Available online at www.sciencedirect.com
ScienceDirect
Stochastic Processes and their Applications 120 (2010) 983–1009
stochastic processes and their applications
www.elsevier.com/locate/spa

Moment bounds for non-linear functionals of the periodogram

Gilles Faÿ*

Laboratoire Paul-Painlevé, Université Lille-1, 59655 Villeneuve-d'Ascq Cedex, France

Received 31 July 2008; received in revised form 22 November 2009; accepted 3 February 2010
Available online 16 February 2010

Abstract

In this paper, we prove the validity of the Edgeworth expansion of the Discrete Fourier transforms of some linear time series. This result is applied to approach moments of non-linear functionals of the periodogram. As an illustration, we give an expression of the mean square error of the slightly modified Geweke and Porter-Hudak estimator of the long memory parameter. We prove that this estimator is rate optimal, extending the result of Giraitis et al. (1997) [12] from Gaussian to linear processes.

Keywords: Linear processes; Discrete Fourier transform; Periodogram; Long range dependence; Geweke and Porter-Hudak (GPH) estimator; Log-periodogram regression

1. Introduction

Many estimators in time series analysis involve non-linear functionals of the periodogram. Examples include the estimation of the innovation variance [4,21,5,11], log-periodogram regression [28,30,26], robust non-parametric estimation of the spectral density [33,19]. Non-linear functionals of the periodogram also play a predominant role in the analysis of long-memory time-series: one of the much widely used estimators of the memory parameter is based on the regression of the log-periodogram ordinates on the log-frequency [10], see also [25,22].

The statistical analysis of such functionals has proved to be a very challenging problem, due to the intricate dependence structure of periodogram ordinates. The first attempts to study these

* Corresponding address: APC, Université Paris-7, 10, rue Alice Domon et Léonie Duquet, 75205 Paris Cedex 13, France. Tel.: +33 1 57 27 69 04; fax: +33 1 57 27 60 71.
E-mail address: gilles.fay@univ-lille1.fr.

0304-4149/$ - see front matter
doi:10.1016/j.spa.2010.02.007

statistics were made under the additional assumption that the underlying process is Gaussian. Because the Fourier transform coefficients are in this case also Gaussian, one may then apply results on non-linear transforms of Gaussian random variables; see for example [31,29,1].

These techniques do not extend to non-Gaussian processes. A first step to weaken this assumption was taken by Chen and Hannan [4], who proved the consistency of an additive functional of the log-periodogram of a linear stationary process, with an application to the estimation of the innovation variance. These techniques were based on the so-called Bartlett [2] expansion; this technique was later improved by Faÿ et al. [8], who proved central limit theorems for these functionals. It used by Velasco [32] to establish the weak consistency of the log-periodogram regression estimate of the long memory parameter for long range dependent linear time series. Edgeworth expansions are used to estimates moments of the functional of the unobservable periodogram of the innovation sequence. Remainder terms can be bounded in probability. The Bartlett expansion is indeed useful to establish limit theorems, but does not in general allow one to determine the moments of these functionals.

An alternative approach has been considered by von Sachs [33] and Janas and Von Sachs [19]. These authors prove the mean-square consistency of a general additive functional of non-linear transforms of the (tapered) periodogram, using Edgeworth expansions of the discrete Fourier transform of the observed time series itself. Janas and von Sachs [19] apply these results to prove the mean-square consistency of a Huberized (peak insensitive) non-parametric spectral estimator. These results rely on the Edgeworth expansion of a triangular array of strongly mixing processes with geometrically mixing coefficient established by Götze and Hipp [13]. The mixing conditions herein are rather stringent, and thus the conclusions reached by Janas and von Sachs [19] are proved under a set of restrictive assumptions, precluding, for instance, their use in a long-memory context.

The main objective of this paper is to develop a method allowing one to compute the moments of functionals of non-linear transforms of the (possibly tapered) periodogram of a *linear* process. These results are based on Edgeworth expansion of a (possibly infinite) triangular array of i.i.d. random variables, obtained earlier in [9] and recalled in Appendix A. The linearity of the process is then crucial. Our results cover short, long and negative memory processes.

The remainder of the paper is organized as follows. In Section 2 we give the assumptions on the linear structure of the time series and define the cumbersome notation related to Edgeworth expansions. In Section 3, we formulate the validity of Edgeworth expansions and moment bounds under the short memory set of hypotheses. As an application, we derive the mean-square consistency of additive functionals of the non-linear transform of the periodogram for a short-memory linear time-series. In Section 4, we follow the same lines, but in a long or negative memory framework, and apply the moment bounds we obtain to control the mean-square error of the Geweke and Porter-Hudak [10] estimator of the fractional difference parameter for a linear process. We work out a modified version of this estimator, where regression is performed on every two periodogram ordinates. This extends the rate optimality property of the Geweke and Porter-Hudak (hereafter, GPH) estimator obtained earlier by Giraitis et al. [12] for Gaussian processes. A Monte-Carlo experiment is run to confirm our results for finite-sample observations. Proofs are postponed to the Appendices.

2. Notations and assumptions

Assume that $X = (X_t)_{t\in\mathbb{Z}}$ is a covariance stationary process that has a spectral density f. For any integer $r \geq 0$, we define the tapered discrete Fourier transform (DFT) and periodogram of order r as

$$d_{r,n}(\lambda) \stackrel{\text{def}}{=} (2\pi n a_r)^{-1/2} \sum_{t=1}^{n} h_{t,n}^r X_t \mathrm{e}^{\mathrm{i}t\lambda}, \qquad I_{r,n}(\lambda) \stackrel{\text{def}}{=} |d_{r,n}(\lambda)|^2 \tag{2.1}$$

where $h_{t,n} \stackrel{\text{def}}{=} 1-\mathrm{e}^{2\mathrm{i}\pi t/n}$ is the data taper introduced in [17] and $a_r \stackrel{\text{def}}{=} n^{-1}\sum_{t=1}^{n} |h_{t,n}|^{2r} = \binom{2r}{r}$ is a normalization factor. Denote $d_{r,n,k} = d_{r,n}(\lambda_k)$ and $I_{r,n,k} = I_{r,n}(\lambda_k)$ the tapered DFT and tapered periodogram evaluated at the Fourier frequencies $\lambda_k \stackrel{\text{def}}{=} \frac{2\pi k}{n}$, $k = 1, \dots, [(n-1)/2]$. Define for $r \in \mathbb{N}$, $D_{r,n}(\lambda)$ the normalized kernel function

$$D_{r,n}(\lambda) \stackrel{\text{def}}{=} (na_r)^{-1/2} \sum_{t=1}^{n} h_{t,n}^r \exp(\mathrm{i}t\lambda) = (na_r)^{-1/2} \sum_{k=0}^{r} \binom{r}{k} (-1)^k D_n(\lambda + \lambda_k) \tag{2.2}$$

where $D_n(\lambda) \stackrel{\text{def}}{=} \sum_{t=1}^{n} \mathrm{e}^{-\mathrm{i}\lambda t}$ denotes the non-symmetric Dirichlet kernel. The latter relation implies that $D_{r,n}(\lambda_k) = 0$ for $k \in \{1, \dots, \tilde{n}\}$, with $\tilde{n} \stackrel{\text{def}}{=} \lfloor (n-2r-1)/2 \rfloor$, so that the tapered Fourier transform is invariant to shift in the mean. As shown in [17], the decay rate of the kernel in the frequency domain increases with the kernel order, namely

$$\forall \lambda \in [-3\pi/2, 3\pi/2], \quad |D_{r,n}(\lambda)| \le \frac{Cn^{1/2}}{(1+n|\lambda|)^{r+1}}. \tag{2.3}$$

This property implies that higher order kernels are more effective to control frequency leakage. Tapering ($r > 0$) allows one to prove accurate bounds on the covariance of the DFT's in the anti-persistent case (negative memory), as recalled in Lemma 8 adapted from [20] (see also the hypotheses of Theorems 6 and 11, for instance). In this case, any spectral estimation of the memory parameter is very sensitive with respect to the leakage from higher frequencies to the null frequency, where the spectral density behaves as a power law with positive index.

If X is a white noise and $r = 0$, the DFT ordinates at different Fourier frequencies are uncorrelated. This property is lost by tapering. More precisely, for $1 \le k \ne j \le \tilde{n}$, $\mathbb{E}[d_{r,n,k} d_{r,n,j}] = 0$, and $\mathbb{E}[d_{r,n,k}\overline{d_{r,n,j}}] \stackrel{\text{def}}{=} (2\pi)^{-1}\varsigma_r(k-j)$, where $\bar{z}$ denotes the complex conjugate of z and ς_r defined in (3.6).

Many statistical applications (see the references given in the Introduction) require one to study weighted sums of non-linear functionals of the periodogram ordinates

$$T_n(X,\phi) = \sum_{k=1}^{K} \beta_{n,k} \phi\left(\frac{I_{r,n,k}}{f(\lambda_k)}\right), \tag{2.4}$$

where $(\beta_{n,k})_{k\in\{1,\dots,K\}}$ is a triangular array of real numbers. If X is a Gaussian white noise, then $(I_{r,n,k})$ are i.i.d and the moments of the sum $T_n(X,\phi)$ can be calculated explicitly. In any other case, the random variables $(I_{r,n,k})_{k\in\{1,\dots,K\}}$ are not independent, and the calculation of the moments of $T_n(X,\phi)$ is a difficult problem. The only attempt to solve it has been made by Janas and von Sachs [19], who proposed a technique to compute moments of order 1 and 2. As already outlined, their results are based on mixing conditions, precluding their use for long-memory processes.

Remark. Sometimes the periodogram ordinates are averaged along blocks of adjacent frequencies. This technique is known as *pooling* and is appropriate to reduce asymptotic variance of the estimators of non-linear functionals of the periodogram (see [25,24]). For simplicity, we will not present any explicit result or application with the pooled periodogram, but the Edgeworth expansion results that follow allow one to derive moment bounds on functionals of tapered *and* pooled periodogram as well.

In this contribution, we focus on strict sense linear processes, *i.e.* it is assumed that

$$X_t = \sum_{j\in\mathbb{Z}} \psi_j Z_{t-j}, \qquad \sum_{j\in\mathbb{Z}} \psi_j^2 < \infty, \tag{2.5}$$

where $(Z_j)_{j\in\mathbb{Z}}$ is a sequence of i.i.d random variables such that $\mathbb{E}[Z_1] = 0$, $\mathbb{E}[Z_1^2] = 1$. In addition, for some $s \geq 3$, $p \geq 1$ and $p' \geq 0$,

(A1) $\mathbb{E}[|Z_1|^s] < \infty$ and $\int_{\mathbb{R}} |t|^{p'} \, |\mathbb{E}[\mathrm{e}^{\mathrm{i}tZ_1}]|^p \, \mathrm{d}t < \infty$.

Remark. Apart from a classical moment condition, (**A1**), suppose that the distribution of the i.i.d. noise is smooth; for example, lattice distributions are forbidden. This condition is stronger than the usual Cramér condition. It ensures that the distributions of the Fourier coefficients of Z are eventually continuous. We need this continuity to bound moments of singular functionals (such as the logarithm) of the periodogram. Note that this condition could be dispensed with, were we concerned with smooth functionals.

Define $\psi(\lambda) = \sum_{j\in\mathbb{Z}} \psi_j \mathrm{e}^{\mathrm{i}j\lambda}$ (the convergence holds in $\mathbb{L}^2([-\pi, \pi], \mathrm{d}x)$) the transfer function of the linear filter $(\psi_j)_{j\in\mathbb{Z}}$ and $f(\lambda) = (2\pi)^{-1}|\psi(\lambda)|^2$ the spectral density of the process X. For an integer $k \in \{1, \ldots, \tilde{n}\}$ such that $f(\lambda_k) \neq 0$, define the normalized DFT $\omega_{r,n,k} \stackrel{\text{def}}{=} \sqrt{2\pi}\; d_{r,n,k}/|\psi(\lambda_k)|$. Let $k_1 < k_2 < \cdots < k_u$ be an ordered u-tuple of such integers in the range $1, \ldots, \tilde{n}$ and write $\mathbf{k} = (k_1, \ldots, k_u)$. Define (the reference to r is suppressed in the notation)

$$\mathbf{S}_n(\mathbf{k}) \stackrel{\text{def}}{=} \left[\mathrm{Re}(\omega_{r,n,k_1}), \mathrm{Im}(\omega_{r,n,k_1}), \ldots, \mathrm{Re}(\omega_{r,n,k_u}), \mathrm{Im}(\omega_{r,n,k_u})\right]. \tag{2.6}$$

With those definitions,

$$I_{n,k,r} = f(\lambda_k)|\omega_{r,n,k}|^2 = f(\lambda_k)\|\mathbf{S}_n(k)\|^2. \tag{2.7}$$

Since X admits the linear representation (2.5), $\mathbf{S}_n(\mathbf{k})$ can be further expressed as a $2u$-dimensional infinite triangular array in the variables $(Z_t)_{t\in\mathbb{Z}}$. Precisely

$$\mathbf{S}_n(\mathbf{k}) = \sum_{j\in\mathbb{Z}} \mathbf{U}_{n,j}(\mathbf{k}) Z_j, \tag{2.8}$$

with

$$\mathbf{U}_{n,j}(\mathbf{k}) \stackrel{\text{def}}{=} (na_r)^{-1/2}\mathbf{F}_n^{-1}(\mathbf{k}) \sum_{t=1}^{n} \psi_{t-j} \mathbf{C}_{n,t}(\mathbf{k}), \tag{2.9}$$

$$\mathbf{C}_{n,t}(\mathbf{k}) \stackrel{\text{def}}{=} \sum_{p=0}^{r} (-1)^p \binom{r}{p} \Big(\cos(t\lambda_{k_1+p}), \sin(t\lambda_{k_1+p}), \ldots, \cos(t\lambda_{k_u+p}), \sin(t\lambda_{k_u+p})\Big)'$$

$$\text{and} \quad \mathbf{F}_n(\mathbf{k}) \stackrel{\text{def}}{=} \mathrm{diag}\big(|\psi(\lambda_{k_1})|, |\psi(\lambda_{k_1})|, \ldots, |\psi(\lambda_{k_u})|, |\psi(\lambda_{k_u})|\big).$$

