

# MONOTONE SPECTRAL DENSITY ESTIMATION

BY DRAGI ANEVSKI

*Lund University*

AND

BY PHILIPPE SOULIER

*Université Paris Ouest Nanterre*

We propose two estimators of a monotone spectral density, that are based on the periodogram. These are the isotonic regression of the periodogram and the isotonic regression of the log-periodogram. We derive pointwise limit distribution results for the proposed estimators for short memory linear processes and long memory Gaussian processes and also that the estimators are rate optimal.

**1. Introduction.** The motivation for doing spectral analysis of stationary time series comes from the need to analyze the frequency content in the signal. The frequency content can for instance be described by the spectral density, defined below, for the process. One could be interested in looking for a few dominant frequencies or frequency regions, which correspond to multimodality in the spectral density. Inference methods for multimodal spectral densities have been treated in [5], using the taut string method. A simpler problem is that of fitting a unimodal spectral density, i.e. the situation when there is only one dominant frequency, which can be known or unknown, corresponding to known or unknown mode, respectively, and leading to the problem of fitting a unimodal spectral density to the data. In this paper we treat unimodal spectral density estimation for known mode. A spectral density that is decreasing on  $[0, \pi]$  is a model for the frequency content in the signal being ordered. A unimodal spectral density is a model for there being one major frequency component, with a decreasing amount of other frequency components seen as a function of the distance to the major frequency.

Imposing monotonicity (or unimodality) means that one imposes a nonparametric approach, since the set of monotone (or unimodal) spectral densities is infinite-dimensional. A parametric problem that is contained in our estimation problem is that of a power law spectrum, i.e. when one assumes that the spectral density decreases as a power function  $f(u) \sim u^{-\beta}$  for  $u \in (0, \pi)$ , with unknown exponent  $\beta$ . Power law spectra seem to have important applications to physics, astronomy and medicine: four different application mentioned in [16] are a) fluctuations in the Earth's rate of rotation cf. [20], b) voltage fluctuations across cell membrane cf. [10], c) time series of impedances of rock layers in boreholes cf. e.g. [13] and d) x-ray time variability of galaxies cf. [17]. We propose to use a nonparametric approach as an alternative to the power law spectrum methods used in these applications. There are (at least) two reasons why this could make sense: Firstly, the reason for using a power function e.g. to model the spectrum in the background radiation is (at best) a theoretical consideration exploiting physical theory and leading to the power function as a good approximation. However, this is a stronger model assumption to impose on the data than merely imposing monotonicity and thus one could imagine a wider range of situations that should be possible to analyze using our methods. Secondly, fitting a power law spectral model to data consists of doing linear regression of the log periodogram; if the data are not very well aligned along a

straight line (after a log-transformation) this could influence the overall fit. A nonparametric approach, in which one assumes only monotonicity, is more robust against possible misfit.

Sometimes one assumes a piecewise power law spectrum, cf. [21], as a model. Our methods are well adapted to these situations when the overall function behaviour is that of a decreasing function.

Furthermore there seem to be instances in the litterature when a monotonically decreasing (or monotonically increasing) spectral density is both implicitly assumed as a model, and furthermore seems feasible: Two examples in [22] (cf. e.g. Figures 20 and 21 in [22]) are e) the wind speed in a certain direction at a certain location measured every 0.025 second (for which a decreasing spectral density seems to be feasible) and f) the daily record of how well an atomic clock keeps time on a day to day basis (which seems to exhibit an increasing spectral density). The methods utilized in [22] are smoothing of the periodogram. We propose to use an order-restricted estimator of the spectral density, and would like to claim that this is better adapted to the situations at hand.

Decreasing spectral densities can arise when one observes a sum of several parametric time series, for instance AR(1) processes with coefficient  $|a| < 1$ ; the interest of the non parametric method in that case is that one does not have to know how many AR(1) are summed up. Another parametric example is an ARFIMA(0,d,0) with  $0 < d < 1/2$ , which has a decreasing spectral density, which is observed with added white noise, or even with added one (or several) AR(1) processes; the resulting time series will have a decreasing spectral density. Our methods are well adapted to this situation, and we will illustrate the nonparametric methods on simulated data from such parametric models.

The spectral measure of a weakly stationary process is the positive measure  $\sigma$  on  $[-\pi, \pi]$  characterized by the relation

$$\text{cov}(X_0, X_k) = \int_{-\pi}^{\pi} e^{ikx} \sigma(dx) .$$

The spectral density, when it exists, is the density of  $\sigma$  with respect to Lebesgue's measure. It is an even nonnegative integrable function on  $[-\pi, \pi]$ . Define the spectral distribution function on  $[-\pi, \pi]$  by

$$\begin{aligned} F(\lambda) &= \int_0^{\lambda} f(u) du , \quad 0 \leq \lambda \leq \pi , \\ F(\lambda) &= -F(-\lambda) , \quad -\pi \leq \lambda < 0 . \end{aligned}$$

An estimate of the spectral density is given by the periodogram

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{k=1}^n X_k e^{-ik\lambda} \right|^2 .$$

The spectral distribution function is estimated by the empirical spectral distribution function

$$F_n(\lambda) = \int_0^{\lambda} I_n(u) du .$$

Functional central limit theorems for  $F_n$  have been established in [4] and [18]. However, since the derivative is not a smooth map, the properties of  $F_n$  do not transfer to  $I_n$ , and furthermore

it is well known that the periodogram is not even a consistent estimate of the spectral density. The standard remedy for obtaining consistency is to use kernel smoothers. This however entails a bandwidth choice, which is somewhat ad hoc. The assumption of monotonicity allows for the construction of adaptive estimators that do not need a pre-specified bandwidth.

We will restrict our attention to the class of non increasing functions.

DEFINITION 1. *Let  $\mathcal{F}$  be the convex cone of integrable, monotone non increasing functions on  $(0, \pi]$ .*

Given a stationary sequence  $\{X_k\}$  with spectral density  $f$ , the goal is to estimate  $f$  under the assumption that it lies in  $\mathcal{F}$ . We suggest two estimators, which are the  $\mathbb{L}^2$  orthogonal projections on the convex cone  $\mathcal{F}$  of the periodogram and of the log-periodogram, respectively.

- (i) The  $\mathbb{L}^2$  minimum distance estimate between the periodogram and  $\mathcal{F}$  is defined as

$$(1) \quad \hat{f}_n = \arg \min_{z \in \mathcal{F}} Q(z) ,$$

with

$$Q(z) = \int_0^\pi (I_n(s) - z(s))^2 ds .$$

This estimator of the spectral density naturally yields a corresponding estimator  $\hat{F}_n$  of the spectral distribution function  $F$ , defined by

$$(2) \quad \hat{F}_n(t) = \int_0^t \hat{f}_n(s) ds .$$

- (ii) The  $\mathbb{L}^2$  minimum distance estimate between the log-periodogram (often called the cepstrum) and the “logarithm of  $\mathcal{F}$ ”, is defined as

$$(3) \quad \tilde{f}_n = \exp \arg \min_{z \in \mathcal{F}} \tilde{Q}(z) ,$$

with

$$\tilde{Q}(z) = \int_0^\pi \{\log I_n(s) + \gamma - \log z(s)\}^2 ds ,$$

where  $\gamma$  is Euler’s constant. To understand the occurrence of the centering  $-\gamma$ , recall that if  $\{X_n\}$  is a Gaussian white noise sequence with variance  $\sigma^2$ , then its spectral density is  $\sigma^2/(2\pi)$  and the distribution of  $I_n(s)/(\sigma^2/2\pi)$  is a standard exponential (i.e. one half of a chi-square with two degrees of freedom), and it is well known that if  $Z$  is a standard exponential, then  $\mathbb{E}[\log(Z)] = -\gamma$  and  $\text{var}(\log Z) = \pi^2/6$ , see e.g. [12]. The log-spectral density is of particular interest in the context of long range dependent time series, i.e. when the spectral density has a singularity at some frequency and might not be square integrable, though it is always integrable by definition. For instance, the spectral density of an ARFIMA(0,  $d$ , 0) process is  $f(x) = \sigma^2|1 - e^{ix}|^{-2d}$ , with  $d \in (-1/2, 1/2)$ . It is decreasing on  $(0, \pi]$  for  $d \in (0, 1/2)$  and not square integrable for  $d \in (1/4, 1/2)$ . In this context, for technical reasons, we will take  $I_n$  to be a step function changing value at the so-called Fourier frequencies  $\lambda_k = 2\pi k/n$ .

