Adaptive semiparametric wavelet estimator and goodness-of-fit test for long memory linear processes

Jean-Marc Bardet, Hatem Bibi

Adaptive semiparametric wavelet estimator and goodness-of-fit test for long memory linear processes

Jean-Marc Bardet[a] and Hatem Bibi[a]

bardet@univ-paris1.fr, hatem.bibi@malix.univ-paris1.fr

[a] SAMM, Université Paris 1, 90 rue de Tolbiac, 75013 Paris, FRANCE.

Abstract

This paper is first devoted to study an adaptive wavelet based estimator of the long memory parameter for linear processes in a general semi-parametric frame. This is an extension of Bardet *et al.* (2008) which only concerned Gaussian processes. Moreover, the definition of the long memory parameter estimator is modified and asymptotic results are improved even in the Gaussian case. Finally an adaptive goodness-of-fit test is also built and easy to be employed: it is a chi-square type test. Simulations confirm the interesting properties of consistency and robustness of the adaptive estimator and test.

1 Introduction

The long-memory processes are now a subject area well studied and often applied (see for instance the book edited by Doukhan *et al*, 2003). The most famous long-memory stationary time series are the fractional Gaussian noises (fGn) with Hurst parameter H and FARIMA(p, d, q) processes. For both these time series, the spectral density f in 0 follows a power law: $f(\lambda) \sim C \, \lambda^{-2d}$ where $H = d + 1/2$ in the case of the fGn. This behavior of the spectral density is generally a definition adopted for a stationary long memory (or long range dependent) process even if this definition requires the existence of a second order moment.

There are a lot of statistical results relative to the estimation of the long memory parameter d. First and main results in this direction were obtained for parametric models with the essential papers of Fox and Taqqu (1986) and Dahlhaus (1989) for Gaussian time series, Giraitis and Surgailis (1990) for linear processes and Giraitis and Taqqu (1999) for non linear functions of Gaussian processes.

However and especially for numerical applications, parametric estimators are not really robust and can induce no consistent estimations. Thus, the research is now rather focused on semiparametric estimators of the d. Different approaches were considered: the famous and seminal R/S statistic (see Hurst, 1951), the log-periodogram estimator (see Moulines and Soulier, 2003), the local Whittle estimator (see Robinson, 1995) or the wavelet based estimator (see Veitch *et al*, 2003, Moulines *et al*, 2007 or Bardet *et al*, 2008). All these estimators require the choice of an auxiliary parameter (frequency bandwidth, scales,...) but adaptive versions of original estimators are generally built for avoiding this choice. In a general semiparametric frame, Giraitis *et al* (1997) obtained the asymptotic lower bound for the minimax risk of estimating d, expressed as a function of the second order parameter of the spectral density expansion around 0. Thus, several adaptive semiparametric are proved to follow an oracle property up to multiplicative logarithm term. But simulations (see for instance Bardet *et al*, 2003 or 2008) show that the most accurate estimators are local Whittle, global log-periodogram or wavelet based estimators.

The use of a wavelet based estimator for estimating d was first proposed in Abry *et al.* (1998) after preliminary studies devoted to selfsimilar processes. Then Bardet *et al.* (2000) provided proofs of the consistency of such an estimator in a Gaussian semiparametric frame. Moulines *et al.* (2007) improved these results,

proved a central limit theorem for the estimator of d and showed that this estimator is rate optimal for the minimax criterion. Finally, Roueff and Taqqu (2009a) established similar results in a semiparametric frame for linear processes. All these papers were obtained using a wavelet analysis based on a discrete multi-resolution wavelet transform, which notably allows to compute the wavelet coefficients with the fast Mallat's algorithm. However, there remains a gap in these papers: in a semiparametric frame the "optimal" scale used for the wavelet analysis is depending on the second order expansion of the spectral density around 0 frequency and these papers consider that the power of the second order expansion is known while this is unknown in practice. Two papers proposed a method for automatically selecting this "optimal" scale in the Gaussian semiparametric frame. Firstly, Veitch *et al.* (2003) using a kind of Chi-square test which provides convincing numerical results but the consistency of this procedure is not established. Secondly, Bardet *et al.* (2008) proved the consistency of a procedure for choosing optimal scales based on the detection of the "most linear part" of the log-variogram graph. In this latter article, the "mother" wavelet is not necessary associated to a multi-resolution analysis: the time consuming is clearly more important but a large choice of continuous wavelet transforms can be chosen and the choice of scales is not restricted to be a power of 2.

The present article is devoted to an extension of the article Bardet *et al.* (2008). Three main improvements are obtained:

1. The semiparametric Gaussian framework of Bardet *et al.* (2008) is extended to a semiparametric framework for linear processes. The same automatic procedure of the selection of the optimal scale is also applied and this leads to adaptive estimators.

2. As in Bardet *et al.* (2008), the "mother" wavelet is not restricted to be associated to a discrete multi-resolution transform. Moreover we modified a little the definition of the sample variance of wavelet coefficients (variogram). The result of both these positions is a multidimensional central limit theorem satisfied by the logarithms of variograms with an extremely simple asymptotic covariance matrix (see (9)) only depending on d and the Fourier transform of the wavelet function. Hence it is easy to compute an adaptive pseudo-generalized least square estimator (PGLSE) of d which is proved to satisfy a CLT with with an asymptotic variance smaller than the one of the adaptive (respectively non-adaptive) ordinary least square estimator of d respectively considered in Bardet *et al.* (2008) and Roueff and Taqqu (2009). Simulations confirm confirm the good performance of this PGLSE.

3. Finally, an adaptive goodness-of-fit test can be built from this PGLSE. It consists in a normalized sum of the squared PGLS-distance between between the PGLS-regression line and the points. We prove that this test statistic converges in distribution to a chi-square distribution. Thus it is a very simple test to be computed since the asymptotic covariance matrix is easy to be approximated. When $d > 0$ this test is a long memory test. Moreover, simulations show that this test provides good properties of consistency under H_0 and reasonable properties of robustness under H_1.

For all these reasons, we can say that this paper is an achievement of the article Bardet *et al.* (2008). Moreover, the adaptive PGLS estimator and test represent an interesting extension of the paper Roueff and Taqqu (2009).

We organized the paper as follows. Section 2 contains the assumptions, definitions and a first multidimensional central limit theorem, while Section 3 is devoted to the construction and consistency of the adaptive PGLS estimator and goodness-of-fit test. In Section 4 we illustrate with Monte-Carlo simulations the convergence of the adaptive estimator and we compare these results to those obtained with other efficient semiparametric estimators; moreover we study the consistency and robustness properties of the adaptive goodness-of-fit test. The proofs are provided in Section 5.

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

For $d < 1/2$ and $d' > 0$, this paper deals with the following semi-parametric framework:

Assumption A(d, d')**:** $X = (X_t)_{t \in \mathbb{Z}}$ *is a zero mean stationary linear process, i.e.*

$$X_t = \sum_{s \in \mathbb{Z}} \alpha(t - s)\xi_s, \quad t \in \mathbb{Z}, \quad where$$

- $(\xi_s)_{s \in \mathbb{Z}}$ is a sequence of independent identically distributed random variables such that the distribution of ξ_0 is symmetric, i.e. $\Pr(\xi_0 > M) = \Pr(\xi_0 < -M)$ for any $M \in \mathbb{R}$, and $\mathbb{E}|\xi_0|^4 < \infty$ with $\mathbb{E}\xi_0 = 0$, $Var\,\xi_0 = 1$ and $\mu_4 := \mathbb{E}\xi_0^4$;

- $(\alpha(t))_{t \in \mathbb{Z}}$ is a sequence of real numbers such that there exist $c_d > 0$ and $c_{d'} \in \mathbb{R}$ satisfying

$$|\widehat{\alpha}(\lambda)|^2 = \frac{1}{\lambda^{2d}}\big(c_d + c_{d'}\lambda^{d'}(1 + \varepsilon(\lambda))\big) \quad for\ any \quad \lambda \in [-\pi, \pi], \tag{1}$$

where $\widehat{\alpha}(\lambda) := \frac{1}{2\pi} \sum_{k \in \mathbb{Z}} \alpha(k)e^{-ik\lambda}$, with $\varepsilon(\lambda) \to 0$ $(\lambda \to 0)$.

As a consequence, the spectral density f of X is such that

$$f(\lambda) = 2\pi\,|\widehat{\alpha}(\lambda)|^2 \tag{2}$$

and satisfies the same kind of expansion than (1). Thus, if $d \in (0, 1/2)$ the process X is a long-memory process, and if $d \leq 0$ a short memory process (see Doukhan *et al.*, 2003).

Now define $\psi : \mathbb{R} \to \mathbb{R}$ the wavelet function. Let $k \in \mathbb{N}^*$. We consider the following assumption on ψ:

Assumption $\Psi(k)$: $\psi : \mathbb{R} \to \mathbb{R}$ is such that

1. the support of ψ is included in $(0, 1)$;

2. $\displaystyle\int_0^1 \psi(t)\,dt = 0$;

3. $\psi \in \mathcal{C}^k(\mathbb{R})$.

Straightforward implications of these assumptions are $\psi^{(j)}(0) = \psi^{(j)}(1) = 0$ for any $0 \leq j \leq k$ and $\widehat{\psi}(u) \sim C\,u^k$ $(u \to 0)$ with C a real number not depending on u.
Define also the Fourier transform of ψ, i.e. $\widehat{\psi}(u) := \int_0^1 \psi(t)\,e^{-iut}dt$. Assumption Assumption $\Psi(k)$ also implies that $\widehat{\psi}$ has a fast decay at infinity.

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 $d(a, b) := \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi(\frac{t-b}{a})Y_t\,dt$. However, a process X satisfying Assumption $A(d, d')$ is a discrete-time process, and from a path $(X_1, \ldots, X_N)$ we define the wavelet coefficients of X by

$$e_N(a, b) := \frac{1}{\sqrt{a}} \sum_{t=1}^{N} X_t \psi(\frac{t-b}{a}) = \frac{1}{\sqrt{a}} \sum_{t=1}^{N} \sum_{s \in \mathbb{Z}} \alpha(t-s)\psi(\frac{t-b}{a})\xi_s \tag{3}$$

for $(a, b) \in \mathbb{N}_+^* \times \mathbb{Z}$. Then,

Property 1. *Under Assumption $A(d, d')$ with $d < 1/2$ and $d' > 0$, and if ψ satisfies Assumption $\Psi(k)$ with $k > d'$, then $(e(a, k))_{b \in \{1, \ldots, N-a\}}$ is a zero mean stationary linear process and*

$$\mathbb{E}(e^2(a, 0)) = 2\pi c_d \left(K_{(\psi, 2d)}\,a^{2d} + \frac{c_{d'}}{c_d}K_{(\psi, 2d-d')}\,a^{2d-d'} \right) + o\big(a^{2d-d'}\big) \quad when \quad a \to \infty, \tag{4}$$

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

$$K_{(\psi, \alpha)} := \int_{-\infty}^{\infty} |\widehat{\psi}(u)|^2\,|u|^{-\alpha}du > 0 \quad for\ all\ \alpha < 1. \tag{5}$$

The proof of this property, like all the other proofs, is provided in Section 5.
Let $(X_1, \ldots, X_N)$ be a sampled path of X satisfying Assumption $A(d, d')$. Property allows an estimation of $2d$ from a log-log regression, as soon as a consistent estimator of $\mathbb{E}(e^2(a, 0))$ is provided. For this and with $1 \leq a < N$, consider the sample variance of the wavelet coefficients,

$$T_N(a) := \frac{1}{N-a} \sum_{k=1}^{N-a} e^2(a, k). \tag{6}$$

3

Remark 1. *In Bardet et al. (2000), (2008) or in Moulines et al. (2007) or Roueff and Taqqu (2009), the considered sample variance of wavelet coefficients is*

$$V_N(a) := \frac{1}{[N/a]} \sum_{k=1}^{[N/a]} e^2(a, ak) \tag{7}$$

(with $a = 2^j$ in case of multiresolution analysis). The definition (6) has a drawback and two advantages with respect to this usual definition (7): it is not adapted to the fast Mallat's algorithm and therefore more time consuming, but it leads to more a simple expression of the asymptotic variance and simulations exhibit that this asymptotic variance is smaller that the one obtained with (7).