To formulate our results, some notations related to Edgeworth expansions are required, which we take from the monograph of Bhattacharya and Rao [3]. For u a positive integer, $\mathbf{v} = (\nu_1, \ldots, \nu_u) \in \mathbb{N}^u$ and $\mathbf{z} = (z_1, \ldots, z_u) \in \mathbb{C}^u$, denote $|\mathbf{v}| = \sum_{i=1}^u \nu_i$, $\mathbf{v}! = \nu_1!\nu_2! \cdots \nu_u!$ and $\mathbf{z}^{\mathbf{v}} = z_1^{\nu_1} z_2^{\nu_2} \cdots z_u^{\nu_u}$. If $1 \le |\mathbf{v}| \le s$, denote $\chi_{n,\mathbf{v}}(\mathbf{k})$ the cumulants of $\mathbf{S}_n(\mathbf{k})$. Then $\chi_{n,\mathbf{v}}(\mathbf{k}) = \kappa_{|\mathbf{v}|} \sum_{j \in \mathbb{Z}} \mathbf{U}^{\mathbf{v}}_{n,j}(\mathbf{k})$ where κ_r denotes the r-th cumulant of Z_1, $r \le s$. Let $\mathbf{V}_n(\mathbf{k}) \stackrel{\text{def}}{=} \operatorname{cov}[\mathbf{S}_n(\mathbf{k})] = \sum_{j \in \mathbb{Z}} \mathbf{U}_{n,j}(\mathbf{k})\mathbf{U}'_{n,j}(\mathbf{k})$. Let $\boldsymbol{\chi} = \{\chi_{\mathbf{v}};\ \mathbf{v} \in \mathbb{N}^u\}$ be a set of real numbers. For any integer $r \ge 2$ and $\mathbf{z} \in \mathbb{C}^u$, define $\chi_r(\mathbf{z}) \stackrel{\text{def}}{=} r! \sum_{|\mathbf{v}|=r} \frac{\chi_{\mathbf{v}} \mathbf{z}^{\mathbf{v}}}{\mathbf{v}!}$. The polynomials $\tilde{P}_r(\mathbf{z}, \boldsymbol{\chi})$ are formally defined for $r \ge 1$ by the identities

$$1 + \sum_{r=1}^{\infty} \tilde{P}_r(\mathbf{z}, \boldsymbol{\chi}) t^r = \exp\Big\{\sum_{r=3}^{\infty} \frac{\chi_r(\mathbf{z})}{r!} t^{r-2}\Big\} = 1 + \sum_{m=1}^{\infty} \frac{1}{m!} \Big(\sum_{r=3}^{\infty} \frac{\chi_r(\mathbf{z})}{r!} t^{r-2}\Big)^m,$$

and we set $\tilde{P}_0 \equiv 0$. Denote $\varphi_{\mathbf{V}}$ the density of a Gaussian r.v in $\mathbb{R}^u$ with zero mean and non-singular covariance matrix $\mathbf{V}$. Define $P_r : \mathbb{R}^u \mapsto \mathbb{R}$ by $P_r(\mathbf{x}, \mathbf{V}, \boldsymbol{\chi}) = \Big[\tilde{P}_r(-D, \boldsymbol{\chi})\Big] \varphi_{\mathbf{V}}(\mathbf{x})$ where, for any polynomial $P(\mathbf{z}) = \sum_{\mathbf{v}} a_{\mathbf{v}} \mathbf{z}^{\mathbf{v}}$, $P(-D)$ is interpreted as a polynomial in the differentiation operator D, $P(-D) = \sum_{\mathbf{v}} a_{\mathbf{v}} (-1)^{|\mathbf{v}|} D^{\mathbf{v}}$, with $D^{\mathbf{v}} = \frac{\partial^{|\mathbf{v}|}}{\partial x_1^{\nu_1} \ldots \partial x_u^{\nu_u}}$, $\mathbf{v} = (\nu_1, \ldots, \nu_u) \in \mathbb{N}^u$. By construction P_r and $\tilde{P}_r$ do not depend on the coefficient $\chi_{\mathbf{v}}$ if $|\mathbf{v}| > r + 2$, and $\tilde{P}_r(\mathrm{i}\mathbf{t}, \boldsymbol{\chi}) \mathrm{e}^{-\mathbf{t}'\mathbf{V}\mathbf{t}/2}$ is the Fourier transform of $P_r(\mathbf{x}, \mathbf{V}, \boldsymbol{\chi})$. Let ξ_Γ be a centered a-dimensional Gaussian vector with covariance matrix Γ and $g : \mathbb{R}^a \to \mathbb{R}$ a measurable mapping. Define $N_s(g) = \int_{\mathbb{R}^a} (1 + \|\mathbf{x}\|^s)^{-1} |g(\mathbf{x})| \, \mathrm{d}\mathbf{x}$ and $\|g\|_\Gamma^2 = \mathbb{E}[g^2(\xi_\Gamma)]$. The Hermite rank of g, $\|g\|_\Gamma^2 < \infty$, with respect to Γ is defined as the smallest integer τ such that there exists a polynomial P of degree τ with $\mathbb{E}[g(\xi_\Gamma) P(\xi_\Gamma)] \ne 0$. We denote $\tau(g, \Gamma)$ the (positive) Hermite rank of $g - \mathbb{E}[g(\xi_\Gamma)]$ with respect to Γ.

3. Moment bounds: Short memory case

In this section we consider short-range dependent processes. For any reals $\alpha, \delta > 0$ and $\beta < \infty$, denote by $\mathcal{G}(\alpha, \beta, \delta)$ the set of real sequences $(\psi_j)_{j \in \mathbb{Z}}$ such that

$$|\psi_0| + \sum_{j \in \mathbb{Z}} |j|^{1/2+\delta} |\psi_j| \le \beta, \tag{3.1}$$

$$\alpha \le \inf_{\lambda \in [-\pi, \pi]} |\psi(\lambda)|. \tag{3.2}$$

Theorem 1. *Assume* (**A1**) *with some integer $s \ge 3$, $p \ge 1$ and $p' = 0$ and assume that $(\psi_j)_{j \in \mathbb{Z}} \in \mathcal{G}(\alpha, \beta, \delta)$ for some $\alpha, \delta > 0$ and $\beta < \infty$. Then, there exists constants C and N (depending only on $s, p, \alpha, \beta, \delta, u$ and the distribution of Z_0) such that, for all $n \ge N$, and all u-tuple* $\mathbf{k}$ *of distinct integers, the distribution of* $\mathbf{S}_n(\mathbf{k})$ *has a density* $q_{n,\mathbf{k}}$ *with respect to Lebesgue's measure on* $\mathbb{R}^{2u}$ *and*

$$\sup_{\mathbf{x} \in \mathbb{R}^{2u}} (1 + \|\mathbf{x}\|^s) \left| q_{n,\mathbf{k}}(\mathbf{x}) - \sum_{r=0}^{s-3} P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\}) \right| \le C n^{-(s-2)/2}. \tag{3.3}$$

Several interesting consequences can be derived from this result. A straightforward integration of the expansion (3.3) yields the following Corollary, which gives an Edgeworth expansion of some moment $\mathbb{E}[g(\mathbf{S}_n(\mathbf{k}))]$ around the centered Gaussian distribution with covariance matrix $\mathbf{V}_n(\mathbf{k})$.

Corollary 2. *There exists a constant C and an integer N (depending only on* $s, p, \alpha, \beta, \delta, u$ *and the distribution of* Z_0*) such that, for any u-tuple of distinct integers* $\mathbf{k}$*,* $n \geq N$ *and measurable function g satisfying* $N_s(g) < \infty$*,*

$$\left| \mathbb{E}[g(\mathbf{S}_n(\mathbf{k}))] - \sum_{r=0}^{s-3} \int_{\mathbb{R}^{2u}} g(\mathbf{x}) P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\}) \, \mathrm{d}\mathbf{x} \right| \leq C \; N_s(g) \; n^{-(s-2)/2}. \qquad (3.4)$$

One can also use Theorem 1 to develop the same moment around the *limiting* Gaussian distribution of $\mathbf{S}_n$. Recalling that $\omega_{r,n,k} = a_r^{-1/2} \sum_{s=0}^{r} \binom{r}{s} (-1)^s \omega_{0,n,k+s}$, we have

$$\lim_{n \to \infty} \mathbf{V}_n(\mathbf{k}) = \mathbf{V}(\mathbf{k})$$

under short memory conditions, where $\mathbf{V}(\mathbf{k})$ is the $2u \times 2u$ matrix defined component-wise by

$$\begin{aligned} &[\mathbf{V}(\mathbf{k})]_{2i-1,2j-1} = [\mathbf{V}(\mathbf{k})]_{2i,2j} = \frac{1}{2} \varsigma_r(k_i - k_j), \\ &[\mathbf{V}(\mathbf{k})]_{2i-1,2j} = [\mathbf{V}(\mathbf{k})]_{2i,2j-1} = 0, \end{aligned} \qquad (3.5)$$

for $i, j = 1, \ldots, u$, with

$$\varsigma_r(l) \stackrel{\text{def}}{=} \begin{cases} 0 & \text{if } |l| > r, \\ a_r^{-1} (-1)^l \binom{2r}{r+l} & \text{if } |l| \leq r. \end{cases} \qquad (3.6)$$

Note that $\mathbf{V}(\mathbf{k}) = \frac{1}{2} \mathbf{I}_{2u}$ if $r = 0$.

Corollary 3. *There exists a constant C and N (depending only on* $s, p, \alpha, \beta, \delta, u$ *and the distribution of* Z_0*), such that for all measurable function g on* $\mathbb{R}^{2u}$ *such that* $N_3(g) < \infty$*, all u-tuple of distinct integers* $\mathbf{k}$*, and any* $n \geq N$*,*

$$\left| \mathbb{E}\left[g(\mathbf{S}_n(\mathbf{k}))\right] - \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \varphi_{\mathbf{V}(\mathbf{k})}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \right| \leq C \left\{ n^{-1/2} N_3(g) + n^{-\tau(g,\mathbf{V}(\mathbf{k}))/2} \; \|g\|_{\mathbf{V}(\mathbf{k})} \right\}. \qquad (3.7)$$

For some functions g, it is possible to sharpen this result by considering higher-order ($s > 3$) expansions and approximating the terms appearing in these expansions. We shall consider mappings $g : \mathbb{R}^{2u} \to \mathbb{R}$ such that

$$\begin{aligned} &g(x_1, \ldots, x_{2u}) = \prod_{j=1}^{u} g_j(x_{2j-1}, x_{2j}) \\ &\text{with } g_j(x, y) = g_j(y, x) = g_j(-x, y), \quad j = 1, \ldots, u. \end{aligned} \qquad (3.8)$$

Recalling (2.7), products of functionals of the periodogram are included in this particular case. Better bounds are obtained by considering frequencies $k_1, \ldots, k_u$ separated by r, so that the asymptotic decorrelation is achieved, $\mathbf{V}(\mathbf{k}) = \frac{1}{2} \mathbf{I}_{2u}$ as in the $r = 0$ case. Under those conditions, the $O(n^{-1/2})$ of Corollary 3 can be improved to $O(n^{-1})$.

Corollary 4. *Under the hypothesis that* $s \geq 4$*, there exists a constant C and N (depending only on* $s, p, \alpha, \beta, \delta, u$ *and the distribution of* Z_0*), such that for all measurable function g satisfying (3.8) and such that* $N_s(g) < \infty$*, all u-tuple of ordered integers* $\mathbf{k}$ *such that*

$k_i < k_{i+1} - r$, and any $n \geq N$,

$$\left|\mathbb{E}\left[g(\mathbf{S}_n(\mathbf{k}))\right] - \int_{\mathbb{R}^{2u}} g(\mathbf{x})\varphi_{\mathbf{I}_{2u}/2}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right| \leq C\left\{n^{-(s-2)/2}N_s(g) + n^{-1}\|(1+\|\mathbf{x}\|^s)g(\mathbf{x})\|_{\mathbf{I}_{2u}}\right\}. \tag{3.9}$$

The proofs of Corollaries 3 and 4 are postponed to Appendix D.

Remark. Pushing to higher orders $s \geq 4$ in Corollary 4 it is sometimes necessary to have $N_s(g) < \infty$ (see the applications below). But it does not improve the $O(n^{-1})$ bound.

To illustrate the results above, we compute bounds for the mean-square error of plug-in estimators of non-linear functionals of the spectral density $\Lambda(f) = \int_0^\pi w(\lambda)G(f(\lambda))\,\mathrm{d}\lambda$, where w is a function of bounded variation and G is a function such that there exists a function H satisfying, for any $x > 0$, $\int_0^\infty |H(xv)|\mathrm{e}^{-v}\,\mathrm{d}v < \infty$ and $\int_{v>0} H(xv)\mathrm{e}^{-v}\,\mathrm{d}v = G(x)$, *i.e.* H is the inverse Laplace transform of the function $t \mapsto G(1/t)/t$. We consider the following estimator

$$\hat{\Lambda}_n = (\pi/\tilde{n})\sum_{k=1}^{\tilde{n}} w(\lambda_k)H(I_{n,k})$$

and put $\Lambda_n = (\pi/\tilde{n})\sum_{k=1}^{\tilde{n}} w(\lambda_k)G(f(\lambda_k))$. Here, $r = 0$ and $I_{n,k} \stackrel{\text{def}}{=} I_{0,n,k}$ is the ordinary periodogram. We assume that the approximation error $\Lambda_n - \Lambda$ may be neglected in comparison with the mean-square error $\mathbb{E}(\hat{\Lambda}_n - \Lambda_n)^2$. These functionals have been studied in [29] in the Gaussian case, and Janas and Von Sachs [19] for non-Gaussian linear process, under rather stringent assumptions (see also [5] and the references therein). The moment bounds we have established allow the extension of Janas and von Sachs's [19]'s result, by relaxing the conditions on the dependence (from $|\psi_j| < C\rho^{|j|}$ for some $\rho \in (0,1)$ to $\sum_{j\in\mathbb{Z}} |j|^{1/2}|\psi_j| < \infty$).