The paper is organized as follows: In Section 2 we derive the algorithms for the estimators  $\hat{f}_n$ ,  $\hat{F}_n$  and  $\tilde{f}_n$ . In Section 3 we derive a lower bound for the asymptotic local minimax risk in monotone spectral density estimation, and show that the rate is not faster than  $n^{-1/3}$ . In Section 4 we derive the pointwise limit distributions for the proposed estimators. The limit distribution of  $\hat{f}_n$  (suitably centered and normalized) is derived for a linear process. The asymptotic distribution is that of the slope of the least concave majorant at 0 of a quadratic function plus a two-sided Brownian motion. Up to constants, this distribution is the so-called Chernoff's distribution, see [8], which turns up in many situations in monotone function estimation, see e.g. [23] for monotone density estimation and [27] for monotone regression function estimation. The limit distribution for  $\tilde{f}_n$  is derived for a Gaussian process, and is similar to the result for  $\hat{f}_n$ . Section 5 contains a simulation study with plots of the estimators. Section 6 contains the proofs of the limit distribution results (Theorems 5 and 6).

**2. Identification of the estimators.** Let  $h$  be a function defined on a compact interval  $[a, b]$ . The least concave majorant  $T(h)$  of  $h$  and its derivative  $T(h)'$  are defined by

$$T(h) = \arg \min \{z : z \geq h, z \text{ concave}\} ,$$

$$T(h)'(t) = \min_{u < t} \max_{v \geq t} \frac{h(v) - h(u)}{v - u} .$$

By definition,  $T(h)(t) \geq h(t)$  for all  $t \in [a, b]$  and it is also clear that  $T(h)(a) = h(a)$ ,  $T(h)(b) = h(b)$ . Since  $T(h)$  is concave, it is everywhere left and right differentiable,  $T(h)'$  as defined above coincides with the left derivative of  $T(h)$  and  $T(h)(t) = \int_a^t T(h)'(s) ds$  (see for instance Hörmander [11, Theorem 1.1.9]). We will also need the following result.

**LEMMA 1.** *If  $h$  is continuous, then the support of the Stieltjes measure  $dT(h)'$  is included in the set  $\{T(h) = h\}$ .*

**PROOF.** Since  $h$  and  $T(h)$  are continuous and  $T(h)(a) - h(a) = T(h)(b) - h(b) = 0$ , the set  $\{T(h) > h\}$  is open. Thus it is a union of open intervals. On such an interval,  $T(h)$  is linear since otherwise it would be possible to build a concave majorant of  $h$  that would be strictly smaller than  $T(h)$  on some smaller open subinterval. Hence  $T(h)'$  is piecewise constant on the open set  $\{T(h) > h\}$ , so that the support of  $dT(h)'$  is included in the closed set  $\{T(h) = h\}$ .  $\square$

The next Lemma characterizes the least concave majorant as the solution of a quadratic optimization problem. For any integrable function  $g$ , define the function  $\bar{g}$  on  $[0, \pi]$  by

$$\bar{g}(t) = \int_0^t g(s) ds .$$

**LEMMA 2.** *Let  $g \in \mathbb{L}^2([0, \pi])$ . Let  $G$  be defined on  $\mathbb{L}^2([0, \pi])$  by*

$$G(f) = \|f - g\|_2^2 = \int_0^\pi \{f(s) - g(s)\}^2 ds .$$

*Then  $\arg \min_{f \in \mathcal{F}} G(f) = T(\bar{g})'$ .*

This result seems to be well known. It is cited e.g. in [15, p. 726] but since we have not found a proof, we give one for completeness.

Let  $G : \mathcal{F} \mapsto \mathbb{R}$  be an arbitrary functional. It is called Gateaux differentiable at the point  $f \in \mathcal{F}$  if the limit

$$G'_f(h) = \lim_{t \rightarrow 0} \frac{G(f + th) - G(f)}{t}$$

exists for every  $h$  such that  $f + th \in \mathcal{F}$  for small enough  $t$ .

PROOF OF LEMMA 2. Denote  $G(f) = \|f - g\|_2^2$  and  $\hat{f} = T(\bar{g})'$ . The Gateaux derivative of  $G$  at  $\hat{f}$  in the direction  $h$  is

$$G'_{\hat{f}}(h) = 2 \int_0^\pi h(t) \{\hat{f}(t) - g(t)\} dt .$$

By integration by parts, and using that  $T(\bar{g})(\pi) - \bar{g}(\pi) = T(\bar{g})(0) - \bar{g}(0) = 0$ , for any function of bounded variation  $h$ , we have

$$(4) \quad G'_{\hat{f}}(h) = -2 \int_0^\pi \{T(\bar{g})(t) - \bar{g}(t)\} dh(t) .$$

By Lemma 1, the support of the measure  $d\hat{f}$  is included in the closed set  $\{T(\bar{g}) = \bar{g}\}$ , thus

$$(5) \quad G'_{\hat{f}}(\hat{f}) = -2 \int_0^\pi \{T(\bar{g})(t) - \bar{g}(t)\} d\hat{f}(t) = 0 .$$

If  $h = f - \hat{f}$ , with  $f$  monotone non increasing, (4) and (5) imply that

$$(6) \quad G'_f(f - \hat{f}) = -2 \int_0^\pi \{T(\bar{g})(t) - \bar{g}(t)\} df(t) \geq 0 .$$

Let  $f \in \mathcal{F}$  be arbitrary and let  $u$  be the function defined on  $[0, 1]$  by  $u(t) = G(\hat{f} + t(f - \hat{f}))$ . Then  $u$  is convex and  $u'(0) = G'_f(f - \hat{f}) \geq 0$  by (6). Since  $u$  is convex, if  $u'(0) \geq 0$ , then  $u(1) \geq u(0)$ , i.e.  $G(f) \geq G(\hat{f})$ . This proves that  $\hat{f} = \arg \min_{f \in \mathcal{F}} G(f)$ .  $\square$

Since  $\hat{f}_n$  and  $\log \tilde{f}_n$  are the minimizers of the  $\mathbb{L}^2$  distance of  $I_n$  and  $\log(I_n) + \gamma$ , respectively, over the convex cone of monotone functions, we can apply Lemma 2 to derive characterizations of  $\hat{f}_n$  and  $\tilde{f}_n$ .

THEOREM 3. Let  $\hat{f}_n, \hat{F}_n$  and  $\tilde{f}_n$  be defined in (1), (2) and (3), respectively. Then

$$\hat{f}_n = T(F_n)' , \quad \hat{F}_n(t) = T(F_n) , \quad \tilde{f}_n = \exp\{T(\tilde{F}_n)'\} ,$$

where

$$F_n(t) = \int_0^t I_n(u) du , \quad \tilde{F}_n(t) = \int_0^t (\log I_n(u) + \gamma) du .$$

Standard and well known algorithms for calculating the map  $y \mapsto T(y)'$  are the pool adjacent violators algorithm (PAVA), the minimum lower set algorithm (MLSA) and the min-max formulas, cf. [24]. Since the maps  $T$  and  $T'$  are continuous operations, in fact the algorithms PAVA and MLSA will be approximations that solve the discrete versions of our problems, replacing the integrals in  $Q$  and  $\tilde{Q}$  with approximating Riemann sums. Note that the resulting estimators are order-restricted means; the discrete approximations entail that these are approximated as sums instead of integrals. The approximation errors are similar to the ones obtained e.g. for the methods in [15] and [1].