The following proposition specifies a central limit theorem satisfied by $\log \widetilde{T}_N(a)$, 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 $(\log \widetilde{T}_N(a_i))_i$ can be established,

Proposition 1. *Under Assumption $A(d, d')$, $d < 1/2$ and $d' > 0$, and if ψ satisfies Assumption $\Psi(k)$ with $k \geq 2$. Define $\ell \in \mathbb{N} \setminus \{0, 1\}$ and $(r_1, \cdots, r_\ell) \in (\mathbb{N}^*)^\ell$. Let $(a_n)_{n \in \mathbb{N}}$ be such that $N/a_N \underset{N \to \infty}{\longrightarrow} \infty$ and $a_N N^{-1/(1+2d')} \underset{N \to \infty}{\longrightarrow} \infty$. Then,*

$$\sqrt{\frac{N}{a_N}} \Big(\log T_N(r_i a_N) - 2d \log(r_i a_N) - \log \big(\frac{c_d}{2\pi} K_{(\psi, 2d)} \big) \Big)_{1 \leq i \leq \ell} \underset{N \to \infty}{\overset{d}{\longrightarrow}} \mathcal{N}_\ell \big(0 \, ; \, \Gamma(r_1, \cdots, r_\ell, \psi, d) \big), \tag{8}$$

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

$$\gamma_{ij} = 4\pi \frac{(r_i r'_j)^{1-2d}}{K_{(\psi, 2d)}^2} \int_{-\infty}^{\infty} \frac{\big|\widehat{\psi}(r_i \lambda)\big|^2 \big|\widehat{\psi}(r_j \lambda)\big|^2}{\lambda^{4d}} d\lambda. \tag{9}$$

3 An adaptive estimator of the memory parameter and an adaptive goodness-of-fit test

The CLT of Proposition 1 is very interesting because it has several consequences. We will see that the (quite) simple expression of the asymptotic covariance matrix is an important advantage compared to the complicated expression of the asymptotic covariance obtained in the case of a multiresolution analysis (see Roueff and Taqqu, 2009a). First it allows to obtain an estimator $\widehat{d}_N$ of d by using an ordinary least square estimation. Hence, define

$$\widehat{d}_N(a_N) := \big(0 \ \tfrac{1}{2}\big) (Z'_{a_N} Z_{a_N})^{-1} Z'_{a_N} \big(\log T_N(r_i a_N) \big)_{1 \leq i \leq \ell} \quad \text{with} \quad Z_{a_N} = \begin{pmatrix} 1 & \log(a_N) \\ 1 & \log(2a_N) \\ \vdots & \vdots \\ 1 & \log(\ell a_N) \end{pmatrix}. \tag{10}$$

Remark 2. *From Proposition 1, it is not possible to chose $(r_1, \ldots, r_\ell)$ for minimizing the asymptotic covariance matrix $\Gamma(r_1, \cdots, r_\ell, \psi, d)$ without knowing the value of d. Hence, in the sequel we will only consider the choice $(r_1, r_2, \cdots, r_\ell) = (1, 2, \ldots, \ell)$.*

Then, it is clear from Proposition 1 that $\widehat{d}_N(a_N)$ converges to d following a central limit theorem with convergence rate $\sqrt{\frac{N}{a_N}}$ when a_N satisfies the condition $a_N N^{-1/(1+2d')} \underset{N \to \infty}{\longrightarrow} \infty$.

However, in practice, d' is unknown. In Bardet *et al.* (2008), an automatic procedure for choosing an "optimal" scale a_N has been proposed. We are going to apply again this procedure after recalling its principle: for $\alpha \in (0, 1)$, define

$$Q_N(\alpha, c, d) = \Big(Y_N(\alpha) - Z_{N^\alpha} \big(\begin{smallmatrix} c \\ 2d \end{smallmatrix} \big)\Big)' \cdot \Big(Y_N(\alpha) - Z_{N^\alpha} \big(\begin{smallmatrix} c \\ 2d \end{smallmatrix} \big)\Big), \quad \text{with} \quad Y_N(\alpha) = \big(\log T_N(i N^\alpha) \big)_{1 \leq i \leq \ell}.$$

$Q_N(\alpha, c, d)$ corresponds to a squared distance between the ℓ points $\big(\log(iN^\alpha)\,,\,\log T_N(iN^\alpha)\big)_i$ and a line. It can be minimized first by defining for $\alpha \in (0,1)$

$$\widehat{Q}_N(\alpha) = Q_N(\alpha, \widehat{c}(N^\alpha), 2\widehat{d}(N^\alpha)) \quad \text{with} \quad \left(\begin{array}{c} \widehat{c}(N^\alpha) \\ 2\widehat{d}(N^\alpha) \end{array} \right) = \left(Z'_{N^\alpha} Z_{N^\alpha} \right)^{-1} Z'_{N^\alpha} Y_N(\alpha);$$

and then define $\widehat{\alpha}_N$ by:

$$\widehat{Q}_N(\widehat{\alpha}_N) = \min_{\alpha \in \mathcal{A}_N} \widehat{Q}_N(\alpha) \quad \text{where} \quad \mathcal{A}_N = \left\{ \frac{2}{\log N}\,,\,\frac{3}{\log N}\,,\,\cdots\,,\,\frac{\log[N/\ell]}{\log N} \right\}.$$

Remark 3. *As it was also claimed in Bardet et al. (2008), in the definition of the set $\mathcal{A}_N$, $\log N$ can be replaced by any sequence negligible with respect to any power law of N. Hence, in numerical applications we will use $10 \log N$ which significantly increases the precision of $\widehat{\alpha}_N$.*

Under the assumptions of Proposition 1, one obtains (see the proof in Bardet *et al.*, 2008),

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

Then define:

$$\widehat{\widetilde{d}}_N := \widehat{d}(N^{\widehat{\alpha}_N}) \quad \text{and} \quad \widehat{\Gamma}_N := \Gamma(1, \cdots, \ell, \widehat{\widetilde{d}}_N, \psi). \tag{11}$$

It is clear that $\widehat{\widetilde{d}}_N \xrightarrow[N \to \infty]{\mathcal{P}} d$ (a convergence rate can also be found in Bardet *et al.*, 2008) and therefore, from the expression of Γ in (9) and its smoothness with respect to the variable d, $\widehat{\Gamma}_N \xrightarrow[N \to \infty]{\mathcal{P}} \Gamma(1, \cdots, \ell, d, \psi)$. Thus it is possible to define a (pseudo)-generalized least square estimator (PGLSE) of d. Before this, define

$$\widetilde{\alpha}_N := \widehat{\alpha}_N + \frac{6 \widehat{\alpha}_N}{(\ell - 2)(1 - \widehat{\alpha}_N)} \frac{\log \log N}{\log N}.$$

For technical reasons (*i.e.* $\Pr(\widetilde{\alpha}_N \leq \alpha^*) \xrightarrow[N \to \infty]{} 0$), which is not satisfied by $\widehat{\alpha}_N$, see Bardet *et al.*, 2008), in the sequel we prefer to consider $\widetilde{\alpha}_N$ rather than $\widehat{\alpha}_N$. Finally, using the usual expression of PGLSE, the adaptive estimators of c and d can be defined as follows:

$$\left(\begin{array}{c} \widetilde{c}_N \\ 2\widetilde{d}_N \end{array} \right) := \left(Z'_{N^{\widetilde{\alpha}_N}} \widehat{\Gamma}_N^{-1} Z_{N^{\widetilde{\alpha}_N}} \right)^{-1} Z'_{N^{\widetilde{\alpha}_N}} \widehat{\Gamma}_N^{-1} Y_N(\widetilde{\alpha}_N). \tag{12}$$

The following theorem provides the asymptotic behavior of the estimator $\widetilde{d}_N$,

Theorem 1. *Under assumptions of Proposition 1, with* $\sigma_d^2(\ell) := \big(0 \ \frac{1}{2}\big)\big(Z'_1\big(\Gamma(1, \cdots, \ell, d, \psi)\big)^{-1} Z_1\big)^{-1}\big(0 \ \frac{1}{2}\big)'$,

$$\sqrt{\frac{N}{N^{\widetilde{\alpha}_N}}}\big(\widetilde{d}_N - d\big) \xrightarrow[N \to \infty]{d} \mathcal{N}(0\,;\,\sigma_d^2(\ell)) \quad \text{and} \quad \forall \rho > \frac{2(1 + 3d')}{(\ell - 2)d'}, \quad \frac{N^{\frac{d'}{1+2d'}}}{(\log N)^\rho} \cdot \big|\widetilde{d}_N - d\big| \xrightarrow[N \to \infty]{\mathcal{P}} 0. \tag{13}$$

Remark 4.   *1. From Gauss-Markov Theorem it is clear that the asymptotic variance of $\widetilde{d}_N$ is smaller or equal to the one of $\widehat{\widetilde{d}}_N$. Moreover $\widetilde{d}_N$ satisfies the CLT (13) which provides confidence intervals that are simple to calculate.*

  2. In the Gaussian case, the adaptive estimator $\widetilde{d}_N$ converge to d with a rate of convergence rate equal to the minimax rate of convergence $N^{\frac{d'}{1+2d'}}$ up to a logarithm factor (see Giraitis et al., 1997). Thus, this estimator can be compared to adaptive log-periodogram or local Whittle estimators (see respectively Moulines and Soulier, 2003, and Robinson, 1995).

  3. Under additive assumptions on ψ (ψ is supposed to have its first m vanishing moments), the estimator $\widetilde{d}_N$ can also be applied to a process X with an additive polynomial trend of degree $\leq m - 1$. Then the tend is "vanished" by the wavelet function and the value of $\widetilde{d}_N$ is the same than without this additive trend. Such robustness property is not possible with an adaptive log-periodogram or local Whittle estimator.