Proposition 5. *Let $(X_t)_{t\in\mathbb{Z}}$ be a sequence satisfying the assumptions of Theorem 1 with some $s \geq 4$. Put $H_1(x_1,x_2) = H(x_1^2+x_2^2)$, $H_2(x_1,x_2,x_3,x_4) = H_1(x_1,x_2)H_1(x_3,x_4)$ and assume that $N_3(H_1^2) < \infty$ and $N_5(H_2) < \infty$. Then, uniformly in $f \in \mathcal{G}(\alpha,\beta,\delta)$*

$$\mathbb{E}[(\hat{\Lambda}_n - \Lambda_n)^2] \leq Cn^{-1}.$$

Sketch of the proof. Applying Corollary 3 to the function $g_{k,f}(x_1,x_2) = H[f(\lambda_k)(x_1^2+x_2^2)]$ and Corollary 4 to $g_{k,j,f}(x_1,x_2,x_3,x_4) = H[f(\lambda_k)(x_1^2+x_2^2)]H[f(\lambda_j)(x_3^2+x_4^2)]$ yield asymptotic expansions for the moments $\mathbb{E}[H^2(I_{n,k})]$ and $\mathbb{E}[H(I_{n,k})H(I_{n,j})]$, which are sufficient to derive the result. The uniformity of the constant C follows from the existence of bounds on $N_3(g_{k,f})$ and $N_4(g_{k,j,f})$, which are uniform with respect to $\psi \in \mathcal{G}(\alpha,\beta,\delta)$. □

4. Moment bounds: Long and negative memory case

4.1. Assumptions and main results

We consider two sets of assumptions, depending on available information on the behavior of the spectral density outside a neighborhood of the zero frequency. Recall that a real valued function ϕ defined in a neighborhood of zero is regularly varying at zero with index $\rho \in \mathbb{R}$ if, for all x and all $t > 0$, $\lim_{x\to 0} \phi(tx)/\phi(x) = t^\rho$. If $\rho = 0$, the function ϕ is said to be slowly varying at zero. Let $\vartheta \in (0, \pi)$, $0 < \delta < 1/2$, $\Delta < \delta$. We say that the linear filter $(\psi_j)_{j\in\mathbb{Z}}$ belongs to the set $\mathcal{F}(\vartheta, \delta, \Delta, \mu)$ if $\sum_{j=-\infty}^{\infty} \psi_j^2 < \infty$ and if there exists $d \in [\Delta, \delta]$ such that $\psi(\lambda)$ is regularly varying at zero with index $-d$ and that

$$\frac{\int_0^\pi \lambda^{2d}|\psi(\lambda)|^2\,\mathrm{d}\lambda}{\min\limits_{0\le|\lambda|\le\vartheta} \lambda^{2d}|\psi(\lambda)|^2} \le \mu, \tag{4.1}$$

$$\forall j \ge 0, \quad \frac{|\psi_j| + \sum\limits_{|t|\ge j}^{\infty} |\psi_{t+1} - \psi_t|}{\min\limits_{0\le\lambda\le\vartheta} \lambda^d|\psi(\lambda)|} \le \mu(1+j)^{d-1}. \tag{4.2}$$

An example is provided by $\psi(\lambda) \overset{\text{def}}{=} \left(1 - \mathrm{e}^{\mathrm{i}\lambda}\right)^{-d}$, the transfer function of the causal fractional integration filter, $\psi_t = \Gamma(t+d)/(\Gamma(d)\Gamma(t+1))$, $t \ge 0$.

Local-to-zero assumptions. We first consider local-to-zero assumptions for which nothing is required outside a neighborhood of the zero frequency, apart from integrability of the spectral density (see [25]). For $\beta > 0$, we say that the sequence $(\psi_j)_{j\in\mathbb{Z}}$ belongs to the set $\mathcal{F}_{\text{local}}(\vartheta, \beta, \delta, \Delta, \mu)$ if $(\psi_j)_{j\in\mathbb{Z}} \in \mathcal{F}(\vartheta, \delta, \Delta, \mu)$ and

$$\forall\lambda \in (0, \vartheta], \quad \frac{|\psi^*(\lambda) - \psi^*(0)|}{\min\limits_{\lambda\in(0,\vartheta]} |\psi^*(\lambda)|} \le \mu\lambda^\beta \tag{4.3}$$

with $\psi^*(\lambda) = (1 - \mathrm{e}^{\mathrm{i}\lambda})^d\psi(\lambda)$, where d is the index of regular variation of ψ. We define $f^* = |\psi^*|^2$. This class is quite general and includes the impulse response of FARIMA filters (see [6] and the references therein) with $\beta = 2$ and $\vartheta = \pi$, but also processes whose spectral density may exhibit singularity outside the zero frequency, such as the Gegenbauer processes. As seen below, under local-to-zero assumptions, the validity of the Edgeworth expansion can only be established for the DFT coefficients in a degenerating neighborhood of zero frequency. This is enough for, say, semi-parametric estimation of the long-memory index by the GPH method.

Global assumptions. In some situations, it is possible to formulate regularity assumptions over the full frequency range $[-\pi, \pi]$ or a subset of it. These assumptions allow one to prove the validity of the Edgeworth expansion for $\mathcal{O}(n)$ frequency ordinates. We say that the sequence (ψ_j) belongs to the set $\mathcal{F}_{\text{global}}(\vartheta, \beta, \delta, \Delta, \mu)$ if $(\psi_j) \in \mathcal{F}_{\text{local}}(\vartheta, \beta, \delta, \Delta, \mu)$ and if in addition, for all $(\lambda, \lambda') \in (0, \vartheta] \times (0, \vartheta]$,

$$\left|\psi^*(\lambda) - \psi^*(\lambda')\right| \le \mu\frac{|\psi^*(\lambda)| \vee |\psi^*(\lambda')|}{|\lambda| \wedge |\lambda'|}|\lambda - \lambda'|. \tag{4.4}$$

Under those assumptions and as in the short-memory case, we are able to prove the validity of the Edgeworth expansion for the DFT's (Theorem 6) and deduce some moment bounds (Corollaries 7, 9 and 10). In comparison with short memory results, note that tapering ($r > 0$) and (**A1**) with $p' \ge s$ are required.

Theorem 6. *Assume* (**A1**) *with some integer* $s \ge 3$, $p \ge 1$ *and* $p' \ge s$. *Let* r *be a positive integer and* $\beta, \delta, \Delta, \mu, \vartheta$ *be constants such that* $0 < \delta < 1/2$, $-r + 1/2 < \Delta \le 0$, $\mu > 0$ *and*

$\vartheta \in (0, \pi]$. *Let* $(m_n)_{n\geq 0}$ *be a non-decreasing sequence. Assume either*

$$(\psi_j)_{j\in\mathbb{Z}} \in \mathcal{F}_{\text{local}}(\vartheta, \beta, \delta, \Delta, \mu) \quad \text{and} \quad \lim_{n\to\infty}\left(\frac{1}{m_n} + \frac{m_n}{n}\right) = 0 \tag{4.5}$$

or

$$(\psi_j)_{j\in\mathbb{Z}} \in \mathcal{F}_{\text{global}}(\vartheta, \beta, \delta, \Delta, \mu) \quad \text{and} \quad m_n \leq \vartheta \tilde{n}. \tag{4.6}$$

Then there exist a constant C *and positive integers* K_0, N_0 *which depend only on* ϑ, β, δ, Δ, μ, *the distribution of* Z_1 *and the sequence* (m_n), *such that for any* $n \geq N_0$ *and* $\mathbf{k} = (k_1, \ldots, k_u)$ *of integers in the range* $\{K_0, \ldots, m_n\}$, *the distribution of* $\mathbf{S}_n(\mathbf{k})$ *has a density* $q_{n,\mathbf{k}}$ *with respect to Lebesgue measure on* $\mathbb{R}^{2u}$ *which satisfies*

$$\sup_{\mathbf{x}\in\mathbb{R}^{2u}} (1 + \|\mathbf{x}\|^s) \left| q_{n,\mathbf{k}}(\mathbf{x}) - \sum_{r=0}^{s-3} P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\}) \right| \leq C n^{-(s-2)/2}. \tag{4.7}$$

If $u = 1$, *one can take* $K_0 = 1$.

Integrating some function g against the density $q_{n,\mathbf{k}}$ and using (4.7) yields the following corollary.

Corollary 7. *Under the assumptions of Theorem 6, there exists a constant* C *and an integer* N *depending only on* ϑ, β, δ, Δ, μ, u, r *and such that, for all* u*-tuple of distinct integers* $\mathbf{k}$ *satisfying* $K_0 \leq \min(\mathbf{k})$, $\max(\mathbf{k}) \leq m_n$ *and any* $n \geq N$, *and all measurable function* g *such that* $N_s(g) < \infty$,

$$\left| \mathbb{E}[g(\mathbf{S}_n(\mathbf{k}))] - \sum_{r=0}^{s-3} \int_{\mathbb{R}^{2u}} g(\mathbf{x}) P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\})\, \mathrm{d}\mathbf{x} \right| \leq C\, N_s(g)\, n^{-(s-2)/2}. \tag{4.8}$$

Similarly to the short-memory case, one could approximate $\mathbb{E}[g(\mathbf{S}_n(\mathbf{k}))]$ using the limiting distribution of $\mathbf{S}_n(\mathbf{k})$ in place of the Gaussian approximation as Corollary 7. Under long-range dependence and for fixed $\mathbf{k}$, the limiting covariance matrix of $\mathbf{S}_n(\mathbf{k})$ fully depends on $\mathbf{k}$ and not only on $(k_2 - k_1, \ldots, k_u - k_{u-1})$. This behavior at "very-low frequencies" as been studied for instance by Hurvich and Beltrao [15]. However, one can control the covariance of the standardized DFT coefficients and then the difference $\mathbf{V}_n(\mathbf{k}) - \mathbf{V}(\mathbf{k})$ thanks to the following lemma.

Lemma 8. *For* $1 \leq k \leq j \leq \vartheta n/\pi - r$ *and* $r \geq 1$, *there exists a constant* C *depending only on* $\vartheta, \beta, \delta, \Delta, \mu$ *such that*

$$\left|\mathbb{E}(\omega_{r,n,k}\omega_{r,n,j})\right| + \left|\mathbb{E}(\omega_{r,n,k}\bar{\omega}_{r,n,j}) - \varsigma_r(k-j)\right| \leq C p(k, j, n, \beta) \tag{4.9}$$

with

$$p(k, j, n, \beta) = \begin{cases} (jk)^{-1/2} + \left(\dfrac{j \vee k}{n}\right)^{\beta} & \text{under (4.5)} \\ (jk)^{-1/2} & \text{under (4.6).} \end{cases} \tag{4.10}$$

This Lemma is adapted from [20] to fit our need of uniformity of the bounds with respect to the function ψ whether it belongs to $\mathcal{F}_{\text{global}}$ or $\mathcal{F}_{\text{local}}$ only. For sake of brevity, its proof is omitted and we refer the interested reader to their paper.

Thus, we can develop the moments around the Gaussian distribution with covariance matrix $\mathbf{V}(\mathbf{k})$ as in the short-memory context. The two following corollaries prove sufficient for our applications. The next corollary is useful for moment bounds on one frequency k.

Corollary 9. *Under the assumptions of Theorem 6, there exist a constant C and a positive integer N_0 which depend only on ϑ, β, δ, Δ, μ, the distribution of Z_1 and the sequence (m_n), such that for any $n \geq N_0$, for any integer k in the range $\{1, \dots, m_n\}$ and any measurable function g on $\mathbb{R}^4$ such that $N_3(g) < \infty$*

$$\left|\mathbb{E}\left[g(\mathbf{S}_n(k))\right] - \int_{\mathbb{R}^2} g(\mathbf{x})\varphi_{\mathbf{I}_2/2}(\mathbf{x})\mathrm{d}\mathbf{x}\right| \leq C\left\{n^{-1/2}N_3(g) + p(k,k,n,\beta)^{-\tau(g,\mathbf{V}(k))/2}\|g(\mathbf{x})\|_{\mathbf{I}_{2u}}\right\}.$$

The next corollary is useful for moment bounds on two frequencies $k < j - r$.

Corollary 10. *Under the assumptions of Theorem 6, there exist a constant C and positive integers $K_1 \geq K_0$, N_0 which depends only on ϑ, β, δ, Δ, μ, the distribution of Z_1 and the sequence (m_n), such that for any $n \geq N_0$ and for any couple $\mathbf{k} = (k, j)$ of integers in the range $\{K_0, \dots, m_n\}$ such that $k < j - r$ and any measurable function g on $\mathbb{R}^4$ verifying (3.8) and such that $N_4(g) < \infty$*

$$\left|\mathbb{E}\left[g(\mathbf{S}_n(\mathbf{k}))\right] - \int_{\mathbb{R}^4} g(\mathbf{x})\varphi_{\mathbf{I}_4/2}(\mathbf{x})\mathrm{d}\mathbf{x}\right| \leq C\left\{n^{-(s-2)/2}N_s(g) + n^{-1/2}p^2(k,j,n,\beta)\|(1+\|\mathbf{x}\|^s)g(\mathbf{x})\|_{\mathbf{I}_4/2}\right\}.$$

4.2. *GPH estimation of the memory parameter*

4.2.1. *Theoretical results*

A very widely used estimator of the memory parameter d was introduced by Geweke and Porter-Hudak [10]. It is obtained from the linear regression of the log-periodogram of the observations using the logarithm of the frequencies as an explanatory variable. In contrast with the Whittle estimator, the GPH is defined explicitly in terms of the log-periodogram ordinates. Much theoretical work has been achieved on this estimator, in stationary or non-stationary contexts (see [7] for a survey of the main results). For instance, Giraitis et al. [12] proved that the GPH of Gaussian X is rate optimal for the quadratic risk and over some classes of spectral densities that are included in our $\mathcal{F}_{\mathrm{local}}$. To compute the risk of the GPH estimator, one needs to compute or approximate moments of the log-periodogram. The log-periodogram is a non-smooth function of the Fourier transform of the observation, which are Gaussian if X is Gaussian. The proof by [12] relies on moment bounds of a non-linear function of Gaussian variables (see [1, 27]); this technique does not extend naturally to non-Gaussian time series. Here, we shall apply the Edgeworth approximations obtained in the preceding section to extend this result to the case of a strong sense linear process.

The next results and the numerical simulation consider a modified version of the GPH estimator. Indeed, as we deal with non-Gaussian series, we need an Edgworth expansion machinery to approach the moments of non-functionals of the DFT's (using Corollaries 9 and 10). Tapering is needed to control the covariance of those DFT's. But tapering is also known to mix adjacent

Fourier frequencies. For the sake of simplicity of exposition, we only consider a taper of order $r = 1$ and write $I_k = I_{1,n,k}$. Skipping every second Fourier frequency allows one to recover some whitening in the periodogram ordinates: the quantities $(I_j/f(\lambda_j), I_k/f(\lambda_k))$, $j < k-1$ are asymptotically uncorrelated for non-degenerating frequencies. It follows that non-diagonal terms in the expansion of the variance of the GPH estimator do not contribute to the MSE.

The classical GPH estimator is obtained by an ordinary least square regression of $\log(I_k)$ on $\log|2\sin(\lambda_k/2)|$ (see [10,25]). With the frequency spacing and taper order r, one regresses on every $r+1$ frequency. For $r = 1$

$$(\hat{d}_m, \hat{C}) \stackrel{\text{def}}{=} \arg\min_{d',C'} \sum_{k=1}^{m} \left\{\log(I_{2k-1}) + 2d' \log|2\sin(\lambda_{2k-1}/2)| - C'\right\}^2$$

where $m = m(n)$ is a bandwidth parameter. Explicitly

$$\hat{d}_m = s_m^{-2} \sum_{k=1}^{m} \nu_k \log(I_{2k-1}), \tag{4.11}$$

with $\nu_k = -2\Big(\log|2\sin(\lambda_{2k-1}/2)| - \frac{1}{m}\sum_{j=1}^{m} \log|2\sin(\lambda_{2j-1}/2)|\Big)$ and $s_m^2 = \sum_{k=1}^{m} \nu_k^2$. We consider $\mathbb{E}[(\hat{d}_m - d)^2]$ the mean square error (MSE) of the GPH estimator. Theorem 11 gives a bound on the MSE which is uniform over a class of long or negative memory linear processes, from which rate optimality can be deduced. As we take a taper of order 1, it covers any range of memory parameter included in $(-1/2, 1/2)$.

