

# Adaptive estimation of the fractional differencing coefficient

ANATOLI IOUDITSKY,<sup>1</sup> ERIC MOULINES<sup>2</sup> and PHILIPPE SOULIER<sup>3</sup>

<sup>1</sup>LMC-IMAG, 51 rue des Mathématiques, B.P. 53, 38041 Grenoble Cedex 9, France

<sup>2</sup>École Nationale des Télécommunications, 46 rue Barrault, 75005 Paris, France

<sup>3</sup>Département de Mathématiques, Université d'Evry Val d'Essonne, 91025 Evry Cedex, France.

E-mail: [soulier@maths.univ-evry.fr](mailto:soulier@maths.univ-evry.fr)

Semi-parametric estimation of the fractional differencing coefficient  $d$  of a long-range dependent stationary time series has received substantial attention in recent years. Some of the so-called local estimators introduced early on were proved rate-optimal over relevant classes of spectral densities. The rates of convergence of these estimators are limited to  $n^{2/5}$ , where  $n$  is the sample size. This paper focuses on the fractional exponential (FEXP) or broadband estimator of  $d$ . Minimax rates of convergence over classes of spectral densities which are smooth outside the zero frequency are obtained, and the FEXP estimator is proved rate-optimal over these classes. On a certain functional class which contains the spectral densities of FARIMA processes, the rate of convergence of the FEXP estimator is  $(n/\log(n))^{1/2}$ , thus making it a reasonable alternative to parametric estimators. As usual in semi-parametric estimation problems, these rate-optimal estimators are infeasible, since they depend on an unknown smoothness parameter defining the functional class. A feasible adaptive version of the broadband estimator is constructed. It is shown that this estimator is minimax rate-optimal up to a factor proportional to the logarithm of the sample size.

*Keywords:* fractional differencing; Gaussian processes; log-periodogram regression; minimax and adaptive estimation

## 1. Introduction

Let  $\{X_k\}_{k \in \mathbb{Z}}$  be a covariance stationary process whose spectral density  $f$  may be written as

$$f(x) = |1 - e^{-ix}|^{-2d} f^*(x), \quad x \in [-\pi, \pi], \quad (1.1)$$

for some  $|d| < \frac{1}{2}$  and a continuous even positive function  $f^*$ . The restriction  $|d| < \frac{1}{2}$  ensures that the process is stationary and invertible. If  $d = 0$  and if  $f^*$  is sufficiently regular, the process is said to be short-range (or weakly) dependent. If  $0 < d < \frac{1}{2}$ , the process is said to be long-range dependent. Its covariance function is not summable and many of its properties differ fundamentally from those of a short-range dependent process (see, for example, Beran 1994). When  $-\frac{1}{2} < d < 0$ , the spectral density has a zero at the zero frequency, yet the process is still invertible: such a situation occurs, for instance, when dealing with the first difference of a less than unit root process (Robinson 1994).

The parameter of greatest interest in such a model is the *fractional differencing* or *memory* parameter  $d$ . There are two natural approaches to estimating  $d$ . In the parametric

approach, a finite-dimensional model is assumed to hold for  $f^*$ . A key example is the FARIMA( $p, d, q$ ) process, introduced by Granger and Joyeux (1980), in which  $f^*$  is the spectral density of a stationary invertible ARMA( $p, q$ ) process. Another example is the fractional exponential model of order  $p$  (FEXP( $p$ )) introduced by Robinson (1994) and Beran (1993), where  $f^*$  is the exponential of a finite-order trigonometric polynomial. The parameters of  $f$ , including  $d$ , may then be estimated using Gaussian maximum likelihood (Dahlhaus 1989) or the Whittle likelihood (Fox and Taquq 1986; Giraitis and Surgailis 1990). In all these cases, the parametric estimator of  $d$  based on a sample of size  $n$  has been shown to be asymptotically normal with the usual parametric rate  $\sqrt{n}$ , if the parametric model (including the model order) is correctly specified. But such a parametric estimator may be inconsistent when the model is misspecified. This drawback provides motivation for semi-parametric estimation of the memory parameter  $d$ . The short-memory parameter  $f^*$  is assumed to belong to an infinite-dimensional class of functions, specified either by its behaviour in a neighbourhood of the zero frequency, or by some regularity condition over the whole frequency range. The former provides motivation for *local methods*, while the latter calls for *global methods*.

Examples of *local* semi-parametric estimators of  $d$  include the log-periodogram (GPH) estimator of Geweke and Porter-Hudak (1983) (see also Robinson 1995b; Hurvich *et al.* 1998), and the Gaussian semi-parametric estimator (Künsch 1987; Robinson 1995a). These estimators have the advantage that, under appropriate technical conditions, they are consistent and asymptotically normal, without any assumption on  $f^*$  (apart from integrability) outside a neighbourhood of the zero frequency. The main drawback of these methods is that the typical rate of convergence of the mean square error of the memory parameter is  $O(n^{-\alpha})$ , where  $\alpha \leq \frac{4}{5}$ , no matter how smooth the spectral density  $f^*$  is in a neighbourhood of the zero frequency (see Giraitis *et al.* 1997; Hurvich *et al.* 1998).

The global (or broadband) approach aims at constructing an estimator of  $f^*$  over the entire frequency range. To that purpose, the use of a truncated Fourier series estimate of  $\log(f^*)$  was suggested by Robinson (1994), and was later investigated by Moulines and Soulier (1999) and Hurvich and Brodsky (2001). Under an appropriate regularity condition, the short-memory component may be expressed as  $f^* = \exp(\sum_{j=0}^{\infty} \theta_j h_j)$ , where  $h_0(x) = 1/\sqrt{2\pi}$  and  $h_j(x) = \cos(jx)/\sqrt{\pi}$ ,  $j \geq 1$ . This expansion has, in general, an infinite number of non-zero coefficients. A nonparametric estimate of  $f^*$  is obtained by truncating this Fourier series and thus estimating only a finite number of Fourier coefficients. A simple solution in practice is to estimate these parameters by linear regression of the log-periodogram, yielding the so-called FEXP estimator proposed by Robinson (1994) and Beran (1993) (see Section 3).

The truncated Fourier series estimator exploits the fact that if the Fourier coefficients of a given function decay rapidly to zero, then this function will be accurately approximated by a low-order trigonometric polynomial. In the present context, it has been shown in Moulines and Soulier (1999) that the rate of convergence of the mean square error of the memory coefficient achieved by the FEXP estimator is  $L^{2/(2\beta+1)} n^{-2\beta/(2\beta+1)}$  whenever

$$\sum_{j \geq p} |\theta_j| \leq L(1+p)^{-\beta}, \quad \text{for some } \beta > 0, L < \infty. \quad (1.2)$$

As outlined by Efromovich (1998a) in the related context of spectral estimation of a short-range dependent process, it is of interest to consider situations where the Fourier coefficients are rapidly decaying at infinity. It has been shown in Moulines and Soulier (1999) that under the assumption

$$\sum_{j=p}^{\infty} |\theta_j| \leq L e^{-\beta p}, \quad \text{for some } \beta > 0, L < \infty, \quad (1.3)$$

the rate  $n^{-1} \log(n)$  can be achieved. This property is of interest since it holds for the FARIMA model, by far the most popular long-memory model. In the latter case, the FEXP semi-parametric estimator is asymptotically nearly as efficient as an estimator based on a fully specified parametric model, while avoiding the theoretical and practical pitfalls of using a misspecified model. Recall also that for these models, local estimators cannot achieve a rate of convergence better than  $O(n^{-4/5})$ . Thus, in situations where  $f^*$  is smooth, the FEXP estimator not only significantly surpasses the local estimator, but should also be preferred to a parametric estimator. The optimality of these rates of convergence is proved in Section 2, where Theorem 2.1 provides a lower bound for the mean square error of any estimator of the memory parameter when the Fourier coefficients of  $\log(f^*)$  satisfies one of the above conditions, as well as more general ones. In Section 3, we generalize the results of Moulines and Soulier (1999) on the mean square error of the FEXP estimator to these general classes, thus showing that it is minimax rate-optimal on a wide range of classes of spectral densities.

Of course, these rates are obtained under the assumption that the smoothness of  $f^*$  is known, which allows the truncation number  $p$  to be selected in an asymptotically optimal way, that is, by balancing the bias and the variance of the estimate for a given sample size  $n$ . It is not reasonable in practice to assume such prior information is available on the spectral density. For instance, if the process  $X$  is a causal invertible FARIMA( $p, d, q$ ) process, then the Fourier coefficients of  $\log(f^*)$  satisfy condition (1.3) for any positive real  $\beta$  such that all the zeros of the AR and MA polynomials fall outside the disc  $|z| \leq e^\beta$ . This clearly cannot be assumed in practice.

The alternative to such arbitrary assumptions on the spectral density is adaptive estimation, which amounts in the present context to selecting the truncation order using a data-driven procedure. This procedure should yield a rate of convergence as close as possible to the optimal minimax rate. To be more specific, it is assumed that  $\log(f^*)$  belongs to a scale of functional classes indexed by a parameter, but the exact value of this parameter is unknown. To each value of the parameter there corresponds a theoretical optimal cut-off sequence  $p_n$ , but since it is unknown, the cutoff must be chosen automatically from the data. One procedure for achieving this goal was suggested by Lepski (1990). We adapt this method to select the truncation number of the trigonometric regression estimate of the short-memory part of the spectral density. The estimator obtained by this method is nearly rate-optimal in the minimax sense, up to a logarithmic loss, which in some cases can be proved unavoidable. This adaptive estimator is presented in Section 4, and upper and lower bounds for adaptive estimation are obtained.

## 2. Minimax lower bounds

Let  $\mathcal{H}$  be the space of even square-integrable functions on  $[-\pi, \pi]$  endowed with the Lebesgue measure. For  $u \in \mathcal{H}$ , define  $\hat{u}(j) = \int_{-\pi}^{\pi} u(x)h_j(x)dx$ , where  $h_0(x) = 1/\sqrt{2\pi}$  and  $h_j(x) = \cos(jx)/\sqrt{\pi}$ ,  $j \geq 1$ . Let  $w = \{w(p)\}_{p \in \mathbb{N}}$  be a non-increasing sequence of positive numbers such that  $\lim_{p \rightarrow \infty} w(p) = 0$ . For such a sequence, we define the functional class  $\mathcal{C}(w)$  as follows

$$\mathcal{C}(w) = \left\{ u \in \mathcal{H} : \forall p \in \mathbb{N}, \sum_{j \geq p} |\hat{u}(j)| \leq 2\sqrt{\pi}w(p) \right\}.$$

The sequence  $w$  controls the rate at which the remainder of the series  $\sum_{j \geq p} |\hat{u}(j)|$  of the Fourier coefficients goes to zero as  $p$  goes to  $\infty$ . Based on the functional class  $\mathcal{C}(w)$ , we may finally define a class of spectral densities. For a sequence  $w$ , and  $\delta \in (0, \frac{1}{2})$ , define

$$\mathcal{F}(w, \delta) = \{f : f(x) = |1 - e^{ix}|^{-2d} \exp(l^*(x)), l^* \in \mathcal{C}(w), |d| < \delta\}.$$

To be able to prove the minimax rates of convergence, it is necessary to make assumptions on the sequence  $w$ .

**Assumption 1.**  $w$  is a positive non-increasing sequence, and there exists a constant  $\tau(w)$  such that

$$\tau(w) := \max_{p \geq 2} \max_{1 \leq q \leq p-1} \frac{w(p) \sum_{j=q}^{p-1} j^{-1}}{w(q)} < \infty.$$

For a sequence satisfying Assumption 1, define

$$a_w(p) = \frac{\tau(w)}{w(p)} \inf_{1 \leq q \leq p-1} \frac{w(q)}{\sum_{j=q}^{p-1} j^{-1}}, \quad (2.1)$$

$$p_n(w) = \min\{p \leq n : a_w(p)w^2(p)p^{-1} \leq n^{-1}\} - 1. \quad (2.2)$$

Assumption 1 implies that  $a_w(p) \geq 1$  for all  $p$ ,  $\lim_{p \rightarrow \infty} w(p) = 0$  and  $\lim_{p \rightarrow \infty} w(p)a_w(p) = 0$ . Thus  $p_n(w)$  is well defined, at least for large enough  $n$ , and tends to infinity with  $n$ . The following technical assumption will be needed.

**Assumption 2.**

$$\limsup_{n \rightarrow \infty} \frac{nw^2(p_n(w))}{p_n(w)} < \infty;$$

$$\forall \delta > 0, \lim_{n \rightarrow \infty} n^{-\delta} a_w(p_n(w)) = 0;$$

$$\exists \rho > 0 \text{ such that } \lim_{n \rightarrow \infty} \frac{p_n^{1+\rho}(w)}{n} = 0.$$

**Theorem 2.1** Under Assumptions 1 and 2, there exists a non-negative constant  $b(w)$  which vanishes if  $\lim_{p \rightarrow \infty} a_w(p) = \infty$  and such that, for any  $0 < \delta < \frac{1}{2}$

$$\liminf_n \inf_{\hat{d}_n} \sup_{f \in \mathcal{F}(w, \delta)} \frac{n}{p_n(w)} E_f[(\hat{d}_n - d)^2] \geq (1 + b(w))^{-1},$$

where the infimum  $\inf_{\hat{d}_n}$  is taken over all possible estimators  $\hat{d}_n$  of  $d$  based on  $n$  observations  $\{X_1, \dots, X_n\}$  of a covariance stationary process  $X = (X_t)_{t \in \mathbb{Z}}$  with spectral density  $f \in \mathcal{F}(w, \delta)$ , and  $E_f$  denotes the expectation with respect to the distribution of the process  $X$ .

**Proof.** Theorem 2.1 is proved by a standard method. A ‘hardest’ one-dimensional parametric family is constructed, to which the so-called van Trees Bayesian information bound (Gill and Levit 1995) is applied. This inequality provides a lower bound for the Bayes risk for any prior distribution verifying some technical conditions (see below). The Bayesian information bound provides a lower bound for the minimax risk. A proper choice of the prior distribution leads to the required minimax rate and lower bound.