Finally it is easy to deduce from the previous pseudo-generalized least square regression an adaptive goodness-of-fit test. It consists on a sum of the PGLS squared distances between the PGLS regression line and the points. More precisely consider the statistic:

$$\widetilde{T}_N = \frac{N}{N^{\widetilde{\alpha}_N}} \left(Y_N(\widetilde{\alpha}_N) - Z_{N^{\widetilde{\alpha}_N}} \left(\begin{smallmatrix} \widetilde{c}_N \\ 2\widetilde{d}_N \end{smallmatrix} \right) \right)' \widehat{\Gamma}_N^{-1} \left(Y_N(\widetilde{\alpha}_N) - Z_{N^{\widetilde{\alpha}_N}} \left(\begin{smallmatrix} \widetilde{c}_N \\ 2\widetilde{d}_N \end{smallmatrix} \right) \right). \tag{14}$$

Then, using the previous results, one obtains:

Theorem 2. *Under assumptions of Proposition 1,*

$$\widetilde{T}_N \xrightarrow[N \to \infty]{d} \chi^2(\ell - 2). \tag{15}$$

This (adaptive) goodness-of-fit test is therefore very simple to be computed and used. In the case where $d > 0$, which can be tested easily from Theorem 1, this test can also be seen as a test of long memory for linear processes. ion

4 Simulations

In the sequel, the numerical consistency and robustness of $\widetilde{d}_N$ are first investigated. Simulation are realized and the results obtained with the estimator $\widetilde{d}_N$ are compared to those obtained with the best known semiparametric long-memory estimators. Finally numerical properties of the test statistic $\widetilde{T}_N$ are also studied.

Remark 5. *Note that all the softwares (in Matlab language) used in this section are available with a free access on* `http://samm.univ-paris1.fr/-Jean-Marc-Bardet`*.*

To begin with, the simulation conditions have to be specified. The results are obtained from 100 generated independent samples of each process belonging to the following "benchmark". The concrete procedures of generation of these processes are obtained from the circulant matrix method in case of Gaussian processes or a truncation of an infinite sum in case of non-Gaussian process (see Doukhan *et al.*, 2003). The simulations are realized for $d = 0, 0.1, 0.2, 0.3$ and 0.4, for $N = 10^3$ and 10^4 and the following processes which satisfy Assumption $A(d, d')$:

1. the fractional Gaussian noise (fGn) of parameter $H = d + 1/2$ (for $0 \leq d < 0.5$) and $\sigma^2 = 1$. A fGn is such that Assumption $A(d, 2)$ holds even if a fGn is generally not studied as a Gaussian linear process;

2. a FARIMA$[p, d, q]$ process with parameter d such that $d \in [0, 0.5)$, $p, q \in \mathbb{N}$. A FARIMA$[p, d, q]$ process is such that Assumption $A(d, 2)$ holds when $\mathbb{E}\xi_0^4 < \infty$ where ξ_0 is the innovation process.

3. the Gaussian stationary process $X^{(d, d')}$, such that its spectral density is

$$f_3(\lambda) = \frac{1}{\lambda^{2d}}(1 + \lambda^{d'}) \quad \text{for } \lambda \in [-\pi, 0) \cup (0, \pi], \tag{16}$$

with $d \in [0, 0.5)$ and $d' \in (0, \infty)$. Therefore the spectral density f_3 is such that Assumption $A(d, d')$ holds and since $X^{(d, d')}$ is a Gaussian process, from the Wold decomposition it is also a linear process.

A "benchmark" which will be considered in the sequel consists of the following particular cases of these processes for $d = 0, 0.1, 0.2, 0.3, 0.4$:

- X_1 : fGn processes with parameters $H = d + 1/2$;

- X_2 : FARIMA$[0, d, 0]$ processes with standard Gaussian innovations;

- X_3 : FARIMA$[0, d, 0]$ processes with innovations following a uniform $\mathcal{U}[-1, 1]$ distribution;

- X_4 : FARIMA$[0, d, 0]$ processes with innovations following a symmetric Burr distribution of parameter $(2, 1)$ (*i.e.* its cumulative distribution function is $F(x) = (1 - \frac{1}{2}(1 + x^2)^{-1}) \mathbb{I}_{x \geq 0} + \frac{1}{2}(1 + x^2) \mathbb{I}_{x < 0}$);

- X_5 : FARIMA$[0, d, 0]$ processes with innovations following a Cauchy distribution;

- X_6 : FARIMA$[1, d, 1]$ processes with standard Gaussian innovations, MA coefficient $\phi = -0.3$ and AR coefficient $\phi = 0.7$;

- X_7 : FARIMA$[1, d, 1]$ processes with innovations following a uniform $\mathcal{U}[-1, 1]$ distribution, MA coefficient $\phi = -0.3$ and AR coefficient $\phi = 0.7$;

- X_8 : $X^{(d,d')}$ Gaussian processes with $d' = 1$.

Note that the processes X_4 and X_5 do not satisfy the condition $\mathbb{E}\xi_0^4$ required in Theorems 1 and 2. However, since we consider the logarithm of wavelet coefficient sample variance and not only the wavelet coefficient sample variance, it should be possible to prove the consistency of $\widetilde{d}_N$ under a condition such as $\mathbb{E}\xi_0^r$ with $r \geq 2$ and perhaps only $r > 0$...

4.1 Comparison of the wavelet based estimator and other estimators

First let us specify the different choices concerning the wavelet based estimator:

Choice of the function ψ**:** as it was said previously, it is not mandatory to use a wavelet function associated with a multi-resolution analysis. We use here the function $\psi(x) = x^3(1-x)^3\big(x^3 - \frac{3}{2}x^2 + \frac{15}{22}x - \frac{1}{11}\big)\mathbb{I}_{x\in[0,1]}$ which satisfies Assumption $\Psi(2)$

Choice of the parameter ℓ**:** This parameter is important to estimate the "beginning" of the linear part of the graph drawn by points $(\log(ia_N), \log T_N(ia_N))_{1\leq i\leq \ell}$ and therefore the data-driven $\widehat{a}_N$. Moreover this parameter is used for the computation of $\widetilde{d}_N$ as the number of regression points. We chose a two step procedure:

1. following a numerical study (not detailed here), $\ell = [2 * log(N)]$ (therefore $\ell = 13$ for $N = 1000$ and $\ell = 18$ for $N = 10000$) seems to be a good choice for the first step: compute $\widehat{\alpha}_n$.

2. for the computation of $\widetilde{d}_N$, we first remark that with the chosen function ψ, $\widehat{\Gamma}_N$ does not seem to depend on d. As a consequence we decide to compute $\sigma_d^2(\ell) = \big(0\ \frac{1}{2}\big)\big(Z_1'\big(\Gamma(1,\cdots,\ell,d,\psi)\big)^{-1}Z_1\big)^{-1}\big(0\ \frac{1}{2}\big)'$ for several values of d and ℓ using classical approximations of the integrals defined in $\Gamma(1,\cdots,\ell,d,\psi)$. The results of these numerical experiments are reported in Figure 2. The conclusion of this numerical experiment is the following: for any $d \in [0, 0.5[$, $\sigma_d^2(\ell)$ is almost not depending on d and decreases when ℓ increases. Therefore we chose for this second step $\ell = N^{1-\widetilde{\alpha}_N}(\log N)^{-1}$: by this way the larger considered scale is $N(\log N)^{-1}$ (which is negligible with respect to N and therefore the CLT 8 holds).
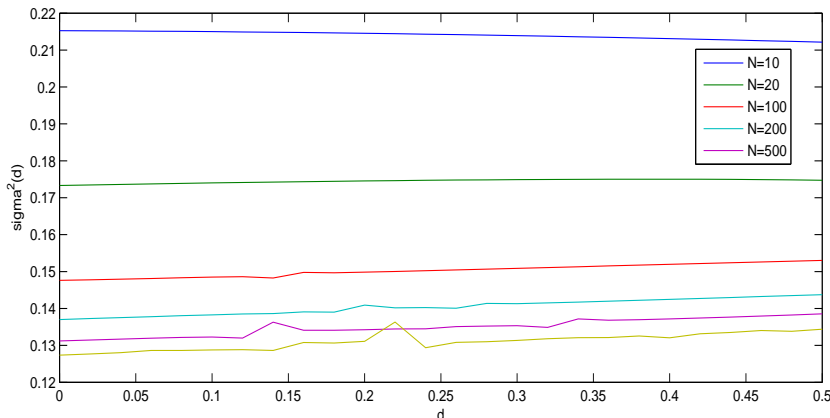
Figure 1: *Graph of the approximated values of $\sigma_d^2(\ell)$ defined in 13 for $d \in [0, 0.5]$ and $N = 10, 20, 50, 100, 200$ and 500.*

Now we consider the previous "benchmark" of processes and apply the estimator $\widetilde{d}_N$ and 2 other semiparametric d-estimators known for their accuracies:

- $\widehat{d}_{MS}$ is the adaptive global log-periodogram estimator introduced by Moulines and Soulier (1998, 2003), also called FEXP estimator, with bias-variance balance parameter $\kappa = 2$;

- $\widehat{d}_R$ is the local Whittle estimator introduced by Robinson (1995). The trimming parameter is $m = N/30$.

Simulation results are reported in Table 1.

Conclusions from Table 1: The wavelet based estimator $\widetilde{d}_N$ numerically shows a convincing convergence rate with respect to the other estimators. Both the "spectral" estimator $\widehat{d}_R$ and $\widehat{d}_{MS}$ provide more stable results almost not sensible to d and the flatness of the spectral density of the process, while the convergence rate of the wavelet based estimator $\widetilde{d}_N$ is more dependent on the spectral density of the process. But, especially in cases of "smooth" spectral densities (fGn and FARIMA$(0, d, 0)$), $\widetilde{d}_N$ is a very accurate semiparametric estimator and is globally more efficient than the other estimators.

Remark 6. *In Bardet et al. (2008) we also compared two adaptive wavelet based estimators (the one defined in Veitch et al., 2003 and the one defined in Bardet et al., 2008) with $\widehat{d}_{MS}$ and $\widehat{d}_R$ (and also with two others defined in Giraitis et al., 2000, and Giraitis et al., 2006, which exhibit worse numerical properties of consistency). We observe that $\sqrt{MSE}$ of $\widetilde{d}_N$ obtained in Table 1 is generally smaller than the one obtained with the estimator defined in Bardet et al. (2008) for two reasons: the choice of the definition (6) of wavelet coefficient sample variance instead of (7) and the choice of a PGLS regression instead of a LS regression.*

Comparison of the robustness of the different semiparametric estimators: To conclude with the numerical properties of the estimators, 3 different processes not satisfying Assumption $A(d, d')$ are considered:

- a Gaussian stationary process with a spectral density $f(\lambda) = \left| |\lambda| - \pi/2 \right|^{-2d}$ for all $\lambda \in [-\pi, \pi] \setminus \{-\pi/2, \pi/2\}$. The local behavior of f in 0 is $f(|\lambda|) \sim (\pi/2)^{-2d} |\lambda|^{-2d}$ with $d = 0$, but the smoothness condition for f in Assumption $A(0, 2)$ is not satisfied.

- a trended Gaussian FARIMA$(0, d, 0)$ with an additive linear trend ($X_t = FARIMA_t + (1 - 2t/n)$ for $t = 1, \cdots, n$ and therefore $mean(X_1, \cdots, X_n) \simeq 0$);

- a Gaussian FARIMA$(0, d, 0)$ with an additive linear trend and an additive sinusoidal seasonal component of period $T = 12$ ($X_t = FARIMA_t + (1 - 2t/n) + \sin(\pi t/6)$ for $t = 1, \cdots, n$ and therefore $mean(X_1, \cdots, X_n) \simeq 0$).

The results of these simulations are given in Table 2.

Conclusions from Table 2: The main advantages of $\widetilde{d}_N$ with respect to $\widehat{d}_{MS}$ and $\widehat{d}_R$ are exhibited in this table: it is robust with respect to smooth trends (or seasonality). Note that the sample mean of $\widehat{d}_{MS}$ and $\widehat{d}_R$ in the case of processes with trend or with trend and seasonality is almost 0.5

4.2   Consistency and robustness of the adaptive goodness-of-fit test:

Tables 1 and 2 provide informations concerning the adaptive goodness-of-fit test. A general conclusion is that the consistency properties of this test are clearly satisfying when N is large enough ($N = 1000$ seems to be too small for using this goodness-of-fit test).

