{\infty} du \frac{|\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)|}{|u|^{D-D'}} \end{aligned}$$

since

$$\int_{-\infty}^{\infty} du \frac{|\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)|}{|u|^{D-D'}} < \infty$$

from Assumption $W(\infty)$. Finally, from (30) and (31), we have $C > 0$ not depending on N such that for all $a_N \in \mathbb{N}^*$,

$$\begin{aligned} & \left| \text{Cov}(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q)) \right. \\ & \left. - \frac{a_N^{1-D} (r_i r_j)^{(1-D)/2}}{K(\psi, D)} \int_{-\pi}^{\pi} du \frac{\widehat{\psi}(ur_i a_N) \overline{\widehat{\psi}(ur_j a_N)}}{|u|^D} e^{-iua_N(r_i p - r_j q)} \right| \quad (32) \\ & \leq C a_N^{-D'}. \end{aligned}$$

It remains to evaluate $a_N^{1-D} \int_{-\pi}^{\pi} du \frac{\widehat{\psi}(ur_i a_N) \overline{\widehat{\psi}(ur_j a_N)}}{|u|^D} e^{-iua_N(r_i p - r_j q)} = \int_{-\pi a_N}^{\pi a_N} du \frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}}{|u|^D} \times e^{-iu(r_i p - r_j q)}$. Thus, if $|r_i p - r_j q| \geq 1$, using an integration by parts,

$$\begin{aligned} & \left| \int_{-\pi a_N}^{\pi a_N} du \frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}}{|u|^D} e^{-iu(r_i p - r_j q)} \right| \\ & = \left| \frac{1}{-i(r_i p - r_j q)} \left[\frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}}{u^D} e^{-iu(r_i p - r_j q)} \right]_{-\pi a_N}^{\pi a_N} \right. \\ & \quad \left. + \frac{1}{i(r_i p - r_j q)} \int_{-\pi a_N}^{\pi a_N} du \frac{\partial}{\partial u} \left(\frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}}{u^D} \right) e^{-iu(r_i p - r_j q)} \right| \quad (33) \\ & \leq \frac{1}{|r_i p - r_j q|} \int_{-\infty}^{\infty} \left(\frac{D}{|u|^{D+1}} |\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}| + \frac{1}{|u|^D} \left| \frac{\partial}{\partial u} (\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}) \right| \right) du \\ & \leq C \frac{1}{|r_i p - r_j q|} \end{aligned}$$

with $C < \infty$ not depending on N since:

- $\widehat{\psi}(\pi r_i a_N) \overline{\widehat{\psi}(\pi r_j a_N)} = \widehat{\psi}(-\pi r_i a_N) \overline{\widehat{\psi}(-\pi r_j a_N)}$ and $\sin(\pi a_N(r_i p - r_j q)) = 0$;
- from Assumption $W(\infty)$, $\limsup_{u \rightarrow 0} u^{-1} |\widehat{\psi}(u)| < \infty$, $\limsup_{u \rightarrow 0} \left| \frac{\partial}{\partial u} \widehat{\psi}(u) \right| < \infty$

$$\implies \limsup_{u \rightarrow 0} u^{-1} \left| \frac{\partial}{\partial u} (\widehat{\psi}(ur_i) \overline{\widehat{\psi}(ur_j)}) \right| < \infty;$$

- from Assumption $W(\infty)$, for all $n \in \mathbb{N}$, $\sup_{u \in \mathbb{R}} (1 + |u|)^n |\widehat{\psi}(u)| < \infty$ and $\sup_{u \in \mathbb{R}} (1 + |u|)^n \left| \frac{\partial}{\partial u} \widehat{\psi}(u) \right| < \infty$.