For the sake of brevity, write  $p_n := p_n(w)$  and define  $d_n := d_n(w) = w(p_n)a_w(p_n)/\tau(w)$ . For  $n \geq 1$ , define the one-dimensional parametric family

$$\mathcal{F}_n = \left\{ f_\theta = \exp \left( \theta \sum_{j=p_n}^{\infty} \alpha_j h_j \right), \theta \in [-d_n, d_n] \right\},$$

where  $\alpha_j = 2\sqrt{\pi}/j$ . Since  $g(x) = -2 \log |1 - e^{ix}| = \sum_{j=1}^{\infty} \alpha_j h_j(x)$  for  $x \neq 0$ , we obtain that

$$\forall x \neq 0, f_\theta(x) = |1 - e^{ix}|^{-2\theta} \exp \left\{ -\theta \sum_{j=1}^{p_n-1} \alpha_j h_j(x) \right\}.$$

By construction,  $d_n \sum_{j=q}^{p_n-1} \alpha_j \leq 2\sqrt{\pi}w(q)$  for all  $q \leq p_n - 1$ , and, for any  $0 < \delta < \frac{1}{2}$ , there exists  $n_\delta < \infty$ , such that, for all  $n \geq n_\delta$ ,  $|d_n| \leq \delta$ , so that  $\mathcal{F}_n \subset \mathcal{F}(w, \delta)$ . Now let  $q$  be a continuously differentiable probability density function on the interval  $[-1, +1]$  such that  $q(-1) = q(1) = q'(-1) = q'(1) = 0$ . Assume in addition that  $q$  has a finite Fisher information  $I_q = \int_{-1}^1 (q'(\theta))^2/q(\theta) d\theta < \infty$ . Define  $q_n(\theta) = d_n^{-1}q(d_n^{-1}\theta)$ . Then, by the Bayesian information bound, we obtain

$$\begin{aligned} \inf_{\hat{d}_n} \sup_{f \in \mathcal{F}(w, \delta)} E_f[(\hat{d}_n - d)^2] &\geq \inf_{\hat{\theta}_n} \sup_{\theta \in [-d_n, d_n]} E_\theta[(\hat{\theta}_n - \theta)^2] \\ &\geq \inf_{\hat{\theta}_n} \int_{-d_n}^{d_n} E_\theta[(\hat{\theta}_n - \theta)^2] q_n(\theta) d\theta \geq \left( \int_{-d_n}^{d_n} I_n(\theta) q_n(\theta) d\theta + I_q d_n^{-2} \right)^{-1}, \end{aligned}$$

where the infimum is taken over all the possible estimators of  $\theta$ ;  $I_n(\theta)$  denotes the Fisher information of  $n$  consecutive samples  $(X_1, \dots, X_n)$  of a Gaussian stationary sequence with spectral density  $f_\theta$  and  $E_\theta$  denotes the expectation under the probability distribution of a

zero-mean Gaussian process with spectral density  $f_\theta$ . To conclude, we need to evaluate the Fisher information  $I_n(\theta)$ .

**Lemma 2.1.** *Under the assumptions of Theorem 2.1,*

$$\int_{-d_n}^{d_n} q_n(\theta) I_n(\theta) d\theta = \frac{n}{p_n} (1 + o(1)).$$

The proof of Lemma 2.1 is postponed to Section 5. It follows from the definitions of  $d_n$  and  $p_n$  that

$$p_n n^{-1} d_n^{-2} \leq \tau^2(w) a_w^{-1}(p_n).$$

Plugging this bound into the Bayesian information bound yields

$$\begin{aligned} \inf_{\hat{d}_n} \sup_{f \in \mathcal{F}(w, \delta)} \frac{n}{p_n} \mathbb{E}_f [(\hat{d}_n - d)^2] &\geq (1 + o(1) + I_q p_n d_n^{-2} n^{-1})^{-1} \\ &\geq (1 + o(1) + I_q \tau^2(w) a_w^{-1}(p_n))^{-1}. \end{aligned}$$

This concludes the proof of Theorem 2.1.  $\square$

**Remark.** It is interesting to note that the hardest one-dimensional family coincides with the family of spectral densities that cancel exactly the  $p_n(a, w)$  first coefficients of the Fourier expansion of the function  $g$ . This is quite natural since this choice of the short-memory component maximally compensates the singularity of  $g$  which carries the useful information on the memory parameter  $d$ .

To illustrate the results above, we now give two examples of bounding sequences  $w$ .

**Example 2.1** *Hyperbolically decaying coefficients.*

For any  $\beta > 0$  and  $L < \infty$ , define  $w_{\beta, L}(p) = L(1 + p)^{-\beta}$ . A straightforward application of the Cauchy–Schwarz inequality shows that there exists a constant  $c_\beta$  such that, for all  $1 \leq q < p$ ,

$$\sum_{j=q}^{p-1} j^{-1} \leq \left( \sum_{j=q}^{p-1} j^{-1+2\beta} \right)^{1/2} \left( \sum_{j=q}^{p-1} j^{-1-2\beta} \right)^{1/2} \leq c_\beta (q+1)^{-\beta} (p+1)^\beta.$$

It follows that Assumption 1 holds with  $\tau(w_{\beta, L}) \leq c_\beta$ . The sequence  $a_{w_{\beta, L}}$  is then bounded. Indeed, by definition, for all  $p \geq 2$ ,

$$a_{w_{\beta, L}}(p) \leq \frac{\tau(w_{\beta, L}) w_{\beta, L}([p/2])}{w_{\beta, L}(p) \sum_{j=[p/2]}^{p-1} j^{-1}} \leq 2.3^\beta \tau(w_{\beta, L}).$$

Finally, it is easily shown that

$$0 < \liminf_n n^{-1/(1+2\beta)} p_n(w_\beta) \leq \limsup_n n^{-1/(1+2\beta)} p_n(w_\beta) < +\infty.$$

Thus, Assumption 2 holds, and we obtain the lower bound

$$\liminf_n \inf_{\hat{d}_n} \sup_{f \in \mathcal{F}(w_{\beta,L}, \delta)} n^{2\beta/(2\beta+1)} E_f[(\hat{d}_n - d)^2] > 0. \tag{2.3}$$

The class  $\mathcal{C}(w_{\beta,L})$  is related to the more classical (in the framework of functional estimation problems) Sobolev classes, defined as

$$\mathcal{S}(\gamma, M) = \left\{ u \in \mathcal{H} : \sum_{j=0}^{\infty} (j+1)^{2\gamma} \hat{u}^2(j) \leq M \right\}, \quad \gamma > 0, M < \infty.$$

The relation

$$\sum_{j \geq p} |\hat{u}(j)| \leq \left( \sum_{j \geq p} (j+1)^{-\beta-1} \right)^{1/2} \left( \sum_{j \geq p} (j+1)^{\beta+1} \hat{u}^2(j) \right)^{1/2}$$

shows that  $\mathcal{C}(w_{\beta,L}) \subset \mathcal{S}(\beta + \frac{1}{2}, M)$ , for large enough  $M$ . The lower bound obtained here (which will be proved to be the minimax rate of convergence) over the classes  $\mathcal{C}(w_{\beta,L})$  and  $\mathcal{S}(\beta + \frac{1}{2}, M)$  is  $n^{-2\beta/(2\beta+1)}$ . This coincides with the rate recently obtained in the related context of pointwise density estimation (Butucea 1999). Note, finally, that similar statements can be obtained if one chooses  $w(p)$  to be a sequence of regular variation of index  $\beta$  at infinity (see Lepski and Levit 1998).

This class is also related to the class  $F(\beta, C_0, K_0, \delta)$  defined in Giraitis *et al.* (1997) as follows:

$$F(\beta, C_0, K_0, \delta) = \{f : f(x) = C|x|^{-2d}(1 + \Delta(x)), C_0^{-1} \leq C \leq C_0, |d| \leq \delta, |\Delta(x)| \leq K_0|x|^\beta, x \in [-\pi, \pi]\}.$$

It is shown in Section 7 that, for  $0 < \beta < 2$ , if  $l^* \in \mathcal{C}(w_{\beta,L}, \delta)$ , then  $f = \exp(dg + l^*) \in F(\beta, C_0, K_0, \delta)$ , for suitably chosen  $C_0$  and  $K_0$ . Moreover, for any  $\beta' > \beta$ , there exist functions in  $\mathcal{C}(w_{\beta,L})$  which are not in  $F(\beta', C'_0, K'_0)$  for any values of  $C'_0$  and  $K'_0$ . Thus, the lower bound proved in Giraitis *et al.* (1997) can be inferred from the bound derived from Theorem 2.1. For  $\beta = 2$ , there is no such inclusion.

**Example 2.2** Exponentially decaying coefficients.

As outlined in Section 1, it is of interest to consider functions whose Fourier coefficients typically decrease faster than a function of regular variation at infinity – see, for example, Efromovich (1998b) and Lepski and Levit (1998) for related problems in functional estimation. To keep the discussion simple, set, for any  $\alpha, \beta > 0$  and  $L < \infty$ ,

$$w_{\alpha,\beta,L}(p) = L e^{-\beta p^\alpha}$$

for all  $p \in \mathbb{N}$ . Using the same trick as above,  $\sum_{j=q}^{p-1} j^{-1}$  can be bounded by

$$\sum_{j=q}^{p-1} j^{-1} \leq \left( \sum_{j=q}^{p-1} j^{-1} e^{2\beta j^\alpha} \right)^{1/2} \left( \sum_{j=q}^{p-1} j^{-1} e^{-2\beta j^\alpha} \right)^{1/2},$$

$$\sum_{j=q}^{p-1} j^{-1} e^{-2\beta j^\alpha} \leq q^{-\alpha} \int_q^\infty x^{\alpha-1} e^{-2\beta x^\alpha} dx = \frac{q^{-\alpha} e^{-2\beta q^\alpha}}{2\beta\alpha} \leq \frac{e^{-2\beta q^\alpha}}{2\beta\alpha},$$

$$\sum_{j=q}^{p-1} j^{-1} e^{2\beta j^\alpha} \leq \int_0^p x^{\alpha-1} e^{2\beta x^\alpha} dx \leq \frac{e^{2\beta p^\alpha}}{2\beta\alpha}.$$

Thus Assumption 1 holds with  $\tau(w_{\alpha,\beta,L}) \leq (2\beta\alpha)^{-1}$ . We now give a lower and an upper bound for the sequence  $a_{w_{\alpha,\beta,L}}$ . The last bound above can be improved by considering separately the cases  $q > p/2$  and  $q \leq p/2$ . Let  $c_{\alpha,\beta}$  denote a generic positive constant which depends only on  $\alpha$  and  $\beta$  and can take different values upon each appearance. If  $q > p/2$ , then

$$\sum_{j=q}^{p-1} j^{-1} e^{2\beta j^\alpha} \leq c_{\alpha,\beta} \left(\frac{p}{2}\right)^{-\alpha} \int_{p/2}^p x^{\alpha-1} e^{2\beta x^\alpha} dx \leq c_{\alpha,\beta} p^{-\alpha} e^{2\beta p^\alpha}.$$

If  $q \leq p/2$ , then the sum is split into two parts:

$$\begin{aligned} \sum_{j=q}^{p-1} j^{-1} e^{2\beta j^\alpha} &\leq \sum_{j=q}^{\lfloor p/2 \rfloor} j^{-1} e^{2\beta j^\alpha} + \sum_{j=\lfloor p/2 \rfloor+1}^{p-1} j^{-1} e^{2\beta j^\alpha} \leq c_{\alpha,\beta} e^{2\beta(p/2)^\alpha} + c_{\alpha,\beta} p^{-\alpha} e^{2\beta p^\alpha} \\ &\leq c_{\alpha,\beta} p^{-\alpha} e^{2\beta p^\alpha} \{1 + p^\alpha e^{\beta p^\alpha} (2^{-\alpha} - 1)\}. \end{aligned}$$

For all  $\alpha > 0$  and  $\beta > 0$ ,  $\lim_{p \rightarrow \infty} p^\alpha e^{\beta p^\alpha (1-2^{-\alpha})} = 0$ , thus we also obtain, if  $q \leq p/2$ , that

$$\sum_{j=q}^{p-1} j^{-1} e^{2\beta j^\alpha} \leq c_{\alpha,\beta} p^{-\alpha} e^{2\beta p^\alpha}.$$

Altogether, we have proved that  $a_{w_{\alpha,\beta,L}}(p) \geq c_{\alpha,\beta} p^{\alpha/2}$ . Now, by definition, we also have that

$$a_{w_{\alpha,\beta,L}}(p) \leq \tau(w_{\alpha,\beta,L}) p e^{\beta(p^\alpha - (p-1)^\alpha)}.$$

It is thus easily seen that

$$\lim_{n \rightarrow \infty} \frac{2\beta p_n(w_{\alpha,\beta,L})^\alpha}{\log(n)} = 1,$$

and thus Assumption 2 holds. Since  $\lim_{p \rightarrow \infty} a_w(p) = \infty$ ,  $b(w) = 0$  and we obtain not only the minimax rate but also an explicit minimax lower bound,

$$\liminf_n n \log^{-1/\alpha}(n) \inf \sup_{\hat{d}_n, f \in \mathcal{F}(w_{\alpha,\beta})} E_f[(\hat{d}_n - d)^2] \geq (2\beta)^{-1/\alpha}.$$

Of particular importance are the classes  $\mathcal{F}(w_{1,\beta,L})$  since they contain the best-known family of parametric fractional models, namely the FARIMA( $p, d, q$ ) models. A causal stable in-



vertible FARIMA( $p, d, q$ ) process is a stationary process with spectral density  $f(x) = |1 - e^{ix}|^{-2d} |P(e^{ix})/Q(e^{ix})|^2$ , where  $P$  and  $Q$  are polynomials of degree  $p$  and  $q$  respectively, without common roots and without roots inside the closed unit disc. If, moreover,  $P$  and  $Q$  have no roots inside a disc of radius  $e^\gamma$  for some  $\gamma > 0$ , then  $f \in \mathcal{F}(w_{1,\beta,L})$  for all  $\beta < \gamma$  and a suitably chosen constant  $L$ . Indeed, if  $P$  is a polynomial of degree  $k$  with complex roots  $z_1, \dots, z_k$  (not necessarily pairwise distinct) such that  $P(0) = 1$ , then it is well known that

$$\log(P(z)) = \sum_{i=1}^k \sum_{j=0}^{\infty} \frac{z_i^{-j}}{j} z^j,$$

the series being convergent for all  $z$  such that  $|z| < \inf_{i=1,\dots,k} |z_i|$ .

It might be difficult to think of a simple model that similarly yields a spectral density with hyperbolically decaying coefficients  $\theta_j$ , but such a spectral density might arise as a consequence of the perturbation of a fractionally differenced process by an independent Gaussian white noise. Indeed, if the observed process is actually  $Y = X + \epsilon$ , where  $X$  is a stationary invertible FARIMA( $p, d, q$ ) process with  $d > 0$  and with spectral density  $f_X(x) = |1 - e^{ix}|^{-2d} f_X^*(x)$ , and  $\epsilon$  is Gaussian white noise with variance  $\sigma^2$ , then the spectral density of  $Y$  is  $f_Y(x) = |1 - e^{ix}|^{-2d} f_Y^*(x)$ , with,

$$f_Y^*(x) = f_X^*(x) + \frac{\sigma^2}{2\pi} |1 - e^{ix}|^{2d},$$

and the regularity of  $f_Y^*$  then depends on  $d$ . More precisely,  $\log(f_Y^*) \in S(\beta, L)$  for any  $\beta < \frac{1}{2} + 2d$  and a suitable choice of  $L$ .

### 3. Minimax upper bounds: the FEXP estimator

We present in this section the so-called FEXP estimator of the fractional differencing parameter presented in Robinson (1994) and Beran (1993), and later studied by Moulines and Soulier (1999) and Hurvich and Brodsky (2001). We will then determine an upper bound for the mean square error of the FEXP estimator of the memory parameter in the restricted case of a *Gaussian* process. Using Theorem 2.1, we will then show that the FEXP estimator is always rate-optimal and in certain cases quasi-efficient.