**3. Lower bound for the local asymptotic minimax risk.** We establish a lower bound for the minimax risk when estimating a monotone spectral density at a fixed point. This result will be proved by looking at parametrized subfamilies of spectral densities in an open set of densities on  $\mathbb{R}^n$ ; the subfamilies can be seen as (parametrized) curves in the set of monotone spectral densities. The topology used will be the one generated by the metric

$$\rho(f, g) = \int_{\mathbb{R}} |f(x) - g(x)| dx$$

for  $f, g$  spectral density functions on  $[-\pi, \pi]$ . Note first that the distribution of a stochastic process is not uniquely defined by the spectral density. To accommodate this, let  $\mathcal{L}_g$  be the set of all laws of stationary processes (i.e. the translation invariant probability distributions on  $\mathbb{R}^\infty$ ) with spectral density  $g$ .

Let  $\epsilon > 0, c_1, c_2$  be given finite constants and let  $t_0 > 0$ , the point at which we want to estimate the spectral density, be given.

**DEFINITION 2.** For each  $n \in \mathbb{Z}$  let  $\mathcal{G}^1 := \mathcal{G}^1(\epsilon, c_1, c_2, t_0)$  be a set of monotone  $C^1$  spectral densities  $g$  on  $[0, \pi]$ , such that

$$(7) \quad \sup_{|t-t_0|<\epsilon} g'(t) < 0,$$

$$(8) \quad c_1 < \inf_{|t-t_0|<\epsilon} g(t) < \sup_{|t-t_0|<\epsilon} g(t) < c_2.$$

**THEOREM 4.** For every open set  $U$  in  $\mathcal{G}^1$  there is a positive constant  $c(U)$  such that

$$\liminf_{n \rightarrow \infty} \inf_{T_n} \sup_{g \in U} \sup_{L \in \mathcal{L}_g} n^{1/3} \mathbb{E}_L[(T_n - g(t_0))^2] \geq c(U),$$

where the infimum is taken over all functions  $T_n$  of the data.

**PROOF.** Let  $k$  be a fixed real valued continuously differentiable function, with support  $[-1, 1]$  such that  $\int k(t) dt = 0, k(0) = 1$  and  $\sup |k'(t)| \leq 1$ . Then, since  $k'$  is continuous with compact support,  $|k'| < C$  for some constant  $C < \infty$ .

For fixed  $h > 0$ , define a parametrized family of spectral densities  $g_\theta$  by

$$g_\theta(t) = g(t) + \theta k\left(\frac{t-t_0}{h}\right).$$

Obviously,  $\{g_\theta\}_{\theta \in \Theta}$  are  $C^1$  functions. Since

$$g'_\theta(t) = g'(t) + \frac{\theta}{h} k'\left(\frac{t-t_0}{h}\right),$$

and since  $k'$  is bounded, we have that, for  $|t-t_0| < \epsilon$ ,  $g'_\theta(t) < 0$  if  $|\theta/h| < \delta$ , for some  $\delta = \delta(C) > 0$ . Thus, in order to make the parametrized spectral densities  $g_\theta$  strictly decreasing in the neighbourhood  $\{t : |t-t_0| < \epsilon\}$ , the parameter space for  $\theta$  should be chosen as

$$\Theta = (-\delta h, \delta h).$$

We will use the van Trees inequality (cf. Gill and Levit [7, Theorem 1]) for the estimand  $g_\theta(t_0) = g(t_0) + \theta$ . Let  $\lambda$  be an arbitrary prior density on  $\Theta$ . Then, for sufficiently small  $\delta$ ,

$\{g_\theta : \theta \in \Theta\} \subset U$  (cf. the definition of the metric  $\rho$ ). Let  $P_\theta$  denote the distribution of a Gaussian process with spectral density  $g_\theta$ , and  $\mathbb{E}_\theta$  the corresponding expectation. Then

$$\begin{aligned} \sup_{g \in U} \sup_{L \in \mathcal{L}_g} \mathbb{E}_L[(T_n - g(t_0))^2] &\geq \sup_{\theta \in \Theta} \mathbb{E}_\theta[(T_n - g_\theta(t_0))^2] \\ &\geq \int_{\Theta} \mathbb{E}_\theta[(T_n - g_\theta(t_0))^2] \lambda(\theta) d\theta . \end{aligned}$$

Then, by the Van Trees inequality, we obtain

$$(9) \quad \int_{\Theta} \mathbb{E}_\theta[(T_n - g_\theta(t_0))^2] \lambda(\theta) d\theta \geq \frac{1}{\int I_n(\theta) \lambda(\theta) d\theta + \tilde{I}(\lambda)} .$$

where

$$I_n(\theta) = \frac{1}{2} \text{tr} (\{M_n^{-1}(g_\theta) M_n(\partial_\theta g_\theta)\}^2)$$

is the Fisher information matrix, cf. [6], with respect to the parameter  $\theta$  of a Gaussian process with spectral density  $g_\theta$ , and for any even nonnegative integrable function  $\phi$  on  $[-\pi, \pi]$ ,  $M_n(\phi)$  is the Toeplitz matrix of order  $n$ :

$$M_n(\phi)_{i,j} = \int_{-\pi}^{\pi} \phi(x) \cos((i-j)x) dx .$$

For any  $n \times n$  nonnegative symmetric matrix  $A$ , define the spectral radius of  $A$  as

$$\rho(A) = \sup\{u^t A u \mid u^t u = 1\} ,$$

where  $u^t$  denotes transposition of the vector  $u$ , so that  $\rho(A)$  is the the largest eigenvalue of  $A$ . Then, for any  $n \times n$  matrix  $B$ , it holds that  $\text{tr}(AB) \leq \rho(A)\text{tr}(B)$ . If  $\phi$  is bounded away from zero, say  $\phi(x) \geq a > 0$  for all  $x \in [-\pi, \pi]$ , then  $\rho(M_n^{-1}(\phi)) \leq a^{-1}$ ; By the Parseval-Bessel inequality,

$$\text{tr}(\{M_n(\phi)\}^2) \leq n \int_{-\pi}^{\pi} \phi^2(x) dx .$$

Thus, if  $g$  is bounded below, then  $I_n(\theta)$  is bounded by some constant times

$$n \int_{-\pi}^{\pi} k^2((t - t_0)/h) dt = nh \int k^2(t) dt .$$

In order to get an expression for  $\tilde{I}(\lambda)$ , let  $\lambda_0$  be an arbitrary density on  $(-1, 1)$ , and define the prior density on  $\Theta = (-\delta h, \delta h)$  as  $\lambda(\theta) = \frac{1}{\delta h} \lambda_0(\frac{\theta}{\delta h})$ . Then

$$\tilde{I}(\lambda) = \int_{-\delta h}^{\delta h} \frac{(\lambda'(\theta))^2}{\lambda(\theta)} d\theta = \frac{1}{\delta^2 h^2} \int_{-1}^1 \frac{\lambda'_0(u)^2}{\lambda_0(u)} du = \frac{I_0}{\delta^2 h^2} .$$

Finally, plugging the previous bounds into (9) yields, for large enough  $n$ ,

$$\sup_{g \in U} \sup_{L \in \mathcal{L}_g} \mathbb{E}_L[(T_n(t_0) - g(t_0))^2] \geq \frac{1}{nhc_3 + I_0 \delta^{-2} h^{-2}} ,$$

which, if  $h = n^{-1/3}$ , becomes

$$\sup_{g \in U} \sup_{L \in \mathcal{L}_g} \mathbb{E}_L[\{T_n(t_0) - g(t_0)\}^2] \geq c_4 n^{-2/3} ,$$

for some positive constant  $c_4$ . This completes the proof of Theorem 4. □

**4. Limit distribution results.** We next derive the limit distributions for  $\hat{f}_n$  and  $\tilde{f}_n$  under general assumptions. The main tools used are local limit distributions for the rescaled empirical spectral distribution function  $F_n$  and empirical log-spectral distribution function  $\tilde{F}_n$  respectively, as well as maximal bounds for the rescaled processes. These will be coupled with smoothness results for the least concave majorant map established in Anevski and Hössjer [1, Theorems 1 and 2]. The proofs are postponed to Section 6.

