

**ADAPTIVE SEMIPARAMETRIC ESTIMATION
OF THE MEMORY PARAMETER***

by

Liudas Giraitis and Peter M Robinson
London School of Economics and Political Science

and

Alexander Samarov
University of Massachusetts-Lowell and MIT

Contents:

Abstract

1. Introduction
2. Lower bound
3. Upper bound
4. Additional lemmas
5. Proofs of lemmas 3.1 and 3.2

References

The Suntory Centre
Suntory and Toyota International Centres
for Economics and Related Disciplines
London School of Economics and Political
Science
Houghton Street
London WC2A 2AE
Tel.: 0171-405 7686

Discussion Paper
No. EM/00/379
January 2000

* The research of the first and second authors was supported by ESRC Grant R000235892. The second author's research was also supported by a Leverhulme Trust Personal Research Professorship. Research of the third author was supported in part by NSF Grant DMS-9626348.

Abstract

In Giraitis, Robinson, and Samarov (1997), we have shown that the optimal rate for memory parameter estimators in semiparametric long memory models with degree of 'local smoothness' β is $n^{-r(\beta)}$, $r(\beta) = \beta/(2\beta + 1)$, and that a log-periodogram regression estimator (a modified Geweke and Porter-Hudak (1983) estimator) with maximum frequency $m = m(\beta) \asymp n^{2r(\beta)}$ is rate optimal. The question which we address in this paper is what is the best obtainable rate when β is unknown, so that estimators cannot depend on β . We obtain a lower bound for the asymptotic quadratic risk of any such adaptive estimator, which turns out to be larger than the optimal nonadaptive rate $n^{-r(\beta)}$ by a logarithmic factor. We then consider a modified log-periodogram regression estimator based on tapered data and with a data-dependent maximum frequency $m = m(\hat{\beta})$, which depends on an adaptively chosen estimator $\hat{\beta}$ of β , and show, using methods proposed by Lepskii (1990) in another context, that this estimator attains the lower bound up to a logarithmic factor. On one hand, this means that this estimator has nearly optimal rate among all adaptive (free from β) estimators, and, on the other hand, it shows near optimality of our data-dependent choice of the rate of the maximum frequency for the modified log-periodogram regression estimator. The proofs contain results which are also of independent interest: one result shows that data tapering gives a significant improvement in asymptotic properties of covariances of discrete Fourier transforms of long memory time series, while another gives an exponential inequality for the modified log-periodogram regression estimator.

Keywords: Long range dependence; semiparametric model; rates of convergence; adaptive bandwidth selection.

JEL classification nos: C13.

© by the authors. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

1 Introduction.

Suppose that we have n observations X_1, \dots, X_n from a stationary, Gaussian time series $\{X_t\}_{t=-\infty}^{\infty}$ with mean μ and spectral density

$$f(\lambda) = \frac{L(\lambda)}{|\lambda|^\alpha}, \quad \lambda \in [-\pi, \pi], \quad \alpha \in (-1, 1), \quad (1.1)$$

and $L(\lambda) \rightarrow C$, $C \in (0, \infty)$, as $\lambda \rightarrow 0$. The memory parameter α determines the behaviour of f near zero and is just a re-expression of the self-similarity parameter $H = (\alpha + 1)/2$ and of the fractional differencing parameter $d = \alpha/2$. X_t is said to exhibit long range dependence when $0 < \alpha < 1$, short range dependence when $\alpha = 0$, and negative dependence when $-1 < \alpha < 0$. There exist several ‘semiparametric’ estimators of α , with f specified only near zero frequency, see, e.g., Geweke and Porter-Hudak (1983), Künsch (1986, 1987), Robinson (1995a,b).