Moreover, if $|r_i p - r_j q| = 0$, by the Cauchy–Schwarz inequality and Property 1, for a_N large enough,

$$|\text{Cov}(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q))| \leq (\mathbb{E}(\tilde{e}^2(r_i a_N, p)) \cdot \mathbb{E}(\tilde{d}^2(r_j a_N, q)))^{1/2} \leq 2. \quad (34)$$

Therefore, using (32), (33), (34) and the inequality $(x + y)^2 \leq 2(x^2 + y^2)$ for all $(x, y) \in \mathbb{R}^2$, we have $C > 0$ such that for a_N large enough,

$$\text{Cov}^2(\tilde{e}(r_i a_N, p), \tilde{e}(r_j a_N, q)) \leq C \left(\frac{1}{(1 + |r_i p - r_j q|)^2} + \frac{1}{a_N^{2D'}} \right). \tag{35}$$

Hence, by (28),

$$\begin{aligned} & |\text{Cov}(\tilde{T}_N(r_i a_N), \tilde{T}_N(r_j a_N))| \\ & \leq C \frac{1}{[N/r_i a_N]} \frac{1}{[N/r_j a_N]} \sum_{p=1}^{[N/r_i a_N]} \sum_{q=1}^{[N/r_j a_N]} \left(\frac{1}{(1 + |r_i p - r_j q|)^2} + \frac{1}{a_N^{2D'}} \right). \end{aligned}$$

But, from unusual comparison of series and integrals,

$$\begin{aligned} \sum_{p=1}^{[N/r_i a_N]} \sum_{q=1}^{[N/r_j a_N]} (1 + |r_i p - r_j q|)^{-2} & \leq \frac{1}{r_i r_j} \int_0^{N/a_N} \int_0^{N/a_N} \frac{du dv}{(1 + |u - v|)^2} \\ & \leq \frac{2}{r_i r_j} \int_0^{N/a_N} \frac{N/a_N dw}{(1 + w)^2} \\ & \leq \frac{2}{r_i r_j} \cdot \frac{N}{a_N}. \end{aligned}$$

As a consequence, if a_N is such that $\limsup_{N \rightarrow \infty} \frac{N}{a_N} \frac{1}{a_N^{2D'}} < \infty$, then $\limsup_{N \rightarrow \infty} \frac{N}{a_N} \times |\text{Cov}(\tilde{T}_N(r_i a_N), \tilde{T}_N(r_j a_N))| < \infty$. More precisely, since this covariance is a sum of positive terms, if $\limsup_{N \rightarrow \infty} \frac{N}{a_N} \frac{1}{a_N^{2D'}} = 0$, then

$$\lim_{N \rightarrow \infty} \frac{N}{a_N} (\text{Cov}(\tilde{S}_N(r_i a_N), \tilde{S}_N(r_j a_N)))_{1 \leq i, j \leq \ell} = \Gamma(r_1, \dots, r_\ell, \psi, D), \tag{36}$$

a non-null (from (29)) symmetric matrix with $\Gamma(r_1, \dots, r_\ell, \psi, D) = (\gamma_{ij})_{1 \leq i, j \leq \ell}$ that can be specified. Indeed, from the previous computations, if $\limsup_{N \rightarrow \infty} \frac{N}{a_N} \frac{1}{a_N^{2D'}} = 0$, then

$$\begin{aligned} \gamma_{ij} & = \lim_{N \rightarrow \infty} \frac{8r_i r_j a_N}{N} \\ & \times \sum_{p=1}^{[N/r_i a_N]} \sum_{q=1}^{[N/r_j a_N]} \left(\frac{(r_i r_j)^{(1-D)/2}}{K(\psi, D)} \int_0^\infty du \frac{\widehat{\psi}(ur_i) \overline{\widehat{\psi}}(ur_j)}{u^D} \cos(u(r_i p - r_j q)) \right)^2 \\ & = \lim_{N \rightarrow \infty} \frac{8(r_i r_j)^{2-D} a_N}{K(\psi, D)^2 N} \end{aligned}$$