Theorem 11. *Under the assumptions of Theorem 6 with $r = 1$, $s \geq 5$, $-1/2 < \delta < \Delta < 1/2$ and conditions (4.5), there exists a constant C which depends only on β, δ, Δ, ϑ, μ and the distribution of Z_1 such that*

$$\mathbb{E}[(\hat{d}_m - d)^2] \leq C \left\{\left(\frac{m}{n}\right)^{2\beta} + \frac{1}{m}\right\}.$$

With m proportional to $n^{2\beta/(2\beta+1)}$, $\mathbb{E}[(\hat{d}_m - d)^2] \leq C n^{-2\beta/(2\beta+1)}$.

Remark. In comparison to this uniform mean-square convergence, the central limit theorem holds for $d \in (-1/2, 1/2)$ with the classical GPH with no tapering (see [25]).

Remark. The condition $s \geq 5$ seems slightly stronger than necessary for bounding the MSE of $\hat{d}$. The reason for this is technical, as it allows the function $h(x_1, \ldots, x_4) = g(x_1, x_2)g(x_3, x_4)$ with $g(\mathbf{x}) = \log(\|\mathbf{x}\|^2) - \bar{\eta}$ to have finite $N_s(h)$ norm (see Corollary 4 and the remark that follows).

Remark. The FEXP estimator is an alternative method for inferring the memory parameter. It is based on a smooth expansion of the regular part of the log-spectral density that holds on some non-degenerating neighborhood $[-\vartheta, \vartheta]$ of the null frequency. It is called a global semi-parametric estimator (see [23]), in contrast with the GPH or local Whittle estimators, which are local semi-parametric estimators. To prove mean-square convergence of the FEXP estimator in the linear case using the results presented here, one needs to verify the hypotheses of Theorem 6 under global conditions (4.6).

Assuming more stringent global conditions on the regularity of the spectral density allows one to evaluate the bias term in the decomposition of the mean squared error. For comparison, we shall use the same set of hypotheses as [16]. Let $\mathcal{F}_0$ the set of the spectral densities f such that $f(\lambda) = |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{-2d} f^*(\lambda)$ where $-1/2 < d < 1/2$, where f^* is even, positive and continuous on $[-\pi, \pi]$ with first derivative $f^{*\prime}(0) = 0$ and second and third derivatives $f^{*\prime\prime}$ and $f^{*\prime\prime\prime}$ bounded in a neighborhood of zero.

Corollary 12. *Assume that the spectral density f of the process X is in $\mathcal{F}_0$. Assume that the sequence $m = m(n)$ is such that $\lim_{n\to\infty} 1/m + m\log(m)/n = 0$. Then*

$$\mathbb{E}(d - \hat{d})^2 = \frac{16\pi^4}{81}\left\{\frac{f^{*\prime\prime}(0)}{f^*(0)}\right\}^2 \frac{m^4}{n^4} + \frac{\pi^2}{24m} + o\left(\frac{m^4}{n^4}\right) + o\left(\frac{1}{m}\right) + O\left\{\frac{m(\log^3 m)}{n^2} + \frac{m^2\log m}{n^{5/2}}\right\} \tag{4.12}$$

uniformly with respect to $f \in \mathcal{F}_0$.

Remark. The stronger assumption $m\log(m)/n \to 0$ (compared to $m/n \to 0$ in previous Theorem) is needed to control the bias term as given by Lemma 1 in [16]. The remainder term $O(m^2 \log m/n^{5/2})$ of the MSE expression comes from the Edgeworth technique and is not present in [16, Theorem 1]. As long as $m = Cn^\gamma$, the whole $O()$ remainder term is negligible compared to the first two terms of the right-hand side of (4.12) for any $\gamma \in (3/4, 5/6)$. In the Gaussian case, $\gamma \in (2/3, 1)$ is enough. In any cases, this remainder term does not affect the MSE if one chooses the optimal value $\gamma = 4/5$.

4.2.2. Monte Carlo results

It follows from Corollary 12 that the MSE of the GPH estimator is asymptotically insensitive to the distribution of the innovation as soon as this distribution satisfies some moment and regularity conditions. We illustrate this statement by a Monte Carlo study. For sample sizes $n = 256, 512, 1024, 2048$ and 4096, we have simulated a thousand realizations of a FARIMA$(1, d, 0)$ process defined by

$$(1 - B)^d (1 - 0.3B) X_t = Z_t$$

where B is the back-shift operator and $(Z_t)_{t\in\mathbb{Z}}$ is a zero mean unit variance i.i.d sequence with the following marginal distributions

(a) Gaussian;
(b) Laplacian;
(c) zero-mean (shifted) Pareto, with $\mathbb{P}(Z_0 \le u) = (1 - (u + 13/6)^{-7})\mathbf{1}_{u\ge -7/6}$.

We first consider taking $d = 0.3, 0, -0.3$. Whereas it is possible to simulate exactly a Gaussian FARIMA(p, d, q) process (*e.g.* computing the covariance structure and using Levinson–Durbin algorithm), there is no general way to do it for non-Gaussian processes. In this Monte-Carlo experiment, the process (X_t) is obtained using either a truncated MA(∞) (if $d < 0$) or a truncated AR(∞) (if $d > 0$) representation, so that the series involved are absolutely summable. For each realization of each process, we evaluate the squared error $(\hat{d}_m - d)$ and define the Monte Carlo MSE as the average of those errors. Although we have not considered the issue of the data-driven choice of the bandwidth, we have focused on the sensitivity with respect to the distribution of Z of the bandwidth m which is optimal in the MSE sense. Fig. 1 illustrates the fact that both the optimal bandwidth and minimal MSE are poorly sensitive to the shock distribution. Similar

Table 1
Optimal bandwidth, variance, bias and MSE for processes (a), (b) and (c) and different sample sizes n. Here $d = -0.3$.

	Innovations	With taper and frequency skipping				Classical GPH			
		$m_n^\star$	Variance	Bias	MSE	$m_n^\star$	Variance	Bias	MSE
	Gaussian	27	0.020	0.104	0.030	43	0.013	0.081	0.020
$n = 256$	Laplace	23	0.025	0.087	0.033	36	0.014	0.072	0.019
	Pareto	27	0.021	0.109	0.033	37	0.014	0.074	0.019
	Gaussian	40	0.012	0.065	0.016	70	0.0074	0.0502	0.0099
$n = 512$	Laplace	44	0.011	0.074	0.017	75	0.0063	0.0629	0.0102
	Pareto	41	0.012	0.070	0.017	70	0.0066	0.0577	0.0099
	Gaussian	74	0.0063	0.0461	0.0084	122	0.0041	0.0393	0.0057
$n = 1024$	Laplace	69	0.0068	0.0497	0.0093	113	0.0044	0.0352	0.0057
	Pareto	70	0.0066	0.0518	0.0092	123	0.0037	0.0429	0.0055
	Gaussian	101	0.0043	0.0263	0.0050	193	0.0023	0.0250	0.0029
$n = 2048$	Laplace	121	0.0039	0.0356	0.0052	200	0.0021	0.0313	0.0031
	Pareto	121	0.0039	0.0327	0.0050	189	0.0022	0.0257	0.0029
	Gaussian	214	0.0019	0.0287	0.0027	332	0.0013	0.0195	0.0016
$n = 4096$	Laplace	224	0.0021	0.0267	0.0028	359	0.0013	0.0213	0.0017
	Pareto	196	0.0022	0.0228	0.0027	326	0.0013	0.0183	0.0016

Table 2
Optimal bandwidth, variance, bias and MSE for processes (a), (b) and (c) and different sample sizes n. Here $d = 0.3$.

	Innovations	With taper and frequency skipping				Classical GPH			
		$m_n^\star$	Variance	Bias	MSE	$m_n^\star$	Variance	Bias	MSE
	Gaussian	26	0.02	0.10	0.03	42	0.012	0.073	0.018
$n = 256$	Laplace	25	0.022	0.107	0.033	42	0.011	0.086	0.018
	Pareto	27	0.018	0.108	0.030	38	0.014	0.069	0.019
	Gaussian	39	0.012	0.075	0.018	64	0.008	0.051	0.011
$n = 512$	Laplace	38	0.013	0.063	0.017	70	0.007	0.054	0.010
	Pareto	43	0.011	0.070	0.016	74	0.0065	0.0595	0.0101
	Gaussian	70	0.0071	0.0520	0.0098	122	0.0045	0.0404	0.0062
$n = 1024$	Laplace	77	0.0060	0.0518	0.0087	120	0.0041	0.0362	0.0054
	Pareto	69	0.0058	0.0543	0.0088	122	0.0040	0.0407	0.0057
	Gaussian	110	0.0043	0.0348	0.0055	185	0.0025	0.0285	0.0033
$n = 2048$	Laplace	122	0.0034	0.0367	0.0048	184	0.0024	0.0234	0.0029
	Pareto	128	0.0036	0.0403	0.0053	188	0.0024	0.0245	0.0030
	Gaussian	189	0.0022	0.0257	0.0028	332	0.0013	0.0218	0.0018
$n = 4096$	Laplace	209	0.0021	0.0275	0.0029	351	0.0012	0.0207	0.0016
	Pareto	198	0.0023	0.0266	0.0031	335	0.0013	0.0199	0.0016

conclusions can be drawn from Tables 1–3. In those tables we display the value of the bias, the variance, and of the mean square error of the GPH at the estimated optimal bandwidth.

For sake of comparison, we have run the more classical GPH with no frequency skipping and no data taper (results are displayed on the right part of the three tables). Though we do not provide a theoretical result similar to Theorem 11 in this case, it seems that this estimator outperforms the GPH with frequency skipping. It can be well understood that both the variance and the bias are smaller for the classical GPH in the following way. Taper induces some bias at low frequency because it mixes adjacent raw periodogram ordinates. This breaks the power-law

Table 3
Optimal bandwidth, variance, bias and MSE for processes (a), (b) and (c) and different sample sizes n. Here $d = 0$.

	Innovations	With taper and frequency skipping				Classical GPH			
		$m_n^\star$	Variance	Bias	MSE	$m_n^\star$	Variance	Bias	MSE
$n = 256$	Gaussian	27	0.020	0.102	0.030	44	0.012	0.073	0.017
	Laplace	27	0.018	0.109	0.030	42	0.011	0.080	0.018
	Pareto	23	0.022	0.085	0.029	45	0.011	0.078	0.018
$n = 512$	Gaussian	43	0.012	0.068	0.017	77	0.0059	0.0571	0.0092
	Laplace	42	0.011	0.074	0.016	75	0.0063	0.0599	0.0098
	Pareto	40	0.011	0.062	0.015	73	0.0064	0.0510	0.0090
$n = 1024$	Gaussian	72	0.0062	0.0490	0.0086	133	0.0035	0.0408	0.0052
	Laplace	75	0.0062	0.0545	0.0091	112	0.0044	0.0340	0.0055
	Pareto	72	0.0061	0.0482	0.0085	125	0.0037	0.0365	0.0050
$n = 2048$	Gaussian	113	0.004	0.031	0.005	203	0.0022	0.0246	0.0028
	Laplace	119	0.0039	0.0358	0.0052	210	0.0022	0.0289	0.0030
	Pareto	108	0.0037	0.0302	0.0046	193	0.0023	0.0252	0.0030
$n = 4096$	Gaussian	192	0.0023	0.0237	0.0029	351	0.0012	0.0195	0.0016
	Laplace	184	0.0024	0.0194	0.0028	348	0.0012	0.0197	0.0016
	Pareto	173	0.0025	0.0209	0.0030	324	0.0013	0.0186	0.0017

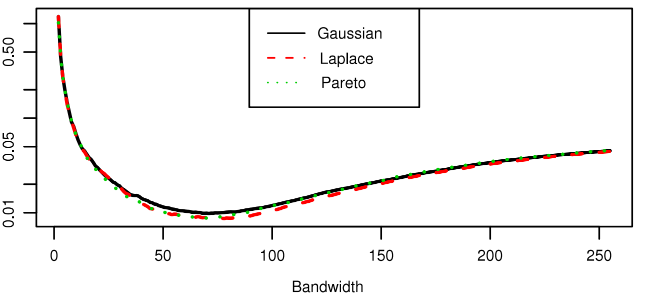

Fig. 1. Comparisons of the MSE versus the bandwidth for the FARIMA processes (a), (b) and (c) with $d = 0.3$. Sample size $n = 1024$.

dependence between spectral density and frequencies. The fact that we drop out every second frequency induces some loss in the variance by a factor 2, even if the periodogram ordinates are not independent. Most arguably, tapering of order 1 for $d \in (-1/2, 1/2)$ may be thought as a condition stronger than what is strictly necessary. However, for lower values of d, tapering does help and enables consistent estimation of d, whereas the classical GPH is highly and positively biased. We have run the same simulation for $d = -0.8$ with a taper order equal to 2. Fig. 3 illustrates the wrong behavior of the classical GPH in this case.

Fig. 2 represents the box-and-whiskers plot of the GPH estimator for two different sample sizes and the three models we are concerned with. Here again, the sensitivity with respect to the distribution of the driving noise is hardly discernible.

Acknowledgements

I am very grateful and indebted to Eric Moulines and Philippe Soulier for their help and the many fruitful discussions we had on this subject some years ago. The two anonymous referees are thanked for their constructive comments.

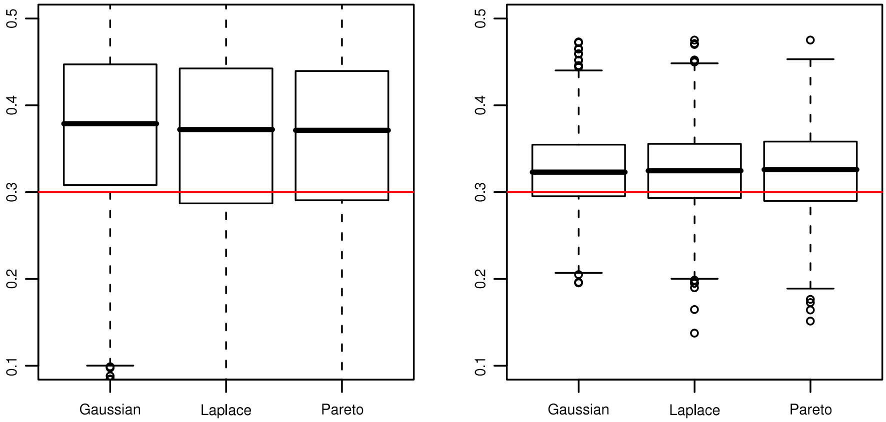

Fig. 2. Box-plot of the GPH estimator for processes (a), (b) and (c) with $d = 0.3$, sample size $n = 512$ (left panel) and $n = 4096$ (right panel).