Consider for $\beta > 0$ the class of spectral densities

$$F(\beta, C_1, C_2, \delta) = \{f : f(\lambda) = c|\lambda|^{-\alpha}(1 + \Delta(\lambda)), \quad 0 < c \leq C_1, \quad (1.2)$$

$$-1 < \alpha < 1 - \delta, \quad |\Delta(\lambda)| \leq C_2|\lambda|^\beta, \quad \lambda \in [-\pi, \pi]\},$$

where C_1, C_2 and $\delta \in (0, 1)$ are independent of β . Of central importance to this paper is the parameter β , whose interpretation we now discuss. It is closely related to the (local-to-zero) smoothness $\sigma > 0$ of $L(\lambda)$ in (1.1) which could be defined as follows. For $0 < \sigma \leq 1$, $L(\lambda)$ has smoothness σ if it satisfies a Lipschitz condition of degree σ around $\lambda = 0$. For $\sigma > 1$, $L(\lambda)$ has smoothness σ if $L(\lambda)$ is s times differentiable around $\lambda = 0$, where $s = [\sigma]$, its s -th derivative satisfying a Lipschitz condition of degree $\sigma - s$ around $\lambda = 0$. Then $\beta = \sigma$ for $\sigma \leq 2$ (noting that $f(\lambda)$ is an even function), whereas $\beta \leq \sigma$ for $\sigma > 2$, with $\beta = \sigma$ if the first s derivatives of $L(\lambda)$ at $\lambda = 0$ are all zero. In general, therefore, for $\sigma > 2$ we have $\beta = 2$ only. This is the case, for example, with fractionally integrated autoregressive moving average processes.

The condition $\delta > 0$ in the definition of class $F(\beta, C_1, C_2, \delta)$ is needed to ensure a finite upper bound for $\text{Var}(X_t) = \int_{-\pi}^{\pi} f(\lambda) d\lambda$, uniformly in $\alpha < 1 - \delta$. Note that if $f(\lambda) = c|\lambda|^{-\alpha}$, $\text{Var}(X_t)$ is not thus upper-bounded if $\alpha < 1$, but it is bounded by $2c\pi^2/\delta$ if $\alpha \in (-1, 1 - \delta)$, $0 < \delta < 1$.

Denote the maximum quadratic risk of an estimator $\tilde{\alpha}$ over $F_0(\beta) = F(\beta, C_1, C_2, 0)$ as

$$R_n(\tilde{\alpha}, F_0(\beta)) = \sup_{f \in F_0(\beta)} E_f[\tilde{\alpha} - \alpha(f)]^2, \quad (1.3)$$

where we write $\alpha(f)$ in place of α in (1.1). In Giraitis, Robinson, and Samarov (1997) (referred to throughout this paper as GRS) we established the following results. First, we showed that, as $n \rightarrow \infty$,

$$\inf_{\tilde{\alpha}} R_n(\tilde{\alpha}, F_0(\beta)) \asymp n^{-2r(\beta)}, \quad (1.4)$$

where

$$r(\beta) = \frac{\beta}{2\beta + 1}, \quad (1.5)$$

and the inf is taken over all possible estimators. Second, we showed that the optimal minimax rate $n^{-2r(\beta)}$ in (1.4) is attained by a modified version of the estimator of Geweke and Porter-Hudak (1983) (hereafter referred to as the GPH estimator).

The question which we address in this paper is what is the best obtainable rate when β is unknown, so that estimators cannot depend on β ? In Section 2 we obtain a lower bound for the

asymptotic quadratic risk of any such adaptive estimator, which turns out to be slower than the optimal nonadaptive rate $n^{-r(\beta)}$ by a logarithmic factor. We then consider a tapered version of the log-periodogram regression estimator, in Section 3. This estimator was proposed by Velasco (1998a, b), as a tapered version of the modified log-periodogram regression estimator of Robinson (1995a). Velasco (1998 a) showed that a data taper can improve estimates of variances and covariances of discrete Fourier transforms given in Theorem 2 of Robinson (1995a). We prove (Lemma 3.1) a slight improvement of Velasco's (1998a) result under somewhat weaker conditions, which allows us to obtain an exponential inequality (Lemma 3.2) for our estimator, which turns out to be an important tool in obtaining the adaptive rate of our estimator and may also be of independent interest. The proofs of these lemmas are reserved for Section 5, following Section 4, which contains three minor lemmas.