$$\begin{aligned} & \times \sum_{m=-\lceil N/d_{ij}a_N \rceil+1}^{\lfloor N/d_{ij}a_N \rfloor-1} \left(\frac{N}{d_{ij}a_N} - |m| \right) \left(\int_0^\infty du \frac{\widehat{\psi}(ur_i)\overline{\widehat{\psi}}(ur_j)}{u^D} \cos(ud_{ij}m) \right)^2 \\ & = \frac{8(r_i, r_j)^{2-D}}{K_{(\psi, D)}^2 d_{ij}} \sum_{m=-\infty}^\infty \left(\int_0^\infty \frac{\widehat{\psi}(ur_i)\overline{\widehat{\psi}}(ur_j)}{u^D} \cos(ud_{ij}m) du \right)^2 \end{aligned}$$

with $d_{ij} = \text{GCD}(r_i; r_j)$. Therefore, the matrix Γ depends only on $r_1, \dots, r_\ell, \psi, D$.

Step 2. Generally speaking, the above result is not sufficient to obtain the central limit theorem

$$\sqrt{\frac{N}{a_N}} (\tilde{T}_N(r_i a_N) - \mathbb{E}(\tilde{e}^2(r_i a_N, 0)))_{1 \leq i \leq \ell} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_\ell(0, \Gamma(r_1, \dots, r_\ell, \psi, D)). \quad (37)$$

However, each $\tilde{T}_N(r_i a_N)$ is a quadratic form of a Gaussian process. Mutatis mutandis, it is exactly the same framework (i.e., a Lindeberg central limit theorem) as that of Bardet (2000), Proposition 2.1 and (37) is satisfied. Moreover, if $(a_n)_n$ is such that $\limsup_{N \rightarrow \infty} \frac{N}{a_N^{1+2D'}} = 0$, then using the asymptotic behavior of $\mathbb{E}(\tilde{e}^2(r_i a_N, 0))$ provided in Property 1,

$$\sqrt{\frac{N}{a_N}} (\mathbb{E}(\tilde{e}^2(r_i a_N, 0)))_{N \rightarrow \infty} \longrightarrow 0.$$

As a consequence, under those assumptions,

$$\sqrt{\frac{N}{a_N}} (\tilde{T}_N(r_i a_N) - 1)_{1 \leq i \leq \ell} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}_\ell(0, \Gamma(r_1, \dots, r_\ell, \psi, D)). \quad (38)$$

Step 3. The logarithm function $(x_1, \dots, x_\ell) \in (0, +\infty)^\ell \mapsto (\log x_1, \dots, \log x_m)$ is \mathcal{C}^2 on $(0, +\infty)^\ell$. As a consequence, using the Delta method, the central limit theorem (10) for the vector $(\log \tilde{T}_N(r_i a_N))_{1 \leq i \leq \ell}$ follows with the same asymptotical covariance matrix $\Gamma(r_1, \dots, r_\ell, \psi, D)$ (because the Jacobian matrix of the function in $(1, \dots, 1)$ is the identity matrix). \square

Proof of Proposition 2. There is a perfect identity between this proof and that of Proposition 1, both being based on the approximations of Fourier transforms provided in the proof of Property 2. \square

Proof of Corollary 3. It is clear that $X'_t = X_t + P_m(t)$ for all $t \in \mathbb{Z}$, with $X = (X_t)_t$ satisfying Propositions 1 and 2. But, any wavelet coefficient of $(P_m(t))_t$ is obviously null from the assumption on ψ . Therefore, the statistic \widehat{T}_N is the same for X and X' . \square

Proof of Proposition 5. Let $\varepsilon > 0$ be a fixed positive real number such that $\alpha^* + \varepsilon < 1$.