Let  $(h_{t,n})_{1 \leq t \leq n}$  be a triangular array of complex numbers. The tapered discrete Fourier transform (DFT) and the tapered periodogram of  $(X_1, \dots, X_n)$  are respectively defined as

$$d_n(x) = \left( 2\pi \sum_{t=1}^n |h_{t,n}|^2 \right)^{-1/2} \sum_{t=1}^n h_{t,n} X_t e^{itx}, \quad I_n(x) = |d_n(x)|^2.$$

Note that the dependence of  $d_n$  and  $I_n$  on the data taper is implicit. These quantities are evaluated at Fourier frequencies, defined as  $x_t = 2\pi t/n$ ,  $1 \leq t \leq [(n-1)/2]$ . Setting  $h_{t,n} = 1$ , for all  $t$  and  $n$ , yields the ordinary periodogram. For the purpose of adaptive estimation (see Section 4), it is necessary to reduce the correlation among DFT ordinates. It has already been observed that tapering is a way to partially achieve this goal; see, for

instance, Velasco (1999a; 1999b). Here we will use the data taper introduced by Hurvich and Chen (2000) (which we refer to as Hurvich’s taper) defined as

$$h_{t,n} = 1 - e^{ix_t}.$$

For both these choices of the coefficient  $h_{t,n}$ ,  $\sum_{t=1}^n h_{t,n} \exp(itx_j) = 0$  for  $1 \leq j \leq [n/2]$ . This is very important since it implies that the mean of  $\{X_t\}$  need not be estimated. This is an advantage of this data taper with respect to Kolmogorov’s taper used by Velasco (1999b), which does not share this property.

Define  $H_n(x) = \sum_{t=1}^n h_{t,n} e^{itx}$  and  $p_{k,j} = \int_{-\pi}^{\pi} H_n(x - x_k) \overline{H_n(x - x_j)} dx$ . It is well known that if no taper is used ( $h_{t,n} = 1$  for all  $t$  and  $n$ ), then  $p_{k,j} = 0$  if  $k \neq j$ . Unfortunately, if Hurvich’s taper is used, we obtain  $p_{k,j} = 0$  if  $k < j - 1$  and  $p_{k,k+1} = -n$ . This implies that the aim of decorrelation of tapered DFTs is not achieved for adjacent Fourier frequencies. It is shown in Theorem 6.1 below that for non-adjacent Fourier frequencies, the correlation between tapered DFTs is significantly less than between non-tapered DFTs. Thus, when the tapered periodogram is used, we will drop half of the frequencies in the regression – see Hurvich *et al.* (2000) for other solutions. This results in a loss of asymptotic efficiency which can be partially compensated by aggregation or pooling of periodogram ordinates. The pooled tapered periodogram is defined as

$$\bar{I}_{n,k} = \begin{cases} \sum_{t=m(k-1)+1}^{mk} I_n(x_t) & 1 \leq k \leq [(n-1)/2m] \text{ (ordinary periodogram),} \\ \sum_{t=m(k-1)+1}^{mk} I_n(x_{2t}), & 1 \leq k \leq [(n-1)/4m] \text{ (tapered periodogram).} \end{cases}$$

For the sake of simplicity, we will let  $K_n = [(n-m)/2m]$  when the ordinary periodogram is used and  $K_n = [(n-2m)/4m]$  when the tapered periodogram is used. For  $1 \leq k \leq K_n$ , let

$$Y_{n,k} = \log(\bar{I}_{n,k}) - \psi(m),$$

where  $\psi(z) = \Gamma'(z)/\Gamma(z)$  is the digamma function (see, for instance, Johnson and Kotz 1970). It is well known that if  $\chi_{2m}^2$  denotes a random variable distributed as a central chi-square with  $2m$  degrees of freedom, then  $E(\log(\chi_{2m}^2)) = \log(2) + \psi(m)$  and  $\text{var}(\log(\chi_{2m}^2)) = \psi'(m)$ . Let  $\hat{d}_{p,n}$ ,  $\hat{\theta}_0, \hat{\theta}_{p-1}$  be the least-squares estimator of  $d, \theta_0, \dots, \theta_{p-1}$ :

$$\hat{d}_{p,n} = \arg \min_{\bar{d}, \bar{\theta}_0, \dots, \bar{\theta}_{p-1}} \sum_{k=1}^{K_n} \left( Y_{n,k} - \bar{d}g(y_k) - \sum_{j=0}^{p-1} \bar{\theta}_j h_j(y_k) \right)^2,$$

where, as above,  $g(x) = -2 \log|1 - e^{ix}|$  and  $y_k = (2k - 1)\pi/2K_n$  are the so-called discrete cosine frequencies. Note that  $\hat{d}_{p,n}$  might be outside the interval  $]-\frac{1}{2}, \frac{1}{2}[$ . Being quadratic in the unknown coefficients, this minimization problem can be solved in closed form. Defining  $\tilde{g}_p^* = g - \sum_{j=0}^{p-1} \tilde{\alpha}_j h_j$ ,  $\tilde{\alpha}_j = (2\pi/K_n) \sum_{k=1}^{K_n} h_j(y_k) g(y_k)$  and  $\tilde{y}_p = (2\pi/K_n) \sum_{k=1}^{K_n} (\tilde{g}_p^*(y_k))^2$ , it is easily seen that  $\hat{d}_{p,n}$  may be expressed as

$$\hat{d}_{p,n} = \frac{2\pi}{K_n \tilde{\gamma}_p} \sum_{k=1}^{K_n} \tilde{g}_p^*(y_k) Y_{n,k}. \tag{3.1}$$

This expression in particular indicates that when the memory parameter is the sole parameter of interest, there is no need to actually compute  $(\hat{\theta}_0, \dots, \hat{\theta}_{p-1})$ , thus reducing the overall computational burden and memory requirements. This is in sharp contrast with Gaussian maximum likelihood or Whittle approximate likelihood techniques, where it is not possible to estimate  $d$  independently of the other parameters.

Recall that the spectral density  $f$  is given by  $f = \exp(dg + l^*)$  with  $l^* = \sum_{j=0}^{\infty} \theta_j h_j$  and define  $\epsilon_{n,k} = Y_{n,k} - \log(f(y_k))$ . Note that for  $j = 0, \dots, p$ ,  $2\pi K_n^{-1} \sum_{k=1}^{K_n} \tilde{g}_p^*(y_k) h_j(y_k) = 0$ , thus (3.1) implies

$$\hat{d}_{p,n} - d = \xi_{n,p} + b_{n,p}(l^*) := \frac{2\pi}{K_n \tilde{\gamma}_p} \sum_{k=1}^{K_n} \tilde{g}_p^*(y_k) \epsilon_{n,k} + \frac{2\pi}{K_n \tilde{\gamma}_p} \sum_{k=1}^{K_n} \tilde{g}_p^*(y_k) l_p^*(y_k), \tag{3.2}$$

where  $l_p^* = \sum_{j=p+1}^{\infty} \theta_j h_j$ . The term  $\xi_{n,p}$  represents a stochastic fluctuation and the bias term  $b_{n,p}(l^*)$  assesses the accuracy of the approximation by the trigonometric regression method. This bias term depends upon the unknown spectral density  $f$  only through  $l_p^*$ , but does not depend on  $d$ . Provided that the Fourier coefficients of  $l^*$  are summable, it is possible to obtain an upper bound for the bias.

**Proposition 3.1.** *Assume that  $\sum_{j=0}^{\infty} |\theta_j| < \infty$  and define  $\theta_p^* = \sum_{j=p}^{\infty} |\theta_j|$ . Then there exists a numerical constant  $c$  (independent of the sequence  $(\theta_j)_{j \in \mathbb{N}}$  and of  $n$ ) such that, for all  $p \leq K_n/5$ ,*

$$|b_{n,p}(l^*)| \leq 2\sqrt{\pi}(1 + c\sqrt{p/K_n})\theta_p^*.$$

If  $l^*$  belongs to some functional class  $\mathcal{C}(w)$ , then the above bound obviously holds with  $\theta_p^*$  replaced by  $w(p)$ . The proof of Proposition 3.1 is in Section 6.

The stochastic term is much more difficult to handle than in the traditional nonparametric regression setting with i.i.d. noise. Here the  $\epsilon_{n,k}$  are neither zero-mean nor independent, as originally observed by Künsch (1986), and their joint distribution depends on the unknown spectral density  $f$ .

However, it is possible to derive minimax and adaptive results when the spectral density  $f$  can be expressed as  $f(x) = |1 - e^{ix}|^{-2d} f^*(x)$  under appropriate smoothness assumptions for the short-memory component  $f^*$ . For  $M > 0$ , define

$$\mathcal{G}(M) = \left\{ u \in \mathbb{C}([-\pi, \pi]), u \in \mathbb{D}([-\pi, \pi] \setminus \{0\}), \max_{x \in [-\pi, \pi] \setminus \{0\}} |u(x)| + |xu'(x)| \leq M \right\}.$$

where  $\mathbb{C}([-\pi, \pi])$  denotes the set of continuous functions on  $[-\pi, \pi]$  and  $\mathbb{D}([-\pi, \pi] \setminus \{0\})$  denotes the set of functions which are differentiable on  $[-\pi, \pi] \setminus \{0\}$ . The assumption that  $xu'(x)$  is bounded was originally introduced by Robinson (1995b). It can actually be weakened to a Lipschitz condition (see Hurvich *et al.* 2000). This class is of interest since it contains the smooth part of the spectral density of an ARFIMA process  $X$  observed with

noise. Recall that the spectral density of such a process can be expressed as  $f_{\xi}^*(x) = f^*(x) + (\sigma^2/2\pi)|1 - e^{ix}|^{2d}$ , where  $f^*$  is a positive and smooth function, say at least twice continuously differentiable. It is then easily seen that  $\log(f_{\xi}^*) \in \mathcal{G}(M)$  for a suitable  $M$ . Since, moreover,  $\log(f_{\xi}^*) \in \mathcal{S}(\beta, L)$  for any  $\beta < \frac{1}{2} + 2d < \frac{3}{2}$  it is clear that this functional class is well suited to our purpose. It is also easily seen that for  $\beta > \frac{3}{2}$ ,  $\nu \geq 1$ ,  $\alpha > 0$ ,  $\gamma > 0$  and  $L > 0$ , the classes  $\mathcal{S}(\beta, L)$ ,  $\mathcal{C}(w_{\nu,L})$  and  $\mathcal{C}(w_{\alpha,\gamma,L})$  are all included in the class  $\mathcal{G}(M)$  for a suitable choice of the constant  $M$ . Now define

$$\sigma_p^2 := \frac{2\pi\psi'(m)}{K_n\tilde{\gamma}_p}. \tag{3.3}$$

$\sigma_p^2$  would be the exact variance of the stochastic term  $\xi_{n,p}$  if the noise  $\epsilon_{n,k}$  were i.i.d. centred with variance  $\psi'(m)$ . The next result states that in spite of the strong correlation between low-frequency periodogram ordinates,  $\sigma_p^2$  is a good approximation of the variance of  $\xi_{n,p}$  when  $\log(f^*) \in \mathcal{G}(M)$ , and that this approximation is uniform if, moreover  $\log(f^*) \in \mathcal{C}(w)$  for some non-increasing sequence  $w$ . Hence we introduce the following class of spectral densities. For any sequence  $w$ , for any real number  $M > 1$  and  $\delta \in (0, 1/2)$ , define

$$\mathcal{H}(M, \delta, w) = \{f = \exp(dg + l^*), l^* \in \mathcal{G}(M) \cap \mathcal{C}(w), |d| \leq \delta\}.$$

**Proposition 3.2.** *For any sequence  $w$  such that  $\lim_{n \rightarrow \infty} w(n) = 0$ , for any real numbers  $M > 1$  and  $\delta \in (0, \frac{1}{2})$ , there exists a constant  $C(M, \delta, w)$  such that*

$$\sup_{f \in \mathcal{H}(M, \delta, w)} |\mathbb{E}_f[\xi_{n,p}]| \leq C(M, \delta, w)\sigma_p^2 \log^s(n), \tag{3.4}$$

$$\sup_{f \in \mathcal{H}(M, \delta, w)} \mathbb{E}_f[\xi_{n,p}^2] \leq C(M, \delta, w)\sigma_p^2 \log^{2s}(n), \tag{3.5}$$

with  $s = 3$  for the ordinary periodogram and  $s = 2$  for the tapered periodogram. Let  $\{p_n\}_{n \in \mathbb{N}}$  be a sequence such that

$$\lim_{n \rightarrow \infty} p_n \log^6(n) K_n^{-1} = 0 \quad \text{if the ordinary periodogram is used,}$$

$$\lim_{n \rightarrow \infty} p_n \log^4(n) K_n^{-1} = 0 \quad \text{if the tapered periodogram is used.}$$

Then

$$\lim_{n \rightarrow \infty} \sup_{f \in \mathcal{H}(M, \delta, w)} |\sigma_{p_n}^{-2} \mathbb{E}_f[\xi_{n,p_n}^2] - 1| = 0.$$

Here, the supremum is evaluated over all Gaussian stationary process  $(X_t)_{t \in \mathbb{Z}}$  with spectral density  $f \in \mathcal{H}(M, \delta, w)$ .

The proof is in Section 6. Propositions 3.1 and 3.2 and the fact, for any sequence  $p_n$  such that  $\lim_{n \rightarrow \infty} p_n = \infty$  and  $\lim_{n \rightarrow \infty} p_n/n = 0$  that  $\lim_{n \rightarrow \infty} p_n \tilde{\gamma}_{p_n} = 4\pi$  yield the following Theorem.

**Theorem 3.1 (Upper bound).** *Let  $w$  be a sequence satisfying Assumptions 1 and 2. Let*

$p_n(w)$  be defined as in Theorem 2.1 and  $\hat{d}_{p_n}(w)$  be the FEXP estimator of the fractional differencing parameter, based either on the tapered or non-tapered periodogram. Recall that  $B(w) := \limsup_{n \rightarrow \infty} n w^2 (p_n(w)) / p_n(w)$  is finite under Assumption 2. If  $B(w) > 0$ , then

$$\limsup_{n \rightarrow \infty} \frac{n}{p_n(w)} \sup_{f \in \mathcal{H}(M, \delta, w)} E_f [(\hat{d}_{p_n}(w) - d)^2] \leq \varrho m \psi'(m) (1 + B(w)).$$

If  $B(w) = 0$ , then

$$\lim_{n \rightarrow \infty} \frac{n}{p_n(w)} \sup_{f \in \mathcal{H}(M, \delta, w)} E_f [(\hat{d}_{p_n}(w) - d)^2] = \varrho m \psi'(m).$$

In both cases,  $\varrho$  is a constant depending on the data taper:  $\varrho = 1$  if the ordinary periodogram is used and  $\varrho = 2$  if the tapered periodogram is used, and the supremum is evaluated over all Gaussian stationary process  $(X_t)_{t \in \mathbb{Z}}$  with spectral density  $f \in \mathcal{H}(M, \delta, w)$ .

**Remarks.**

(i) The increased value of the constant  $\varrho$  when Hurvich’s taper is used, reflects the fact that in this case, half of the Fourier frequencies are dropped from the regression; weaker loss (but at the expense of some additional technicalities) can be obtained by modifying the pooled periodogram (see Hurvich *et al.* 2000).

(ii) The upper bounds presented above hold for smaller classes of spectral density than the lower bounds of Theorem 2.1. Nevertheless, the lower bounds on these classes are the same, since the one-dimensional parametric family  $\mathcal{F}_n$  used to obtain the lower bound belongs to  $\mathcal{H}(w, \delta, M)$  for all  $M > 1$  and sufficiently large  $n$ . Thus the FEXP estimator is minimax rate-optimal in all classes  $\mathcal{H}(M, \delta, w)$ , for all sequences  $w(p)$  satisfying Assumptions 1 and 2.