We also would like to know the behavior of the test statistic under the assumption H_1. We are going to study the case of a process which does not satisfy either the stationarity condition either the relation (1) also verified by the spectral density. Hence 3 particular cases are considered:

1. a process X denoted MFARIMA and defined as a succession of two independent Gaussian FARIMA processes. More precisely, we consider $X_t = FARIMA(0, 0.1, 0)$ for $t = 1, \cdots, n/2$ and $X_t = FARIMA(0, 0.4, 0)$ for $t = n/2 + 1, \cdots, n$.

2. a process X denoted MGN and defined by the increments of a multifractional Brownian motion (introduced in Peltier and Lévy-Vehel, 1995). Using the harmonizable representation, define $Y = (Y_t)_t$ such that

$$Y_t = C(t) \int_{\mathbb{R}} \frac{e^{itx} - 1}{|x|^{H(t) + 1/2}} W(dx)$$

where $W(dx)$ is a complex-valued Gaussian noise with variance dx and $H(\cdot)$ is a function (the case $H(\cdot) = H$ with $H \in (0, 1)$ is the case of fBm), $C(\cdot)$ i a function. Here we consider the functions $H(t) = 0.5 + 0.4 \sin(t/10)$ and $C(t) = 1$. Then $X_t = Y_{t+1} - Y_t$ for $t \in \mathbb{Z}$. X is not a stationary process but "locally" behaves as a fGn with a parameter $H(t)$ (therefore depending on t).

8

$N = 10^3 \longrightarrow$

Model	$\sqrt{\overline{MSE}}$	$d=0$	$d=0.1$	$d=0.2$	$d=0.3$	$d=0.4$
X_1	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.089	0.091	0.096	0.090	0.100
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.102	0.114	0.116	0.106	0.102
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.047**	**0.046**	**0.042**	**0.052**	**0.047**
	$\widetilde{p}_n$	0.85	0.76	0.78	0.76	0.64
X_2	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.091	0.094	0.086	0.091	0.099
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.107	0.105	0.112	0.110	0.097
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.048**	**0.050**	**0.053**	**0.061**	**0.074**
	$\widetilde{p}_n$	0.82	0.82	0.75	0.73	0.67
X_3	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.092	0.094	0.080	0.099	0.096
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.113	0.113	0.100	0.112	0.095
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.052**	**0.071**	**0.063**	**0.077**	**0.092**
	$\widetilde{p}_n$	0.84	0.72	0.75	0.67	0.51
X_4	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.088	0.079	0.079	0.093	0.104
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.096	0.100	0.103	0.097	0.095
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.051**	**0.066**	**0.056**	**0.061**	**0.064**
	$\widetilde{p}_n$	0.84	0.78	0.78	0.75	0.66
X_5	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	**0.069**	**0.067**	**0.077**	0.121	0.143
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.072	0.078	0.093	**0.087**	**0.074**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.073	0.069	0.083	**0.087**	0.120
	$\widetilde{p}_n$	0.73	0.69	0.68	0.74	0.64
X_6	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	**0.096**	**0.091**	**0.090**	**0.086**	**0.093**
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.111	0.102	0.100	0.101	0.101
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.153	0.146	0.144	0.158	0.147
	$\widetilde{p}_n$	0.52	0.47	0.48	0.39	0.50
X_7	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	**0.085**	**0.096**	**0.086**	**0.093**	0.098
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.106	0.116	0.097	0.099	**0.092**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.155	0.150	0.56	0.147	0.157
	$\widetilde{p}_n$	0.60	0.55	0.49	0.52	0.41
X_8	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	**0.097**	**0.104**	**0.097**	**0.094**	**0.101**
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.120	0.116	0.117	0.113	0.110
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.179	0.189	0.177	0.175	0.176
	$\widetilde{p}_n$	0.75	0.75	0.68	0.66	0.67

$N = 10^4 \longrightarrow$

Model	$\sqrt{\overline{MSE}}$	$d=0$	$d=0.1$	$d=0.2$	$d=0.3$	$d=0.4$
X_1	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.032	0.029	0.031	0.031	0.036
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.028	0.028	**0.029**	0.029	0.032
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.016**	**0.027**	0.034	**0.025**	**0.022**
	$\widetilde{p}_n$	0.97	0.93	0.97	0.94	0.97
X_2	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.034	0.030	0.029	0.032	0.028
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.027	0.027	0.029	0.028	**0.023**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.026**	**0.019**	**0.019**	**0.019**	0.025
	$\widetilde{p}_n$	0.95	0.97	0.98	0.96	0.94
X_3	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.034	0.034	0.033	0.030	0.031
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.029	0.028	0.028	0.028	**0.029**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.027**	**0.017**	**0.016**	**0.022**	0.030
	$\widetilde{p}_n$	0.93	0.96	0.97	0.93	0.92
X_4	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.029	0.060	0.036	0.031	0.031
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.025	0.027	0.029	0.031	0.029
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	**0.016**	**0.020**	**0.021**	**0.015**	**0.023**
	$\widetilde{p}_n$	0.95	0.91	0.97	0.92	0.91
X_5	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.093	**0.046**	0.039	0.073	0.047
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	**0.040**	**0.046**	0.035	0.032	**0.024**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.056	0.071	**0.027**	**0.025**	**0.024**
	$\widetilde{p}_n$	0.85	0.88	0.93	0.86	0.85
X_6	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.031	0.032	0.033	0.032	0.029
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	**0.029**	**0.028**	**0.028**	**0.028**	**0.028**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.045	0.044	0.046	0.044	0.041
	$\widetilde{p}_n$	0.96	0.93	0.89	0.93	0.90
X_7	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	0.030	0.031	0.037	0.030	0.029
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	**0.027**	**0.027**	**0.032**	**0.028**	**0.027**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.049	0.044	0.050	0.048	0.046
	$\widetilde{p}_n$	0.94	0.91	0.88	0.87	0.86
X_8	$\sqrt{\overline{MSE}}\ \widehat{d}_{MS}$	**0.038**	0.040	**0.040**	**0.035**	0.037
	$\sqrt{\overline{MSE}}\ \widehat{d}_{R}$	0.039	**0.038**	**0.040**	0.036	**0.035**
	$\sqrt{\overline{MSE}}\ \widetilde{d}_{N}$	0.085	0.083	0.086	0.087	0.085
	$\widetilde{p}_n$	0.92	0.94	0.94	0.95	0.93

Table 1: *Comparison of the different long-memory parameter estimators for processes of the benchmark. For each process and value of d and N, $\sqrt{\overline{MSE}}$ are computed from 100 independent generated samples. Here $\widetilde{p}_n = \frac{1}{n}\#\big(\widetilde{T}_N < q_{\chi^2(\ell-2)}(0.95)\big)$: this is the frequency of acceptation of the adaptive goodness-of-fit test.*

Model	$\sqrt{MSE}$	$d=0$	$d=0.1$	$d=0.2$	$d=0.3$	$d=0.4$
GARMA$(0,d,0)$	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.089	0.091	0.123	0.132	0.166
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.112	0.111	0.119	**0.106**	**0.106**
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.041**	**0.076**	**0.114**	0.142	0.180
	$\widetilde{p}_{n}$	0.82	0.78	0.63	0.59	0.46
Trend	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.548	0.411	0.292	0.190	0.142
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.499	0.394	0.279	0.167	0.091
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.044**	**0.052**	**0.056**	**0.060**	**0.065**
	$\widetilde{p}_{n}$	0.83	0.81	0.80	0.73	0.64
Trend + Seasonality	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.479	0.347	0.233	0.142	0.112
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.499	0.393	0.279	0.167	**0.091**
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.216**	**0.215**	**0.215**	**0.217**	0.185
	$\widetilde{p}_{n}$	0.35	0.26	0.18	0.21	0.18

$N = 10^3 \longrightarrow$

Model	$\sqrt{MSE}$	$d=0$	$d=0.1$	$d=0.2$	$d=0.3$	$d=0.4$
GARMA$(0,d,0)$	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.031	0.035	0.039	0.049	0.062
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.028	**0.031**	**0.030**	**0.030**	**0.034**
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.023**	0.053	0.052	0.058	0.060
	$\widetilde{p}_{n}$	0.96	0.94	0.93	0.91	0.88
Trend	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.452	0.286	0.167	0.096	0.056
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.433	0.308	0.191	0.100	0.051
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.014**	**0.016**	**0.016**	**0.021**	**0.028**
	$\widetilde{p}_{n}$	0.99	0.97	0.97	0.95	0.93
Trend + Seasonality	$\sqrt{MSE}\ \widehat{d}_{MS}$	0.471	0.307	0.196	0.123	0.076
	$\sqrt{MSE}\ \widehat{d}_{R}$	0.432	0.305	0.191	0.100	0.052
	$\sqrt{MSE}\ \widetilde{d}_{N}$	**0.044**	**0.069**	**0.047**	**0.042**	**0.045**
	$\widetilde{p}_{n}$	0.83	0.81	0.76	0.78	0.82

$N = 10^4 \longrightarrow$

Table 2: *Robustness of the different long-memory parameter estimators. For each process and value of d and N, $\sqrt{MSE}$ are computed from 100 independent generated samples. Here $\widetilde{p}_{n} = \frac{1}{n}\#\big(\widetilde{T}_{N} < q_{\chi^2(\ell-2)}(0.95)\big)$: this is the frequency of acceptation of the adaptive goodness-of-fit test.*

Model	$N = 10^3$	$N = 10^4$
MFARIMA	$\widetilde{p}_{n} = 0.58$	$\widetilde{p}_{n} = 0.87$
MGN	$\widetilde{p}_{n} = 0.18$	$\widetilde{p}_{n} = 0.08$
MFGN	$\widetilde{p}_{n} = 0.02$	$\widetilde{p}_{n} = 0.04$

Table 3: *Robustness of the adaptive goodness-of-fit test with $\widetilde{p}_{n} = \frac{1}{n}\#\big(\widetilde{T}_{N} < q_{\chi^2(\ell-2)}(0.95)\big)$ the frequency of acceptation of the adaptive goodness-of-fit test.*

3. a process X denoted MFGN and defined by the increments of a multiscale fractional Brownian motion (introduced in Bardet and Bertrand, 2007). Let $Z = (Z_t)_t$ be such that

$$Z_t = \int_{\mathbb{R}} \sigma(x)\, \frac{e^{itx} - 1}{|x|^{H(x)+1/2}} W(dx)$$

where $W(dx)$ is a complex-valued Gaussian noise with variance dx, $H(\cdot)$ and $\sigma(\cdot)$ are piecewise constant functions. Here we consider the functions $H(x) = 0.9$ for $0.001 \leq x \leq 0.04$ and $H = 0.1$ for $0.04 \leq x \leq 3$. Then $X_t = Z_{t+1} - Z_t$ for $t \in \mathbb{Z}$ and X is a Gaussian stationary process which can be written as a linear process behaving as a fGn of parameter 0.9 for low frequencies (large time) and as a fGn of parameter 0.1 for high frequencies (small time).