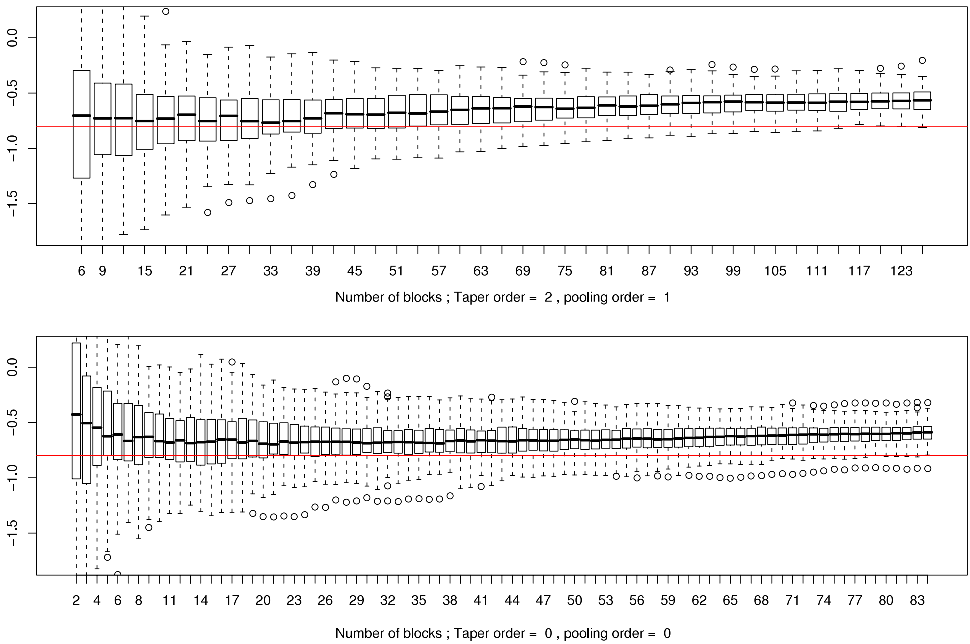

Fig. 3. With $d < 0.5$, the classical GPH (bottom panel) is biased, whereas the modified GPH (top panel) with taper order equal to 3 and skipping of two frequencies out of three is mean-square convergent (see Theorem 11). Sample size is $n = 256$, innovations are Laplacian random variables.

Appendix A. Edgeworth expansion for triangular arrays

In this section we recall the theorem established in [9]. Let $(Z_t)_{t\in\mathbb{Z}}$ be an i.i.d sequence and $(\mathbf{U}_{n,j})_{j\in\mathbb{Z},n\in\mathbb{N}}$ an array of vectors in $\mathbb{R}^u$, where u is an integer. Define $\mathbf{S}_n = \sum_{j\in\mathbb{Z}} \mathbf{U}_{n,j} Z_j$ and let $\mathbf{V}_n = \sum_{j\in\mathbb{Z}} \mathbf{U}_{n,j} \mathbf{U}'_{n,j}$. For $\mathbf{v} \in \mathbb{N}^u$, $2 \le |\mathbf{v}| \le s$, denote $\chi_{n,\mathbf{v}}$ the cumulants of $\mathbf{S}_n$. Then

$\chi_{n,\mathbf{v}} = \kappa_{|\mathbf{v}|} \sum_{j\in\mathbb{Z}} \mathbf{U}^{\mathbf{v}}_{n,j}$. where κ_r denotes the r-th cumulant of Z_1, $r \le s$. Consider the following assumptions.

(**B1**) There exist positive constants v_* and v^* such that

$$v_* \le \liminf_n v_{\min}[\mathbf{V}_n] \le \limsup_n v_{\max}[\mathbf{V}_n] \le v^*$$

where $v_{\min}[\mathbf{V}_n]$ (resp. $v_{\max}[\mathbf{V}_n]$) is the smallest (resp. the largest) eigenvalue of $\mathbf{V}_n$.

(**B2**) There exist positive constants η, c_0, a sequence $(M_n)_{n\in\mathbb{N}}$ of positive numbers, and a sequence $(J_n)_{n\in\mathbb{N}}$ of subsets of $\mathbb{Z}$, such that, for all $n \ge 0$

$$\sup_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}\| \le M_n \tag{A.1}$$

$$\lim_{n\to\infty} M_n = 0 \tag{A.2}$$

$$\operatorname{card}(J_n) \le c_0 M_n^{-2} \quad \text{and} \quad \frac{\sum_{j\in J_n} \|\mathbf{U}_{n,j}\|^2}{\sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}\|^2} \ge \eta. \tag{A.3}$$

(**B3**) There exist $\zeta \ge 1$ and a sequence $(M_n)_{n\in\mathbb{N}}$ satisfying (A.1) such that

$$\sup_{n\ge 0} M_n^{\zeta} \sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}\| < \infty.$$

Theorem 13 (*[9]*). *Let $s \ge 3$, and $p' \ge 0$ be integers and $p \ge 1$ be a real number. Assume* (**A1**) (s, p, p'), (**B1**) *and* (**B2**). *Assume in addition either* (**B3**) *or $p' \ge s$ in* (**A1**) (s, p, p'). *Then, there exist a constant C and an integer N (depending only on the distribution of Z_1, and the constants appearing in the assumptions) such that, for all $n \ge N$, the distribution of $\mathbf{S}_n$ has a density q_n with respect to Lebesgue measure on $\mathbb{R}^u$ which satisfies*

$$\sup_{\mathbf{x}\in\mathbb{R}^u} (1 + \|\mathbf{x}\|^s) \left| q_n(\mathbf{x}) - \sum_{r=0}^{s-3} P_r(\mathbf{x}, \mathbf{V}_n, \{\chi_{n,\mathbf{v}}\}) \right| \le C \sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}\|^s. \tag{A.4}$$

Appendix B. Proof of Theorem 1

The proof consists in checking that assumptions (**B1**), (**B2**) and (**B3**) hold uniformly with respect to $\psi \in \mathcal{G}(\alpha, \beta, \delta)$ and $\mathbf{k}$ for $\mathbf{U}_{n,j}$'s of the form (2.9). To prove (**B1**), write $\mathbf{V}_n(\mathbf{k}) = \mathbf{V}(\mathbf{k}) + \mathbf{W}_n(\mathbf{k})$, with $\mathbf{V}(\mathbf{k})$ defined in (3.5). Define $\|\mathbf{W}\|_1 = \max_{1\le i\le v} \sum_{j=1}^{v} |w_{i,j}|$ for any matrix $\mathbf{W} = (w_{i,j})_{1\le i,j\le v}$. Similarly to Hannan [14, p. 54], we have under (3.1)

$$\|\mathbf{W}_n(\mathbf{k})\|_1 \le C(\alpha, \beta) n^{-1}. \tag{B.1}$$

The matrices $\mathbf{V}(\mathbf{k})$ have the following algebraic property.

Lemma 14. *There exist two positive constants v_* and v^* such that*

$$2v_* \le \inf v_{\min}[\mathbf{V}(\mathbf{k})] \le \sup v_{\max}[\mathbf{V}(\mathbf{k})] \le 2v^* \tag{B.2}$$

where the infimum and supremum are taken over all the u-tuples of distinct integers in $\mathbb{N}^u$.

Proof. Noting that $\mathrm{trace}[\mathbf{V}(\mathbf{k})] = u$,

$$v_{\max}[\mathbf{V}(\mathbf{k})] \leq \mathrm{trace}[\mathbf{V}(\mathbf{k})]/2u = 1/2. \tag{B.3}$$

Take $v^* = 1/4$. Recall that $k_1 < \cdots < k_u$. Note that, for any $n \geq 2k_u + 2r + 1$, $\mathbf{V}(\mathbf{k})$ is the covariance matrix of

$$\sqrt{2\pi}(c^Y_{r,n,k_1}, s^Y_{r,n,k_1}, \ldots, c^Y_{r,n,k_u}, s^Y_{r,n,k_u})$$

with $c^Y_{r,n,k} = (2\pi a_r n)^{-1/2}\sum_{t=1}^n h^r_{t,n} Y_t \cos(t\lambda_k)$ and $s^Y_{r,n,k} = (2\pi a_r n)^{-1/2}\sum_{t=1}^n h^r_{t,n} Y_t \sin(t\lambda_k)$ the sine and cosine transform of a unit-variance zero-mean Gaussian white noise $(Y_n)_{n\in\mathbb{Z}}$. Recall that

$$c^Y_{r,n,k} = a_r^{-1/2}\sum_{l=0}^{r}(-1)^l\binom{r}{l}c^Y_{0,n,k+l} \quad \text{and} \quad s^Y_{r,n,k} = a_r^{-1/2}\sum_{l=0}^{r}(-1)^l\binom{r}{l}s^Y_{0,n,k+l}. \tag{B.4}$$

The random variables $c_{0,n,k}$ and $s_{0,n,k}$, $k = 1, \ldots, [(n-1)/2]$ are centered i.i.d Gaussian with variance $1/4\pi$. Assume that $\mathbf{V}(\mathbf{k})$ is not invertible. It yields that for some $2u$-tuple of reals $(\alpha_1, \beta_1, \ldots, \alpha_u, \beta_u) \neq (0, 0, \ldots, 0, 0)$,

$$\sum_{j=1}^{u}(\alpha_j c_{r,n,k_j} + \beta_j s_{r,n,k_j}) \overset{\mathbb{L}^2}{=} 0.$$

Then by (B.4), there exists a linear combination of $c_{0,n,k}$'s and $s_{0,n,k}$'s that is equal to zero. c_{0,n,k_u+r} and s_{0,n,k_u+r} appear in this combination with coefficients $a_r^{-1/2}(-1)^r\alpha_u$ and $a_r^{-1/2}(-1)^r\beta_u$, respectively. It follows from the independence and non-degeneracy of the $c_{0,n,k}$'s and $s_{0,n,k}$'s that $\alpha_u = \beta_u = 0$. Iterating the argument yields the contradiction $\alpha_u = \beta_u = \alpha_{u-1} = \beta_{u-1} = \cdots = \alpha_1 = \beta_1 = 0$. Thus for any u-tuple $\mathbf{k}$ of distinct integers

$$v_{\min}[\mathbf{V}(\mathbf{k})] > 0. \tag{B.5}$$

It remains to prove that $v_{\min}[\mathbf{V}(\mathbf{k})]$ is bounded away from zero uniformly in $\mathbf{k}$. Define

$$K_u = \{\mathbf{k} = (k_1, \ldots, k_{u'}) \in \mathbb{N}^{u'}, 1 \leq u' \leq u, 0 < k_{i+1} - k_i \leq r\}.$$

Note now that by (3.5), $v_{\min}[\mathbf{V}(\mathbf{k})]$ is a function of the vector $(k_2 - k_1, k_3 - k_2, \ldots, k_u - k_{u-1})$, thus taking finitely many different values on K_u. From the this remark and (B.5),

$$v_1 \overset{\text{def}}{=} \inf_{\mathbf{k}\in K_u} v_{\min}[\mathbf{V}(\mathbf{k})] > 0 \tag{B.6}$$

since the infimum is taken on a finite set of positive values. Consider now a u-tuple $\mathbf{k}$ that does not belong to K_u; In this case, for some $i \in \{1, \ldots, u-1\}$, $k_{i+1} - k_i > r$, and then $\mathbf{k}$ may be partitioned as $L \geq 2$ blocks of indexes $(\mathbf{k}_1, \ldots, \mathbf{k}_L)$ such that all the $\mathbf{k}_i$'s belong to K_u and, for all $i \in \{1, \ldots, L-1\}$, $\min \mathbf{k}_{i+1} - \max \mathbf{k}_i > r$. Let l_i denote the length of the block $\mathbf{k}_i$, $i = 1, \ldots, L$. By this construction and (3.5), the matrix $\mathbf{V}(\mathbf{k})$ has a block-diagonal structure

$$\mathbf{V}(\mathbf{k}) = \begin{pmatrix} \mathbf{V}(\mathbf{k}_1) & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{V}(\mathbf{k}_L) \end{pmatrix}.$$

Using (B.6),

$$v_{\min}[\mathbf{V}(\mathbf{k})](v_{\max}[\mathbf{V}(\mathbf{k})])^{2u-1} \geq \det[\mathbf{V}(\mathbf{k})] = \prod_{i=1}^{L} \det[\mathbf{V}(\mathbf{k}_i)] \geq \prod_{i=1}^{L} v_1^{2l_i} = v_1^{2u}. \tag{B.7}$$

We conclude from (B.7) and (B.3) that

$$v_{\min}[\mathbf{V}(\mathbf{k})] \geq v_1^{2u} 2^{2u-1} =: v_2 > 0. \tag{B.8}$$

(B.2) follows from (B.6) and (B.8) with $v_* = \frac{1}{2}\min(v_1, v_2)$. □

Proof of Theorem 1. By (B.1) and Lemma 14, (**B1**) holds with v_* and v^* of Lemma 14, for some N_0, $n \geq N_0$ and uniformly in $\mathbf{k}$, α and β. With (B.3),

$$\left| \sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2 - u \right| = |\mathrm{trace}[\mathbf{V}_n(\mathbf{k})] - u| \leq C(\alpha, \beta) n^{-1/2}. \tag{B.9}$$

Prove now that (**B2**) is verified. Since f is bounded away from zero and $\sum_{j\in\mathbb{Z}} |\psi_j| \leq \beta < \infty$, (A.1) and (A.2) are verified with $M_n \stackrel{\text{def}}{=} C(r)\beta\alpha^{-1/2} n^{-1/2}$. Put $J_n = \{j, |j| < 2n\}$. Then $\mathrm{card}(J_n) \leq c_0 M_n^{-2}$ for some c_0 depending only on r, α, β and

$$\begin{aligned} \sum_{|j|\in\mathbb{Z}\setminus J_n} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2 &\leq C(r)\alpha^{-2} n^{-1} \sum_{|j|\geq 2n} \left(\sum_{t=1}^{n} |\psi_{t+j}| \right)^2 \\ &\leq C(r)\alpha^{-2} \sum_{|j|\geq 2n} \sum_{t=1}^{n} \psi_{t+j}^2 \leq C(r)\alpha^{-2} \sum_{|j|\geq n} |j|\psi_j^2. \end{aligned} \tag{B.10}$$

Under (3.1), $|\psi_j| \leq \beta |j|^{-1/2-\delta}$ so that

$$\sum_{|j|\geq n} |j|\psi_j^2 \leq \beta n^{-2\delta} \sum_{|j|\geq n} |j|^{1/2+\delta} |\psi_j| \leq \beta^2 n^{-2\delta}. \tag{B.11}$$

For any $\epsilon > 0$ and large enough n, $\sum_{|j|\geq n} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2 \leq \epsilon$ uniformly in $\mathbf{k}$ and $\psi \in \mathcal{G}(\alpha, \beta, \delta)$. (A.3) follows from (B.9)–(B.11). Finally,

$$\begin{aligned} \sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\| &\leq C(r)\alpha^{-1/2} n^{-1/2} \sum_{t=1}^{n} \sum_{j\in\mathbb{Z}} |\psi_{t+j}| = C(r)\alpha^{-1/2} n^{1/2} \sum_{j\in\mathbb{Z}} |\psi_j| \\ &\leq C(r)\alpha^{-1/2}\beta n^{1/2} = C(r)^2 \alpha^{-1}\beta^2 M_n^{-1} \end{aligned}$$

so that (**B3**) holds with $\zeta = 1$. □

Appendix C. Proof of Theorem 6

The proof of Theorem 6 consists in checking that assumptions (**B1**), (**B2**) hold uniformly.