(iii) By definition of  $p_n(w)$ ,  $n w^2 (p_n(w)) / p_n(w) \geq a_w^{-1}(p_n(w))$ . Thus  $B(w) = 0$  implies that  $b(w) = 0$ . Moreover, the function  $m \psi'(m)$  decreases in  $m$ , taking values  $\pi^2/6$  at  $m = 1$ , 1.289 at  $m = 2$  and 1.185 at  $m = 3$ , etc., and tends to 1 as  $m \rightarrow \infty$ . Thus, we have also proved that the lower bounds obtained in Theorem 2.1 for these classes are sharp in any class  $\mathcal{F}(w, \delta)$  such that  $B(w) = 0$ . Over such a class the FEXP estimator (with a rectangular taper) is not only rate-optimal but also quasi-efficient in the sense that, choosing the pooling number  $m$  appropriately, its loss of efficiency can be arbitrarily close to 1. This is the case, for instance, for the classes  $\mathcal{F}(w_{\alpha, \beta, L}, \delta)$  where we have, choosing  $p_n = (\log(n)/2\beta)^{1/\alpha}$  and for the rectangular taper:

$$\lim_{n \rightarrow \infty} \frac{n}{\log^{1/\alpha}(n)} \sup_{f \in \mathcal{F}(w_{\alpha, \beta, L}, \delta)} E_f [(\hat{d}_{p_n} - d)^2] = m \psi'(m) (2\beta)^{-1/\alpha}.$$

For these classes, as mentioned above, the restriction  $\log(f^*) \in \mathcal{G}(M)$  is not needed to obtain the upper bounds.

More generally, we have proved the following corollary.

**Corollary 3.1.** *Let  $w$  be a sequence such that Assumptions 1 and 2 hold and  $\lim_{n \rightarrow \infty} n w^2 (p_n(w)) / p_n(w) = 0$ . Then*

$$\liminf_{n \rightarrow \infty} \sup_{\hat{d}_n} \sup_{f \in \mathcal{H}(M, \delta, w)} \frac{n}{P_n(w)} E_f[(\hat{d}_n - d)^2] = 1.$$

We conclude this section with a brief discussion of when the GPH estimator (Geweke and Porter-Hudak 1983) should be preferred to the FEXP estimator. Since we proved above that for  $0 < \beta \leq 2$ ,  $\mathcal{F}(w_{\beta, L}, \delta) \subset F(\beta, C_0, K_0, \delta)$ , the GPH estimator, being rate-optimal over the whole class  $F(\beta, C_0, K_0, \delta)$ , is thus also rate-optimal over the class  $\mathcal{F}(w_{\beta, L}, \delta)$ . The rate alone cannot thus help to decide which estimator should be preferred. This obviously depends upon the kind of assumptions the practitioner is willing to accept to model the data at hand. If one suspects that the spectral density outside a neighbourhood of zero frequency is likely to be very irregular, then the local method should be privileged. On the other hand, if one is ready to accept a parametric model for  $f^*$  (which usually implies smoothness of  $f^*$  over all the frequency range), then the FEXP estimator is more appropriate, because it can adapt to a much wider class of regularity, especially in its adaptive version described below.

#### 4. Adaptive estimation

The rate-optimal estimators of the previous section are infeasible, since they depend on an unknown smoothness parameter defining the functional class. Thus it is necessary to select a truncation order  $\hat{p}$  automatically from the data. Classical model selection techniques have been adapted in the context of estimation of  $d$ . In the parametric framework, Beran *et al.* (1998) have proposed a method to select the order of a FARIMA( $p, d, q$ ) model, and Kokoszka and Bhansali (1999) have empirically studied a semi-parametric estimator of  $d$  which is related to the FEXP estimator in that it fits a parametric model of order  $p$ , then lets  $p$  tend to infinity with the sample size, and uses Akaike's information criterion (AIC) to select  $p$  from the data. In our opinion, these methods have two important shortcomings. The main one is perhaps that one has to decide a priori if the model is parametric or semi-parametric to choose the order selection criterion. The AIC criterion should be used in the semi-parametric setting, provided that the model is 'truly' nonparametric – for example, does not belong to any finite-order parametric class. On the other hand, the AIC yields inconsistent estimates of the order in the parametric setting; in this case the Bayesian information criterion (BIC) or other order selection techniques should be used. The second shortcoming is that these methods do not provide risk bounds such as the mean square error for the estimator of  $d$ .

The method of adaptive estimation under consideration in this paper is due to Lepski (1990). It was first developed in the context of a signal perturbed by white noise and was later extended to many statistical situations. This method has none of the above mentioned shortcomings. No prior assumption has to be made on the true order of the model (finite or infinite) and it provides a risk bound for the estimator. The theoretical cost of this method is that it necessitates an exponential inequality (Lemma 4.1 below) which can be proved (at the time of writing) only for a Gaussian process. We now describe this method.

Recall that  $\sigma_p^2 = 2\pi\psi'(m)/(K_n\hat{\gamma}_p)$  is the leading term in the variance of  $\hat{d}_p$ . Let  $\epsilon_n =$

$\log^{-4}(n)$ . For  $1 \leq p \leq \epsilon_n K_n$ , we determine the estimators  $\hat{d}_p = \hat{d}_{p,n}$  using the FEXP estimator based on the tapered periodogram. Let  $\kappa > 0$  (the value of which is specified in the statement of Theorem 4.1). Following the terminology introduced by Lepski (1990), the truncation index  $p$  is said to be *admissible* if

$$\text{for all } q \text{ such that } p \leq q \leq \epsilon_n K_n, \quad |\hat{d}_q - \hat{d}_p| \leq \kappa \sqrt{\log(K_n)} \sigma_q.$$

Let  $\hat{p}$  be the smallest admissible integer. The adaptive estimator is defined as

$$\tilde{d}_n = \hat{d}_{\hat{p}}. \tag{4.1}$$

For a spectral density  $f = \exp(dg + l^*)$  such that  $l^* = \sum_{j=0}^{\infty} \theta_j h_j$  with  $\sum_{j=0}^{\infty} |\theta_j| < \infty$ , recall that  $\theta_p^* = \sum_{j=p}^{\infty} |\theta_j|$  and define

$$p_n^*(f) = \max\{p : 1 \leq p \leq K_n, \sigma_p \sqrt{\log(K_n)} \leq 2\sqrt{\pi} \theta_p^*\} + 1. \tag{4.2}$$

The key result of this section is the following theorem.

**Theorem 4.1.** *Let  $\kappa > 6$ ,  $M > 1$  and  $0 < \delta < \frac{1}{2}$ , and let  $w$  be a sequence such that  $\lim_{n \rightarrow \infty} w(n) = 0$ . For any  $f \in \mathcal{H}(M, \delta, w)$ , there exists a constant  $C(M, \delta, \kappa, w)$  such that*

$$E_f[(\hat{d}_{\hat{p}} - d)^2] \leq 2(1 + \kappa)^2 \log(K_n) \sigma_{p_n^*(f)}^2 + C(M, \delta, \kappa, w) K_n^{-1}.$$

Since  $p_n^*(f)$  depends upon the spectral density  $f$ , this proves that this estimator is adaptive to the target function  $f$  (cf. Barron *et al.* 1999, pp. 360ff.). In particular, if the true spectral density has only finitely many coefficients, say  $p^*$ , then, for sufficiently large  $n$ ,  $p_n^* \leq p^* + 1$  and the mean square error of the adaptive FEXP estimator is of order  $\log(n)/n$ . The above bound can be extended to a functional class defined by a sequence  $w$ . Define

$$p_n^*(w) = \max\{p : 1 \leq p \leq \epsilon_n K_n, \sigma_p \sqrt{\log(K_n)} \leq w(p)\} + 1.$$

Obviously, if  $f \in \mathcal{H}(M, \delta, w)$ , then  $p_n^*(f) \leq p_n^*(w)$ , thus we have:

**Corollary 4.1.** *Let  $\kappa > 6$ ,  $M > 1$  and  $0 < \delta < \frac{1}{2}$ , and let  $w$  be a sequence such that  $\lim_{n \rightarrow \infty} w(n) = 0$ . There exists a constant  $C(M, \delta, \kappa, w)$  such that*

$$\sup_{f \in \mathcal{H}(M, \delta, w)} E_f[(\hat{d}_{\hat{p}} - d)^2] \leq 2(1 + \kappa)^2 \log(K_n) \sigma_{p_n^*(w)}^2 + c(M, \delta, \kappa, w) K_n^{-1}.$$

For the classes defined by the sequences  $w_{\beta,L}$  and  $w_{\alpha,\beta,L}$ , we obtain the following asymptotic upper bounds.

$$\limsup_n \sup_{f \in \mathcal{H}(w_{\beta,L}, \delta, M)} \left( \frac{n}{\log(n)} \right)^{2\beta/(2\beta+1)} E_f[(\hat{d}_{\hat{p}} - d)^2] \leq C(\beta, L),$$

$$\limsup_n \sup_{f \in \mathcal{F}(w_{\alpha,\beta,L}, \delta)} \frac{n}{\log^{1+1/\alpha}(n)} E_f[(\hat{d}_{\hat{p}} - d)^2] \leq C(\alpha, \beta, L).$$

It appears that the adaptive FEXP estimator achieves the optimal rate on the classes  $\mathcal{F}(w_{\beta,L}, \delta)$  and  $\mathcal{F}(w_{\alpha,\beta,L}, \delta)$  up to a factor  $\log(n)$ . The next theorem shows that this loss is unavoidable in the case  $\mathcal{F}(w_{\beta}0, L, \delta)$ .

**Theorem 4.2.** *Let  $\beta_0 > \beta > 0$ . Assume that  $\hat{d}_n$  is a data-driven estimator of  $d$  that achieves the minimax rate of convergence on the class  $\mathcal{F}(w_{\beta_0}, L, \delta)$ . Then*

$$\liminf_n \inf_{\hat{d}_n} \sup_{f \in \mathcal{F}(w_{\beta,L}, \delta)} \left( \frac{n}{\log(n)} \right)^{2\beta/(2\beta+1)} \mathbb{E}_f[(\hat{d}_n - d)^2] > 0.$$

We have not been able either to prove that this loss is unavoidable in the classes  $\mathcal{F}(w_{\alpha,\beta,L}, \delta)$ , or to obtain a better rate of convergence for an adaptive estimator. Thus it is still an open question to find the optimal adaptive rate of convergence in the classes  $\mathcal{F}(w_{\alpha,\beta,L}, \delta)$ .

In spite of this extra logarithmic loss, which anyhow is shared by the adaptive version of the GPH estimator (Giraitis *et al.* 2000), our conclusion is that if the data show no evidence of a singularity in the spectral density outside the zero frequency, the FEXP estimator should be preferred to the GPH estimator, since it can take advantage of the smoothness of the spectral density outside a neighbourhood of the zero frequency. In any case, it should be preferred to a parametric estimator, since it avoids any assumptions on the order of the model, and it is still consistent in the case of a misspecified model. Simulations (not reported here) support this claim. When the underlying model is truly parametric (say FARIMA( $p, d, q$ ), with known  $p$  and  $q$ ), a parametric estimator such as Whittle's minimum contrast estimator is better than any semi-parametric estimator. The differences are, however, in general rather small when the order of the parametric model is large. When the model order is unknown and is determined automatically by using, say, the AIC or BIC criterion, as usually suggested in the literature (without theoretical justifications), parametric and semi-parametric approaches yield comparable results (with a slight advantage to the semi-parametric model when the upper bound – used in the AIC and the BIC – for the model order is large). In the presence of even a small amount of additive noise, the parametric estimator is inconsistent.

We now prove Theorems 4.1 and 4.2.

**Proof of Theorem 4.1.** For brevity, we write  $p_n^* = p_n^*(f)$ . Write  $\mathbb{E}_f[(\hat{d}_{\hat{p}} - d)^2] = M_-^2 + M_+^2$ , with

$$M_-^2 = \mathbb{E}_f[(\hat{d}_{\hat{p}} - d)^2 \mathbf{1}_{\{\hat{p} \leq p_n^*\}}],$$

$$M_+^2 = \mathbb{E}_f[(\hat{d}_{\hat{p}} - d)^2 \mathbf{1}_{\{\hat{p} > p_n^*\}}].$$

Since by definition  $\hat{p}$  is admissible, if  $\hat{p} \leq p_n^*$ , then  $|\hat{d}_{\hat{p}} - d_{p_n^*}| \leq \kappa \sqrt{\log(K_n)} \sigma_{p_n^*}$ . Thus we obtain the following bound for  $M_-$ :



$$\begin{aligned} M_- &\leq \mathbb{E}_f^{1/2}[(\hat{d}_{\hat{p}} - d_{p_n^*})^2 \mathbf{1}_{\{\hat{p} \leq p_n^*\}}] + \mathbb{E}_f^{1/2}[(\hat{d}_{p_n^*} - d)^2] \\ &\leq \kappa \sqrt{\log(K_n)} \sigma_{p_n^*} + \mathbb{E}_f^{1/2}[(\hat{d}_{p_n^*} - d)^2]. \end{aligned}$$

Applying Propositions 3.1 and 3.2 and the definition of  $p_n^*$ , we obtain that

$$\mathbb{E}_f^{1/2}[(\hat{d}_{p_n^*} - d)^2] \leq \sigma_{p_n^*} \sqrt{\log(K_n)} (1 + O(\sqrt{\epsilon_n})),$$

and thus

$$M_- \leq (1 + \kappa) \sigma_{p_n^*} \sqrt{\log(K_n)} (1 + O(\sqrt{\epsilon_n})). \quad (4.3)$$

The term  $M_+$  requires substantially more attention. By definition,  $\hat{p} - 1$  is not admissible. Thus, there exists an integer  $p \geq \hat{p}$  such that

$$\kappa \sqrt{\log(K_n)} \sigma_p < |\hat{d}_p - \hat{d}_{\hat{p}-1}| \leq |\hat{d}_p - d| + |d - \hat{d}_{\hat{p}-1}|,$$

and this implies that either  $|\hat{d}_p - d| > \kappa \sqrt{\log(K_n)} \sigma_p / 2$  or  $|d - \hat{d}_{\hat{p}-1}| > \kappa \sqrt{\log(K_n)} \sigma_p / 2$ . Since, for each fixed  $n$ ,  $\sigma_p$  increases with  $p$ , the last inequality implies that  $|d - \hat{d}_{\hat{p}-1}| \geq \kappa \sqrt{\log(K_n)} \sigma_{\hat{p}-1} / 2$ . If  $\hat{p} > p_n^*$ , then  $\hat{p} - 1 \geq p_n^*$ , and we obtain

$$\begin{aligned} P_f(\hat{p} > p_n^*) &\leq P_f\left(\exists p \geq p_n^*, |\hat{d}_p - d| > \frac{\kappa}{2} \sqrt{\log(K_n)} \sigma_p\right) \\ &\leq \sum_{p=p_n^*}^{\epsilon_n K_n} P_f\left(|\hat{d}_p - d| > \frac{\kappa}{2} \sqrt{\log(K_n)} \sigma_p\right). \end{aligned}$$