4.1. *The limit distribution for the estimator  $\hat{f}_n$ .*

ASSUMPTION 1. *The process  $\{X_i, i \in \mathbb{Z}\}$  is linear with respect to an i.i.d. sequence  $\{\epsilon_i, i \in \mathbb{Z}\}$  with zero mean and unit variance, i.e.*

$$(10) \quad X_k = \sum_{j=0}^{\infty} a_j \epsilon_{k-j},$$

where the sequence  $\{a_j\}$  satisfies

$$(11) \quad \sum_{j=1}^{\infty} (j^{1/2}|a_j| + j^{3/2}a_j^2) < \infty.$$

REMARK. Condition (11) is needed to deal with remainder terms and apply the results of [18] and [3]. It is implied for instance by the simpler condition

$$(12) \quad \sum_{j=1}^{\infty} j^{3/4}|a_j| < \infty.$$

It is satisfied by most usual linear time series such as causal invertible ARMA processes.

The spectral density of the process  $\{X_i\}$  is given by  $f(u) = (2\pi)^{-1} \left| \sum_{j=0}^{\infty} a_j e^{iu} \right|^2$ . Unfortunately, there is no explicit condition on the coefficients  $a_j$  that implies monotonicity of  $f$ , but the coefficients  $a_j$  are not of primary interest here.

The limiting distribution of the estimator will be expressed in terms of the so-called Chernoff's distribution, i.e. the law of a random variable  $\zeta$  defined by  $\zeta = \arg \max_{s \in \mathbb{R}} \{W(s) - s^2\}$ , where  $W$  is a standard two sided Brownian motion. See [8] for details about this distribution.

THEOREM 5. *Let  $\{X_i\}$  be a linear process such that (10) and (11) hold and  $\mathbb{E}[\epsilon_0^8] < \infty$ . Assume that its spectral density  $f$  belongs to  $\mathcal{F}$ . Assume  $f'(t_0) < 0$  at the fixed point  $t_0$ . Then, as  $n \rightarrow \infty$ ,*

$$n^{1/3}(\hat{f}_n(t_0) - f(t_0)) \xrightarrow{\mathcal{L}} 2\{-\pi f^2(t_0) f'(t_0)\}^{1/3} \zeta.$$

4.2. *The limit distributions for the estimator  $\tilde{f}_n$ .* In this section, in order to deal with the technicalities of the log-periodogram, we make the following assumption.

ASSUMPTION 2. *The process  $\{X_k\}$  is Gaussian. Its spectral density  $f$  is monotone on  $(0, \pi]$  and can be expressed as  $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$ , with  $|d| < 1/2$  and  $f^*$  is bounded above and away from zero and there exists a constant  $C$  such that for all  $x, y \in (0, \pi]$ ,*

$$|f(x) - f(y)| \leq C \frac{|x - y|}{x \wedge y}.$$

REMARK. This condition is usual in the long memory literature. Similar conditions are assumed in [25, Assumption 2], [19, Assumption 2], [26, Assumption 1] (with a typo). It is used to derive covariance bounds for the discrete Fourier transform ordinates of the process, which yield covariance bounds for non linear functionals of the periodogram ordinates in the Gaussian case. It is satisfied by usual long memory processes such as causal invertible ARFIMA  $(p, d, q)$  processes with possibly an additive independent white noise or AR(1) process.

Recall that  $\tilde{f}_n = \exp \arg \min_{f \in \mathcal{F}} \int_0^\pi \{\log f(s) - \log I_n(s) + \gamma\}^2 ds$  where  $\gamma$  is Euler's constant and  $I_n$  is the periodogram, defined here as a step function:

$$I_n(t) = I_n(2\pi[nt/2\pi]/n) = \frac{2\pi}{n} \left| \sum_{k=1}^n X_k e^{i2k\pi[nt/2\pi]/n} \right|^2 .$$

THEOREM 6. *Let  $\{X_i\}$  be a Gaussian process that satisfies Assumption 2. Assume  $f'(t_0) < 0$  at the fixed point  $t_0 \in (0, \pi)$ . Then, as  $n \rightarrow \infty$ ,*

$$n^{1/3} \{\log \tilde{f}_n(t_0) - \log f(t_0)\} \xrightarrow{\mathcal{L}} 2 \left( \frac{-\pi^4 f'(t_0)}{3f(t_0)} \right)^{1/3} \zeta .$$

COROLLARY 7. *Under the assumptions of Theorem 6,*

$$n^{1/3} \{\tilde{f}_n(t_0) - f(t_0)\} \xrightarrow{\mathcal{L}} 2 \{-\pi^4 f^2(t_0) f'(t_0)/3\}^{1/3} \zeta .$$

REMARK. This is the same limiting distribution as in Theorem 5, up to the constant  $3^{-1/3}\pi > 1$ . Thus the estimator  $\tilde{f}_n$  is less efficient than the estimator  $\hat{f}_n$ , but the interest of  $\tilde{f}_n$  is to be used when long memory is suspected, i.e. the spectral density exhibits a singularity at zero, and the assumptions of Theorem 5 are not satisfied.

**5. Simulations and finite sample behaviour of estimators.** In this section we apply the nonparametric methods on simulated time series data of sums of parametric models. The algorithms used for the calculation of  $\hat{f}_n$  and  $\tilde{f}_n$  are the discrete versions of the estimators  $\hat{f}$ ,  $\tilde{f}_n$ , that are obtained by doing isotonic regression of the data  $\{(\lambda_k, I_n(\lambda_k)), k = 1, \dots, [(n-1)/2]\}$  where  $\lambda_k = 2\pi k/n$ . For instance the discrete version  $\hat{f}_n^d$  of  $\hat{f}_n$  is calculated as

$$\hat{f}_n^d = \arg \min_{z \in \mathcal{F}} \sum_{k=1}^n (I_n(\lambda_k) - z(\lambda_k))^2 .$$

Note that the limit distribution for  $\tilde{f}_n$  is stated for the discrete version  $\tilde{f}_n^d$ . The simulations were done in R, using the “fracdiff” package. The code is available from the corresponding author upon request.

EXAMPLE 1. The first example consists of sums of several AR(1) processes. Let  $\{X_k\}$  be a stationary AR(1) process, i.e. for all  $k \in \mathbb{Z}$ ,

$$X_k = aX_{k-1} + \epsilon_k ,$$

with  $|a| < 1$ . This process has spectral density function  $f(\lambda) = (2\pi)^{-1}\sigma^2|1 - ae^{i\lambda}|^{-2}$  for  $-\pi \leq \lambda \leq \pi$ , with  $\sigma^2 = \text{var}(\epsilon_1^2)$  and thus  $f$  is decreasing on  $[0, \pi]$ . If  $X^{(1)}, \dots, X^{(p)}$  are

independent AR(1) processes with coefficients  $a_j$  such that  $|a_j| < 1, j = 1, \dots, p$ , and we define the process  $X$  by

$$X_k = \sum_{j=1}^p X_k^{(j)}$$

then  $X$  has spectral density  $f(\lambda) = (2\pi)^{-1} \sum_{j=1}^p \sigma_j^2 |1 + a_j e^{i\lambda}|^{-2}$  which is decreasing on  $[0, \pi]$ , since it is a sum of decreasing functions. Assuming that we do not know how many AR(1) processes are summed, we have a nonparametric problem: estimate a monotone spectral density. Figure 1 shows a plot of the periodogram, the true spectral density and the nonparametric estimators  $\hat{f}_n$  and  $\tilde{f}_n$  for simulated data from a sum of three independent AR(1) processes with  $a_1 = 0.5, a_2 = 0.7, a_3 = 0.9$ . Figure 2 shows the pointwise means and 95% confidence intervals of  $\hat{f}_n$  and  $\tilde{f}_n$  for 1000 realizations.