I. First, a bound of $\Pr(\widehat{\alpha}_N \leq \alpha^* + \varepsilon)$ is provided. Indeed,

$$\Pr(\widehat{\alpha}_N \leq \alpha^* + \varepsilon) \geq \Pr\left(\widehat{Q}_N(\alpha^* + \varepsilon/2) \leq \min_{\alpha \geq \alpha^* + \varepsilon \text{ and } \alpha \in \mathcal{A}_N} \widehat{Q}_N(\alpha)\right)$$

$$\begin{aligned} &\geq 1 - \Pr\left(\bigcup_{\alpha \geq \alpha^* + \varepsilon \text{ and } \alpha \in \mathcal{A}_N} \widehat{Q}_N(\alpha^* + \varepsilon/2) > \widehat{Q}_N(\alpha)\right) \tag{39} \\ &\geq 1 - \sum_{k=\lceil(\alpha^* + \varepsilon) \log N\rceil}^{\log[N/\ell]} \Pr\left(\widehat{Q}_N(\alpha^* + \varepsilon/2) > \widehat{Q}_N\left(\frac{k}{\log N}\right)\right). \end{aligned}$$

But, for $\alpha \geq \alpha^* + 1$,

$$\Pr(\widehat{Q}_N(\alpha^* + \varepsilon/2) > \widehat{Q}_N(\alpha)) = \Pr(\|P_N(\alpha^* + \varepsilon/2) \cdot Y_N(\alpha^* + \varepsilon/2)\|^2 > \|P_N(\alpha) \cdot Y_N(\alpha)\|^2)$$

with $P_N(\alpha) = I_\ell - A_N(\alpha) \cdot (A'_N(\alpha) \cdot A_N(\alpha))^{-1} \cdot A_N(\alpha)$ for all $\alpha \in (0, 1)$, that is, $P_N(\alpha)$ is the matrix of an orthogonal projection on the orthogonal subspace (in \mathbb{R}^ℓ) generated by $A_N(\alpha)$ (and I_ℓ is the identity matrix in \mathbb{R}^ℓ). From the expression for $A_N(\alpha)$, it is obvious that for all $\alpha \in (0, 1)$,

$$P_N(\alpha) = P = I_\ell - A \cdot (A' \cdot A)^{-1} \cdot A,$$

with the matrix

$$A = \begin{pmatrix} \log(r_1) & 1 \\ \vdots & \vdots \\ \log(r_\ell) & 1 \end{pmatrix}$$

as in Proposition 3. Thus,

$$\begin{aligned} &\Pr(\widehat{Q}_N(\alpha^* + \varepsilon/2) > \widehat{Q}_N(\alpha)) \\ &= \Pr(\|P \cdot Y_N(\alpha^* + \varepsilon/2)\|^2 > \|P \cdot Y_N(\alpha)\|^2) \\ &= \Pr\left(\left\|P \cdot \sqrt{\frac{N}{N\alpha^* + \varepsilon/2}} Y_N(\alpha^* + \varepsilon/2)\right\|^2 > N^{\alpha - (\alpha^* + \varepsilon/2)} \left\|P \cdot \sqrt{\frac{N}{N\alpha}} Y_N(\alpha)\right\|^2\right) \\ &\leq \Pr(V_N(\alpha^* + \varepsilon/2) > N^{(\alpha - (\alpha^* + \varepsilon/2))/2}) + \Pr(V_N(\alpha) \leq N^{-(\alpha - (\alpha^* + \varepsilon/2))/2}) \end{aligned}$$

with $V_N(\alpha) = \|P \cdot \sqrt{\frac{N}{N\alpha}} Y_N(\alpha)\|^2$ for all $\alpha \in (0, 1)$. From Proposition 1, for all $\alpha > \alpha^*$, the asymptotic law of $P \cdot \sqrt{\frac{N}{N\alpha}} Y_N(\alpha)$ is a Gaussian law with covariance matrix $P \cdot \Gamma \cdot P'$. Moreover, the rank of the matrix $P \cdot \Gamma \cdot P'$ is $\ell - 2$ (this is the rank of P) and there exists $0 < \lambda_-$, not depending on N , such that $P \cdot \Gamma \cdot P' - \lambda_- P \cdot P'$ is a non-negative matrix ($0 < \lambda_- < \min\{\lambda \in \text{Sp}(\Gamma)\}$). As a consequence, for large enough N ,