We applied the test statistic to 100 independent replications of both these processes. The results of this simulation are proposed in Table 3. We observed that the processes MGN and MFGN are clearly rejected with the adaptive goodness-of-fit test. However, the test is not able to reject the process MFARIMA which does not satisfy the Assumption of the Theorem 2. The reason is that the test does an average of the behavior of the sample and in the case of changes (it is such the case for MFARIMA) it is the average LRD parameter which is estimated (an average of 0.30 for $\widetilde{d_N}$ and a standard deviation 0.03 are obtained).

5 Proofs

First, we will use many times the following lemma:

Lemma 1. *If ψ satisfies Assumption $\Psi(k)$ with $k \geq 1$, then there exists $C_\psi \geq 0$ such that for all $\lambda \in \mathbb{R}$,*

$$\left| \frac{1}{a} \sum_{k=1}^{a} \psi\left(\frac{k}{a}\right) e^{-i\lambda \frac{k}{a}} - \int_0^1 \psi(t) e^{-i\lambda t} dt \right| \leq C_\psi \frac{(1 + |\lambda|^k)}{a^k}. \tag{17}$$

Proof of Lemma 1. This proof is easily established from a mathematical induction on k when $\lambda \in [-\pi, \pi]$. Then since we consider 2π-periodic functions (of λ) the result can be extended to $\mathbb{R}$. $\square$

Proof of Property 1. First, it is clear that for $a \in \mathbb{N}^*$, $(e(a,b))_{1 \leq b \leq N-a}$ is a centered linear process. It is a stationary process because X is a stationary process and clearly $\sum_{k=1}^{N} \frac{1}{a} \psi^2\left(\frac{k-b}{a}\right) < \infty$.
Now following similar computations to those performed in Bardet *et al.* (2008), we obtain for $a \in \mathbb{N}^*$,

$$\mathbb{E}(e^2(a,0)) \quad = \quad \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times \left| \frac{1}{a} \sum_{k=1}^{a} \psi\left(\frac{k}{a}\right) e^{-i\frac{k}{a}u} \right|^2 du.$$

Now, since $\sup_{u \in \mathbb{R}} |\widehat{\psi}(u)| < \infty$, for a large enough,

$$\left| \mathbb{E}(e^2(a,0)) - \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \quad \leq \quad 2 \sup_{u \in \mathbb{R}} |\widehat{\psi}(u)| C_\psi \frac{1}{a^k} \int_{-a\pi}^{a\pi} (1 + |u|^k) f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du$$

Under Assumption $\Psi(k)$ for any $k \geq 1$, $\sup_{u \in \mathbb{R}} (1 + u^n) |\widehat{\psi}(u)| < \infty$ for all $n \in \mathbb{N}$. Therefore, since there exists $c_a > 0$ satisfying $f(\lambda) \leq c_a \lambda^{-2d}$ for all $\lambda \in [-\pi, \pi]$, for all $d < 1/2$,

$$\int_{-a\pi}^{a\pi} (1 + |u|^k) f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du \leq \left(c_a \int_{-\infty}^{\infty} (1 + |u|^k) u^{-2d} |\widehat{\psi}(u)|^2 du \right) a^{2d},$$

and thus there exists $C > 0$ (not depending on a) such that for a large enough,

$$\left| \mathbb{E}(e^2(a,0)) - \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) \times |\widehat{\psi}(u)|^2 du \right| \quad \leq \quad C a^{2d-k}. \tag{18}$$

Following the same reasoning, for any $n \geq 0$, there exists $C(n) > 0$ (not depending on a) such that for a large enough,

$$\left| \int_{-a\pi}^{a\pi} f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du - \int_{-\infty}^{\infty} f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du \right| \leq C(n) a^{-n}. \tag{19}$$

Finally, from Assumption $A(d,d')$, we obtain the following expansion:

$$\int_{-\infty}^{\infty} f\left(\frac{u}{a}\right) |\widehat{\psi}(u)|^2 du \quad = \quad 2\pi \int_{-\infty}^{\infty} \left(c_d \left(\frac{u}{a}\right)^{-2d} + c_{d'} \left(\frac{u}{a}\right)^{d'-2d} + \left(\frac{u}{a}\right)^{d'-2d} \varepsilon\left(\frac{u}{a}\right) \right) |\widehat{\psi}(u)|^2 du$$

$$= \quad 2\pi c_d K_{(\psi,2d)} a^{-2d} + 2\pi c_{d'} K_{(\psi,2d-d')} a^{2d-d'} + o(a^{2d-d'}) \tag{20}$$

using the definition (5) of $K_{(\psi,\alpha)}$ and because $\lim_{\lambda \to 0} \varepsilon(\lambda) = 0$ and applying Lebesgue Theorem. Then, using (18), (19) and (20), we obtain that

$$\left| \mathbb{E}(e^2(a,0)) - 2\pi c_d K_{(\psi,2d)} a^{-2d} + 2\pi c_{d'} K_{(\psi,2d-d')} a^{2d-d'} \right| \quad \leq \quad o(a^{2d-d'}) + C a^{2d-k}. \tag{21}$$

When $k > d'$, it implies (4). $\square$

Proof of Theorem 1. We decompose this proof in 4 steps. First define the normalized wavelet coefficients of X by:

$$\widetilde{e}_N(a,b) := \frac{e(a,b)}{\sqrt{\mathbb{E}e^2(a,1)}} \quad \text{for } a \in \mathbb{N}^* \text{ and } b \in \mathbb{Z}, \tag{22}$$

11

and the normalized sample variance of wavelet coefficients:

$$\widetilde{T}_N(a) := \frac{1}{N-a} \sum_{k=1}^{N-a} \widetilde{e}^2(a,k). \tag{23}$$

Step 1 We prove in this part that $N\mathrm{Cov}\,(\widetilde{T}_N(r\,a_N), \widetilde{T}_N(r'\,a_N))$ converges to the asymptotic covariance matrix $\Gamma(\ell_1, \cdots, r_\ell, \psi, d)$ defined in (9). First for $\lambda \in \mathbb{R}$, denote

$$S_a(\lambda) := \frac{1}{a} \sum_{t=1}^{a} \psi(\frac{t}{a}) e^{i\lambda t/a}.$$

Then for $a \in \mathbb{N}^*$ and $b = 1, \cdots, N-a$, since ψ is $(0,1]$-supported function,

$$
\begin{aligned}
\sum_{t=1}^{N} \alpha(t-s)\psi(\frac{t-b}{a}) &= \sum_{t=0}^{a} \psi(\frac{t}{a}) \int_{-\pi}^{\pi} \widehat{\alpha}(\lambda) e^{i\lambda(t-s+b)} d\lambda \\
&= \int_{-\pi}^{\pi} a S_a(a\lambda)\widehat{\alpha}(\lambda) e^{i(b-s)\lambda} d\lambda \\
&= \int_{-a\pi}^{a\pi} S_a(\lambda)\widehat{\alpha}(\frac{\lambda}{a}) e^{i(b-s)\frac{\lambda}{a}} d\lambda.
\end{aligned} \tag{24}
$$

Thus,

$$
\begin{aligned}
\mathrm{Cov}\,(\widetilde{T}_N(a), \widetilde{T}_N(a')) &= \frac{1}{N-a}\frac{1}{N-a'} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} \mathrm{Cov}\,(\widetilde{e}^2(a,b), \widetilde{e}^2(a',b')) \\
&\simeq \frac{(a\,a')^{-2d}(c_d\,K_{(\psi,2d)})^{-2}}{4\pi^2(N-a)(N-a')} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} \mathrm{Cov}\,(e^2(a,b), e^2(a',b')).
\end{aligned} \tag{25}
$$

But,

$$
\begin{aligned}
\mathrm{Cov}\,(e^2_{(a,b)}, e^2_{(a',b')}) &= \frac{1}{a\,a'} \sum_{t_1,t_2,t_3,t_4=1}^{N} \sum_{s_1,s_2,s_3,s_4 \in \mathbb{Z}} \Big(\prod_{i=1}^{2} \alpha(t_i-s_i)\psi(\frac{t_i-b}{a})\Big)\Big(\prod_{i=1}^{2} \alpha(t_i-s_i)\psi(\frac{t_i-b'}{a'})\Big)\mathrm{Cov}\,(\xi_{s_1}\xi_{s_2}, \xi_{s_3}\xi_{s_4}) \\
&= C_1 + C_2,
\end{aligned} \tag{26}
$$

since there are only two nonvanishing cases: $s_1 = s_2 = s_3 = s_4$ (Case 1 => C_1), $s_1 = s_3 \neq s_2 = s_4$ and $s_1 = s_4 \neq s_2 = s_3$ (Case 2 => C_2).

* *Case 1:* in such a case, $\mathrm{Cov}\,(\xi_{s_1}\xi_{s_2}, \xi_{s_3}\xi_{s_4}) = \mu_4 - 1$ and

$$
\begin{aligned}
C_1 &= \frac{\mu_4 - 1}{a\,a'} \sum_{s \in \mathbb{Z}} \Big| \sum_{t=1}^{N} \alpha(t-s)\psi(\frac{t-b}{a})\Big|^2 \Big| \sum_{t=1}^{N} \alpha(t-s)\psi(\frac{t-b'}{a'})\Big|^2 \\
C_1 &= (\mu_4 - 1)\,a\,a' \lim_{M \to \infty} \int_{[-\pi,\pi]^4} d\lambda d\lambda' d\mu d\mu' e^{i[b(\lambda-\lambda')+b'(\mu-\mu')]} \\
&\qquad \times \sum_{s=-M}^{M} e^{is[(\lambda-\lambda')+(\mu-\mu')]} S_a(a\lambda)\widehat{\alpha}(\lambda)\overline{S_a(a\lambda')\widehat{\alpha}(\lambda')} S_{a'}(a'\mu)\widehat{\alpha}(\mu)\overline{S_{a'}(a'\mu')\widehat{\alpha}(\mu')}
\end{aligned}
$$

using the relation (24). From usual asymptotic behavior of Dirichlet kernel, for $g \in \mathcal{C}^1_{2\pi}((-\pi,\pi))$, $\lim_{M \to \infty} \int_{-\pi}^{\pi} D_M(z)g(x+z)dz = g(x)$ uniformly in x with $D_M(z) = \frac{1}{2\pi}\frac{\sin\left((2M+1)z/2\right)}{\sin\left(z/2\right)} = \frac{1}{2\pi} \sum_{k=-M}^{M} e^{ikz}$. Therefore with h a $\mathcal{C}^1$ function 2π-periodic for each component,

$$\lim_{M \to \infty} \int_{[-\pi,\pi]^4} 2\pi D_M((\lambda-\lambda')+(\mu-\mu'))h(\lambda,\lambda',\mu,\mu')d\lambda d\lambda' d\mu d\mu' = 2\pi \int_{[-\pi,\pi]^3} h(\lambda'-\mu+\mu', \lambda', \mu, \mu')d\lambda' d\mu d\mu';$$

12

Therefore,

$$C_1 = 2\pi \left(\mu_4 - 1\right) a\,a' \int_{[-\pi,\pi]^3} d\lambda' d\mu d\mu' e^{i(\mu-\mu')(b'-b)}$$

$$\times S_a(a(\lambda'-\mu+\mu'))\widehat{\alpha}(\lambda'-\mu+\mu')\overline{S_a(a\lambda')\widehat{\alpha}(\lambda')}S_{a'}(a'\mu)\widehat{\alpha}(\mu)\overline{S_{a'}(a'\mu')\widehat{\alpha}(\mu')}. \quad (27)$$