EXAMPLE 2. The second example is a sum of an ARFIMA(0,d,0) process and an AR(1) process. Let  $X^{(1)}$  be an ARFIMA(0,d,0)-process with  $0 < d < 1/2$ . This has a spectral density  $(2\pi)^{-1} \sigma_1^2 |1 - e^{i\lambda}|^{-2d}$ . If we add an independent AR(1)-process  $X^{(2)}$  with coefficient  $|a| < 1$  the resulting process  $X = X^{(1)} + X^{(2)}$  will have spectral density  $f(\lambda) = (2\pi)^{-1} \sigma_1^2 |1 - e^{i\lambda}|^{-2d} + (2\pi)^{-1} \sigma_2^2 |1 - ae^{i\lambda}|^{-2}$  on  $[0, \pi]$ , and thus the resulting spectral density  $f$  will be a monotone function on  $[0, \pi]$ . As above, if an unknown number of independent processes is added we have a nonparametric estimation problem. Figure 3 shows a plot of the periodogram, the true spectral density and the nonparametric estimators  $\hat{f}_n$  and  $\tilde{f}_n$  for simulated time series data from a sum of an ARFIMA(0,d,0)-process with  $d = 0.2$  and an AR(1)-process with  $a = 0.5$ . Figure 4 shows the pointwise means and 95% confidence intervals of  $\hat{f}_n$  and  $\tilde{f}_n$  for 1000 realizations.

Table 1 shows mean square root of sum of squares errors (comparing with the true function), calculated on 1000 simulated samples of the times series of Example 1. Table 2 shows the analog values for Example 2. Both estimators  $\hat{f}_n$  and  $\tilde{f}_n$  seem to have good finite sample properties. As indicated by the theory  $\tilde{f}_n$  seems to be less efficient than  $\hat{f}_n$ .

MISE	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
$I_n$	9.59	12.96	13.67	14.25
$\hat{f}_n$	6.38	5.48	4.76	2.95
$\tilde{f}_n$	9.11	8.52	7.27	4.26

TABLE 1  
MISE values for Example 1

MISE	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$
$I_n$	1.80	1.99	2.02	2.07
$\hat{f}_n$	0.710	0.520	0.432	0.305
$\tilde{f}_n$	1.12	0.803	0.659	0.472

TABLE 2  
MISE values for Example 2

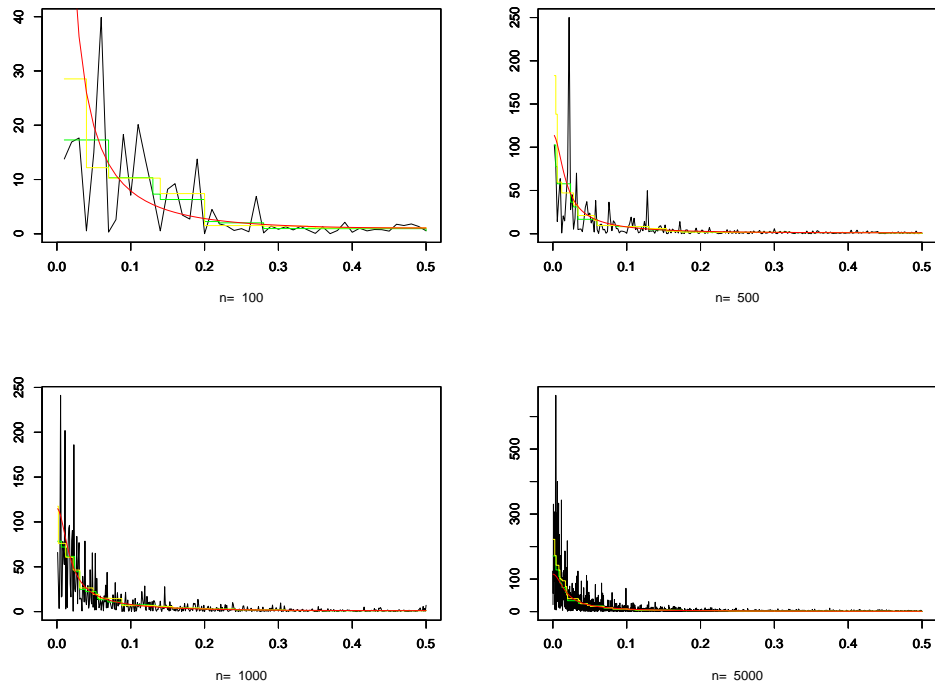


FIG 1. The spectral density (red), the periodogram (black), the estimates  $\hat{f}_n$  (green) and  $\tilde{f}_n$  (yellow), for  $n=100, 500, 1000$ , and  $5000$  data points, for Example 1.

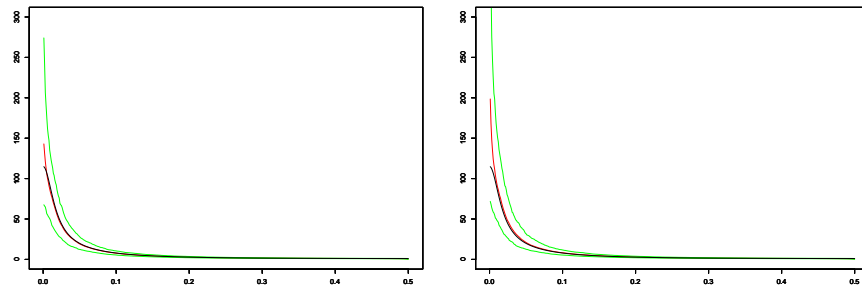


FIG 2. Left plot: Spectral density (black), pointwise mean of estimates  $\hat{f}_n$  (red) and 95% confidence intervals (green). Right plot: Spectral density (black), pointwise mean of the estimates  $\tilde{f}_n$  (red) and 95% confidence intervals (green), for  $n=1000$  data points, for Example 1.

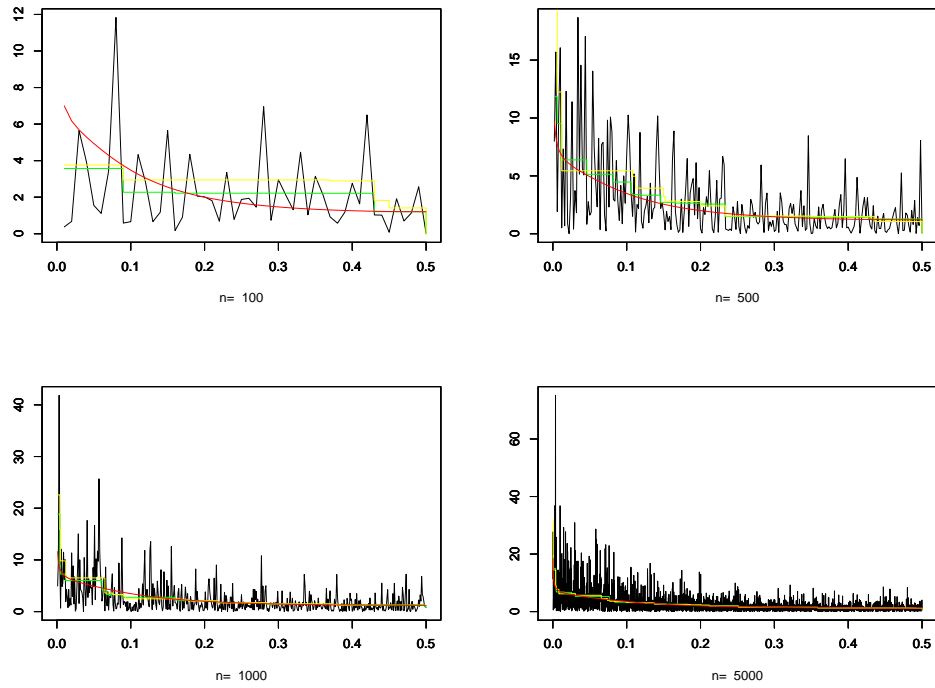


FIG 3. The spectral density (red), the periodogram (black), the estimates  $\hat{f}_n$  (green) and  $\tilde{f}_n$  (yellow), for  $n=100, 500, 1000$ , and  $5000$  data points, for Example 2.