$$\begin{aligned} \Pr(V_N(\alpha) \leq N^{-(\alpha - (\alpha^* + \varepsilon/2))/2}) &\leq 2 \cdot \Pr(V_- \leq N^{-(\alpha - (\alpha^* + \varepsilon/2))/2}) \\ &\leq \frac{1}{2^{\ell/2 - 2} \Gamma(\ell/2)} \cdot \left(\frac{N}{\lambda_-}\right)^{-(\ell/2 - 1)((\alpha - (\alpha^* + \varepsilon/2))/2)}, \end{aligned}$$

with $V_- \sim \lambda_- \cdot \chi^2(\ell - 2)$. Moreover, from the Markov inequality,

$$\begin{aligned} \Pr(V_N(\alpha^* + \varepsilon/2) > N^{(\alpha - (\alpha^* + \varepsilon/2))/2}) &\leq 2 \cdot \Pr(\exp(\sqrt{V_+}) > \exp(N^{(\alpha - (\alpha^* + \varepsilon/2))/4})) \\ &\leq 2 \cdot \mathbb{E}(\exp(\sqrt{V_+})) \cdot \exp(-N^{(\alpha - (\alpha^* + \varepsilon/2))/4}) \end{aligned}$$

with $V_+ \sim \lambda_+ \cdot \chi^2(\ell - 2)$ and $\lambda_+ > \max\{\lambda \in \text{Sp}(\Gamma)\} > 0$. As $\mathbb{E}(\exp(\sqrt{V_+})) < \infty$ does not depend on N , we obtain that there exists some $M_1 > 0$, not depending on N , such that for large enough N ,

$$\Pr(\widehat{Q}_N(\alpha^* + \varepsilon/2) > \widehat{Q}_N(\alpha)) \leq M_1 \cdot N^{-(\ell/2-1)((\alpha - (\alpha^* + \varepsilon/2))/2)}$$

and, therefore, the inequality (39) becomes, for N large enough,

$$\begin{aligned} \Pr(\widehat{\alpha}_N \leq \alpha^* + \varepsilon) &\geq 1 - M_1 \cdot \sum_{k=[(\alpha^* + \varepsilon) \log N]}^{\log[N/\ell]} N^{-((\ell-2)/4)((k/(\log N)) - (\alpha^* + \varepsilon/2))} \\ &\geq 1 - M_1 \cdot \log N \cdot N^{-((\ell-2)/12)\varepsilon}. \end{aligned} \tag{40}$$

II. Second, a bound for $\Pr(\widehat{\alpha}_N \geq \alpha^* - \varepsilon)$ is provided. Following the above arguments and notation,

$$\begin{aligned} \Pr(\widehat{\alpha}_N \geq \alpha^* - \varepsilon) &\geq \Pr\left(\widehat{Q}_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*}\varepsilon\right) \leq \min_{\alpha \leq \alpha^* - \varepsilon \text{ and } \alpha \in \mathcal{A}_N} \widehat{Q}_N(\alpha)\right) \\ &\geq 1 - \sum_{k=2}^{[(\alpha^* - \varepsilon) \log N] + 1} \Pr\left(\widehat{Q}_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*}\varepsilon\right) > \widehat{Q}_N\left(\frac{k}{\log N}\right)\right) \end{aligned} \tag{41}$$

and, as above,

$$\begin{aligned} &\Pr\left(\widehat{Q}_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*}\varepsilon\right) > \widehat{Q}_N(\alpha)\right) \\ &= \Pr\left(\left\|P \cdot \sqrt{\frac{N}{N^{\alpha^* + ((1 - \alpha^*)/(2\alpha^*))\varepsilon}}} Y_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*}\varepsilon\right)\right\|^2\right. \\ &\quad \left.> N^{\alpha - (\alpha^* + ((1 - \alpha^*)/(2\alpha^*))\varepsilon)} \left\|P \cdot \sqrt{\frac{N}{N^\alpha}} Y_N(\alpha)\right\|^2\right). \end{aligned} \tag{42}$$