The key element in the construction of our estimator is a data-dependent selection of the maximum frequency used, $m = m(\hat{\beta})$, which depends on an adaptively chosen $\hat{\beta}$, obtained using a modification of the procedure proposed by Lepskii (1990) in a different nonparametric setting. Informally, $\hat{\beta}$ is defined as the largest β for which the log-periodogram regression estimator using $m = m(\beta)$ is not significantly different from all such estimators using $m(\gamma)$, $\gamma < \beta$. The procedure can be also interpreted by graphing the estimator versus a grid of values of β together with a variable-width band around it: $\hat{\beta}$ is chosen as the largest β on the grid for which the corresponding estimator stays within the band for all $\gamma < \beta$. See (3.6)-(3.8) below for the precise definition.

The memory parameter estimator considered in Section 3 achieves nearly optimal rate of convergence in the class $F(\beta, C_1, C_2, \delta)$. Clearly, $F(\beta, C_1, C_2, \delta)$ includes all classes $F(\beta', C_1, C_2, \delta)$, $\beta' \geq \beta$. Therefore if a particular density f belongs to $F(\beta', C_1, C_2, \delta) \cap F(\beta, C_1, C_2, \delta)$, the rate of convergence of the estimator will be determined by β' , and it will be better than in case of $f \in F(\beta, C_1, C_2, \delta)$ such that $f \notin F(\beta', C_1, C_2, \delta)$ when $\beta' > \beta$. Summarising, in the case of a particular density f the rate of convergence is determined by the largest β for which the inequality in (1.2) holds.

In Section 3 we show that our adaptive estimator attains the lower bound obtained in Section 2 up to a logarithmic factor. This means, on one hand, that this estimator is nearly rate - optimal among all possible adaptive estimators, and, on the other hand, that our data-dependent choice of m is also nearly optimal for the log-periodogram regression estimator. The technique of the proof, the idea of which also comes from Lepskii (1990), requires one to assume that though unknown β does not exceed a known finite maximum value $\beta^* \in (0, \infty)$.

2 Lower bound

This section is devoted to establishing the following lower bound.

Theorem 2.1 *Uniformly in $\beta \leq \beta^*$, the sequence $\{\phi_n(\beta) = (\log n/n)^{r(\beta)}\}$ gives the lower bound to the asymptotic minimax risk for the class $F(\beta) = F(\beta, C_1, C_2, \delta)$, $\beta \leq \beta^*$, that is for some $\underline{C} > 0$*

$$\liminf_{n \rightarrow \infty} \inf_{\tilde{\alpha}} \sup_{\beta \leq \beta^*} \phi_n^{-2}(\beta) R_n(\tilde{\alpha}, F(\beta)) \geq \underline{C}. \quad (2.1)$$

PROOF OF THEOREM 2.1: Let $0 < \beta_1 < \beta_2 \leq \beta^*$. As in GRS (see also Hall and Welsh (1984)), let $f_0(\lambda) = 1$, $\lambda \in [-\pi, \pi]$, be the spectral density of white noise, and define a sequence of 'perturbed' spectral densities $f_n(\lambda)$ exactly as in formulae (2.3)-(2.5) in GRS but with $\delta_n = (\gamma \log n/n)^{1/(2\beta_1+1)}$, where $\gamma > 0$ will be chosen later. We have $\alpha(f_0) = 0$ and

$$\alpha(f_n) = \kappa \delta_n^{\beta_1} = \tau \phi_n(\beta_1), \quad (2.2)$$

with $\tau = \kappa\gamma^{\beta_1/(2\beta_1+1)}$, for some $\kappa > 0$. Clearly, $f_0 \in F(\beta_2, C_1, C_2)$.

The following two lemmas are proved exactly as in GRS.