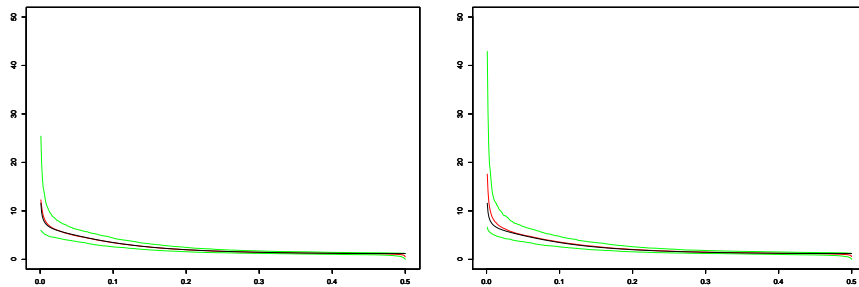


FIG 4. Left plot: Spectral density (black), pointwise mean of estimates  $\hat{f}_n$  (red) and 95% confidence intervals (green). Right plot: Spectral density (black), pointwise mean of the estimates  $\tilde{f}_n$  (red) and 95% confidence intervals (green), for  $n=1000$  data points, for Example 2.

**6. Proof of Theorems 5 and 6.** Let  $J_n$  be the integral of the generic preliminary estimator of the spectral density, that is the integral of  $I_n$  or of  $\log(I_n)$ , let  $K$  denote  $F$  or the primitive of  $\log f$ , respectively, and write

$$(13) \quad J_n(t) = K(t) + v_n(t) ,$$

Let  $d_n \downarrow 0$  be a deterministic sequence and define the rescaled process and rescaled centering

$$(14) \quad \tilde{v}_n(s; t_0) = d_n^{-2} \{v_n(t_0 + sd_n) - v_n(t_0)\} ,$$

$$(15) \quad g_n(s) = d_n^{-2} \{K(t_0 + sd_n) - K(t_0) - K'(t_0)d_n s\} .$$

Consider the following conditions.

(AH1) There exists a stochastic process  $\tilde{v}(\cdot; t_0)$  such that

$$(16) \quad \tilde{v}_n(\cdot; t_0) \xrightarrow{\mathcal{L}} \tilde{v}(\cdot; t_0) ,$$

in  $D(-\infty, \infty)$ , endowed with the topology generated by the supnorm metric on compact intervals, as  $n \rightarrow \infty$ .

(AH2) For each  $\epsilon, \delta > 0$  there is a finite  $\tau$  such that

$$(17) \quad \limsup_{n \rightarrow \infty} P \left( \sup_{|s| \geq \tau} \left| \frac{\tilde{v}_n(s; t_0)}{g_n(s)} \right| > \epsilon \right) < \delta ,$$

$$(18) \quad P \left( \sup_{|s| \geq \tau} \left| \frac{\tilde{v}(s; t_0)}{s^2} \right| \epsilon \right) < \delta .$$

(AH3) There is a constant  $A < 0$  such that for each  $c > 0$ ,

$$(19) \quad \lim_{n \rightarrow \infty} \sup_{|s| \leq c} |g_n(s) - As^2| = 0 ;$$

(AH4) For each  $a \in \mathbb{R}$  and  $c, \epsilon > 0$

$$(20) \quad P(\tilde{v}(s; t_0)(s) - \tilde{v}(0; t_0) + As^2 - as \geq \epsilon|s| \text{ for all } s \in [-c, c]) = 0 .$$

If there exists a sequence  $d_n$  such that these four conditions hold, then, defining the process  $y$  by  $y(s) = \tilde{v}(s; t_0) + As^2$ , by Anevski and Hössjer [1, Theorems 1 and 2], as  $n \rightarrow \infty$ , it holds that

$$(21) \quad d_n^{-1} \{T(J_n)'(t_0) - K'(t_0)\} \xrightarrow{\mathcal{L}} T(y)'(0) ,$$

where  $T(y)'(0)$  denotes the slope at zero of the smallest concave majorant of  $y$ .

6.1. *Proof of Theorem 5.* The proof consists in checking Conditions (AH1) - (AH4) with  $J_n = F_n$  and  $K = F$ .

- It is proved in Lemma 8 below that (16) holds with  $d_n = n^{1/3}$  and  $\tilde{v}(\cdot; t_0)$  the standard two sided Brownian motion times  $\sqrt{\pi^2/6}$ .
- If  $f'(t_0) < 0$ , then (19) holds with  $A = \frac{1}{2}f'(t_0)$  and  $d_n \downarrow 0$  an arbitrary deterministic sequence.

- Lemma 9 shows that (17) holds and the law of iterated logarithm yields that (18) holds for the two-sided Brownian motion.
- Finally, (20) also holds for the two sided Brownian motion.

Thus (21) holds with the process  $y$  defined by

$$y(s) = \frac{1}{2}f'(t_0)s^2 + \sqrt{2\pi}f(t_0)W(s) .$$

The scaling property of the Brownian motion yields the representation of  $T(y)'(0)$  in terms of Chernoff's distribution.  $\square$

LEMMA 8. *Assume the process  $\{X_n\}$  is given by (10), that (11) holds and that  $\mathbb{E}[\epsilon_0^8] < \infty$ . If  $d_n = n^{-1/3}$ , then the sequence of processes  $\tilde{v}_n(\cdot; t_0)$  defined in (14) converges weakly in  $C([-c, c])$  to  $\sqrt{2\pi}f(t_0)W$  where  $W$  is a standard two sided Brownian motion.*

PROOF. For clarity, we omit  $t_0$  in the notation. Write

$$\tilde{v}_n(s) = \tilde{v}_n^\epsilon(s) + R_n(s)$$

with

$$(22) \quad \tilde{v}_n^\epsilon(s) = d_n^{-2} \int_{t_0}^{t_0+d_n s} f(u) \{I_n^\epsilon(u) - 1\} du, \quad I_n^\epsilon(u) = \frac{1}{n} \left| \sum_{k=1}^n \epsilon_k e^{iku} \right|^2,$$

$$(23) \quad R_n = d_n^{-2} \int_{t_0}^{t_0+d_n s} r_n(u) du, \quad r_n(u) = I_n(u) - f(u)I_n^\epsilon(u) .$$

Note that  $(2\pi)^{-1}I_n^\epsilon$  is the periodogram for the white noise sequence  $\{\epsilon_i\}$ . We first treat the remainder term  $R_n$ . Denote  $\mathcal{G} = \{g : \int_{-\pi}^{\pi} g^2(u)f^2(u) du < \infty\}$ . Equation (5.11) (with a typo in the normalization) in [18] states that if (11) and  $\mathbb{E}[\epsilon_0^8] < \infty$  hold, then

$$(24) \quad \sqrt{n} \sup_{g \in \mathcal{G}} \int_{-\pi}^{\pi} g(x)r_n(x) dx = o_P(1) .$$