Lemma 15. *There exist integers N_0, K_0, and $v_* > 0$, $v^* > 0$ (depending only on ϑ, β, δ, Δ, μ) such that, for all $n \geq N_0$, we have,*

(1) *for all u-tuple $\mathbf{k}$ of distinct integers, $1 \leq \min \mathbf{k} \leq \max \mathbf{k} \leq m_n$,*

$$v_{\max}[\mathbf{V}_n(\mathbf{k})] \leq v^*; \tag{C.1}$$

(2) *for all integer k,* $1 \le k \le m_n$

$$v_* \le v_{\min}[\mathbf{V}_n(k)]; \tag{C.2}$$

(3) *for all u-tuple* $\mathbf{k}$ *of distinct integers,* $K_0 \le \min \mathbf{k} \le \max \mathbf{k} \le m_n$,

$$v_* \le v_{\min}[\mathbf{V}_n(\mathbf{k})]. \tag{C.3}$$

Proof. As in Appendix B, we put $\mathbf{V}_n(\mathbf{k}) = \mathbf{V}(\mathbf{k}) + \mathbf{W}_n(\mathbf{k})$ where $\mathbf{V}(\mathbf{k})$ is defined in (3.5). Applying Lemma 8, we obtain

$$\|\mathbf{W}_n(\mathbf{k})\|_1 \le C(\vartheta, \beta, \delta, \Delta, \mu) \begin{cases} \dfrac{1}{k_1} + \left(\dfrac{m_n}{n}\right)^{\beta} \log\left(\dfrac{m_n}{n}\right) & \text{under (4.5)}, \\ \dfrac{1}{k_1} & \text{under (4.6)}. \end{cases} \tag{C.4}$$

(C.1) follows immediately. The proof of (C.3) follows by picking N_0, K_0 large enough. For (C.2), it remains to prove that for any integer k, $1 \le k \le K_0$, $\mathbf{V}_n(k)$ converges to a positive definite matrix $\widetilde{\mathbf{V}}(k)$ and that this convergence is uniform w.r.t to ψ, for $\psi \in \mathcal{F}_{\text{local}}(\vartheta, \beta, \delta, \Delta, \mu)$ or $\psi \in \mathcal{F}_{\text{global}}(\vartheta, \beta, \delta, \Delta, \mu)$. What follows is an adaptation of Iouditsky et al. [18, Lemma 7.3]. Write

$$\mathbb{E}[|\omega_{r,n,k}|^2] = \frac{1}{f(\lambda_k)} \left(\int_{|\lambda| \le \vartheta\pi} + \int_{|\lambda| > \vartheta\pi} \right) |D_{r,n}(\lambda - \lambda_k)|^2 f(\lambda) \mathrm{d}\lambda =: A_1 + A_2 \tag{C.5}$$

where $D_{r,n}$ is defined in (2.2). For $n \ge 4\pi K_0/\vartheta$, $1 \le k \le K_0$ and $|\lambda| \ge \vartheta\pi$, $|n(\lambda - \lambda_k)| \ge n\vartheta/2$. Using (2.3) and (4.1), we get

$$A_2 \le \frac{C\lambda_k^{2d}}{\lambda_k^{2d} f(\lambda_k)} n^{-2r-1} \int_{|\lambda| > \vartheta\pi} \lambda^{2d} f(\lambda) \mathrm{d}\lambda \le Cn^{-2r}. \tag{C.6}$$

By change of variable,

$$A_1 = \frac{n^{2d}|1 - \mathrm{e}^{-\mathrm{i}\lambda_k}|^{2d}}{f^*(\lambda_k)} \int_{|\lambda| \le n\vartheta} \left| n^{-1/2} D_{r,n}(\lambda/n - \lambda_k) \right|^2 n^{-2d} |1 - \mathrm{e}^{-\mathrm{i}\lambda/n}|^{-2d} f^*\left(\frac{\lambda}{n}\right) \mathrm{d}\lambda.$$

Write $\lim_{n\to\infty} n^{-1/2} D_{r,n}(\lambda/n) = \frac{1}{\sqrt{2\pi a_r}} \int_0^1 (1 - \mathrm{e}^{2\mathrm{i}\pi s})^r \mathrm{e}^{-\mathrm{i}s\lambda} \mathrm{d}s =: \hat{h}_r(\lambda)$. By Riemann approximation, it can be seen that $|n^{-1/2} D_{r,n}(\lambda/n) - \hat{h}_r(\lambda)| \le C(1 + |\lambda|)/n$. Note also that $|\hat{h}_r(\lambda)| \le C|\lambda|^{-r-1}$. Then

$$\begin{aligned} &\left| A_1 - \frac{n^{2d}|1 - \mathrm{e}^{\mathrm{i}\lambda_k}|^{2d}}{f^*(\lambda_k)} \times \int_{-n\vartheta}^{n\vartheta} \left| \hat{h}_r(\lambda - 2\pi k) \right|^2 n^{-2d} |1 - \mathrm{e}^{\mathrm{i}\lambda/n}|^{-2d} f^*\left(\frac{\lambda}{n}\right) \mathrm{d}\lambda \right| \\ &\quad \le Ck^{2d} n^{-r} \le Cn^{-r}. \end{aligned} \tag{C.7}$$

Here and in the following, C is a generic constant which depends only on $\vartheta, \beta, \delta, \Delta, \mu, r$ and K_0. For $|\lambda| \le n\vartheta$, using (4.5),

$$\begin{aligned} &\frac{f^*(0)}{f^*(\lambda_k)} \left| n^{-2d} |1 - \mathrm{e}^{\mathrm{i}\lambda/n}|^{-2d} - |\lambda|^{-2d} \right| + \frac{|f^*(\frac{\lambda}{n}) - f^*(0)|}{f^*(\lambda_k)} n^{-2d} |1 - \mathrm{e}^{\mathrm{i}\lambda/n}|^{-2d} \\ &\quad \le C \left\{ \left| n^{-2d} |1 - \mathrm{e}^{\mathrm{i}\lambda/n}|^{-2d} - |\lambda|^{-2d} \right| + n^{-2d} |1 - \mathrm{e}^{\mathrm{i}\lambda/n}|^{-2d} \left| \frac{\lambda}{n} \right|^{\beta'} \right\} \end{aligned} \tag{C.8}$$

with $\beta' = \beta \wedge 1$. For $x \in [-\pi, \pi]$, $\frac{2}{\pi}|x| \le |e^{ix} - 1| = |2 \sin \frac{x}{2}| \le |x|$ and $\|e^{ix} - 1| - |x\| \le x^2/2$. Also, for any $\upsilon \in \mathbb{R}$, $x > 0$, $y > 0$, $|x^\upsilon - y^\upsilon| \le |\upsilon|(x^{\upsilon-1} \vee y^{\upsilon-1})|x - y|$. Using those relations, write, for $\lambda \in [-n\pi, n\pi]$,

$$\left|n^{-2d}|1 - e^{i\lambda/n}|^{-2d} - |\lambda|^{-2d}\right| \le Cn^{-2d}\left|\frac{\lambda}{n}\right|^{-2d-1}\left||1 - e^{i\lambda/n}| - |\frac{\lambda}{n}|\right| \le Cn^{-1}|\lambda|^{-2d+1}.$$

Then

$$\begin{aligned}&\int_{-n\vartheta}^{n\vartheta} |\hat{h}_r(2\pi k - \lambda)|^2 \left|n^{-2d}|1 - e^{i\lambda/n}|^{-2d} - |\lambda|^{-2d}\right| d\lambda \\ &\le Cn^{-1}\int_{-n\vartheta}^{n\vartheta} |\hat{h}_r(2\pi k - \lambda)|^2|\lambda|^{-2d+1}d\lambda \\ &\le Cn^{-1}\int_{-n\vartheta}^{n\vartheta} |\hat{h}_r(2\pi k - \lambda)|^2|\lambda|^{2r}d\lambda \le Cn^{-1} \qquad (C.9)\end{aligned}$$

and

$$\begin{aligned}&\int_{-n\vartheta}^{n\vartheta} |\hat{h}_r(2\pi k - \lambda)|^2 n^{-2d}|1 - e^{i\lambda/n}|^{-2d}\left|\frac{\lambda}{n}\right|^{\beta'} d\lambda \\ &\le n^{-\beta'}\int_{-n\vartheta}^{n\vartheta} |\hat{h}_r(2\pi k - \lambda)|^2|\lambda|^{-2d+\beta'}d\lambda \\ &\le Cn^{-\beta'}. \qquad (C.10)\end{aligned}$$

Gathering (C.5), (C.6), (C.7), (C.8), (C.9), (C.10) yields

$$\left|\mathbb{E}[|\omega_{r,n,k}|^2] - \frac{(2\pi)^{2d}k^{2d}f^*(0)}{f^*(\lambda_k)}\int_{\infty}^{+\infty}\left|\hat{h}_r(\lambda - 2\pi k)\right|^2|\lambda|^{-2d}d\lambda\right| \le Cn^{-\beta'}.$$

Similar arguments leads to

$$\left|\mathbb{E}[\omega_{r,n,k}^2] - \frac{(2\pi)^{2d}k^{2d}f^*(0)}{f^*(\lambda_k)}\int_{\infty}^{+\infty}\hat{h}_r(\lambda - 2\pi k)\hat{h}_r(\lambda + 2\pi k)|\lambda|^{-2d}d\lambda\right| \le Cn^{-\beta'}.$$

Defining the scalar product $(u, v)_d = \int_{\mathbb{R}} u(\lambda)v(\lambda)|\lambda|^{-2d}d\lambda$, Then $\det \mathbf{V}_n(k)$ is uniformly approximated by the Gram determinant of the functions $\hat{h}_r(\lambda - 2k\pi)$ and $\hat{h}_r(\lambda + 2k\pi)$ associated with the product $(\cdot, \cdot)_d$ and then is a continuous function of $\eta_k(d) := \lim_{n\to\infty}\mathbb{E}[|\omega_{r,n,k}|^2]$ and $\eta'_k(d) := \lim_{n\to\infty}\mathbb{E}[\omega_{r,n,k}^2]$. The whole set of functions $\hat{h}_r(\lambda + 2j\pi)$, $j \in \mathbb{Z}$ is linearly independent, so that those determinant are positive. Using continuity of η_k and η'_k w.r.t. d, the infimum on the compact set $[-\Delta, \delta]$ and the minimum over $k = 1, \ldots, K_0$ is positive too, which concludes the proof. □

Lemma 16. *There exists a constant C (depending only on ϑ, β, δ, Δ, μ, r) such that for all $k \in \{1, \ldots, \tilde{n}\}$,*

$$\frac{1}{\sqrt{nf(\lambda_k)}}\left|\sum_{t=1}^{n} h_{t,n}^r \psi_{t+j} e^{it\lambda_k}\right| \le Cn^{-1/2}. \qquad (C.11)$$

Proof. The main tool of the proof is the bound (2.3) and the technique is the same as the one used in the proof of Lemma 8. Decompose

$$|\psi(\lambda_k)|^{-1}\frac{1}{\sqrt{2\pi a_r n}}\sum_{t=1}^{n} h_{t,n}^r \psi_{t-j} e^{it\lambda_k} = |\psi(\lambda_k)|^{-1}\int_{-\pi}^{\pi}\psi(\lambda)e^{ij\lambda}D_{n,r}(\lambda_k-\lambda)d\lambda$$

into

$$A_1 = |\psi(\lambda_k)|^{-1}\left(\int_{-\pi}^{-\vartheta}+\int_{\vartheta}^{\pi}\right)\psi(\lambda)e^{ij\lambda}D_{n,r}(\lambda_k-\lambda)d\lambda,$$

$$A_2 = |\psi(\lambda_k)|^{-1}\psi^*(0)\int_{-\vartheta}^{\vartheta}(1-e^{i\lambda})^{-d}e^{ij\lambda}D_{n,r}(\lambda_k-\lambda)d\lambda,$$

$$A_3 = |\psi(\lambda_k)|^{-1}\int_{-\vartheta}^{\vartheta}(1-e^{i\lambda})^{-d}(\psi^*(\lambda)-\psi^*(0))e^{ij\lambda}D_{n,r}(\lambda_k-\lambda)d\lambda.$$

By Eq. (2.3), if $|\lambda| \in [\vartheta,\pi]$, $|D_{n,r}(\lambda_k-\lambda)| \le Cn^{-1/2-r}$. Note that $n^{-1}\lambda_k^d = n^{-1}\lambda_k^{-1}\lambda_k^{d+1} \le 1/(2\pi k)$. (4.1) implies that $|A_1| \le Cn^{1/2-r}k^{-1} \le Cn^{-1/2}$. Consider A_2. Since $\int_{-\pi}^{\pi} D_{n,r}(\lambda)d\lambda = 0$,

$$A_2 = \int_{-\vartheta}^{\vartheta}\Delta(\lambda,\lambda_k)D_{n,r}(\lambda_k-\lambda)d\lambda,$$

$$\Delta(\lambda,\lambda_k) = \left((1-e^{i\lambda})^{-d}-(1-e^{i\lambda_k})^{-d}\right)e^{ij\lambda}|\psi(\lambda_k)|^{-1}.$$

Decompose this integral on the intervals $[-\vartheta,-\lambda_k/2]$, $[-\lambda_k/2,\lambda_k/2]$, $[\lambda_k/2,2\lambda_k]$ and $[2\lambda_k,\vartheta]$. If $\lambda \in [-\lambda_k/2,\lambda_k/2]$, then $|D_{n,r}(\lambda_k-\lambda)| \le C\sqrt{n}k^{-r-1}$ and $|\Delta(\lambda,\lambda_k)| \le C\left(|\lambda|^{-d}\lambda_k^d+1\right)$. Hence:

$$\left|\int_{-\lambda_k/2}^{\lambda_k/2}\Delta(\lambda,\lambda_k)D_{n,r}(\lambda_k-\lambda)d\lambda\right| \le Ck^{-r}n^{-1/2}.$$

If $\lambda \in [\lambda_k/2, 2\lambda_k]$, then $|\Delta(\lambda,\lambda_k)| \le C\left(\lambda_k^{-1}|\lambda-\lambda_k|+1\right)$. Since $\int_{-\lambda_k/2}^{\lambda_k}(1+n|\lambda|)^{-r-1}d\lambda \le Cn^{-1}$, we have

$$\left|\int_{\lambda_k/2}^{2\lambda_k}\Delta(\lambda,\lambda_k)\ D_{n,r}(\lambda_k-\lambda)d\lambda\right| \le Cn^{-1/2}.$$