* *Case 2:* in such a case, with $s_1 \neq s_2$, $\mathrm{Cov}\left(\xi_{s_1}\xi_{s_2}, \xi_{s_1}\xi_{s_2}\right) = 1$ and

$$C_2 = \frac{2}{a\,a'} \sum_{(s,s')\in\mathbb{Z}^2, s\neq s'} \sum_{t_1=1}^{N} \alpha(t_1-s)\psi(\frac{t_1-b}{a}) \sum_{t_2=1}^{N} \alpha(t_2-s)\psi(\frac{t_2-b'}{a'}) \sum_{t_3=1}^{N} \alpha(t_3-s')\psi(\frac{t_3-b}{a}) \sum_{t_4=1}^{N} \alpha(t_4-s')\psi(\frac{t_4-b'}{a'})$$

$$= -\frac{2C_1}{\mu_4-1} + \frac{1}{a\,a'} \sum_{(s,s')\in\mathbb{Z}^2} \sum_{t_1=1}^{N} \alpha(t_1-s)\psi(\frac{t_1-b}{a}) \sum_{t_2=1}^{N} \alpha(t_2-s)\psi(\frac{t_2-b'}{a'}) \sum_{t_3=1}^{N} \alpha(t_3-s')\psi(\frac{t_3-b}{a}) \sum_{t_4=1}^{N} \alpha(t_4-s')\psi(\frac{t_4-b'}{a'})$$

$$C_2 = -\frac{2C_1}{\mu_4-1} + 2\,a\,a' \lim_{M\to\infty} \lim_{M'\to\infty} \int_{[-\pi,\pi]^4} d\lambda d\lambda' d\mu d\mu' e^{i[b(\lambda-\mu)-b'(\lambda'-\mu')]}$$

$$\times \sum_{s=-M}^{M} \sum_{s'=-M'}^{M'} e^{is(\lambda'-\lambda)+is'(\mu'-\mu)} S_a(a\lambda)\widehat{\alpha}(\lambda)\overline{S_{a'}(a'\lambda')\widehat{\alpha}(\lambda')}S_a(a\mu)\widehat{\alpha}(\mu)\overline{S_{a'}(a'\mu')\widehat{\alpha}(\mu')}$$

$$= -\frac{2C_1}{\mu_4-1} + 8\pi^2\,a\,a' \int_{[-\pi,\pi]^2} e^{i(\lambda-\mu)(b-b')} S_a(a\lambda)\overline{S_{a'}(a'\lambda)}S_a(a\mu)\overline{S_{a'}(a'\mu)} \times \left|\widehat{\alpha}(\lambda)\right|^2 \left|\widehat{\alpha}(\mu)\right|^2 d\lambda d\mu,$$

using the asymptotic behaviors of two Dirichlet kernels.

Now we have to compute $\sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} (C_1 + C_2)$. In both cases ($C_1$ and C_2), one again obtains a function of a Dirichlet kernel:

$$F_N(a,a',v) := \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} e^{i\,v\,(b-b')} = e^{iv(a-a')/2}\frac{\sin((N-a)v/2)\,\sin((N-a')v/2)}{\sin^2(v/2)}. \quad (28)$$

For a continuous function $h : [-\pi,\pi] \mapsto \mathbb{R}$,

$$\lim_{N\to\infty} \frac{1}{N} \int_{-\pi}^{\pi} h(v)F_N(a,a',v)dv = \lim_{N\to\infty} \frac{1}{N^2} \int_{-\pi N}^{\pi N} h(\frac{v}{N})F_N(a,a',\frac{v}{N})dv = 4h(0) \int_{-\infty}^{\infty} \frac{\sin^2(v/2)}{v^2}dv = 2\pi h(0),$$

thanks to Lebesgue Theorem and with $a/N \to 0$ ($N \to 0$). Then, from (27),

$$N \frac{1}{N-a} \frac{1}{N-a'} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} C_1 \sim 4\pi^2 \left(\mu_4-1\right)aa' \int_{[-\pi,\pi]^2} d\lambda' d\mu' |S_a(a\lambda')|^2 |S_{a'}(a'\mu')|^2 |\widehat{\alpha}(\lambda')|^2 |\widehat{\alpha}(\mu')|^2$$

$$\sim 4\pi^2 \left(\mu_4-1\right) \int_{-a\pi}^{a\pi} |S_a(\lambda)|^2 |\widehat{\alpha}(\lambda/a)|^2 d\lambda \int_{-a'\pi}^{a'\pi} |S_a(\mu)|^2 |\widehat{\alpha}(\mu/a')|^2 d\mu$$

$$\Longrightarrow N \frac{(aa')^{-2d}(c_d K_{(\psi,2d)})^{-2}}{4\pi^2(N-a)(N-a')} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} C_1 \underset{N\to\infty}{\longrightarrow} \left(\mu_4-1\right) \quad (29)$$

$$\text{and} \quad \frac{N}{a_N} \frac{(ra_N r'a_N)^{-2d}(c_d K_{(\psi,2d)})^{-2}}{4\pi^2(N-ra_N)(N-r'a_N)} \sum_{b=1}^{N-ra_N} \sum_{b'=1}^{N-r'a_N} C_1 \underset{N\to\infty}{\longrightarrow} 0, \quad (30)$$

using the same arguments than in Property 1 since $a_N \to \infty$ (and therefore $a \to \infty$ and $a' \to \infty$).

Moreover, if we consider that $a = r a_N$ and $a' = r' a_N$,

$$N \frac{1}{N-a} \frac{1}{N-a'} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} C_2 \sim 16\pi^3 a a' \int_{-\pi}^{\pi} |S_a(a\lambda)|^2 |S_{a'}(a'\lambda)|^2 |\widehat{\alpha}(\lambda)|^4 d\lambda - \frac{2N}{\mu_4 - 1} \frac{1}{N-a} \frac{1}{N-a'} \sum_{b=1}^{N-a} \sum_{b'=1}^{N-a'} C_1$$

$$\sim 16\pi^3 r r' a_N \int_{-a_N \pi}^{a_N \pi} |S_{r a_N}(r\lambda)|^2 |S_{r' a_N}(r'\lambda)|^2 |\widehat{\alpha}(\lambda/a_N)|^4 d\lambda - \frac{2N}{\mu_4 - 1} \frac{1}{N-r a_N} \frac{1}{N-r' a_N} \sum_{b=1}^{N-r a_N} \sum_{b'=1}^{N-r' a_N} C_1$$

$$\implies \frac{N}{a_N} \frac{(r \, r' \, a_N^2)^{-2d}(c_d K_{(\psi,2d)})^{-2}}{4\pi^2(N-r a_N)(N-r' a_N)} \sum_{b=1}^{N-r a_N} \sum_{b'=1}^{N-r' a_N} C_2 \underset{N\to\infty}{\longrightarrow} 4\pi \frac{(rr')^{1-2d}}{K_{(\psi,2d)}^2} \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(r\lambda)|^2 |\widehat{\psi}(r'\lambda)|^2}{\lambda^{4d}} d\lambda,$$

always using the same trick than in Property 1 since $a \to \infty$ and $a' \to \infty$. Therefore, with (30), one deduces that:

$$\frac{N}{a_N} \operatorname{Cov}\left(\widetilde{T}_N(r\,a_N), \widetilde{T}_N(r'\,a_N)\right) \underset{N\to\infty}{\longrightarrow} 4\pi \frac{(rr')^{1-2d}}{K_{(\psi,2d)}^2} \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(r\lambda)|^2 |\widehat{\psi}(r'\lambda)|^2}{\lambda^{4d}} d\lambda. \tag{31}$$

Note that if $r = r'$ then $\dfrac{N}{r\,a_N} \operatorname{Var}\left(\widetilde{T}_N(r\,a_N)\right) \underset{N\to\infty}{\longrightarrow} \sigma_\psi^2(d) = 64\pi^5 \dfrac{K_{(\psi*\psi,4d)}}{K_{(\psi,2d)}^2}$ only depending on ψ and d.

Step 2 We prove here that if the distribution of the innovations $(\xi_t)_t$ is such that there exists $r > 0$ satisfying $\mathbb{E}(e^{r\xi_0}) \leq \infty$ (condition so-called the Cramèr condition), then for any $a \in \mathbb{N}^*$, $(\widetilde{T}_N(r_i a_N))_{1\leq i\leq \ell} = \left(\frac{1}{N-r_i a_N} \sum_{k=1}^{N-r_i a_N} \widetilde{e}^2(r_i a_N, k)\right)_{1\leq i\leq \ell}$ satisfies a central limit theorem. Such theorem is implied by the proof that $\sqrt{\frac{N}{a_N}} \sum_{i=1}^{\ell} \frac{u_i}{N-r_i a_N} \sum_{k=1}^{N-r_i a_N} \widetilde{e}^2(r_i a_N, k)$ asymptotically follows a Gaussian distribution for any $(u_i)_{1\leq i\leq \ell} \in \mathbb{R}^\ell$. For establishing this result we are going to adapt a proof of Giraitis (1985) which shows central limit theorems for function of linear processes using a decomposition with Appell polynomials. Indeed we specified that X is a two-sided linear process and therefore martingale type results as in Wu (2002) or Furmanczyk (2007) can not be applied. Moreover, since a_N is a sequence depending on N it is required to prove a central limit theorem for triangular arrays. Unfortunately the recent paper of Roueff and Taqqu (2009) dealing with central limit theorems for arrays of decimated linear processes, and which can be applied to establish a multidimensional central limit for the variogram of wavelet coefficients associated to a multi-resolution analysis can not be applied here because in this paper this variogram is defined as in (7) with coefficients taken every n/n_j ($\simeq a_N$ with our notation) and the mean of n_j (N/a_N with our notation) coefficients is considered (and the convergence rate is $\sqrt{n_j}$). Our definition of the wavelet coefficient variogram (6) is an average of $N - a_N$ terms and the convergence rate is N/a_N. Then we chose to adapt the results and method of Giraitis (1985). More precisely, consider the case $\ell = 1$. For $a > 0$, $(\widetilde{e}(a,b))_{1\leq b\leq N-a}$ is a stationary linear process satisfying assumptions of the paper of Giraitis (called X_t in this article). Now we consider $H_2(x) = x^2 - 1$ the second-order Hermite polynomial and would like to prove that

$$\left(\frac{N}{a_N}\right)^{1/2} \frac{1}{N-a_N} \sum_{b=1}^{N-a_N} \left(\widetilde{e}^2(a_N, b) - 1\right) \simeq \left(\frac{1}{N a_N}\right)^{-1/2} \sum_{b=1}^{N-a_N} H_2(\widetilde{e}(a_N, b)) \underset{N\to\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\left(0, \sigma_\psi^2(d)\right).$$

Now since the distribution of ξ_0 is supposed to satisfy the Cramèr condition, following the proof of Proposition 6 (Giraitis, 1985), define $S_N^{(n)} = \sum_{b=1}^{N-a_N} A_n^{(a_N)}(\widetilde{e}(a_N, b))$ where $A_n^{(a_N)}$ is the Appell polynomial of degree n corresponding to the probability distribution of $\widetilde{e}(a_N, \cdot)$. We are going to prove that the cumulants of order $k \geq 3$ are such as

$$\chi\left(S_N^{(n(1))}, \ldots, S_N^{(n(k))}\right) = o\left((N a_N)^{k/2}\right) \tag{32}$$

for any $n(1), \cdots, n(k) \geq 2$ (the computation of the cumulant of order 2 is induced by Step 1 of this proof). Indeed, $\chi\left(S_N^{(n(1))}, \ldots, S_N^{(n(k))}\right) = \sum_{\gamma \in \Gamma_0(T)} d_\gamma I_\gamma(N)$ where $\Gamma_0(T)$ is the set of possible diagrams and the definition of $I_\gamma(N)$ is provided in (34) of Giraitis (1985).