Now, in the case $a_N = N^\alpha$ with $\alpha \leq \alpha^*$, the sample variance of wavelet coefficients is biased. In this case, from the relation of Corollary 1 under Assumption A1',

$$(Y_N(\alpha))_{1 \leq i \leq \ell} = \left(\frac{C_{D'} K(\psi, D - D')}{f^*(0) K(\psi, D)} (i N^\alpha)^{-D'} (1 + o_i(1))\right)_{1 \leq i \leq \ell} + \left(\sqrt{\frac{N^\alpha}{N}} \cdot \varepsilon_N(\alpha)\right)_{1 \leq i \leq \ell},$$

with $o_i(1) \rightarrow 0$ when $N \rightarrow \infty$ for all i and $\mathbb{E}(Z_N(\alpha)) = 0$. As a consequence, for large enough N ,

$$\begin{aligned} & \left\| P \cdot \sqrt{\frac{N}{N^\alpha}} Y_N(\alpha) \right\|^2 \\ &= \left\| P \cdot \varepsilon_N(\alpha) \right\|^2 + N^{(\alpha^* - \alpha)/\alpha^*} \left\| P \cdot \left(\frac{C_{D'} K(\psi, D - D')}{f^*(0) K(\psi, D)} i^{-D'} (1 + o_i(1)) \right)_{1 \leq i \leq \ell} \right\|^2 \\ &\geq D \cdot N^{(\alpha^* - \alpha)/\alpha^*}, \end{aligned}$$

with $D > 0$, because the vector $(i^{-D'})_{1 \leq i \leq \ell}$ is not in the orthogonal subspace of the subspace generated by the matrix A . Relation (42) then becomes

$$\begin{aligned} & \Pr\left(\widehat{Q}_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*} \varepsilon\right) > \widehat{Q}_N(\alpha)\right) \\ &\leq \Pr\left(\left\| P \cdot \sqrt{\frac{N}{N^{\alpha^* + ((1 - \alpha^*)/(2\alpha^*))\varepsilon}}} Y_N\left(\alpha^* + \frac{1 - \alpha^*}{2\alpha^*} \varepsilon\right) \right\|^2 \right. \\ &\quad \left. \geq D \cdot N^{\alpha - (\alpha^* + ((1 - \alpha^*)/(2\alpha^*))\varepsilon)} \cdot N^{(\alpha^* - \alpha)/\alpha^*}\right) \\ &\leq \Pr(V_+ \geq D \cdot N^{((1 - \alpha^*)/(2\alpha^*)) (2(\alpha^* - \alpha) - \varepsilon)}) \\ &\leq M_2 \cdot N^{-(\ell/2 - 1)((1 - \alpha^*)/(2\alpha^*))\varepsilon}, \end{aligned}$$

with $M_2 > 0$, because $V_+ \sim \lambda_+ \cdot \chi^2(\ell - 2)$ and $\frac{1 - \alpha^*}{2\alpha^*} (2(\alpha^* - \alpha) - \varepsilon) \geq \frac{1 - \alpha^*}{2\alpha^*} \varepsilon$ for all $\alpha \leq \alpha^* - \varepsilon$. Hence, from the inequality (41), for large enough N ,

$$\Pr(\widehat{\alpha}_N \geq \alpha^* - \varepsilon) \geq 1 - M_2 \cdot \log N \cdot N^{-(\ell/2 - 1)((1 - \alpha^*)/(2\alpha^*))\varepsilon}. \tag{43}$$