If $\lambda \in [2\lambda_k,\vartheta]$ (and similarly on $[-\vartheta,-\lambda_k/2]$), we use that $|\Delta(\lambda,\lambda_k)| \le C(\lambda^{-d}\lambda_k^d+1)$ and $|D_{n,r}(\lambda-\lambda_k)| \le n^{-1/2-r}|\lambda-\lambda_k|^{-1-r}$. Hence,

$$\left|\int_{2\lambda_k}^{\vartheta}\Delta(\lambda,\lambda_k)D_{n,r}(\lambda_k-\lambda)d\lambda\right| \le Cn^{-1/2-r}\int_{\lambda_k}^{\infty}\left(\lambda^d\lambda_k^{-d}+1\right)\lambda^{-1-r}d\lambda \le Cn^{-1/2}k^{-r}.$$

Consider A_3. Under (4.3), we have

$$|A_3| \le C\lambda_k^d\int_{-\vartheta}^{\vartheta}|\lambda|^{-d+\beta}|D_{n,r}(\lambda-\lambda_k)|d\lambda.$$

Decompose this integral as above. If $\lambda \in [-\lambda_k/2,\lambda_k/2]$, proceeding as above:

$$\lambda_k^d\int_{-\lambda_k/2}^{\lambda_k/2}|\lambda|^{-d+\beta}|D_{n,r}(\lambda-\lambda_k)|d\lambda \le Cn^{-1/2}k^{-r}\lambda_k^{\beta}.$$

If $\lambda_k \in [\lambda_k/2, 2\lambda_k]$, $\lambda_k^d |\lambda|^{-d+\beta} \leq C\lambda_k^\beta$, and $\int_{\lambda_k/2}^{2\lambda_k} |D_{n,r}(\lambda - \lambda_k)| \mathrm{d}\lambda \leq Cn^{-1/2}$. Hence:

$$\lambda_k^d \int_{\lambda_k/2}^{2\lambda_k} |\lambda|^{-d+\beta} |D_{n,r}(\lambda - \lambda_k)| \mathrm{d}\lambda \leq Cn^{-1/2} \lambda_k^\beta.$$

Finally, if $\lambda \in [2\lambda_k, \vartheta]$ (and similarly, if $\lambda \in [-\vartheta, -\lambda_k/2]$), we have as above:

$$\lambda_k^d \int_{2\lambda_k}^{\vartheta} \lambda^{-d+\beta} \lambda^{-1-r} \mathrm{d}\lambda \leq C\lambda_k^d n^{-1/2-r} \int_{\lambda_k}^{\vartheta} \lambda^{-d-1-r} \mathrm{d}\lambda = Cn^{-1/2} k^{-r}. \quad \square$$

Lemma 17. *There exists a constant C (depending only on ϑ, β, δ, Δ, μ, r) such that for all $k \in \{1, \dots, \tilde{n}\}$,*

$$\frac{1}{\sqrt{nf(\lambda_k)}} \left| \sum_{t=1}^{n} h_{t,n}^r \psi_{t+j} \mathrm{e}^{\mathrm{i}t\lambda_k} \right| \leq Cn^{-1/2}\, \lambda_k^{d-1} (1 + |j|)^{d-1} \leq Cn^{-1/2} ((1 + |j|)/n)^{d-1}. \tag{C.12}$$

Proof. By applying the definition of the weights $h_{t,n}^r$ and summation by parts, we have:

$$\sum_{t=1}^{n} h_{t,n}^r \psi_{t+j} \mathrm{e}^{\mathrm{i}t\lambda} = \sum_{p=0}^{r} (-1)^p \binom{r}{p} \sum_{t=1}^{n} \psi_{t+j} \mathrm{e}^{\mathrm{i}t(\lambda+\lambda_p)},$$

$$\sum_{t=1}^{n} \psi_{t+j} \mathrm{e}^{\mathrm{i}t\lambda_k} = \sum_{t=1}^{n-1} \left\{ \left(\sum_{u=1}^{t} \mathrm{e}^{\mathrm{i}u\lambda_k} \right) (\psi_{t+j} - \psi_{t+j+1}) + \left(\sum_{u=1}^{n} \mathrm{e}^{\mathrm{i}u\lambda_k} \right) \psi_{n+j} \right\}.$$

For all $y \in (0, \pi)$ and all $\ell \in \mathbb{N}^*$, $\left| \sum_{u=1}^{\ell} \mathrm{e}^{\mathrm{i}uy} \right| \leq 2/y$. The proof follows from condition (4.2). $\square$

Proceed now with the proof of Theorem 6. If $|j| \geq n$, then $((1 + |j|)/n)^{d-1} \leq 1$. Hence by Lemma 16, for some constant C which depends only on β, δ, Δ, ϑ, μ, r and the distribution of Z_1,

$$\forall\, j, n, \mathbf{k}, \quad M_{n,j} \stackrel{\text{def}}{=} Cn^{-1/2} \left(1 \wedge ((1 + |j|)/n)^{\delta-1} \right) \geq \|\mathbf{U}_{n,j}(\mathbf{k})\|.$$

Note that

$$M_n \stackrel{\text{def}}{=} \sup_{j \in \mathbb{Z}} M_{n,j} = Cn^{-1/2}.$$

Then (A.1) and (A.2) hold uniformly in $\mathbf{k}$. By Lemma 15, Eq. (C.3) or (C.2), we have

$$\sum_j \|\mathbf{U}_{n,j}(\mathbf{k})\|^2 = \mathrm{trace}[\mathbf{V}_n(\mathbf{k})] \geq v_* > 0.$$

Finally, define for any $\gamma \geq 1$ the set $J_n = \{j \in \mathbb{Z}, |j| \leq \gamma n\}$. Then $\mathrm{card}(J_n) \leq c_0 M_n^{-2}$ and

$$\frac{\sum_{j \in \mathbb{Z} \setminus J_n} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2}{\sum_{j \in \mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2} \leq \frac{\sum_{j \in \mathbb{Z} \setminus J_n} M_{n,j}^2}{\sum_{j \in \mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2} \leq C(v_*)^{-1} n^{1-2\delta} \sum_{|j| \geq \gamma n} j^{2\delta-2} \leq C(v_*)^{-1} \gamma^{2\delta-1}.$$

Choosing γ large enough yields (A.3) uniformly.

Appendix D. Proofs of Corollaries 3, 4, 9 and 10

Proof of Corollary 3. By the triangle inequality, the LHS of inequality (3.7) is bounded by

$$\left| \mathbb{E}\left[g(\mathbf{S}_n(\mathbf{k})) \right] - \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \varphi_{\mathbf{V}_n(\mathbf{k})}(\mathbf{x})\, \mathrm{d}\mathbf{x} \right| + \left| \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \left\{ \varphi_{\mathbf{V}_n(\mathbf{k})}(\mathbf{x}) - \varphi_{\mathbf{V}(\mathbf{k})}(\mathbf{x}) \right\} \mathrm{d}\mathbf{x} \right|.$$

By Corollary 2 with $s = 3$, the first term of the previous display is bounded by $Cn^{-1/2} N_3(g)$. For A a matrix, denote $\rho(A)$ its spectral radius. Denote $\mathbf{I}_a$ the a-dimensional identity matrix. To bound the second term, note that $\rho(\mathbf{V}_n(\mathbf{k}) - \mathbf{V}(\mathbf{k})) \le C(\alpha, \beta) n^{-1}$ by (B.1) and that $\tau(g, \mathbf{V}(\mathbf{k})) \ge 1$ by definition, then apply the following lemma which is an easy adaptation of Soulier [27, Theorem 2.1]. □

Lemma 18. *Let Γ be a u-dimensional positive matrix. There exists $\epsilon > 0$ and a constant C such that, for all symmetric positive matrix Γ' verifying $\rho(\Gamma'^{-1} - \Gamma^{-1}) < \epsilon$, and for all measurable functions g on $\mathbb{R}^u$ satisfying $\|g\|_\Gamma^2 < \infty$, we have*

$$\left| \int_{\mathbb{R}^u} g(\mathbf{x}) \left\{ \varphi_{\Gamma'}(\mathbf{x}) - \varphi_{\Gamma}(\mathbf{x}) \right\} \mathrm{d}\mathbf{x} \right| \le C \rho^{\tau(g,\Gamma)/2}(\Gamma' - \Gamma) \|g\|_\Gamma.$$

Proof of Corollary 4. The LHS of (3.9) is bounded by $A_1 + A_2 + A_3 + A_4$ with

$$A_1 = \left| \mathbb{E}\left[g(\mathbf{S}_n(\mathbf{k})) \right] - \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \sum_{r=0}^{s-3} P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\})\, \mathrm{d}\mathbf{x} \right|,$$

$$A_2 = \left| \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \left\{ \varphi_{\mathbf{V}_n(\mathbf{k})}(\mathbf{x}) - \varphi_{\mathbf{I}_{2u}/2}(\mathbf{x}) \right\} \mathrm{d}\mathbf{x} \right|,$$

$$A_3 = \left| \int_{\mathbb{R}^{2u}} g(\mathbf{x}) P_1(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\})\, \mathrm{d}\mathbf{x} \right|,$$

$$A_4 = \left| \int_{\mathbb{R}^{2u}} g(\mathbf{x}) \sum_{r=2}^{s-3} P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\})\, \mathrm{d}\mathbf{x} \right|,$$

$A_4 = 0$ if $s = 4$. Using (3.8), we get $\tau(g, \mathbf{I}_{2u}/2) = 2$. It follows, as in the proof of Corollary 3 that A_2 is bounded by $Cn^{-1}\|g\|_{\mathbf{V}(\mathbf{k})}$, whereas A_1 is bounded by $Cn^{-(s-2)/2} N_s(g)$. Write shortly $P_r(\mathbf{x}, \mathbf{V}_n(\mathbf{k}), \{\chi_{n,\mathbf{v}}(\mathbf{k})\}) = R_r(\mathbf{x}) \varphi_{\mathbf{V}_n(\mathbf{k})}$, where R_r is a polynomial of order $r+2$ (the dependence w.r.t $\mathbf{V}_n(\mathbf{k})$ and $\{\chi_{n,\mathbf{v}}(\mathbf{k})\}$ is omitted in this notation). Note also that

$$\begin{aligned} |\chi_{n,\mathbf{v}}(\mathbf{k})| &\le |\kappa_{|\mathbf{v}|}| \sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\|^{|\mathbf{v}|} \le |\kappa_{|\mathbf{v}|}| M_n^{|\mathbf{v}|-2} \left(\sum_{j\in\mathbb{Z}} \|\mathbf{U}_{n,j}(\mathbf{k})\|^2 \right) \\ &\le |\kappa_{|\mathbf{v}|}| M_n^{|\mathbf{v}|-2} \mathrm{trace}[\mathbf{V}_n(\mathbf{k})] \le C |\kappa_{|\mathbf{v}|}| M_n^{|\mathbf{v}|-2} \end{aligned} \tag{D.1}$$

where $M_n \le C(\alpha, \beta) n^{-1/2}$. Then, the coefficients of R_r are $O(n^{-(r/2)})$ uniformly in $\mathbf{k}$ and ψ since they involve $\chi_{n,\mathbf{v}}(\mathbf{k})$'s with $|\mathbf{v}| = r$ and elements of $\mathbf{V}_n^{-1}(\mathbf{k})$ (for details, see [3]). Let $\mathbf{F}_n(\mathbf{k}) \stackrel{\text{def}}{=} (\mathbf{V}_n^{-1}(\mathbf{k}) - \mathbf{V}^{-1}(\mathbf{k}))/2$ and write

$$\begin{aligned} &\int g(\mathbf{x}) \varphi_{\mathbf{V}_n(\mathbf{k})}(\mathbf{x}) R_r(\mathbf{x})\, \mathrm{d}\mathbf{x} \\ &\quad = \left| \frac{\det \mathbf{V}(\mathbf{k})}{\det \mathbf{V}_n(\mathbf{k})} \right|^{1/2} \int g(\mathbf{x}) R_r(\mathbf{x}) \exp\{-\mathbf{x}' \mathbf{F}_n(\mathbf{k}) \mathbf{x}\} \varphi_{\mathbf{V}(k)}(\mathbf{x})\, \mathrm{d}\mathbf{x}. \end{aligned} \tag{D.2}$$

By (B.1), $\|\mathbf{F}_n(\mathbf{k})\|_1 \le Cn^{-1}$ and $|\det(\mathbf{V}_n(\mathbf{k}))^{-1/2} - \det(\mathbf{I}_{2u}/2)^{-1/2}| \le Cn^{-1}$ uniformly so that $A_4 \le Cn^{-1}\|(1+\|\mathbf{x}\|^s)g(\mathbf{x})\|_{2\mathbf{I}_{2u}/3}$. We can derive this way that $A_3 \le Cn^{-1/2}$, which is not enough. Improving this bound requires some care and uses the symmetries of g. Actually, R_1 is a sum of polynomials which are odd with respect to one or three components. Write

$$|\exp\{-\mathbf{x}'\mathbf{F}_n(\mathbf{k})\mathbf{x}\} - 1 + \mathbf{x}'\mathbf{F}_n(\mathbf{k})\mathbf{x}| \le Cn^{-2}\|\mathbf{x}\|^4 \exp\left\{Cn^{-1}\|\mathbf{x}\|^2\right\} \tag{D.3}$$

and notice that $\{1 - \mathbf{x}'\mathbf{F}_n(\mathbf{k})\mathbf{x}\}R_1(\mathbf{x})$ is a sum of polynomials of the form $\prod_i r_i(x_{2i-1}, x_{2i})$, each of them being odd with respect to at least one variable. Consider a typical term odd with respect to x_1, say. Using (3.8)

$$\int_{\mathbb{R}^{2u}} g(\mathbf{x}) \prod_i r_i(x_{2i-1}, x_{2i})\varphi_{\mathbf{I}_{2u}/2}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^2} g_1(x_1, x_2) r_1(x_1, x_2)\varphi_{\mathbf{I}_2/2}(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$
$$\times \int_{\mathbb{R}^{2u-2}} \prod_{i>1} g_i(x_{2i-1}, x_{2i}) r_i(x_{2i-1}, x_{2i})\varphi_{\mathbf{I}_{2u-2}/2}(\mathbf{x})\,\mathrm{d}x_3 \cdots \mathrm{d}x_{2u} = 0,$$

since the first integral vanishes. Hence, $\int_{\mathbb{R}^4} h(\mathbf{x})R_1(\mathbf{x})(\mathbf{x}'\mathbf{F}_n(\mathbf{k})\mathbf{x})\varphi_{\mathbf{I}_4}(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$. Gathering (D.1)–(D.3), $A_3 \le Cn^{-2}$. □