In the case of Gaussian diagrams, $I_\gamma(N) = o\left((N a_N)^{k/2}\right)$, since this case is induced by the Gaussian case and the second order moments.

If γ is a non Gaussian diagram, *mutatis mutandis*, we are going to follow the notation and proof of Lemma 2 of Giraitis (1985). Note first from Step 1, we can write:

$$\widetilde{e}(a,b) = \sum_{s\in\mathbb{Z}} \beta_a(b-s) \xi_s \quad \text{with} \quad \beta_a(s) = \frac{\sqrt{a}}{\sqrt{\mathbb{E}e^2(a,b)}} \int_{-\pi}^{\pi} S_a(a\lambda)\widehat{\alpha}(\lambda) e^{i\lambda s} d\lambda. \tag{33}$$

14

Then for $u \in [-\pi, \pi]$,

$$
\begin{aligned}
\widehat{\beta}_a(u) & = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} \beta_a(s) e^{-isu} \\
& = \frac{\sqrt{a}}{2\pi \sqrt{\mathbb{E}e^2(a,b)}} \lim_{m \to \infty} \int_{-\pi}^{\pi} \sum_{s=-m}^{m} S_a(a\lambda) \widehat{\alpha}(\lambda) e^{is(\lambda-u)} d\lambda \\
& = \frac{\sqrt{a}}{\sqrt{\mathbb{E}e^2(a,b)}} S_a(au) \widehat{\alpha}(u),
\end{aligned}
$$

with the asymptotic behavior of Dirichlet kernel. Now, in case a/ of Lemma 2 of Giraitis (1985), consider the diagram $V_1 = \{(1,1), (2,1), (3,1)\}$ and assume that for the rows L_j of the array T, $j = 1, \cdots, k$ $(k \geq 3)$, $|V_1 \cap L_j| \geq 1$ for at least 3 different rows L_j. Then the inequality (39) can be repeated, and on the hyperplane x_{V_1}, a part of the integral (34) provides

$$
\Big| \int_{\{x_{11}+x_{21}+x_{31}=0\} \cap [-\pi,\pi]^3} dx_{11} dx_{21} dx_{31} \prod_{j=1}^{3} D_N((x_{j1} + \cdots + x_{jn(j)}) \widehat{\beta}_a(x_{j1}) \Big| \leq C \, \alpha_1(u_1) \alpha_2(u_2) \alpha_3(u_3),
$$

with $u_i = x_{i2} + \cdots + x_{in(i)}$ and the same expressions of α_i provided in Giraitis (1985). It remains to bound $\alpha_i(u)$. But, with the same approximations as in the proof of Property 1, for a and N large enough

$$
\begin{aligned}
\alpha_1(u) & = \int_{-\pi}^{\pi} |\widehat{\beta}_a(u) \, D_N(x+u)| dx \sim \sqrt{2\pi} \frac{1}{\sqrt{a}} \int_{-a\pi}^{a\pi} \Big| \frac{\widehat{\psi}(x)}{|x|^d} \Big| \, \Big| D_N \Big(\frac{x}{a} + u \Big) \Big| du \\
& \leq 2\sqrt{a} \sup_{x \in \mathbb{R}} \Big\{ \frac{|\widehat{\psi}(x)|}{|x|^d} \Big\} \int_{-\pi}^{\pi} |D_N(x+u)| dx \\
& \leq 2C \sup_{x \in \mathbb{R}} \Big\{ \frac{|\widehat{\psi}(x)|}{|x|^d} \Big\} \sqrt{a} \, \log N,
\end{aligned}
$$

since there exists $C > 0$ such as $\int_{-\pi}^{\pi} |D_N(x+u)| dx \leq C \log N$ for any $u \in [-\pi, \pi]$. Now for $i = 2, 3$, a and N large enough,

$$
\begin{aligned}
\alpha_i^2(u) & = \|\widehat{\beta}_{a_N}(\cdot) \, D_N(u + \cdot)\|_2^2 \\
& \leq 2 \int_{-a_N\pi}^{a_N\pi} \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}} D_N^2 \Big(\frac{x}{a_N} + u \Big) du \\
& \leq 2C \sup_{x \in \mathbb{R}} \Big\{ \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}} \Big\} a_N \int_{-\pi}^{\pi} |D_N^2(x+u)| dx \\
& \leq C' \sup_{x \in \mathbb{R}} \Big\{ \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}} \Big\} N a_N.
\end{aligned}
$$

Then $\alpha_1(u_1) \alpha_2(u_2) \alpha_3(u_3) = o((Na_N)^{3/2})$.

For the $k-3$ other terms, a result corresponding to Lemma 1 of Giraitis (1985) can also be obtained. Indeed, for a and N large enough,

$$
\begin{aligned}
\|g_{N,j}\|_2^2 & = \int_{[-\pi,\pi]^{n(j)}} dx \, D_N^2(x_1 + \cdots + x_{n(j)}) \prod_{i=1}^{n(j)} |\widehat{\beta}_a(x_i)|^2 \\
& \leq C \int_{[-a\pi,a\pi]^{n(j)}} dx \, D_N^2 \Big(\frac{1}{a}(x_1 + \cdots + x_{n(j)}) \Big) \prod_{i=1}^{n(j)} \frac{|\widehat{\psi}(x_i)|^2}{|x_i|^{2d}} \\
& \leq C \, \big| \sup_{x \in \mathbb{R}} \Big\{ \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}} \Big\} \big|^{n(j)} a \, \|D_N(\cdot)\|_2^2 \\
& \leq C' \, N a_N
\end{aligned}
$$

with $C' \geq 0$ not depending on N and a_N. Thus $\|g_{N,j}\|_2 \leq C (Na_N)^{1/2}$ with $C \geq 0$. Using the same reasoning, there also exists $C' \geq 0$ such as $\|g'_{N,j}\|_2 \leq C (Na_N)^{1/2}$ for $j \geq 2$ while $\|g'_{N,1}\|_2 = O(\sqrt{a_N} \log N) =$

$o((Na_N)^{1/2})$. As a consequence, for γ such as $|V_1 \cap L_j| \geq 1$ for at least 3 different rows L_j, and more generally with $|V_1| \geq 3$,

$$I_\gamma(N) = o\big((Na_N)^{k/2}\big). \tag{34}$$

For other γ, it remains to bound the function $h(u_1, u_2)$ defined in Giraitis (1985, p. 32) as follows (with $x = x_{11} + x_{12}$) and with $u_1 + u_2 \neq 0$:

$$
\begin{aligned}
h(u_1, u_2) &= \Big(\int_{-\pi}^{\pi} \big| \widehat{\beta}_{a_N}(-x)\, D_N(u_1 + x)\, D_N(u_2 - x) \big| dx \Big) \Big(\int_{-\pi}^{\pi} \big| \widehat{\beta}_{a_N}(x) \big|^2 dx \Big) \\
&\leq \Big| \sup_{x \in \mathbb{R}} \big\{ \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}} \big\} \Big| a_N \Big(\int_{-\pi}^{\pi} \big| D_N(u_1 + x)\, D_N(u_2 - x) \big| dx \Big) \Big(2\pi \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(x)|^2}{|x|^{2d}}\, dx \Big).
\end{aligned}
$$

But

$$
\begin{aligned}
\int_{-\pi}^{\pi} \big| D_N(u_1 + x)\, D_N(u_2 - x) \big| dx &\leq 2 \int_{-2\pi N}^{2\pi N} \Big| \frac{\sin(x)}{x} \frac{\sin(\frac{N}{2}(u_1 + u_2) - x)}{\sin(\frac{1}{2}(u_1 + u_2) - \frac{x}{N})} \Big| dx \\
&\leq \begin{cases} C\log N \big| \sin(\frac{1}{2}(u_1 + u_2)) \big|^{-1} & \text{if } |u_1 + u_2| \geq (N\log N)^{-1} \\ C\,N & \text{if } |u_1 + u_2| < (N\log N)^{-1} \end{cases}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\|h(u_1, u_2)\|_2^2 = \int_{[-\pi,\pi]^2} h^2(u_1, u_2) du_1 du_2 &\leq C a_N^2 \Big(\log^2 N \int_{(N\log N)^{-1}}^{\pi} (\sin x)^{-2}\, dx + N^2 \int_0^{(N\log N)^{-1}} dx \Big) \\
&\leq C a_N^2 \big(N\log^3 N + N\log N \big),
\end{aligned}
$$

and hence $\|h(u_1, u_2)\|_2 = o(Na_N)$. Finally, (34) holds for all γ and it implies (32).

If $\ell > 1$, the same proof can be repeated from the linearity properties of cumulants. Thus, $(\widetilde{T}_N(r_i a_N))_{1 \leq i \leq \ell}$ satisfies the following central limit:

$$\sqrt{\frac{N}{a_N}} \big(\widetilde{T}_N(r_i a_N) - 1 \big)_{1 \leq i \leq \ell} \xrightarrow[N \to \infty]{d} \mathcal{N}\big(0, \Gamma(r_1, \cdots, r_\ell, \psi, d) \big), \tag{35}$$

with $\Gamma(r_1, \cdots, r_\ell, \psi, d) = (\gamma_{ij})_{1 \leq i, j \leq \ell}$ given in (9).

Step 3 Now we extend the central limit obtained in Step 2 for linear processes with an innovation distribution satisfying a Cramèr condition ($\mathbb{E}\big(e^{r\xi_0} \big) < \infty$) to the weaker condition $\mathbb{E}\xi_0^4 < \infty$ using a truncation procedure. Thus assume now that $\mathbb{E}\xi_0^4 < \infty$. Let $M > 0$ and define $\xi_t^- = \xi_t \, \mathbb{I}_{|\xi| \leq M}$ and $\xi_t^+ = \xi_t \, \mathbb{I}_{|\xi| > M}$, $\widetilde{e}^-(a, b) = \sum_{s \in \mathbb{Z}} \beta_a(b-s)\, \xi_s^-$ and $\widetilde{e}^+(a, b) = \sum_{s \in \mathbb{Z}} \beta_a(b-s)\, \xi_s^+$ using (33). Clearly $\widetilde{e}(a, b) = \widetilde{e}^+(a, b) + \widetilde{e}^-(a, b)$. We are going to prove that (35) holds. For this, we begin by writing

$$\widetilde{T}_N(r_i a_N) - 1 = \frac{1}{N - r_i a_N} \Big(\sum_{b=1}^{N - r_i a_N} \big(\widetilde{e}^-(r_i a_N, b) \big)^2 - 1 \big) - 2\widetilde{e}^+(r_i a_N, b)\widetilde{e}^-(r_i a_N, b) + \big(\widetilde{e}^+(r_i a_N, b) \big)^2 \Big) \tag{36}$$