Define the set  $\tilde{\mathcal{G}} = \{k_n(\cdot, s)f : n \in \mathbf{N}, s \in [-c, c]\}$ . Since  $f$  is bounded, we have that  $\int k_n^2(u, s)f^2(u) du < \infty$ , so  $\tilde{\mathcal{G}} \subset \mathcal{G}$  and we can apply (24) on  $\tilde{\mathcal{G}}$ , which shows that  $R_n$  converges uniformly (over  $s \in [-c, c]$ ) to zero. We next show that

$$(25) \quad \tilde{v}_n^\epsilon(s) \xrightarrow{\mathcal{L}} \sqrt{2\pi}f(t_0)W(s) ,$$

as  $n \rightarrow \infty$ , on  $C(\mathbb{R})$ , where  $W$  is a standard two sided Brownian motion. Since  $\{\epsilon_k\}$  is a white noise sequence, we set  $t_0 = 0$  without loss of generality. Straightforward algebra yields

$$(26) \quad \tilde{v}_n^\epsilon(s) = d_n^{-2} \{\hat{\gamma}_n(0) - 1\}F(d_n s) + S_n(s)$$

with

$$\hat{\gamma}_n(0) = n^{-1} \sum_{j=1}^n \epsilon_j^2, \quad S_n(s) = \sum_{k=2}^n C_k(s)\epsilon_k,$$

$$C_k(s) = d_n^{3/2} \sum_{j=1}^{k-1} \alpha_j(s)\epsilon_{k-j}, \quad \alpha_j(s) = d_n^{-1/2} \int_{-d_n s}^{d_n s} f(u) e^{iju} du .$$

Since  $\{\epsilon_j\}$  is a white noise sequence with finite fourth moment, it is easily checked that

$$(27) \quad n \operatorname{var}(\hat{\gamma}_n(0)) = \operatorname{var}(\epsilon_0^2),$$

$$\sup_{s \in [-c, c]} d_n^{-2} \int_0^{d_n s} f(u) du |\hat{\gamma}_n(0) - 1| = O_P(d_n^{-1} n^{-1/2}) = O_P(\sqrt{d_n}).$$

so that the first term in (26) is negligible. It remains to prove that the sequence of processes  $S_n$  converges weakly to a standard Brownian motion. We prove the convergence of finite dimension distribution by application of the Martingale central limit Theorem, cf. for instance Hall and Heyde [9, Corollary 3.1]. It is sufficient to check the following conditions

$$(28) \quad \lim_{n \rightarrow \infty} \sum_{k=2}^n \mathbb{E}[C_k^2(s)] = 2\pi f^2(0)s,$$

$$(29) \quad \lim_{n \rightarrow \infty} \sum_{k=2}^n \mathbb{E}[C_k^4(s)] = 0.$$

By the Parseval-Bessel identity, we have

$$\sum_{j=-\infty}^{\infty} \alpha_j^2(s) = 2\pi d_n^{-1} \int_{-d_n s}^{d_n s} f^2(u) du \sim 4\pi f^2(0)s.$$

Since  $\alpha_0(s) \sim 2f(0)\sqrt{d_n}$ , this implies that

$$\sum_{k=2}^n \mathbb{E}[C_k^2(s)] = \sum_{j=1}^{n-1} (1 - j/n) \alpha_j^2(s) \sim \sum_{j=1}^{\infty} \alpha_j^2(s) \sim 2\pi f^2(0)s.$$

This proves Condition (28). For the asymptotic negligibility condition (29), we use Rosenthal's inequality (cf. Hall and Heyde [9, Theorem 2.12]),

$$E[C_k^4] \leq cst n^{-2} \sum_{j=1}^{k-1} \alpha_j^4(s) + cst n^{-2} \left( \sum_{j=1}^{k-1} \alpha_j^2(s) \right)^2 = O(n^{-2}),$$

implying  $\sum_{k=1}^n \mathbb{E}[C_k^4(s)] = O(n^{-1})$ , which proves (29). To prove tightness, we compute the fourth moment of the increments of  $S_n$ . Write

$$S_n(s) - S_n(s') = n^{-1/2} \sum_{k=1}^n \sum_{j=1}^{k-1} \alpha_j(s, s') \epsilon_{k-j} \epsilon_k,$$

with

$$\alpha_j(s, s') = d_n^{-1/2} \int_{d_n s'}^{d_n s} f(u) e^{iju} du + d_n^{-1/2} \int_{-d_n s}^{-d_n s'} f(u) e^{iju} du.$$

By Parseval inequality, it holds that

$$\sum_{j=1}^n \alpha_j^2(s, s') \leq C|s - s'|.$$

Applying again Rosenthal inequality, we obtain that  $\mathbb{E}[|S_n(s) - S_n(s')|^4]$  is bounded by a constant times

$$n^{-1} \sum_{j=1}^n \alpha_j^4(s, s') + \left( \sum_{j=1}^n \alpha_j^2(s, s') \right)^2 \leq C|s - s'|^2.$$

Applying [2, Theorem 15.6] concludes the proof of tightness.  $\square$

LEMMA 9. *For any  $\delta > 0$  and any  $\kappa > 0$ , there exists  $\tau$  such that*

$$(30) \quad \limsup_{n \rightarrow \infty} \mathbb{P} \left( \sup_{|s| \geq \tau} \frac{|\tilde{v}_n(s)|}{|s|} > \kappa \right) \leq \delta.$$

PROOF. Without loss of generality, we can assume that  $f(t_0) = 1$ . Recall that  $\tilde{v}_n = \tilde{v}_n^{(\epsilon)} + R_n$  and  $\tilde{v}_n^{(\epsilon)}(s) = F(d_n s) \zeta_n + S_n(s)$ , where  $\tilde{v}_n^{(\epsilon)}$  and  $R_n$  are defined in (22) and (23),  $\zeta_n = d_n^{-2}(\hat{\gamma}_n(0) - 1)$  and  $S_n$  is defined in (26). Then

$$\begin{aligned} \mathbb{P} \left( \sup_{s \geq \tau} \frac{|\tilde{v}_n(s)|}{s} > \kappa \right) &\leq \mathbb{P} \left( \sup_{s \geq \tau} |\zeta_n| F(d_n s) / s > \kappa/3 \right) \\ &\quad + \mathbb{P} \left( \sup_{s \geq \tau} \frac{|S_n(s)|}{s} > \kappa/3 \right) + \mathbb{P} \left( \sup_{s \geq \tau} \frac{|R_n(s)|}{s} > \kappa/3 \right). \end{aligned}$$

The spectral density is bounded, so  $F(d_n s)/s \leq C d_n$  for all  $s$ . Since  $\text{var}(\zeta_n) = O(d_n^{-1})$ , by (27) and Bienayme-Chebyshev inequality, we get

$$\mathbb{P} \left( \sup_{s \geq \tau} |\zeta_n| F(d_n s) / s > \kappa \right) \leq O(d_n^{-1} d_n^2) = O(d_n).$$

Let  $\{s_j, j \geq 0\}$  be an increasing sequence. Then we have the bound

$$\begin{aligned} \mathbb{P} \left( \sup_{s \geq s_0} \frac{|S_n(s)|}{s} > \kappa \right) &\leq \sum_{j=0}^{\infty} \mathbb{P}(|S_n(s_j)| > \kappa s_j) \\ &\quad + \sum_{j=1}^{\infty} \mathbb{P} \left( \sup_{s_{j-1} \leq s \leq s_j} |S_n(s) - S_n(s_{j-1})| > \kappa s_{j-1} \right). \end{aligned}$$

From (28), we know that  $\text{var}(S_n(s)) = O(s)$ . Thus

$$\sum_{j=0}^{\infty} \mathbb{P}(|S_n(s_j)| > \kappa s_j) \leq cst \kappa^{-2} \sum_{j=0}^{\infty} s_j^{-1}.$$

Thus if the series  $s_j^{-1}$  is summable, this sum can be made arbitrarily small by choosing  $s_0$  large enough. It was shown in the proof of Lemma 8 that

$$\mathbb{E}[|S_n(s) - S_n(s')|^4] \leq C|s - s'|^2.$$

By Billingsley [2, Theorem 15.6] (or more specifically Ledoux and Talagrand [14, Theorem 11.1]), this implies that

$$\mathbb{P}\left(\sup_{s_{j-1} \leq s \leq s_j} |S_n(s) - S_n(s_{j-1})| > \kappa s_{j-1}\right) \leq \frac{C(s_j - s_{j-1})^2}{\kappa^2 s_{j-1}^2}.$$

Thus choosing  $s_j = (s_0 + j)^\rho$  for some  $\rho > 1$  implies that the series is convergent and

$$\mathbb{P}\left(\sup_{s \geq s_0} \frac{|S_n(s)|}{s} > \kappa\right) = O(s_0^{-1}),$$

which is arbitrarily small for large  $s_0$ .