Lemma 2.1 *For all sufficiently large n ,*

- (i) $f_n \in F(\beta_1, C_1, C_2)$ and
- (ii) $\int_{-\pi}^{\pi} (f_n(\lambda) - f_0(\lambda))^2 d\lambda \leq K\gamma \log n/n$ for some constant $K > 0$.

As in GRS, denote by P_n and P_0 the probability measures on R^n generated by n observations $\mathbf{X} = (X_1, \dots, X_n)$ of the Gaussian stationary sequence with the same mean μ and spectral densities f_n and f_0 respectively, denote by E_n and E_0 the corresponding expectations, and by $\Lambda_n = \log \frac{dP_n}{dP_0}(\mathbf{X})$ the log likelihood ratio.

Lemma 2.2 *There exist finite positive constants K_1 and K_2 such that for all sufficiently large n*

- (i) $m_n := E_n \Lambda_n \leq K_1 \gamma \log n$;
- (ii) $\sigma_n^2 := E_n (\Lambda_n - m_n)^2 \leq K_2 \gamma \log n$.

From Lemmas 2.1 and 2.2, we have, as in (2.6) in GRS, that for any event A and any $a > 0$

$$P_n\{A\} \leq e^a P_0\{A\} + \frac{M \log^2 n}{a^2}, \quad (2.3)$$

with

$$M = (\gamma K_1)^2 + \gamma K_2. \quad (2.4)$$

Denoting $T_n = \phi_n^{-1}(\beta_1)\tilde{\alpha}$, we have, using (2.2), for any $\epsilon > 0$

$$\sup_{\beta \leq \beta^*} \phi_n^{-2}(\beta) R_n(\tilde{\alpha}, F(\beta)) \geq \frac{1}{2} \{E_0[\phi_n^{-1}(\beta_2)(\tilde{\alpha} - \alpha(f_0))]^2 + E_n[\phi_n^{-1}(\beta_1)(\tilde{\alpha} - \alpha(f_n))]^2\} \geq$$

$$\frac{\epsilon^2}{2} [\phi_n^{-1}(\beta_2)\phi_n(\beta_1)]^2 P_0\{|T_n| \geq \epsilon\} + \frac{1}{2} E_n[(T_n - \tau)^2 1\{|T_n| < \epsilon\}].$$

Using now (2.3) and choosing $\epsilon < \tau/2$ and

$$a = \log([\phi_n^{-1}(\beta_2)\phi_n(\beta_1)]^2), \quad (2.5)$$

we find that $\sup_{\beta \leq \beta^*} \phi_n^{-2}(\beta) R_n(\tilde{\alpha}, F(\beta))$ is lower-bounded by

$$\begin{aligned} & \frac{\epsilon^2}{2} [\phi_n^{-1}(\beta_2)\phi_n(\beta_1)]^2 \exp(-a) [P_n\{|T_n| \geq \epsilon\} - \frac{M \log^2 n}{a^2}] + \frac{1}{2} (\tau - \epsilon)^2 P_n\{|T_n| < \epsilon\} \\ & \geq \frac{1}{2} \min\{\epsilon^2, (\tau - \epsilon)^2\} - \frac{\epsilon^2}{2} \frac{M \log^2 n}{a^2} \\ & \geq \frac{\epsilon^2}{2} \left(1 - \frac{M \log^2 n}{a^2}\right). \end{aligned}$$

Using now (2.4) and the fact that a , as defined in (2.5), satisfies for all large enough n the inequality $a \geq D \log n$ with $D = \frac{\beta_2 - \beta_1}{(2\beta_1 + 1)(2\beta_2 + 1)}$, the last expression is lower-bounded by

$$\frac{\epsilon^2}{2} \left(1 - \frac{((\gamma K_1)^2 + \gamma K_2)}{D^2}\right) \geq \frac{\epsilon^2}{4},$$

on choosing $0 < \gamma < \frac{(K_2^2 + 2K_1^2 D^2)^{1/2} - K_2}{2K_1^2}$. □

3 Upper bound.

In this section we establish an upper bound for adaptive estimation, and present an estimator which attains it. To define our adaptive estimator we employ a further modification of the GPH estimator beyond that proposed in Robinson (1995a) and GRS, by using a tapered discrete Fourier transform (DFT), as do Hurvich and Ray (1995), Velasco (1998a,b) in a similar context. Let