We first prove that $\big(\widetilde{T}_N^-(r_i a_N) - 1 \big)_{1 \leq i \leq \ell} = \big(\frac{1}{N - r_i a_N} \sum_{b=1}^{N - r_i a_N} \big(\widetilde{e}^-(r_i a_N, b) \big)^2 - 1 \big)_{1 \leq i \leq \ell}$ also satisfies (35). Indeed, $\big(\widetilde{e}^-(r_i a_N, b) \big)$ is a linear process with innovations (ξ_t^-) satisfying the Cramèr condition and it is obvious that $\Big(\frac{\mathbb{E}\big(\widetilde{e}(r_i a_N, b) \big)^2}{\mathbb{E}\big(\widetilde{e}^-(r_i a_N, b) \big)^2} \Big)^{1/2} \widetilde{e}^-(r_i a_N, b)_{b,i}$ has exactly the same distribution than $\widetilde{e}(r_i a_N, b)_{b,i}$. Therefore it remains to prove that $\sqrt{\frac{N}{a_N}} \Big(\frac{\mathbb{E}\big(\widetilde{e}(r_i a_N, b) \big)^2}{\mathbb{E}\big(\widetilde{e}^-(r_i a_N, b) \big)^2} - 1 \Big)$ converges to 0. We have $\mathbb{E}\big(\widetilde{e}(r_i a_N, b) \big)^2 = \big(\sum_{s \in \mathbb{Z}} \beta_a^2(s) \big) \mathbb{E}(\xi_0)^2 = 1$ and $\mathbb{E}\xi_0^2 = 1$ (from Property 1). Then

$$\Big| \frac{\mathbb{E}\big(\widetilde{e}^-(r_i a_N, b) \big)^2}{\mathbb{E}\big(\widetilde{e}(r_i a_N, b) \big)^2} - 1 \Big| \leq 2 \big(\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b) \big)^2 \big)^{1/2} + \mathbb{E}\big(\widetilde{e}^+(r_i a_N, b) \big)^2.$$

We have $\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b) \big)^2 = \big(\sum_{s \in \mathbb{Z}} \beta_a^2(s) \big) \mathbb{E}(\xi_0^+)^2 = \mathbb{E}(\xi_0^+)^2$ from previous arguments and since we assume that the distribution of ξ_0 is symmetric. But using Hölder's and Markov's inequalities $\mathbb{E}(\xi_0^+)^2 \leq (\mathbb{E}\xi_0^4)^{1/2}(\Pr(|\xi_0| > M))^{1/2} \leq (\mathbb{E}\xi_0^4)\, M^{-2}$. Hence, there exists $C > 0$ not depending on M and N,

$$\sqrt{\frac{N}{a_N}} \Big| \frac{\mathbb{E}\big(\widetilde{e}^-(r_i a_N, b) \big)^2}{\mathbb{E}\big(\widetilde{e}(r_i a_N, b) \big)^2} - 1 \Big| \leq \frac{C}{M} \sqrt{N} a_N \xrightarrow[N \to \infty]{} 0$$

when $M = N$ (for instance). Therefore $\big(\widetilde{T}_N^-(r_i\, a_N) - 1\big)_{1 \leq i \leq \ell}$ satisfies the CLT (35).
From (36), it remains to prove that

$$\sqrt{\frac{N}{a_N}}\,\frac{1}{N - r_i a_N}\Big(\sum_{b=1}^{N-r_i a_N} -2\widetilde{e}^+(r_i a_N, b)\widetilde{e}^-(r_i a_N, b) + \big(\widetilde{e}^+(r_i a_N, b)\big)^2\Big) \xrightarrow[N \to \infty]{\mathcal{P}} 0.$$

From Markov's and Hölder inequalities, this is implied when $\sqrt{\frac{N}{a_N}}\,\Big(\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b)\big)^2 + 2\sqrt{\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b)\big)^2}\Big) \xrightarrow[N \to \infty]{} 0$
with $\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b)\big)^2 = 1$. Using $\mathbb{E}\big(\widetilde{e}^+(r_i a_N, b)\big)^2 \leq \big(\mathbb{E}\xi_0^4\big) M^{-2}$ obtained above, we deduce that this statement
holds when $M = N$ (for instance). As a consequence, from (36), the CLT (35) holds even if the distribution
of ξ_0 is only symmetric and such that $\mathbb{E}\xi_0^4 < \infty$.

Step 4 It remains to apply the Delta-method to (35) with function $(x_1, \cdots, x_\ell) \mapsto (\log x_1, \cdots, \log x_\ell)$:

$$\sqrt{\frac{N}{a_N}}\,\big(\log\big(T_N(r_i\, a_N)\big) - \log(\mathbb{E}e^2(a_N, 1))\big)_{1 \leq i \leq \ell} \xrightarrow[N \to \infty]{d} \mathcal{N}\big(0\,,\, \Gamma(r_1, \cdots, r_\ell, \psi, d)\big),$$

With $\mathbb{E}e^2(a_N, 1)$ provided in Property 1, we obtain

$$\log \mathbb{E}e^2(a_N, 1) = 2d\,\log(a_N) + \log\big(\frac{c_d K_{(\psi, 2d)}}{2\pi}\big) + \frac{c_{d'} K_{(\psi, 2d-d')}}{2\pi\, a_N^{d'}}\big(1 + o(1)\big)$$

Therefore, when $\sqrt{\frac{N}{a_N}}\,\frac{1}{a_N^{d'}} \xrightarrow[N \to \infty]{} 0$, *i.e.* $N^{\frac{1}{1+2d'}} = o(a_N)$, the CLT (8) holds. $\qquad\square$

Proof of Theorem 1. Here we use Theorem 1 of Bardet *et al.* (2008) where it was proved that the CLT (8) is
still valid when a_N is replaced by $N^{\widetilde{\alpha}_N}$. Then, since $\widetilde{d}_N = \widetilde{M}_N\, Y_N(\widetilde{\alpha}_N)$ with $\widetilde{M}_N = \big(0\ 1/2\big)\big(Z_1'\widehat{\Gamma}_N^{-1}Z_1\big)^{-1}Z_1'\widehat{\Gamma}_N^{-1}$
we deduce that $\sqrt{N/N^{\widetilde{\alpha}_N}}\,\big(\widetilde{d}_N - d\big)$ is asymptotically Gaussian with asymptotic variance the limit in probability
of $\widetilde{M}_N\,\Gamma(1, \ldots, \ell, d, \psi)\,\widetilde{M}_N'$, that is σ^2.
The right hand side relation of (13) is also an obvious consequence of Theorem 1 of Bardet *et al.* (2008). $\quad\square$

Proof of Theorem 2. The theory of linear models can be applied: $Z_{N^{\widetilde{\alpha}_N}}\begin{pmatrix}\widetilde{c}_N \\ 2\widetilde{d}_N\end{pmatrix}$ is an orthogonal projector
of $Y_N(\widetilde{\alpha}_N)$ on a subspace of dimension 2, therefore $Y_N(\widetilde{\alpha}_N) - Z_{N^{\widetilde{\alpha}_N}}\begin{pmatrix}\widetilde{c}_N \\ 2\widetilde{d}_N\end{pmatrix}$ is an orthogonal projector of
$Y_N(\widetilde{\alpha}_N)$ on a subspace of dimension $\ell - 2$. Moreover, using the CLT (8) where a_N is replaced by $N^{\widetilde{\alpha}_N}$, we
deduce that $\sqrt{N/N^{\widetilde{\alpha}_N}}\widehat{\Gamma}_N^{-1}Y_N(\widetilde{\alpha}_N)$ asymptotically follows a Gaussian distribution with asymptotic covariance
matrix I_ℓ (identity matrix). Hence form Cochran Theorem we deduce (15). $\qquad\square$

References

[1] Abry, P., Flandrin, P., Taqqu, M.S. and Veitch, D. (2003). Self-similarity and long-range dependence through the wavelet lens, In P. Doukhan, G. Oppenheim and M.S. Taqqu editors, *Long-range Dependence: Theory and Applications*, Birkhäuser.

[2] Bardet, J.-M. and Bertrand, P. (2007) Definition, properties and wavelet analysis of multiscale fractional Brownian motion, *Fractals* **15**, 73–87.

[3] Bardet, J.M., Bibi H. and Jouini, A. (2008). Adaptive wavelet-based estimator of the memory parameter for stationary Gaussian processes, *Bernoulli*, 14, 691-724.

[4] Bardet, J.M., Lang, G., Moulines, E. and Soulier, P. (2000). Wavelet estimator of long range-dependant processes, *Statist. Infer. Stochast. Processes*, 3, 85–99.

[5] Bardet, J.M., Lang, G., Oppenheim, G., Philippe, A., Stoev, S. and Taqqu, M.S. (2003). Semiparametric estimation of the long-range dependence parameter: a survey. In *Theory and applications of long-range dependence*, Birkhäuser Boston, 557-577.

[6] Bhansali, R. Giraitis L. and Kokoszka, P.S. (2006). Estimation of the memory parameter by fitting fractionally-differenced autoregressive models, *J. Multivariate Analysis*, 97, 2101–2130.

[7] Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes, *Ann. Statist.*, 17, 1749-1766.

[8] Doukhan, P., Oppenheim, G. and Taqqu M.S. (Editors) (2003). *Theory and applications of long-range dependence*, Birkhäuser.

[9] Fox, R. and Taqqu, M.S. (1986). Large-sample properties of parameter estimates for strongly dependent Gaussian time series. *Ann. Statist.* 14, 517-532.

[10] Furmanczyk, K. (2007). Some remarks on the central limit theorem for functionals of linear processes under short-range dependence. *Probab. Math. Statist.*, 27, 235–245.

[11] Giraitis, L. (1985). The central limit theorem for functionals of a linear process. *Lith. Mat. J.*, 25, 43–57.

[12] 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.

[13] Giraitis, L., Robinson P.M., and Samarov, A. (2000). Adaptive semiparametric estimation of the memory parameter. *J. Multivariate Anal.*, 72, 183–207.

[14] Giraitis, L. and Surgailis, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its applications to the asymptotic normality of Whittle estimate. *Prob. Th. and Rel. Field.* 86, 87-104.

[15] Giraitis, L. and Taqqu, M.S. (1999). Whittle estimator for finite-variance non-Gaussian time series with long memory. *Ann. Statist.* 27, 178–203.

[16] 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.

[17] Moulines, E. and Soulier, P. (2003). Semiparametric spectral estimation for fractionnal processes, In P. Doukhan, G. Openheim and M.S. Taqqu editors, *Theory and applications of long-range dependence*, 251–301, Birkhäuser, Boston.

[18] R. Peltier and J. Lévy Véhel (1995) Multifractional Brownian Motion: definition and preliminary results. Preprint INRIA, available on `http://hal.inria.fr/docs/00/07/40/45/PDF/RR-2645.pdf`.

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

[20] F. Roueff and M. S. Taqqu (2009a). Asymptotic normality of wavelet estimators of the memory parameter for linear processes. *J. Time Ser. Anal.*, 30, 534–558.

[21] F. Roueff and M. S. Taqqu (2009b). Central limit theorems for arrays of decimated linear processes. *Stochastic Process. Appl.*, 119, 3006–3041.

[22] Taqqu, M.S., and Teverovsky, V. (1996). Semiparametric graphical estimation technics for long memeory data, *Lectures Notes in Statistics*, 115, 420–432.

[23] Veitch, D., Abry, P. and Taqqu, M.S. (2003). On the Automatic Selection of the Onset of Scaling, *Fractals*, 11, 377–390.

[24] Wu, W.B. (2002). Central limit theorems for functionals of linear processes and their applications. *Statist. Sinica*, 12, 635–649.