To deal with the remainder term  $R_n$ , we prove that  $\mathbb{P}(\sup_{s \geq s_0} |R_n(s)|/s > s_0) = o_P(1)$  by the same method as that used for  $S_n$ . Thus we only need to obtain a suitable bound for the increments of  $R_n$ . By definition of  $R_n$ , we have, for  $s < s'$ ,

$$R_n(s') - R_n(s) = d_n^{-2} \int_{t_0 + d_n s}^{t_0 + d_n s'} f(u) r_n(u) du.$$

Since  $f$  is bounded, by Hölder's inequality, we get

$$\mathbb{E}[|R_n(s') - R_n(s)|^2] \leq \|f\|_\infty^2 n(s' - s) \int_{t_0 + d_n s}^{t_0 + d_n s'} \mathbb{E}[r_n^2(u)] du.$$

Under (11), it is known (see e.g. Brockwell and Davis [3, Theorem 10.3.1]) that

$$\mathbb{E}[r_n^2(u)] \leq Cn^{-1}.$$

Hence,

$$\mathbb{E}[|R_n(s') - R_n(s)|^2] \leq Cd_n n(s' - s)^2.$$

The rest of the proof is similar to the proof for  $S_n$ . This concludes the proof of (30).  $\square$

6.2. *Sketch of Proof of Theorem 6.* The proof consists in checking Conditions (AH1)-(AH4) with  $J_n$  and  $K_n$  now defined by  $J_n(t) = \int_0^t \{\log I_n(s) + \gamma\} ds$  and  $K(t) = \int_0^t \log f(2\pi[ns/2\pi]/n) ds$ . Let  $\lambda_k = 2k\pi/n$  denote the so-called Fourier frequencies. For  $t \in [0, \pi]$ , denote  $k_n(t) = [nt/2\pi]$ . Denote

$$\xi_k = \log\{I_n(\lambda_k)/f(\lambda_k)\} + \gamma$$

where  $\gamma$  is Euler's constant. Then

$$v_n(t) = J_n(t) - K(t) = \frac{2\pi}{n} \sum_{j=1}^{k_n(t)} \xi_j + (t - \lambda_{k_n(t)})\xi_{k_n(t)}.$$

The log-periodogram ordinates  $\xi_j$  are not independent, but sums of log-periodogram ordinates, such as the one above, behave asymptotically as sums of independent random variables with zero mean and variance  $\pi^2/6$  (cf. [26]), and bounded moments of all order. Thus, for  $t_0 \in (0, \pi)$ , the process  $\tilde{v}_n(s; t_0) = d_n^{-2}\{v_n(t_0 + d_n s) - v_n(t_0)\}$  with  $d_n = n^{-1/3}$  converges weakly in  $D(-\infty, \infty)$  to the two-sided Brownian motion with variance  $2\pi^4/3$ . It can be shown by using the moment bounds of [26] that (17) holds. Finally, if  $f$  is differentiable at  $t_0$ , it is easily seen that  $d_n^{-2}(K(t_0 + d_n s) - K(t_0) - d_n s J'_b(t_0))$  converges to  $\frac{1}{2}As^2$  with  $A = f'(t_0)/f(t_0)$ .  $\square$

**Acknowledgements.** We kindly thank Sir David Cox for suggesting the problem of estimating a spectral density under monotonicity assumptions. We would also like to thank the Associate Editor and Referees for their valuable comments.

## References.

- [1] Dragi Anevski and Ola Hössjer. A general asymptotic scheme for inference under order restrictions. *The Annals of Statistics*, 34(4):1874–1930, 2006.
- [2] Patrick Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1968.
- [3] Peter J. Brockwell and Richard A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, 1991.
- [4] Rainer Dahlhaus. Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17(4):1749–1766, 1989.
- [5] P. Laurie Davies and Arne Kovac. Densities, spectral densities and modality. *The Annals of Statistics*, 32(3):1093–1136, 2004.
- [6] Kacha Dzhaparidze. *Parameter estimation and hypothesis testing in spectral analysis of stationary time series*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
- [7] Richard D. Gill and Boris Y. Levit. Applications of the Van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, 1(1-2):59–79, 1995.
- [8] Piet Groeneboom and Jon A. Wellner. Computing Chernoff’s distribution. *Journal of Computational and Graphical Statistics*, 10(2):388–400, 2001.
- [9] Peter Hall and Chris C. Heyde. *Martingale limit theory and its application*. Academic Press, New York, 1980.
- [10] A.V. Holden. *Models of the Stochastic Activity of Neurons*. Springer Verlag, 1976.
- [11] Lars Hörmander. *Notions of convexity*. Birkhäuser, Boston, 2007.
- [12] Clifford M. Hurvich and Rohit Deo and Julia Brodsky. The mean squared error of Geweke and Porter-Hudak’s estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 19(1):19–46, 1998.
- [13] C. Kerner and P.E. Harris. Scattering attenuation in sediments modeled by ARMA processes - validation of simple Q modeles. *Geophysics* Vol. 53, 1813–1826, 1994.
- [14] Michel Ledoux and Michel Talagrand. *Probability in Banach spaces. Isoperimetry and processes*. Springer, Berlin, 1991.
- [15] Enno Mammen. Estimating a smooth monotone regression function. *The Annals of Statistics*, 19(2):724–740, 1991.
- [16] Emma J. McCoy, Andrew T. Walden and Donald B. Percival. Multitaper spectral estimation of power law processes. *IEEE Transactions of Signal Processing*, Vol. 46No. 3, 655–668, March 1998.
- [17] I. McHardy and B. Czerny. Fractal X-ray time variability and spectral invariance of the Seyfert galaxy NGC 5507 *Nature*, Vol. 325No. 3, 696–698, 1987.
- [18] Thomas Mikosch and Rimantas Norvaiša. Uniform convergence of the empirical spectral distribution function. *Stochastic Processes and their Applications*, 70(1):85–114, 1997.
- [19] Eric Moulines and Philippe Soulier. Broadband log-periodogram regression of time series with long-range dependence. *The Annals of Statistics*, 27(4):1415–1439, 1999.
- [20] W.H. Munk and G.J.F Macdonald. *The rotation of the Earth*. Cambridge University Press, 1975.
- [21] Donald B. Percival. Characterization of frequency stability: Frequency-domain estimation of stability measures. *IEEE, Proceedings*, Vol. 79, 961–972, 1991.
- [22] Donald B. Percival and Andrew T. Walden. *Spectral analysis for physical applications*. Cambridge University Press, 1993.
- [23] B. L. S. Prakasa Rao. Estimation of a unimodal density. *Sankhyā Ser. A*, 31:23–36, 1969.
- [24] Tim Robertson, F. T. Wright, and Richard L. Dykstra. *Order restricted statistical inference*. John Wiley & Sons Ltd., Chichester, 1988.
- [25] Peter M. Robinson. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics*, 23(3):1048–1072, 1995.
- [26] Philippe Soulier. Moment bounds and central limit theorem for functions of Gaussian vectors. *Statistics & Probability Letters*, 54(2):193–203, 2001.
- [27] F T. Wright. The asymptotic behavior of monotone regression estimates. *The Annals of Statistics*, 9(2):443–448, 1981.