Since  $\hat{d}_p - d = \xi_{n,p} + b_{n,p}$ ,  $|\hat{d}_p - d| > (\kappa/2) \sqrt{\log(K_n)} \sigma_p$  implies that  $|\xi_{n,p}| > (\kappa/2) \sqrt{\log(K_n)} \sigma_p - |b_{n,p}|$ . Applying Proposition 3.1 and the definition of  $p_n^*$ , we obtain, for  $p \geq p_n^*$  and for all  $f \in \mathcal{F}(w, \delta)$ ,

$$|b_{n,p}| \leq (1 + O(\sqrt{\epsilon_n})) \sigma_p \sqrt{\log(K_n)}.$$

Above and in the following, the term  $O(\sqrt{\epsilon_n})$  is uniform with respect to  $1 \leq p \leq \sqrt{\epsilon_n} K_n$ . Thus, for  $p \geq p_n^*$ ,  $|\hat{d}_p - d| > \kappa \sqrt{\log(K_n)} \sigma_p / 2$  implies that  $|\xi_{n,p}| \geq (\kappa/2 - 1 - O(\sqrt{\epsilon_n})) \sqrt{\log(K_n)} \sigma_p$ . This yields

$$P_f(\hat{p} > p_n^*) \leq \sum_{p=p_n^*}^{\epsilon_n K_n} P\left(|\xi_{n,p}| \geq \left(\frac{\kappa}{2} - 1 - O(\sqrt{\epsilon_n})\right) \sqrt{\log(K_n)} \sigma_p\right).$$

We now bound separately the probabilities in the above sum. Let  $\beta_{n,k} = \sqrt{2\pi/K_n} \tilde{\gamma}_p \tilde{g}_p^*(y_k)$ ,  $1 \leq k \leq K_n$ . Then  $\sum_{k=1}^{K_n} \beta_{n,k}^2 = 1$  and, applying (6.11) in Section 6 below, we obtain

$$\sup_{1 \leq p \leq \epsilon_n K_n} \max_{1 \leq k \leq K_n} |\beta_{n,k}| \leq C_g \log^{-1}(K_n), \quad (4.4)$$

where  $C_g$  depends only on the function  $g$ . With this notation, we can write

$$\xi_{n,p} = (\psi'(m))^{-1/2} \sigma_p \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k},$$

and

$$\begin{aligned}
 P_f \left( |\xi_{n,p}| > \left( \frac{\kappa}{2} - 1 - O(\sqrt{\epsilon_n}) \right) \sqrt{\log(K_n) \sigma_p} \right) \\
 = P_f \left( \left| \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k} \right| > \left( \frac{\kappa}{2} - 1 - O(\sqrt{\epsilon_n}) \right) \sqrt{\frac{\log(K_n)}{\psi'(m)}} \right).
 \end{aligned}$$

Set  $x = (\kappa/2 - 1 - O(\sqrt{\epsilon_n}))\sqrt{\log(K_n)/\psi'(m)}$ . For any  $\alpha > 0$ , the Markov inequality yields

$$P_f \left( \left| \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k} \right| > x \right) 2 \leq e^{-\alpha x} E_f \left[ \exp \left( \alpha \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k} \right) \right] \tag{4.5}$$

We now need the following exponential inequality.

**Lemma 4.1.** *Let  $\rho > 0$ ,  $\delta \in (0, \frac{1}{2})$ ,  $M > 1$  and let  $w$  be a sequence such that  $\lim_{n \rightarrow \infty} w(n) = 0$ . There exist a real  $\eta(\rho) > 0$  which depends only on  $\rho$ , and a constant  $C(M, \delta, \rho, w)$ , such that, for any real number  $\alpha$  and any triangular array of real numbers  $(\beta_k)_{1 \leq k \leq K_n}$  satisfying  $\sum_{k=1}^{K_n} \beta_k^2 = 1$  and  $|\alpha| \max_{1 \leq k \leq K_n} |\beta_k| \leq \eta(\rho)$ ,*

$$\sup_{f \in \mathcal{H}(M, \delta, w)} E_f \left[ \exp \left( \alpha \sum_{k=1}^{K_n} \beta_k \epsilon_{n,k} \right) \right] \leq C(M, \delta, \rho, w) e^{\alpha^2 \psi'(m)(1+\rho)/2}.$$

Using Lemma 4.1 and setting  $\alpha = x/\psi'(m)(1 + \rho)$ , (4.5) becomes

$$P_f \left( \left| \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k} \right| > x \right) \leq C(M, \delta, \rho, w) e^{-x^2/2\psi'(m)(1+\rho)}.$$

This choice of  $\alpha$  is valid for large enough  $n$  thanks to (4.4). Thus we obtain, for any  $\rho > 0$ , for  $n$  large enough (depending on  $M, \delta, \rho$  and  $w$ ) and for all  $f \in \mathcal{H}(M, \delta, w)$ ,

$$P_f \left( |\xi_{n,p}| > \left( \frac{\kappa}{2} - 1 - O(\sqrt{\epsilon_n}) \right) \sqrt{\log(K_n) \sigma_p} \right) \leq C(M, \delta, \rho, w) K_n^{-(\kappa/2 - 1 - O(\sqrt{\epsilon_n}))^2/2(1+\rho)}.$$

Applying Hölder’s inequality, we obtain, for  $a, b > 0$  such that  $a^{-1} + b^{-1} = 1$ ,

$$E_f [(\hat{d}_{\hat{p}} - d)^2 \mathbf{1}_{\{\hat{p} > p_n^*\}}] \leq \left( \sum_{p_n^* \leq p < \epsilon_n K_n} E_f [(\hat{d}_p - d)^{2a}] \right)^{1/a} P_f^{1/b}(\hat{p} > p_n^*).$$

**Lemma 4.2.** *For any real  $a \geq 1$ , there exists a constant  $C(M, \delta, a)$  such that*

$$\sup_{1 \leq p \leq \epsilon_n K_n} \sup_{f \in \mathcal{H}(\delta, M, w)} E_f [(\hat{d}_p - d)^{2a}] \leq C(M, \delta, a, w). \tag{4.6}$$

We finally obtain that for large enough  $n$  (still depending on  $M, \delta, \rho$  and  $\kappa$ )

$$\sup_{f \in \mathcal{H}(\delta, M, w)} E_f[(\hat{d}_{\hat{p}} - d)^2 \mathbf{1}_{\{\hat{p} > p_n^*\}}] \leq c(M, \delta, \rho, \kappa, b, w) K_n^{1 - (\kappa/2 - 1 - \sqrt{\epsilon_n})^2 / 2b(1 + \rho)}.$$

Since it has been assumed that  $\kappa > 6$ , the choice of  $\rho$  and  $b$  can be made in such a way that  $(\kappa/2 - 1 - O(\sqrt{\epsilon_n}))^2 / 2b(1 + \rho) > 2$ . Such a choice finally yields  $M^+ \leq C(M, \delta, \kappa, w) K_n^{-1}$  for all  $f \in \mathcal{H}(\delta, M, w)$  and for  $n$  greater than some  $n_0(M, \delta, \kappa, w)$  which depends only on  $M, \delta, \kappa$  and  $w$ . Gathering the bounds for  $M_-$  and  $M_+$  concludes the proof of Theorem 4.1.  $\square$

**Proof of Theorem 4.2.** The proof of Theorem 4 is a straightforward application of the following constrained risk inequality:

**Proposition 4.1. (Brown and Low 1996).** *Let  $Z$  have distribution with density  $f_{\theta_0}$  or  $f_{\theta_1}$  with respect to a measure  $\mu$ . Define  $q = f_{\theta_1}/f_{\theta_0}$  and  $I = E_{\theta_0}[q^2(Z)]$ . Let  $R(\delta, \theta_i) = E_{\theta_i}[(\theta_i - \delta)^2]$  be the risk of an estimator  $\delta$  based on  $Z$ . If  $R(\theta_0, \delta) \leq \epsilon^2$  and  $I\epsilon^2/(\theta_1 - \theta_0)^2 < 1$ , then*

$$R(\theta_1, \delta) \geq (\theta_1 - \theta_0)^2 (1 - I^{1/2}\epsilon|\theta_1 - \theta_0|^{-1})^2.$$

Let  $\mathcal{C}_0$  be a class of spectral densities that contains the spectral density of white noise and upon which the optimal rate of convergence for the estimation of  $d$  is  $\epsilon_n$ . Let  $(q_n)_{n \in \mathbb{N}}$  be a non-decreasing sequence of integers and  $(\theta_n)_{n \in \mathbb{N}}$  be a decreasing sequence of reals, and define  $f_n = \exp\{\theta_n \sum_{j=q_n}^{\infty} \alpha_j h_j\}$ . Let  $X$  be a Gaussian process with spectral density  $f_n$  and let  $p_n$  be the density of  $X_1, \dots, X_n$ . Let  $g_n$  be the density of  $n$  independent  $N(0, 1)$  random variables  $Z_1, \dots, Z_n$ . Let  $I_n = E[\{p_n(Z_1, \dots, Z_n)/g_n(Z_1, \dots, Z_n)\}^2]$  and assume that  $\xi := \limsup_n I_n \epsilon_n^2 \theta_n^{-2} < 1$ . Let  $\hat{d}_n$  be a data-driven estimator (based on  $X_1, \dots, X_n$ ) which achieves the optimal minimax rate of convergence on the class  $\mathcal{C}_0$ . Then, applying Proposition 4.1, we obtain, for large enough  $n$ ,

$$E[(\hat{d}_n - \theta_n)^2] \geq \theta_n^2 (1 - \epsilon_n I_n^{1/2} \theta_n^{-1})^2.$$

Consider now a class of spectral densities  $\mathcal{C}$  such that for large enough  $n$ ,  $f_n \in \mathcal{C}$ . Then

$$\liminf_n \sup_{f \in \mathcal{C}} \theta_n^{-2} E_f[(\hat{d}_n - d)^2] \geq (1 - \sqrt{\xi})^2.$$

We now set  $\mathcal{C}_0 = \mathcal{F}(w_{\beta} 0, L, \delta)$  and  $\mathcal{C} = \mathcal{F}(w_{\beta, L}, \delta)$ , with  $\beta_0 > \beta > 0$ . Note that in that case,  $\epsilon_n = n^{\beta_0/(2\beta_0+1)}$ . Define  $q_n$  as the smallest integer greater than  $(n/\log(n))^{1/(2\beta+1)}$  and  $\theta_n = \gamma(\log(n)/n)^{\beta/(2\beta+1)}$ , where  $\gamma$  will be chosen later. Using the same kind of computations as in the proof of Theorem 2.1 and Lemma 2.1, it is easily shown that  $f_n \in \mathcal{F}(w_{\beta, L}, \delta)$  for large enough  $n$  and for small enough  $\gamma > 0$ , and that  $I_n = \exp\{2n\theta_n^2 q_n^{-1}(1 + o(1))\} = n^{2\gamma^2}(1 + o(1))$ . Thus

$$\frac{I_n \epsilon_n^2}{\theta_n^2} = n^{2\gamma^2 - 2\beta_0/(2\beta_0+1) + 2\beta/(2\beta+1)} \log(n)^{-2\beta/(2\beta+1)} (1 + o(1)).$$

Choosing  $0 < \gamma^2 \leq 2\beta_0/(2\beta_0+1) - 2\beta/(2\beta+1)$  yields  $\limsup_n I_n \epsilon_n^2 \theta_n^{-2} = 0$ . This concludes the proof of Theorem 4.2.  $\square$

### 5. Proof of Lemma 2.1

In the rest of this paper,  $c, C, \dots, c(\cdot)$  denote generic constants that may take different values at each appearance and depend only on their arguments when any. For brevity, write  $p_n := p_n(w)$ . The Fisher information quantity for  $n$  consecutive samples  $\{X_1, \dots, X_n\}$  of a zero-mean Gaussian process, under a regular parametric model  $(f_\theta, \theta \in \mathbb{R})$  may be expressed as

$$I_n(\theta) = \frac{1}{2} \text{tr}((\Sigma_\theta^{-1} \Sigma'_\theta)^2),$$

where  $\Sigma_\theta = (\Sigma_\theta(i, j))_{1 \leq i, j \leq n}$  is the Toeplitz matrix associated with  $f_\theta$ , that is,

$$\Sigma_\theta(j, k) = \int_{-\pi}^{\pi} f_\theta(x) e^{i(j-k)x} dx = \begin{cases} \sqrt{2\pi} \hat{f}_\theta(0) & \text{if } j = k, \\ \sqrt{\pi} \hat{f}_\theta(|j - k|) & \text{if } j \neq k, \end{cases}$$

and  $\Sigma'_\theta$  is its derivative with respect to  $\theta$ .

Recall now that  $f_\theta = \exp(\theta g_{p_n(a,w)}^*)$ , where  $g_p^* = \sum_{s=p}^\infty \alpha_s h_s$ . Note that with this normalization, we have  $\Sigma_\theta = 2\pi J_n + \Delta_n$ , where  $J_n$  is the  $n \times n$  identity matrix, and we must prove that the spectral radius of  $\Delta_n$ , say  $\rho_n^*$ , tends to zero as  $n$  tends to infinity. Standard results on the spectral radius of Toeplitz matrices yield the bound

$$\rho_n^* \leq 2|\sqrt{2\pi} \hat{f}_\theta(0) - 2\pi| + 2\sqrt{\pi} \sum_{k=1}^n |\hat{f}_\theta(k)|.$$

Since the functions  $h_j$  form an orthonormal sequence, the Parseval–Bessel identity implies

$$\|g_p^*\|_2^2 = \int_{-\pi}^{\pi} (g_p^*(x))^2 dx = \sum_{j=p} \alpha_j^2 \leq 8\pi p^{-1}.$$

Since  $g \in L^q([-\pi, \pi])$  for all  $q \geq 1$ , it follows from Katznelson (1968, Theorem 11.1.5, p. 50), that  $\lim_{p \rightarrow \infty} \|g_p^*\|_q = 0$  (where  $\|\cdot\|_q$  denotes the  $L^q$ -norm). Let  $q > 2$  and  $0 < \epsilon < \frac{1}{2}$ . Applying Hölder’s inequality, we obtain

$$\begin{aligned} \int_{-\pi}^{\pi} |g_p^*(x)|^q dx &= \int_{-\pi}^{\pi} |g_p^*(x)|^{2-2\epsilon} |g_p^*(x)|^{q+2\epsilon-2} dx \\ &\leq \left( \int_{-\pi}^{\pi} |g_p^*(x)|^2 dx \right)^{1-\epsilon} \left( \int_{-\pi}^{\pi} |g_p^*(x)|^{(q+2\epsilon-2)/\epsilon} dx \right)^\epsilon \leq c_1(q, \epsilon) p^{-1+\epsilon}. \end{aligned} \tag{5.1}$$

Now, for large enough  $n$ ,  $|\theta| \leq d_n$  and  $p \leq n$ , we also have

$$|\theta g_p^*| \leq d_n g + 2d_n \sum_{j=1}^p j^{-1} \leq \frac{g}{4} + 1. \tag{5.2}$$

Equations (5.1), (5.2) and the Cauchy–Schwarz inequality imply that, for all  $\epsilon > 0$  and all integers  $k > 1$ , there exists a constant  $c_2(k, \epsilon)$  such that, for large enough  $n$ , for all  $|\theta| \leq d_n$  and for all  $p \leq n$ ,

$$\int_{-\pi}^{\pi} |g_p^*(x)|^k e^{\theta g_p^*(x)} dx \leq c_2(k, \epsilon) p^{-1/2+\epsilon}. \tag{5.3}$$