Proof of Corollaries 9 and 10. As those Corollaries are the counterparts of Corollaries 3 and 4 in a long memory context, we only give the necessary adaptations from the preceding proofs. From Lemma 8, $\rho(\mathbf{V}_n(k) - \mathbf{V}(k)) \le Cp(k, k, n, \beta)$, $\|\mathbf{F}_n(\mathbf{k})\|_1 \le Cp(k, j, n, \beta)$ and

$$|\det(\mathbf{V}_n(\mathbf{k}))^{-1/2} - \det(\mathbf{V}(\mathbf{k}))^{-1/2}| \le Cp(k, j, n, \beta).$$

The LHS of (D.3) is now bounded by $p^2(k, j, n, \beta)\|\mathbf{x}\|^4 \exp\left\{Cp(k, j, n, \beta)\|\mathbf{x}\|^2\right\}$. The term A_3 is then bounded by

$$Cn^{-1/2}p^2(k, j, n, \beta)\int_{\mathbb{R}^4} \|\mathbf{x}\|^5 h(\mathbf{x}) \exp\{-\|\mathbf{x}\|^2(1 + Cp(k, j, n, \beta))\}\,\mathrm{d}\mathbf{x}.$$

If $m = o(n)$ and $K_1 > 2C$, then for large enough n and $K_1 \le k < j - r \le m - r$, the integral is uniformly bounded. Thus $A_3 \le Cn^{-1/2}p^2(k, j, n, \beta)$ whereas $A_1 \le Cp^2(k, j, n, \beta)$. □

Appendix E. Proof of Theorem 11

In the sequel, C denotes a constant which depends only on β, δ, ϑ, μ and the distribution of Z_1 and whose value may change upon each appearance. Note first that $|\nu_k| = O(\log(k))$, $s_m^2 = 4m(1 + o(1))$ (see for instance [25], or [16]). Define $L(\lambda) = \log(f^*(\lambda)/f^*(0))$. Since $\psi \in \mathcal{F}(\vartheta, \beta, \delta, \Delta, \mu)$, there exists a constant C such that

$$\forall k \in \{1, \ldots, m\}, \quad |L(\lambda_k)| \le C|\lambda_k|^\beta. \tag{E.1}$$

Let $\bar{\eta}$ denote $\mathbb{E}(\log \|\mathbf{Y}\|^2)$, where $\mathbf{Y}$ is a centered Gaussian random vector with covariance matrix $\mathbf{I}_2/2$. Define $\eta_k = \log(I_k/f(\lambda_k)) - \bar{\eta}$, $1 \le k \le m$. With these notations and since $\sum_{k=1}^m \nu_k = 0$, (4.11) yields

$$\hat{d}_m = d + s_m^{-2}\sum_{k=1}^m \nu_k\eta_k + s_m^{-2}\sum_{k=1}^m \nu_k L(\lambda_k) =: d + W_m + b_m. \tag{E.2}$$

The mean-square error of the GPH writes $\mathbb{E}((\hat{d}_m - d)^2) = \mathbb{E}W_m^2 + 2b_m\mathbb{E}W_m + b_m^2$. Applying (E.1) and the Cauchy–Schwartz inequality,

$$|b_m| \le C s_m^{-2} \sum_{k=1}^{m} |\nu_k| \lambda_k^{\beta} \le C(m/n)^{\beta}. \tag{E.3}$$

Thus, to prove Theorem 11, we only need to show that $\mathbb{E}[W_m^2] \le Cm^{-1}$. We now compute $\mathbb{E}[W_m^2]$. Let $\ell = \ell(m)$ be a non-decreasing sequence of integers such that $1 \le \ell \le m$ and define $W_{1,m} = s_m^{-2} \sum_{k=1}^{\ell} \nu_k \eta_k$ and $W_{2,m} = W_m - W_{1,m}$. We first give a bound for $\mathbb{E}[W_{1,m}^2]$. Note that

$$\mathbb{E}[W_{1,m}^2] \le \ell s_m^{-4} \sum_{k=1}^{\ell} \nu_k^2 \mathbb{E}[\eta_k^2]. \tag{E.4}$$

For $\mathbf{x} \in \mathbb{R}^2$, define $g(\mathbf{x}) = \log(\|\mathbf{x}\|^2) - \bar{\eta}$. Then $\eta_k = g(\mathbf{S}_{n,k})$ and $N_3(g^2) < \infty$. For $(x_1, \ldots, x_4) \in \mathbb{R}^2$, define $h(x_1, \ldots, x_4) = g(x_1, x_2) g(x_3, x_4)$. Then $\eta_k \eta_j = h(\mathbf{S}_{n,(k,j)})$, h has property (3.8) and

$$N_5(h) = \int_{\mathbb{R}^4} \frac{g((x_1, x_2)) g((x_3, x_4))}{1 + \|\mathbf{x}\|^5} \le 4(N_{5/2}(g))^2 \, \mathrm{d}\mathbf{x} < \infty$$

where we have used $4(1 + (a^2 + b^2)^{s/2}) \ge (1 + |a|^{s/2})(1 + |b|^{s/2})$. Note that $N_4(h) = N_2(g) = +\infty$, which motivates the expansion up to order $s = 5$. Let $\sigma^2 \stackrel{\text{def}}{=} \operatorname{var}(\log \|\mathbf{Y}\|^2) = \pi^2/6$. Applying Corollaries 9 and 10 respectively to the functions g, h, we get for some integer l_0 and any k, j such that $l_0 \le k < j \le m$,

$$|\mathbb{E}[\eta_k^2] - \sigma^2| \le C(\beta, \delta, \vartheta, \mu) \left\{ k^{-1} + (k/n)^{\beta} + n^{-1/2} \right\} \tag{E.5}$$

$$\left| \mathbb{E}[\eta_k \eta_j] \right| \le C(\beta, \delta, \vartheta, \mu) \left\{ k^{-2} + (j/n)^{2\beta} + n^{-1} \right\}. \tag{E.6}$$

(E.4) and (E.5) yield $\mathbb{E}[W_{1,m}^2] \le C\ell^2 m^{-2}$. We now bound $\mathbb{E}[W_{2,m}^2]$:

$$\mathbb{E}[W_{2,m}^2] = s_m^{-4} \sum_{k=\ell+1}^{m} \nu_k^2 \mathbb{E}[\eta_k^2] + 2 s_m^{-4} \sum_{\ell < k < j \le m} \nu_k \nu_j \mathbb{E}[\eta_k \eta_j].$$

Using (E.5) and (E.6), we obtain

$$\begin{aligned} \left| \mathbb{E}[W_{2,m}^2] - s_m^{-2} \sigma^2 \right| &\le C(\beta, \delta, \vartheta, \mu) s_m^{-4} \sum_{k=\ell+1}^{m} \nu_k^2 \left(k^{-1} + (k/n)^{\beta} + n^{-1/2} \right) \\ &\quad + C(\beta, \delta, \vartheta, \mu) s_m^{-4} \sum_{\ell < k < j \le m} \nu_k \nu_j \left(k^{-2} + (j/n)^{2\beta} + n^{-1} \right) \\ &= C(\beta, \delta, \vartheta, \mu) s_m^{-2} \left\{ 1 + O\left(\ell^{-1/2} + m^{1/2} l^{-3/2} + m^{2\beta+1} n^{-2\beta} + m/n \right) \right\}. \end{aligned} \tag{E.7}$$

Choosing $\ell \le m$ such that $\ell^2 = o(m)$ and $m = o(\ell^3)$ (for instance $\ell = [m^{\eta}]$ with $1/3 < \eta < 1/2$) yields $\mathbb{E}[W_m^2] = O(m^{-1})$. This bound and (E.3) conclude the proof of Theorem 11.

Appendix F. Proof of Corollary 12

Hurvich et al. [16] have shown (4.12) in the Gaussian case. The deterministic “bias” term

$$b_m = s_m^{-2} \sum_{k=1}^{m} \nu_k \{\log(f^*(\lambda_k)) - \log(f^*(0))\}$$

is treated by their Lemma 1, up to some multiplicative constant due to the fact that we regress on every two frequencies. We have here

$$b_m = -\frac{4\pi^2}{9} \frac{f^{*''}(0)}{f^*(0)} \frac{m^2}{n^2} \{1 + o(1)\}. \tag{F.1}$$

We use the same decomposition (E.2) as in the proof of Theorem 11, but as the hypotheses of the Corollary implies the global assumptions (4.6), we apply Corollaries 9 and 10 with $p(k, j, n, \beta) \leq k^{-1}$. Then the bound (E.7) reduces to

$$\mathbb{E}[W_{2,m}^2] = \frac{\pi^2}{6 s_m^2} \left\{1 + O\left(\ell^{-1/2} + m^{1/2}\ell^{-3/2} + m/n\right)\right\}.$$

Choosing ℓ such that $\ell^2 = o(m)$ and $m = o(\ell^3)$, and using $s_m^2 = 4m(1 + o(1))$ yields the variance term of the MSE, namely

$$\mathbb{E}[W_m^2] = \frac{\pi^2}{24m}(1 + o(1)). \tag{F.2}$$

Treat now the cross term $2b_n\mathbb{E}[W_m]$. Denote by $\mathbb{E}^0[W_m]$ the expectation of W_m under the Gaussian assumption. Then (see [16] Lemma 8)

$$\mathbb{E}^0[W_m] = O\Big(\frac{\log^3 m}{m}\Big).$$

Using Corollary 9, we get $\left|\mathbb{E}[\eta_k] - \mathbb{E}^0[\eta_k]\right| \leq C(n^{-1/2} + k^{-1})$. It follows that

$$\mathbb{E}[W_m] = O\Big(\frac{\log^3 m}{m} + n^{-1/2}\log m\Big). \tag{F.3}$$

Corollary 12 follows from (F.1)–(F.3).

References

[1] M. Arcones, Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors, Ann. Probab. 22 (4) (1994) 2243–2274.
[2] M.S. Bartlett, An Introduction to Stochastic Processes, Cambridge University Press, 1955.
[3] R.N. Bhattacharya, R. Ranga Rao, Normal Approximation and Asymptotic Expansions, 1st edition, Wiley, 1976.
[4] Z.-G. Chen, E.J. Hannan, The distribution of periodogram ordinates, J. Time Ser. Anal. 1 (1980) 73–82.
[5] R.S. Deo, W.W. Chen, On the integral of the squared periodogram, Stoch. Proc. Appl. 85 (1) (2000) 159–176.
[6] P. Doukhan, G. Oppenheim, M.S. Taqqu (Eds.), Theory and Applications of Long-Range Dependence, Birkhäuser Boston Inc., Boston, MA, 2003.
[7] G. Faÿ, E. Moulines, F. Roueff, M. Taqqu, Estimators of long-memory: Fourier versus wavelets, J. Econometrics (2008) arXiv:0801.4329v1 [math.ST].
[8] G. Faÿ, E. Moulines, Ph. Soulier, Non linear functionals of the periodogram, J. Time Ser. Anal. 23 (5) (2002) 523–553.
[9] G. Faÿ, E. Moulines, Ph. Soulier, Edgeworth expansions for linear statistics of possibly long range dependent linear processes, Statist. Probab. Lett. 66 (3) (2004) 275–288.
[10] J. Geweke, S. Porter-Hudak, The estimation and application of long memory time series models, J. Time Ser. Anal. 4 (1983) 221–238.

[11] M.S. Ginovian, Asymptotically efficient nonparametric estimation of nonlinear spectral functionals, in: Proceedings of the Eighth Vilnius Conference on Probability Theory and Mathematical Statistics, Part I 2002, vol. 78, 2003, pp. 145–154.

[12] L. Giraitis, P. Robinson, A. Samarov, Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence, J. Time Ser. Anal. 18 (1997) 49–61.

[13] F. Götze, C. Hipp, Asymptotic expansions for sum of weakly dependent random vectors, Z. Wahrscheinlichkeitstheorie Verwandte Geb. 64 (1983) 211–239.

[14] E.J. Hannan, Time Series Analysis, in: Methuen's Monographs on Applied Probability and Statistics, Methuen& Co. Ltd, London, 1960.

[15] C.M. Hurvich, K.I. Beltrao, Asymptotics for the low-frequency ordinates of the periodogram of a long-memory time series, J. Time Ser. Anal. 14 (5) (1993) 455–472.

[16] C.M. Hurvich, R. Deo, J. Brodsky, The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series, J. Time Ser. Anal. 19 (1) (1998) 19–46.

[17] C.W. Hurvich, W.W. Chen, An efficient taper for potentially overdifferenced long-memory time series, J. Time Ser. Anal. 21 (2000) 155–180.

[18] A. Iouditsky, E. Moulines, Ph. Soulier, Adaptive estimation of the fractional differencing coefficient, Bernoulli 7 (5) (2001) 699–731.

[19] D. Janas, R. von Sachs, Consistency for non-linear functions of the periodogram of tapered data, J. Time Ser. Anal. 16 (1995) 585–606.

[20] G. Lang, P. Soulier, Empirical spectral process, Application to the estimation of long memory, 2002 (unpublished) URL: http://www.tsi.enst.fr/˜soulier/papers/uclt.ps.

[21] Y.H. Lee, S. Cho, W.C. Kim, B.U. Park, On estimating integrated squared spectral density derivatives, J. Statist. Plann. Inference 48 (2) (1995) 165–184.

[22] E. Moulines, Ph. Soulier, Broad band log-periodogram regression of time series with long range dependence, Ann. Statist. 27 (3) (1999) 1415–1539.

[23] E. Moulines, Ph. Soulier, Semiparametric spectral estimation for fractional processes, in: P. Doukhan, G. Oppenheim, M.S. Taqqu (Eds.), Theory and Applications of Long-range Dependence, Birkhäuser, 2003, pp. 527–556.

[24] P.M. Robinson, Gaussian semiparametric estimation of long range dependence, Ann. Statist. 24 (1995) 1630–1661.

[25] P.M. Robinson, Log-periodogram regression of time series with long range dependence, Ann. Statist. 23 (1995) 1043–1072.

[26] Katsumi Shimotsu, Peter C.B. Phillips, Pooled log periodogram regression, J. Time Ser. Anal. 23 (1) (2002) 57–93.

[27] Ph. Soulier, Moment bounds and central limit theorem for functions of Gaussian vectors, Statist. Probab. Lett. 54 (2) (2001) 193–203.

[28] M. Taniguchi, On estimation of parameters of Gaussian stationary processes, J. Appl. Probab. 16 (1979) 575–591.

[29] M. Taniguchi, On estimation of the integrals of certain functions of spectral density, J. Appl. Probab. 17 (1980) 73–83.

[30] M. Taniguchi, Higher Order Asymptotic Theory for Time Series Analysis, in: Lecture Notes in Statistics, vol. 68, Springer-Verlag, 1991.

[31] M.S. Taqqu, Law of the iterated logarithm for sums of nonlinear functions of Gaussian variables that exhibit long range dependence, Z. Wahrscheinlichkeitstheorie Verw. Geb. 40 (1977) 203–238.

[32] C. Velasco, Non-Gaussian log-periodogram regression, Econometric Theory 16 (2000) 44–79.

[33] R. von Sachs, Peak-insensitive non-parametric spectrum estimation, J. Time Ser. Anal. 15 (4) (1994) 453–474.