Inequalities (40) and (43) imply that

$$\Pr(|\widehat{\alpha}_N - \alpha| \geq \varepsilon) \xrightarrow[N \rightarrow \infty]{} 0.$$

□

Proof of Theorem 1. The central limit theorem of (16) can be established from the following arguments. First, $\Pr(\widehat{\alpha}_N > \alpha^*) \xrightarrow[N \rightarrow \infty]{} 1$. Following the previous proof, for all $\varepsilon > 0$,

$$\Pr(\widehat{\alpha}_N \geq \alpha^* - \varepsilon) \geq 1 - M_2 \cdot \log N \cdot N^{-(\ell/2 - 1)((1 - \alpha^*)/(2\alpha^*))\varepsilon}.$$

Consequently, if $\varepsilon_N = \lambda \cdot \frac{\log \log N}{\log N}$ with $\lambda > \frac{2}{(\ell - 2)D'}$, then

$$\Pr(\widehat{\alpha}_N \geq \alpha^* - \varepsilon_N) \geq 1 - M_2 \cdot \log N \cdot N^{-\lambda((\ell - 2)D'/2) \cdot (\log \log N)/\log N}$$

$$\begin{aligned} &\geq 1 - M_2 \cdot (\log N)^{1-\lambda((\ell-2)D'/2)} \\ &\implies \Pr(\widehat{\alpha}_N + \varepsilon_N \geq \alpha^*) \xrightarrow{N \rightarrow \infty} 1. \end{aligned}$$

Now, from Corollary 4, $\widehat{D}'_N \xrightarrow{\mathcal{P}}_{N \rightarrow \infty} D'$. Therefore, $\Pr(\widehat{D}'_N \leq \frac{4}{3}D') \xrightarrow{N \rightarrow \infty} 1$. Thus, with $\lambda \geq \frac{9}{4(\ell-2)D'}$, $\Pr(\tilde{\alpha}_N + (\varepsilon_N - \frac{3}{(\ell-2)\widehat{D}'_N} \cdot \frac{\log \log N}{\log N}) \geq \alpha^*) \xrightarrow{N \rightarrow \infty} 1$, which implies $\Pr(\tilde{\alpha}_N > \alpha^*) \xrightarrow{N \rightarrow \infty} 1$. Second, for $x \in R$,

$$\begin{aligned} \lim_{N \rightarrow \infty} \Pr\left(\sqrt{\frac{N}{N^{\widehat{\alpha}_N}}}(\tilde{D}_N - D) \leq x\right) &= \lim_{N \rightarrow \infty} \Pr\left(\sqrt{\frac{N}{N^{\widehat{\alpha}_N}}}(\tilde{D}_N - D) \leq x \cap \tilde{\alpha}_N > \alpha^*\right) \\ &\quad + \lim_{N \rightarrow \infty} \Pr\left(\sqrt{\frac{N}{N^{\widehat{\alpha}_N}}}(\tilde{D}_N - D) \leq x \cap \tilde{\alpha}_N \leq \alpha^*\right) \\ &= \lim_{N \rightarrow \infty} \int_{\alpha^*}^1 \Pr\left(\sqrt{\frac{N}{N^\alpha}}(\tilde{D}_N - D) \leq x\right) f_{\widehat{\alpha}_N}(\alpha) d\alpha \\ &= \lim_{N \rightarrow \infty} \Pr(Z_\Gamma \leq x) \cdot \int_{\alpha^*}^1 f_{\widehat{\alpha}_N}(\alpha) d\alpha \\ &= \Pr(Z_\Gamma \leq x) \end{aligned}$$

with $f_{\widehat{\alpha}_N}(\alpha)$ the probability density function of $\widehat{\alpha}_N$ and $Z_\Gamma \sim \mathcal{N}(0; (A' \cdot A)^{-1} \cdot A' \cdot \Gamma \cdot A \cdot (A' \cdot A)^{-1})$.