We now have all the necessary ingredients to determine an upper bound for  $\rho_n^*$ . For  $r \in \mathbb{N}$ , write  $\phi_{n,r} := [g_{p_n}^*]^r$ . Using

$$\left| e^x - \sum_{j=0}^k \frac{x^j}{j!} \right| \leq e^{\max(x,0)} \frac{|x|^{k+1}}{(k+1)!},$$

together with (5.2) and (5.3), we obtain, for large enough  $n$ , for all  $|\theta| \leq d_n$  and  $p \leq n$ ,

$$\begin{aligned} \rho_n^* &\leq 2 \sum_{k=0}^n \left| \int_{-\pi}^{\pi} (e^{\theta g_{p_n}^*(x)} - 1) e^{ikx} dx \right| \\ &\leq 2 \sum_{r=1}^s \frac{\theta^r}{r!} \sum_{k=0}^n \left| \int_{-\pi}^{\pi} \phi_{n,r}(x) e^{ikx} dx \right| + 2 \frac{c_2(s+1, \epsilon) n \theta^{s+1} p_n^{-1/2+\epsilon}}{(s+1)!} \\ &\leq 4\sqrt{\pi} d_n \sum_{k=p_n}^n \alpha_k + 2 \sum_{r=2}^s \frac{\sqrt{\pi} d_n^r}{r!} \sum_{k=0}^n |\hat{\phi}_{n,r}(k)| + 2 \frac{c_2(s+1, \epsilon) n \theta^{s+1} p_n^{-1/2+\epsilon}}{(s+1)!}. \end{aligned}$$

Applying the Cauchy–Schwarz and Parseval–Bessel inequalities and using (5.1), we obtain, for any  $\epsilon > 0$ ,

$$\begin{aligned} \sum_{k=0}^n |\hat{\phi}_{n,r}(k)| &\leq \sqrt{n} \left( \sum_{k=0}^n [\hat{\phi}_{n,r}(k)]^2 \right)^{1/2} \\ &\leq \sqrt{n} \left\{ \int_{-\pi}^{\pi} \phi_{n,2r}(x) dx \right\}^{1/2} \leq c_1^{1/2}(2r, \epsilon) \sqrt{n} p_n^{-1/2+\epsilon}. \end{aligned}$$

As a result, we obtain

$$\rho_n^* = O(d_n \log(n) + \sqrt{n} d_n^2 p_n^{-1/2+\epsilon} + n d_n^{s+1} p_n^{-1/2+\epsilon}).$$

Assumption 2 implies that, for all  $\delta > 0$ , there exists a constant  $c(a, w, \delta)$  such that, for large enough  $n$ ,  $d_n \leq c(a, w, \delta) p_n^{1/2} n^{-1/2+\delta}$ . Assumption 2 now yields

$$\begin{aligned} d_n \log(n) &\leq c(a, w, \delta, \rho) n^{-1/2+(1+\rho)/2+\delta}, \\ \sqrt{n} d_n^2 p_n^{-1/2+\epsilon} &\leq c(a, w, \delta, \rho) n^{-1/2+2\delta+\epsilon/(1+\rho)}, \\ n d_n^{s+1} p_n^{-1/2+\epsilon} &\leq c(a, w, \delta, \rho) n^{-s/2+1/2+\delta(s+1)+\epsilon/(1+\rho)}. \end{aligned}$$

If  $\epsilon$  and  $\delta$  are chosen small enough and  $s$  is chosen large enough, then all the above terms are  $o(1)$  and thus  $\rho_n^* = o(1)$ .

For large enough  $n$ ,  $\rho_n^* := \rho(\Delta_n) < 2\pi$  (where  $\rho(\cdot)$  denotes the spectral radius), and  $\Sigma_\theta^{-1} = (2\pi)^{-1} J_n + E_n$ , with  $E_n := \sum_{k=1}^\infty (2\pi)^{-k} \Delta_n^k$ . Hence,  $\rho(E_n) \leq c \rho_n^*$ , for some finite constant  $c$ . Using the bound for symmetric matrices  $|\text{tr}(A^2 B)| \leq \rho(B) \text{tr}(A^2)$ , we obtain

$$\begin{aligned} 8\pi^2 I_n(\theta) &= \frac{1}{8\pi^2} \text{tr}[(\Sigma'_\theta + 2\pi\Sigma'_\theta E_n)^2] = \{\text{tr}(\Sigma'^2_\theta) + 4\pi\text{tr}(\Sigma'^2_\theta E_n) + 4\pi^2\text{tr}(\Sigma'^2_\theta E_n^2)\} \\ &= \text{tr}(\Sigma'^2_\theta)(1 + O(\rho_n^*)). \end{aligned}$$

We now compute the entries of  $\Sigma'_\theta$ . Using the same technique as above, we obtain

$$\Sigma'_\theta(i, j) = \sqrt{\pi}\alpha_{|i-j|}\mathbf{1}_{\{|i-j|>p_n\}} + \sum_{r=1}^s \frac{\sqrt{\pi}\theta^r}{r!} \hat{\phi}_{n,r+1}(|i-j|) + O(d_n^{s+1} p_n^{-1/2+\epsilon}),$$

where the  $O(d_n^{s+1} p_n^{-1/2+\epsilon})$  term is uniform with respect to  $|\theta| \leq d_n$ . Using straightforward algebra, we may show that

$$\sum_{|i-j|>p_n} \alpha_{|i-j|}^2 = 2n \sum_{k=p_n}^{n-1} + 1 \left( \frac{1-k}{n} \right) \alpha_k^2 = 8\pi n p_n^{-1} (1 + o(1)).$$

Applying the Parseval–Bessel inequality, we have

$$\sum_{1 \leq i, j \leq n} \hat{\phi}_{n,r+1}^2(|i-j|) \leq 2n \sum_{s=0}^{n-1} \hat{\phi}_{n,r+1}^2(s) \leq 2n \int_0^\pi \phi_{n,2(r+1)}(x) dx \leq 2c_1(2(r+1), \epsilon) n p_n^{-1+\epsilon}.$$

Hence, for all  $|\theta| \leq d_n$ ,

$$\sum_{1 \leq i, j \leq n} \left( \sum_{r=1}^s \frac{\sqrt{\pi}\theta^r}{r!} \hat{\phi}_{n,r}(|i-j|) \right)^2 \leq c n p_n^{-1} \sum_{r=1}^s d_n^{2r} p_n^\epsilon$$

for some finite constant  $c$ . Choosing, as above,  $\epsilon$  sufficiently small and  $s$  sufficiently large, we obtain

$$\text{tr}(\Sigma'^2_\theta) = 2\pi n \left( \sum_{k=p_n}^n (1 - k/n) \alpha_k^2 \right) (1 + o(1)) = 8\pi^2 n p_n^{-1} (1 + o(1)),$$

where the  $o(1)$  is uniform for  $|\theta| \leq d_n$ . This concludes the proof of Lemma 2.1.

## 6. Proofs of Propositions 3.1 and 3.2

**Proof of Proposition 3.1.** For notational convenience, we use the following convention. We identify any function  $\varphi$  with the  $K_n$ -dimensional vector  $(\varphi(y_1), \dots, \varphi(y_{K_n}))^\top$ . We define a scalar product on  $\mathbb{R}^{K_n}$  as follows: for any vectors  $u, v$  in  $\mathbb{R}^{K_n}$ ,

$$\langle u, v \rangle = \frac{2\pi}{K_n} \sum_{k=1}^{K_n} u_k v_k.$$

With this notation, we have  $\langle h_j, h_{j'} \rangle = \delta_{j,j'}$  for all  $0 \leq j, j' < K_n$ , where  $\delta$  is the Kronecker delta. Define  $g_p^* = \sum_{j=p}^\infty \alpha_j h_j$ . Then  $\tilde{g}_p^* = g_p^* - \sum_{j=0}^{p-1} \langle h_j, g_{K_n}^* \rangle h_j$ . It is easily seen that

$$\begin{aligned} \langle \tilde{g}_p^*, l_p^* \rangle &= \langle g_p^*, l_p^* \rangle - \sum_{j=0}^{p-1} \langle h_j, g_{K_n}^* \rangle \langle h_j, l_p^* \rangle \\ &= \sum_{j=p}^{K_n-1} \alpha_j \langle h_j, l_p^* \rangle + \langle g_{K_n}^*, l_p^* \rangle - \sum_{j=0}^{p-1} \langle h_j, g_{K_n}^* \rangle \langle h_j, l_p^* \rangle, \end{aligned} \tag{6.1}$$

$$= \sum_{j=p}^{K_n-1} \alpha_j \langle h_j, l_p^* \rangle + \sum_{j=p}^{K_n-1} \langle g_{K_n}^*, h_j \rangle \langle h_j, l_p^* \rangle. \tag{6.2}$$

For  $j \in \{1, \dots, K_n\}$  and  $j' \geq K_n$ ,

$$\langle h_j, h_{j'} \rangle = \begin{cases} 1 & j - j' = 0 \pmod{2K_n}, \\ -1 & j + j' = 0 \pmod{2K_n}, \\ 0 & \text{otherwise.} \end{cases}$$

This implies the bounds

$$|\langle h_j, g_{K_n}^* \rangle| \leq \alpha_{2K_n - j} \leq \alpha_{K_n}, \quad j \in \{0, \dots, K_n - 1\}, \tag{6.3}$$

$$\sum_{j=0}^{K_n-1} |\langle h_j, l_p^* \rangle| \leq \theta_p^*. \tag{6.4}$$

Plugging these bounds into (6.2) yields

$$|\langle \tilde{g}_p^*, l_p^* \rangle| \leq 2\sqrt{\pi} p^{-1} \left( \frac{1+p}{K_n} \right) \theta_p^*.$$

We must now give a lower and an upper bound and an asymptotic equivalent for  $p\tilde{\gamma}_p$ . Note that  $\tilde{g}_p = \sum_{j=p}^{K_n-1} \alpha_j h_j + \sum_{j=p}^{K_n-1} \langle g_{K_n}^*, h_j \rangle h_j$ . Thus,

$$\left( \sum_{j=p}^{K_n-1} \alpha_j^2 \right)^{1/2} - \left( \sum_{j=p}^{K_n-1} \langle g_{K_n}^*, h_j \rangle^2 \right)^{1/2} \leq \|\tilde{g}_p\| \leq \left( \sum_{j=p}^{K_n-1} \alpha_j^2 \right)^{1/2} + \left( \sum_{j=p}^{K_n-1} \langle g_{K_n}^*, h_j \rangle^2 \right)^{1/2}.$$

Applying (6.3), we have  $\sum_{j=p}^{K_n-1} \langle g_{K_n}^*, h_j \rangle^2 \leq 4\pi/K_n$  and

$$\left( p \sum_{j=p}^{K_n-1} \alpha_j^2 \right)^{1/2} - 2\sqrt{\pi} \left( \frac{p}{K_n} \right)^{1/2} \leq (p\tilde{\gamma}_p)^{1/2} \leq \left( p \sum_{j=p}^{K_n-1} \alpha_j^2 \right)^{1/2} + 2\sqrt{\pi} \left( \frac{p}{K_n} \right)^{1/2}.$$

This yields, for  $1 \leq p \leq K_n$ ,

$$4\pi \left( 1 - 2\sqrt{\frac{p}{K_n}} \right) \leq p\tilde{\gamma}_p \leq 4\pi p \sum_{j=p}^{\infty} j^{-2} + 12\pi \frac{p}{K_n}. \tag{6.5}$$

Both bounds always hold, but the lower bound is meaningful only if  $p \leq K_n/4$ . This implies that if  $p$  is a non-decreasing sequence of integers such that  $p_n \leq K_n$  and  $\lim_{n \rightarrow \infty} (1/p_n + p_n/K_n) = 0$ , then  $\lim_{n \rightarrow \infty} p_n \tilde{\gamma}_{p_n} = 4\pi$ . The proof of Proposition 3.1 finally follows from  $\tilde{\gamma}_p b_{n,p} = \langle \tilde{g}_p^*, l_p^* \rangle$ .  $\square$

To prove Proposition 3.2 and Lemmas 4.1 and 4.2, we need covariance bounds for the ordinary and tapered discrete Fourier transforms computed at the Fourier frequencies. These bounds were obtained in Moulines and Soulier (1999) in the case of the ordinary DFTs and in Hurvich *et al.* (2000) in the case of the tapered DFTs.

**Theorem 6.1.** *Let  $M > 1$  and  $\delta \in (0, \frac{1}{2})$ . Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary Gaussian process with spectral density  $f = e^{dg+l^*}$ , where  $l^* \in \mathcal{G}(M)$  and  $|d| \leq \delta$ . For  $k = 1, \dots, \tilde{n}$ , define  $\zeta_{n,k} = d_n(x_k)/\sqrt{f(x_k)}$ . Then, there exists a constant  $C(M, \delta)$  such that*

$$|E_f[\zeta_{n,k} \overline{\zeta_{n,k}}] - 1| + |E_f[\zeta_{n,k} \zeta_{n,k}]| \leq C(M, \delta)p(k), \tag{6.6}$$

$$|E_f[\zeta_{n,k} \overline{\zeta_{n,j}}]| + |E_f[\zeta_{n,k} \zeta_{n,j}]| \leq C(M, \delta)p(d, k, j), \tag{6.7}$$

with  $p(k) = \log(1+k)/k$  and  $p(d, k, j) = \log(j)k^{-|d|}j^{|d|-1}$  for  $j > k$  and for the ordinary DFTs, and with  $p(k) = k^{-1}$  and  $p(d, k, j) = k^{-1}(j-k)^{-2}(j/k)^{|d|}$  for  $j > k+1$  and for the tapered DFTs.

Since the log-periodogram ordinates are functionals of the Gaussian vector  $[\zeta_{n,1}, \dots, \zeta_{n,K}]$ , moment bounds for log-periodogram ordinates can be obtained from expressions for the moment bounds of functions of Gaussian vectors, such as the covariance inequality of Arcones (1994, Lemma 1) later extended by Soulier (1998, Theorem 1). Noting that the function defined on  $\mathbb{R}^{2m}$  as  $x \rightarrow \log(|x|^2/2)$  has Hermite rank 2 (as defined in Arcones 1994), we obtain the following corollary.

**Corollary 6.1.** *Let  $M > 1$ ,  $\delta \in (0, \frac{1}{2})$  and let  $w$  be a sequence such that  $\lim_{n \rightarrow \infty} w(n) = 0$ . Let  $\{X_t\}_{t \in \mathbb{Z}}$  be a stationary Gaussian process with spectral density  $f \in \mathcal{H}(M, \delta, w)$ . Then there exists a constant  $C(M, \delta, w)$  such that, for all  $1 \leq k < j \leq K_n$ ,*

$$|E_f[\epsilon_k]| + |E_f[\epsilon_k^2] - \psi'(m)| \leq C(M, \delta, w)p(k), \tag{6.8}$$

$$|\text{cov}_f(\epsilon_k, \epsilon_j)| \leq C(M, \delta, w)p^2(d, k, j). \tag{6.9}$$

We will make also repeated use of the following technical result which is easily checked and whose proof is omitted.

**Lemma 6.1.** *For all  $\delta \in (0, \frac{1}{2})$ , there exists a constant  $C(\delta)$  such that, for any integers  $K > L > 0$  and any real  $|d| < \delta$ ,*

$$\sum_{j=L}^K \sum_{k=L}^{j-1} p^2(d, j, k) \leq \begin{cases} C(\delta)\log^3(K), & \text{for the ordinary DFTs} \\ C(\delta)L^{-1}, & \text{for the tapered DFTs.} \end{cases} \tag{6.10}$$

**Proof of Proposition 3.2.** We first need to obtain an upper bound for  $\tilde{g}_p^*(y_k)$ . Recall that  $\tilde{g}_p^* = g_p^* - \sum_{j=0}^{p-1} \langle h_j, g_K^* \rangle h_j$  and that  $|\langle h_j, g_K^* \rangle| \leq \alpha_K$ . Hence,

$$|\tilde{g}_p^*(y_k)| \leq |g_p^*(y_k)| + 2pK^{-1}.$$



Since  $g_p^* = g - \sum_{j=1}^{p-1} a_j h_j$ , we obtain, for all  $1 \leq p \leq K$ ,

$$\max_{1 \leq k \leq K} |g_p^*(y_k)| \leq \max_{1 \leq k \leq K} |g(y_k)| + 2 \sum_{j=1}^{p-1} j^{-1} \leq c \log(K),$$

for some finite constant  $c$ . Combining these two bounds yields, under the stated assumptions,

$$\max_{1 \leq p \leq \epsilon_n K} \max_{1 \leq k \leq K} |\tilde{g}_p^*(y_k)| \leq c \log(K). \tag{6.11}$$

Define  $\nu_k = 2\pi \tilde{g}_p^*(y_k)/(K_n \tilde{\gamma}_p)$ . Then  $\sum_{k=1}^{K_n} \nu_k^2 = \sigma_p^2 / \psi'(m)$  and (6.11) implies that  $|\nu_k| \leq c \log(K_n) \sigma_p^2 / \psi'(m)$ . With this notation, we obtain

$$\begin{aligned} E_f[\xi_{n,p}^2] &= (E_f[\xi_{n,p}])^2 + \text{var}_f(\xi_{n,p}) \\ &= (E_f[\xi_{n,p}])^2 + \sum_{k=1}^{K_n} \nu_k^2 \text{var}_f(\epsilon_{n,k}) + 2 \sum_{1 \leq k < j \leq K_n} \nu_k \nu_j \text{cov}_f(\epsilon_{n,k}, \epsilon_{n,j}). \end{aligned}$$

Applying (6.8) and (6.11), we obtain

$$|E_f[\xi_{n,p}]| \leq C(M, \delta, w) \sigma_p^2 \log(K_n) \sum_{k=1}^{K_n} p(k) \leq C(M, \delta, w) \sigma_p^2 \log^\nu(n),$$

with  $\nu = 6$  for the ordinary periodogram and  $\nu = 4$  for the tapered periodogram. Applying (6.8) again, we obtain

$$\sum_{k=1}^{K_n} \nu_k^2 \text{var}(\epsilon_{n,k}) = \sigma_p^2 + C(M, \delta, w) \sigma_p^4 \log^2(K_n) \sum_{k=1}^{K_n} p(k) = \sigma_p^2 (1 + C(M, \delta, w) \sigma_p^2 \log^\nu(K_n))$$

with  $\nu = 4$  ( $\nu = 3$ ) when the ordinary (tapered) periodogram is used. Finally, applying (6.9), Lemma 6.1 and (6.11), we obtain

$$\begin{aligned} \sum_{1 \leq k < j \leq K_n} \nu_k \nu_j \text{cov}_f(\epsilon_{n,k}, \epsilon_{n,j}) &\leq C(M, \delta, w) \sigma_p^4 \log^2(K_n) \\ \sum_{1 \leq k < j \leq K_n} p^2(d, k, j) &\leq C(M, \delta, w) \sigma_p^4 \log^\nu(n). \end{aligned}$$

with  $\nu = 5$  ( $\nu = 2$ ) when the ordinary (tapered) periodogram is used. This concludes the proof of Proposition 3.2.  $\square$

## 7. Proofs

**Proof of Lemma 4.1.** Let  $a \in (1, 2]$  and suppose  $b$  is such that  $1/a + 1/b = 1$ . For brevity, denote  $|\beta|_\infty = \max_{k \in \{1, \dots, K_n\}} |\beta_k|$ . By Hölder's inequality, we have, for any positive integer  $L \leq K_n$ ,

$$\begin{aligned} \mathbb{E}_f \left[ \alpha \exp \left\{ \sum_{k=1}^{K_n} \beta_k \epsilon_k \right\} \right] &\leq \mathbb{E}_f^{1/b} \left[ \exp \left\{ b\alpha \sum_{k=1}^L \beta_k \epsilon_k \right\} \right] \mathbb{E}_f^{1/a} \left[ \exp \left\{ a\alpha \sum_{k=L+1}^{K_n} \beta_k \epsilon_k \right\} \right] \\ &=: B(\alpha)^{1/b} A^{1/a}(\alpha) \end{aligned}$$

We bound these two terms separately. If  $2b|\alpha| |\beta|_\infty \leq 1$ , Lemma 7.3 below implies that

$$B(\alpha) \leq C(M, \delta, w, L). \quad (7.1)$$

For  $\xi \in \mathbb{R}^{2m}$ , define  $\phi(\xi) = \log(|\xi|^2/2) - \psi(m)$  and

$$\Phi(\xi_{L+1}, \dots, \xi_{K_n}) = \exp \left\{ \alpha \sum_{k=L+1}^{K_n} \beta_k \phi(\xi_k) \right\}.$$

Set  $d_n(x_k) := C_n(x_k) + iS_n(x_k)$  and let  $\Gamma_n$  be the covariance matrix of the  $2m(K_n - L)$ -dimensional Gaussian vector

$$U_n = \left( \frac{C_n(x_{mL+2})}{\sqrt{f(y_{L+1})}}, \frac{S_n(x_{mL+2})}{\sqrt{f(y_{L+1})}}, \dots, \frac{C_n(x_{mK_n})}{\sqrt{f(y_{K_n})}}, \frac{S_n(x_{mK_n})}{\sqrt{f(y_{K_n})}} \right)^T.$$

Theorem 6.1 and Lemma 6.1 imply that, for any  $1 < a \leq 2$  and  $\epsilon = \frac{1}{2}$ , say, we can choose  $L = L(M, \delta)$  large enough for the assumptions of Lemma 7.1 below to hold. Hence, we obtain

$$A(\alpha) = \mathbb{E}_f[\Phi(U_n)] \leq C(a, \epsilon) \prod_{k=L+1}^{K_n} \|e^{a\alpha\beta_k\phi}\|_a,$$

where the norm  $\|\cdot\|_a$  is defined in Lemma 7.1. Now applying Lemma 7.2 below, we obtain, for all  $\alpha$  such that  $a^2|\alpha| |\beta|_\infty \leq \eta_0(a)$ ,

$$\|\exp(a\alpha\beta_{n,k}\phi)\|_a^a = \exp(-a^2\alpha\psi(m))\mathbb{E}[Y^{a^2\alpha\beta_k}] \leq \exp(a^2\beta_k^2\psi'(m)a^5/2),$$

where  $\eta_0$  and  $Y$  are defined in Lemma 7.2. We now obtain that

$$A(\alpha) \leq C(a) \prod_{k=L+1}^{K_n} e^{\alpha^2\beta_k^2\psi'(m)a^4/2} \leq C(a)e^{\alpha^2\psi'(m)^2a^4/2}, \quad (7.2)$$

where the last inequality is a consequence of the fact that  $\sum_{k=1}^{K_n} \beta_k^2 = 1$ . In conclusion, if  $|\alpha| |\beta|_\infty \leq (1/2b) \wedge (\eta_0(a)/a^2)$ , gathering (7.1) and (7.2), we obtain

$$E_f \left[ \alpha \exp \left\{ \sum_{k=1}^{K_n} \beta_k \epsilon_k \right\} \right] \leq C(M, \delta, w, a) \exp(\alpha^2 \psi'(m)^2 a^4 / 2).$$

Setting  $a^4 = 1 + \rho$  and  $\eta(\rho) = (1/2b) \wedge (\eta_0(a)/a^2)$  concludes the proof of Lemma 4.  $\square$

**Lemma 7.1.** *Let  $X$  be a  $k$ -dimensional centred Gaussian vector with covariance matrix  $\Gamma$ . Let  $\Phi : \mathbb{R}^k \rightarrow \mathbb{R}$  be a function such that  $\int_{\mathbb{R}^k} \Phi^2(x) \exp(-\frac{1}{2}x^T x) dx < \infty$ ; and, for  $1 < a \leq 2$ , write  $\|\Phi\|_a = (2\pi)^{-k/2} \int_{\mathbb{R}^k} \Phi^a(x) \exp(-\frac{1}{2}x^T x) dx$ . If  $\text{tr}(\{\Gamma - I_k\}^2) \leq (a-1)^2(1-\epsilon)$ , then there exists a constant  $c(a, \epsilon)$  which depends only on  $a$  and  $\epsilon$  but not on  $k$ , such that*

$$|E[\Phi(X)]| \leq c(a, \epsilon) \|\Phi\|_a$$

**Proof.** The proof of Lemma 7.1 is an adaptation of the proof of Lemma 4.2 in Giraitis *et al.* (2000). For  $a \in (1, 2]$ , denote by  $b$  the real number in  $[2, \infty[$  such that  $a^{-1} + b^{-1} = 1$ . Applying Hölder's inequality, we obtain

$$E[\Phi(X)] = (2\pi)^{-k/2} |\Gamma|^{-1/2} \int_{\mathbb{R}^k} \Phi(x) \exp(-\frac{1}{2a}x^T x) \exp(-\frac{1}{2}x^T (\Gamma^{-1} - I_k/a)x) dx, \quad (7.3)$$

$$|E[\Phi(X)]| \leq |\Gamma|^{-1/2} \|\Phi\|_a \left| (2\pi)^{-k/2} \int_{\mathbb{R}^k} \exp(-\frac{b}{2}x^T (\Gamma^{-1} - I_k/a)x) dx \right|^{1/b}. \quad (7.4)$$

The last term on the right-hand side of the latter equation is finite provided that  $\Gamma^{-1} - I_k/a$  is invertible. Let  $\Delta = \Gamma - I_k$ . Then

$$\Gamma^{-1} - \frac{I_k}{a} = \Gamma^{-1} \left( I_k - \frac{\Gamma}{a} \right) = \Gamma^{-1} \left( \left( 1 - \frac{1}{a} \right) I_k - \frac{\Delta}{a} \right) = b^{-1} \Gamma^{-1} (I_k - (b-1)\Delta).$$

Under the assumptions in the statement of the lemma,  $I_k - (b-1)\Delta$  is invertible and thus  $(\Gamma^{-1} - I_k/a)^{-1} = b\Gamma(I_k - (b-1)\Delta)^{-1}$ . Thus

$$(2\pi)^{-k/2} \int_{\mathbb{R}^k} \exp(-\frac{b}{2}x^T (\Gamma^{-1} - I_k/a)x) dx = b^{-k/2} \left| \Gamma^{-1} - \frac{I_k}{a} \right|^{-1/2}.$$

Finally, we obtain

$$\begin{aligned} |E[\Phi(X)]| &\leq |\Gamma|^{-1/2} b^{-k/2b} \left| \Gamma^{-1} - \frac{I_k}{a} \right|^{-1/2b} \|\Phi\|_a = |\Gamma|^{-1/2a} |I_k - (b-1)\Delta|^{-1/2b} \|\Phi\|_a \\ &= |I_k + \Delta|^{-1/2a} |I_k - (b-1)\Delta|^{-1/2b} \|\Phi\|_a. \end{aligned}$$

Let  $\lambda_i$ ,  $1 \leq i \leq k$  denote the eigenvalues of  $\Delta$ . Then

$$\begin{aligned}
 |\log(|I_k + \Delta|^{-1/2a}|I_k - (b-1)\Delta|^{-1/2b})| &= \left| -\frac{1}{2a} \sum_{i=1}^k \log(1 + \lambda_i) - \frac{1}{2b} \sum_{i=1}^k \log(1 - (b-1)\lambda_i) \right| \\
 &\leq c(\epsilon) \sum_{i=1}^k \lambda_i^2 \leq 2c(\epsilon)(b-1)^2 \text{tr}(\{\Gamma - I_k\}^2) \leq 2c(\epsilon),
 \end{aligned}$$

since by assumption  $(b-1)^2 \text{tr}(\{\Gamma - I_k\}^2) \leq 1$ . This concludes the proof of Lemma 7.1.  $\square$

**Lemma 7.2.** *Let  $Y$  be a  $\Gamma(m, 1)$  random variable. For any  $a > 1$ , there exists a real  $\eta_0(a) > 0$  depending only on  $a$  such that, for all  $s \geq -\eta_0(a)$ ,*

$$\log(\mathbb{E}[Y^s]) \leq s\psi(m) + a\psi'(m) \frac{s^2}{2}. \tag{7.5}$$

**Proof.** By concavity of the digamma function, we obtain, for  $s > -m$ ,

$$\log(\mathbb{E}[Y^s]) \leq s\psi(m) + \frac{s^2}{2} \psi'((m+s) \vee m).$$

If  $s > 0$ , then (7.5) holds. If  $s < 0$ , then  $\psi'(m+s) > \psi'(m)$ . Nevertheless, by continuity, for any  $a > 1$ , there exists an  $\eta_0(a) > 0$  such that if  $s > -\eta_0(a)$ , then  $\psi'(m+s) \leq a\psi'(m)$ .  $\square$

**Lemma 7.3.** *Let  $M > 1$ ,  $\delta \in (0, \frac{1}{2})$ ,  $w$  be a sequence such that  $\lim nw(n) = 0$  and  $L$  be a positive integer. There exists a constant  $C(M, \delta, w)$  such that if  $r_1, \dots, r_L$  are real numbers satisfying  $\max_{1 \leq k \leq L} |r_k| \leq \frac{1}{2}$ , then*

$$\sup_{f \in \mathcal{H}(M, \delta, w)} \mathbb{E}_f \left[ \exp \left\{ \sum_{k=1}^L r_k \epsilon_k \right\} \right] \leq C(M, \delta, w, L). \tag{7.6}$$

**Proof.** Pooling creates a technical difficulty due to the fact that the regression is based on the frequencies  $y_k$ . We first get rid of this problem. Let  $J_k = \{2m(k-1) + 2, \dots, 2mk\}$  ( $1 \leq k \leq L$ ),  $u_s = C_n(x_s)/f(x_s)$  and  $v_s = S_n(x_s)/f(x_s)$  ( $1 \leq s \leq 2mL$ ), and  $T_{n,L} = \exp\{\sum_{k=1}^L r_k \epsilon_k\}$ . With this notation, we can write

$$T_{n,L} = \prod_{k=1}^L \left( e^{-\psi(m)} \sum_{s \in J_k} \frac{f(x_s)}{f(y_k)} (u_s^2 + v_s^2) \right)^{r_k}.$$

Thus it is easily seen that there exists a constant  $C(M, \delta, L)$  such that

$$T_{n,L} \leq C(M, \delta, L) \prod_{k=1}^L \left( \sum_{s \in J_k} (u_s^2 + v_s^2) \right)^{r_k}.$$

Since the weights  $r_k$  satisfy  $|r_k| \leq 1/2$ , it also holds that

$$T_{n,L} \leq C(M, \delta, L) \prod_{k=1}^L \left( \left( \sum_{s \in J_k} (u_s^2 + v_s^2) \right)^{1/2} + \left( \sum_{s \in J_k} (u_s^2 + v_s^2) \right)^{-1/2} \right).$$

Let  $V_{n,L}$  be the  $2mL$ -dimensional Gaussian vector  $V_{n,L} = (u_{2s}, v_{2s})_{1 \leq k \leq L}$ . For  $\underline{z} = (z_1, \dots, z_m) \in \mathbb{R}^{2m}$ , write  $\|\underline{z}\|^2 = \sum_{i=1}^m z_i^2$ , and for  $\underline{z}_1, \dots, \underline{z}_L$ , define

$$\Phi(\underline{z}_1, \dots, \underline{z}_L) = \prod_{i=1}^L \left( \|\underline{z}_i\| + \frac{1}{\|\underline{z}_i\|} \right).$$

What we must now prove is that

$$\sup_n \sup_{f \in \mathcal{H}(M, \delta, w)} E_f[\Phi(V_{n,L})] \leq C(M, \delta, w). \tag{7.7}$$

Let  $\sum_{n,L}(f)$  be the covariance matrix of  $V_{n,L}$ . Since  $E[\Phi(\xi)] < \infty$  if  $\xi$  is a  $2mL$ -dimensional vector with i.i.d. standard Gaussian components, to prove (7.7), it is sufficient to prove that

$$\inf_n \inf_{f \in \mathcal{H}(M, \delta, w)} \lambda_{\min}(\sum_{n,L}(f)) = \lambda_* > 0, \tag{7.8}$$

where  $\lambda_{\min}(A)$  denotes the smallest eigenvalue of the matrix  $A$ . Basically, the argument for proving (7.8) is that it holds for any  $f \in \mathcal{H}(M, \delta, w)$ , and that  $\mathcal{H}(M, \delta, w)$  is isomorphic to  $[-\delta, \delta] \times \mathcal{C}(w) \cap \mathcal{G}(M)$  which is a compact set for the product topology on  $\mathbb{R} \times \mathbb{C}([-\pi, \pi])$ . We now give some details. We follow the arguments in the proof of Lemma 3 in Moulines and Soulier (1999), which hold for one function  $f$ , and mention where uniformity with respect to  $f$  is used.