To prove the second part of (16), we infer from above that

$$\Pr\left(\alpha^* < \tilde{\alpha}_N < \alpha^* + \frac{3}{(\ell-2)\widehat{D}'_N} \cdot \frac{\log \log N}{\log N} + \mu \cdot \frac{\log \log N}{\log N}\right) \xrightarrow{N \rightarrow \infty} 1$$

with $\mu > \frac{12}{\ell-2}$. Therefore, $\nu < \frac{4}{(\ell-2)D'} + \frac{12}{\ell-2}$,

$$\Pr(N^{\alpha^*} < N^{\tilde{\alpha}_N} < N^{\alpha^*} \cdot (\log N)^\nu) \xrightarrow{N \rightarrow \infty} 1.$$

This inequality and the previous central limit theorem result in the fact that for all $\rho > \nu/2$ and $\varepsilon > 0$,

$$\begin{aligned} \Pr\left(\frac{N^{D'/(1+2D')}}{(\log N)^\rho} \cdot |\tilde{D}_N - D| > \varepsilon\right) &= \Pr\left(\frac{N^{1/2(\widehat{\alpha}_N - \alpha^*)}}{(\log N)^\rho} \cdot \sqrt{\frac{N}{N^{\widehat{\alpha}_N}}} |\tilde{D}_N - D| > \varepsilon\right) \\ &\xrightarrow{N \rightarrow \infty} 0. \end{aligned}$$

□

Acknowledgements

The authors are very grateful to anonymous referees for many relevant suggestions and corrections that strongly improved the content of the paper.

References

- Abry, P., Veitch, D. and Flandrin, P. (1998). Long-range dependent: Revisiting aggregation with wavelets. *J. Time Ser. Anal.* **19** 253–266. [MR1626734](#)
- Andrews, D.W.K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica* **72** 569–614. [MR2036733](#)
- Bardet, J.M. (2000). Testing for the presence of self-similarity of Gaussian time series having stationary increments. *J. Time Ser. Anal.* **21** 497–516. [MR1794484](#)
- Bardet, J.M. and Bertrand, P. (2007). Identification of the multiscale fractional Brownian motion with biomechanical applications. *J. Time Ser. Anal.* **28** 1–52. [MR2332850](#)
- Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long range-dependant processes. *Statist. Inference Stochast. Processes* **3** 85–99. [MR1819288](#)
- Bhansali, R., Giraitis, L. and Kokoszka, P.S. (2006). Estimation of the memory parameter by fitting fractionally-differenced autoregressive models. *J. Multivariate Anal.* **97** 2101–2130. [MR2301628](#)
- Devore, P. and Lorentz, G. (1993). *Constructive Approximations*. New York: Springer.
- Doukhan, P., Oppenheim, G. and Taqqu, M.S., eds. (2003). *Theory and Applications of Long-Range Dependence*. Boston: Birkhäuser. [MR1956041](#)
- Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inf. Theory* **38** 910–917. [MR1162229](#)
- Giraitis, L., Robinson, P.M. and Samarov, A. (1997). Rate optimal semi-parametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.* **18** 49–61. [MR1437741](#)
- Giraitis, L., Robinson, P.M. and Samarov, A. (2000). Adaptive semiparametric estimation of the memory parameter. *J. Multivariate Anal.* **72** 183–207. [MR1740636](#)
- 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** 155–187. [MR2345656](#)
- Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractional processes. In *Theory and Applications of Long-Range Dependence* (P. Doukhan, G. Openheim and M.S. Taqqu, eds.) 251–301. Boston: Birkhäuser. [MR1956053](#)
- Robinson, P.M. (1995). Gaussian semi-parametric estimation of long range dependence. *Ann. Statist.* **23** 1630–1661. [MR1370301](#)
- Veitch, D., Abry, P. and Taqqu, M.S. (2003). On the automatic selection of the onset of scaling. *Fractals* **11** 377–390. [MR2028959](#)

Received October 2006 and revised February 2008