We first prove that  $\sum_{n,L}(f)$  converges uniformly to a positive definite matrix  $\sum_L(d)$ . Recall that  $H_n(x) = \sum_{t=1}^n h_{t,n} e^{itx}$  and  $\zeta_{n,k} = d_n(x_k) / \sqrt{f(x_k)}$ . Since  $\sum_{t=1}^n |h_{t,n}|^2 = 2n$ , we obtain

$$\begin{aligned} E_f[\zeta_{n,k} \zeta_{n,j}] &= \frac{1}{4\pi n \sqrt{f(x_k) f(x_j)}} \int_{-\pi}^{\pi} H_n(x + x_k) H_n(x_j - x) f(x) dx \\ &= \frac{|1 - e^{ix_k}|^d |1 - e^{ix_j}|^d}{4\pi \sqrt{f^*(x_k) f^*(x_j)}} \int_{-n\pi}^{n\pi} n^{-1} H_n\left(x_k + \frac{y}{n}\right) n^{-1} H_n\left(x_j - \frac{y}{n}\right) |1 - e^{iy/n}|^{-2d} f^*\left(\frac{y}{n}\right) dy. \end{aligned}$$

Note that  $n^{-1} |H_n(x/n)| \leq C(1 + |x|)^{-1}$  for all  $x \in [-3n\pi/2, 3n\pi/2]$ , and that, for all  $x \in [-n\pi, n\pi]$ ,

$$\lim_{n \rightarrow \infty} n^{-1} H_n\left(\frac{x}{n}\right) = \int_0^1 (1 - e^{2i\pi s}) e^{isx} ds = \frac{2\pi(1 - e^{ix})}{x(x + 2\pi)} =: \hat{h}(x).$$

Applying the dominated convergence theorem, we thus have, uniformly with respect to  $f \in \mathcal{H}(M, \delta, w)$ ,

$$\lim_{n \rightarrow \infty} E_f[\zeta_{n,k} \zeta_{n,j}] = 2^{2d-2} \pi^{2d-1} k^d j^d \int_{-\infty}^{\infty} \hat{h}(2\pi k + y) \hat{h}(2\pi j - y) |y|^{-2d} dy =: \varsigma_{k,j}(d).$$

The non-singularity of the matrix  $\sum_L(d) = (\zeta_{k,j}(d))_{1 \leq k,j \leq L}$  now follows from the fact that the functions  $\hat{h}_k : x \rightarrow \hat{h}(2k\pi + x)$ ,  $k \in \mathbb{Z}$ , are linearly independent. Moreover, by continuity and compactness,  $\inf_{|d| \leq \delta} \lambda_{\min}(\sum_L(d)) > 0$ . Thus there exists an integer  $n_0(M, \delta, w)$  such that

$$\inf_{n \geq n_0} \inf_{f \in \mathcal{H}(M, \delta, w)} \lambda_{\min}(\sum_{n,L}(f)) > 0.$$

For each  $n \leq n_0(M, \delta, w)$ , and each  $f \in \mathcal{H}(M, \delta, w)$ , it is clear that  $\lambda_{\min}(\sum_{n,L}(f)) > 0$  (cf. Lemma 3 in Moulines and Soulier 1999). Hence by compactness we also have,

$$\inf_{n \leq n_0} \inf_{f \in \mathcal{H}(M, \delta, w)} \lambda_{\min}(\sum_{n,L}(f)) > 0.$$

This proves that (7.8) holds, and this concludes the proof of Lemma 7. □

**Proof of Lemma 4.2.** Using the previous notation,  $\hat{d}_p - d$  can be decomposed as

$$\hat{d}_p - d = \frac{\sigma_p}{\sqrt{\psi'(m)}} \sum_{k=1}^{K_n} \beta_{n,k} \epsilon_{n,k} + b_{n,p}(I^*) =: \frac{\sigma_p}{\sqrt{\psi'(m)}} \zeta_{n,p} + b_{n,p}(I^*).$$

Hence there exists a constant  $C(a)$  such that, for all  $p \leq \epsilon_n K_n$ ,

$$E^{1/a}[|\hat{d}_p - d|^a] \leq \frac{\sigma_p}{\sqrt{\psi'(m)}} E^{1/a}[|\zeta_{n,p}|^a] + |b_{n,p}(I^*)| \leq C(a) E^{1/2a}[e^{\zeta_{n,p}}] + |b_{n,p}(I^*)|.$$

Applying Proposition 3, we obtain  $|b_{n,p}(I^*)| \leq c\theta_0^*$ . For large enough  $n$ , (4.4) implies that the assumptions of Lemma 4.1 hold, thus

$$E[e^{\zeta_{n,p}}] \leq C(M, \delta, w).$$

This concludes the proof of Lemma 4. □

For the sake of completeness, and because we did not find a relevant reference, we give a proof of the fact that a function  $h$  in  $\mathcal{C}(w_{\beta,L})$  is Hölder continuous with exponent  $\beta$ , for  $0 < \beta < 2$ . Let  $h$  be an even function with summable Fourier coefficients  $\theta_j$ ,  $j \geq 0$ . By definition,  $h \in \mathcal{C}(w_{\beta,L})$  implies that  $\theta_p^* \leq L(1+p)^{-\beta}$ . For  $0 < x \leq 1$ , let  $n_x$  be the greatest integer smaller than or equal to  $1/x$ . Then we can write

$$h(x) - h(0) = \sum_{j=1}^{\infty} \theta_j(\cos(jx) - 1) = \sum_{j=1}^{n_x} \theta_j(\cos(jx) - 1) + \sum_{j=n_x+1}^{\infty} \theta_j(\cos(jx) - 1).$$

The last term above is easily bounded:

$$\left| \sum_{j=1}^{\infty} \theta_j(\cos(jx) - 1) \right| \leq 2\theta_{n_x}^* \leq 2L(1+n_x)^{-\beta} \leq 2Lx^\beta.$$

To bound the first term, note that if  $0 < j < n_x$ , then  $0 < 1 - \cos(jx) \leq (jx)^2/2 \leq jx/2$ . If  $0 < \beta < 1$ , we then obtain

$$\left| \sum_{j=1}^{n_x} \theta_j \cos(jx) \right| \leq \frac{x}{2} \sum_{j=1}^{n_x} j |\theta_j| \leq \frac{x}{2} \sum_{j=1}^{n_x} \theta_j^* \leq \frac{Lx}{2} \sum_{j=1}^{n_x} (1+j)^{-\beta} \leq \frac{Lx n_x^{1-\beta}}{2(1-\beta)} \leq \frac{Lx^\beta}{2(1-\beta)}.$$

If  $1 \leq \beta < 2$ , then

$$\left| \sum_{j=1}^{n_x} \theta_j \cos(jx) \right| \leq x^2 \sum_{j=1}^{n_x} \frac{j^2}{2} |\theta_j| \leq x^2 \sum_{j=1}^{n_x} j \theta_j^* \leq Lx^2 \sum_{j=1}^{n_x} (1+j)^{1-\beta} \leq \frac{Lx^2 n_x^{2-\beta}}{2-\beta} \leq \frac{Lx^\beta}{2-\beta}.$$

For  $0 < \beta \leq 2$ , let  $\mathcal{H}(\beta, C)$  denote the class of even functions  $h$  such that, for all  $x \in [0, 2\pi]$ ,  $|h(x) - h(0)| \leq C|x|^\beta$  for some constant  $C$ . To prove that  $\mathcal{C}(w_{\beta,L})$  is not included in  $\mathcal{H}(\beta', C)$  for any  $\beta' > \beta$  and any  $C$ , and that the class  $\mathcal{C}(w_{2,L})$  is strictly greater than the class  $\mathcal{H}(2, C)$ , we need only consider the Weierstrass-type functions  $h_\beta$  defined as  $h_\beta(x) = \sum_{k=0}^{\infty} 2^{-\beta k} \cos(2^k x)$ . It is easily checked that, for all  $0 < \beta < 2$ ,  $h_\beta \in \mathcal{C}(w_{\beta,L})$  for  $L = (1 - 2^{-\beta})^{-1}$ . Thus  $h_\beta \in \mathcal{H}(\beta, C)$  for some constant  $C$ . It can be also easily checked that, for  $0 < \beta < 2$ ,  $\lim_{x \rightarrow 0} x^{-\beta} h_\beta(x) > 0$ , which implies that  $h_\beta \notin \mathcal{H}(\beta', C)$  for any  $\beta' > \beta$  and any  $C$ . For  $\beta = 2$ , it can be shown that

$$\lim_{x \rightarrow 0} \frac{h_2(x)}{x^2 \log(1/x)} > 0,$$

which implies that  $h_2 \notin \mathcal{H}(2, C)$  for any  $C$ .

## Acknowledgements

The authors thank Professor Alexander Tsybakov who pointed out a mistake in an earlier version of the paper and the referees for their constructive comments.

## References

- Arcones, M.A. (1994) Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.*, **15**, 2243–2274.
- Barron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, **113**, 301–413.
- Beran, J. (1993) Fitting long memory models by generalized regression. *Biometrika*, **80**, 817–822.
- Beran, J. (1994) *Statistics for Long Memory Processes*. London: Chapman & Hall.
- Beran, J., Bhansali, R.J. and Ocker, D. (1998) On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika*, **85**, 921–934.
- Butucea, C. (1999) Exact adaptive pointwise estimation on Sobolev classes of densities. *C.R. Acad. Sci. Paris Sér. I Math.*, **329**(6), 535–540.
- Brown, D.L. and Low, M.G. (1996) A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, **24**, 2524–2535.
- Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Ann. Statist.*, **17**, 1749–1766.
- Efromovich, S. (1998a) Data-driven efficient estimation of the spectral density. *J. Amer. Statist. Assoc.*, **93**, 762–769.

- Efromovich, S. (1998b) On global and pointwise estimation. *Bernoulli*, **4**, 273–282.
- Fox, R. and Taqqu, M.S. (1986) Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.*, **14**, 517–532.
- Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221–238.
- Gill, R.D. and Levit, B.Y. (1995) Applications of the van Trees inequality: A Bayesian Cramér–Rao bound. *Bernoulli*, **1**, 59–79.
- Giraitis, L. and Surgailis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle’s estimate. *Probab. Theory Related Fields*, **86**, 87–104.
- Giraitis, L., Robinson, P.M. and Samarov, A. (1997) Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Ser. Anal.*, **18**, 49–61.
- Giraitis, L., Robinson, P.M. and Samarov, A. (2000) Adaptive rate-optimal estimation of the long memory parameter. *J. Multivariate Anal.*, **72**, 183–207.
- Granger, C.W.J. and Joyeux, R. (1980) An introduction to long memory time series and fractional differencing. *J. Time Ser. Anal.*, **1**, 15–30.
- Hurvich, C.M. and Brodsky, J. (2001) Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. *J. Time Ser. Anal.*, **22**, 221–249.
- Hurvich, C.M. and Chen, W.W. (2000) An efficient taper for potentially overdifferentiated long-memory time series. *J. Time Ser. Anal.*, **21**, 155–180.
- Hurvich, C.M., Deo, R. and Brodsky, J. (1998) The mean squared error of Geweke and Porter-Hudak’s estimator of a long-memory time-series. *J. Time Ser. Anal.*, **19**, 19–46.
- Hurvich, C.M., Moulines, E. and Soulier, P. (2000) The FEXP estimator for non Gaussian, potentially non stationary processes. Prépublication 133, l’Université d’Evry Val d’Essonne. To appear in *Stochastic Process. Appl.*
- Johnson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions I*. New York: Wiley.
- Katznelson, I. (1968) *An Introduction to Harmonic Analysis*. New York: Dover.
- Kokoszka, P.S. and Bhansali, R.J. (1999) Estimation of the long memory parameter by fitting fractionally differenced autoregressive models. Preprint.
- Künsch, H.R. (1986) Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.*, **23**, 1025–1030.
- Künsch, H.R. (1987) Statistical aspects of self-similar processes. In Yu.A. Prohorov and V.V. Sazonov (eds), *Proceedings of the 1st World Congress of the Bernoulli Society*, Vol. 1, pp. 67–74. Utrecht: VNU Science Press.
- Lepski, O. (1990) One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, **35**, 459–470.
- Lepski, O. and Levit, B.Y. (1998) Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, **7**, 123–156.
- Moulines, E. and Soulier, P. (1999) Broadband log-periodogram regression of time series with long-range dependence. *Ann. Statist.*, **27**, 1415–1439.
- Robinson P.M. (1994) Time series with long range dependence. In C.A. Sims (ed), *Advances in Econometrics. Papers from the Sixth World Congress of the Econometric Society*. Cambridge: Cambridge University Press.
- Robinson P.M. (1995a) Log-periodogram regression of time series with long range dependence. *Ann. Statist.*, **23**, 1043–1072.



- Robinson P.M. (1995b) Gaussian semiparametric estimation of long range dependence. *Ann. of Statist.*, **23**, 1630–1661.
- Soulier, P. (2001) Some new bounds and a central limit theorem for functions of Gaussian vectors. Prépublication 71, l'Université d'Evry Val d'Essonne. *Statist. Probab. Lett.*, **54**(2), 193–203.
- Velasco, C. (1999a) Gaussian semiparametric estimation of non-stationary time series. *J. Time Ser. Anal.*, **20**, 87–127.
- Velasco, C. (1999b) Non-stationary log-periodogram regression. *J. Econometrics*, **91**, 325–371.

Received September 1999 and revised May 2001