$$w_h(\lambda) = (2\pi \sum_{t=1}^n h_t^2)^{-1/2} \sum_{t=1}^n h_t X_t e^{i\lambda t}, \quad (3.1)$$

where the sequence h_t is given by the cosine-bell taper

$$h_t = \frac{1}{2}(1 - \cos \lambda t), \quad t = 1, \dots, n, \quad (3.2)$$

for $\lambda_t = 2\pi t/n$. Define

$$\hat{\alpha}_m = -\frac{\sum_{j \in I(m)} \nu_j \log I_h(\lambda_j)}{\sum_{j \in I(m)} \nu_j^2}, \quad (3.3)$$

where

$$\nu_j = \log j - \frac{1}{p} \sum_{k \in I(m)} \log k,$$

$I_h(\lambda) = |w_h(\lambda)|^2$ and the sum $\sum_{j \in I(m)}$ is taken over $I(m) = \{j : j = l + 3k, k = 1, \dots, p\}$, where $p = [(m - l)/3]$, ($[a]$ is the integer part of a). (In the expression (3.1) for the log-periodogram regression estimator in our previous paper, GRS, the factors ν_j were erroneously omitted in the numerator sum due to a typographical error.) Here m is a bandwidth number, indicating the greatest frequency employed, and $l < m$ is a trimming number, $l + 2$ being the number of low frequencies discarded. Robinson (1995a) and GRS have used the estimator with the untapered DFT,

$$w(\lambda) = \left(\frac{1}{2\pi n}\right)^{1/2} \sum_{t=1}^n X_t e^{it\lambda}$$

in place of $w_h(\lambda)$, so that $h_t \equiv 1$, and with the summation over all $j \in [l + 1, m]$ in (3.3). The estimator (3.3), which tapers the modified GPH estimator of Robinson (1995a), was proposed by Velasco (1998a,b), for a different purpose. The motivation for the trimming in (3.3) is to produce sufficiently small autocorrelation between the $w_h(\lambda_j)$ and $w_h(\lambda_k)$, for $j \neq k$, (see Lemma 3.1) so as to enable an exponential inequality for $\hat{\alpha}_m$ (see Lemma 3.2). The motivation for omitting about 2/3 of the frequencies λ_j between λ_l and λ_m is suggested by the identity

$$w_h(\lambda_j) = -\frac{1}{\sqrt{6}}\{w(\lambda_{j-1}) - 2w(\lambda_j) + w(\lambda_{j+1})\}, \quad 2 \leq j \leq n - 2, \quad (3.4)$$

indicating non-negligible correlation between $w_h(\lambda_j)$ and $w_h(\lambda_k)$, $|j - k| \leq 2$. The basic motivation for an estimator of type (3.3), as in GPH, Robinson (1995a) and GRS, comes from approximating the logarithm of (1.1), and least squares regression of log periodogram ordinates on log frequencies. This works in the estimators of GPH, Robinson (1995a) and GRS due to the approximate independence of the $w(\lambda_j)$, but this property is only achieved for the $w_h(\lambda_j)$ by the omission of frequencies (though of course (3.4) implies that in fact all the $w(\lambda_j)$, $j = 1, \dots, m$ are used in (3.3)). A disadvantage of this device is that, for given m , the variance of the estimate $\hat{\alpha}_m$ is approximately tripled. This could be alleviated (but not completely corrected) by the pooling method employed in Robinson (1995a).

One desirable feature which $\hat{\alpha}_m$ preserves is invariance to location-shift in the X_t , due to (3.4) and the location-invariance of the $w(\lambda_j)$, $1 \leq j \leq n-1$; thus no mean correction is required, irrespective of whether or not μ is known.

The notation $\hat{\alpha}_m$ stresses the importance of the choice of bandwidth m . For given β (such as $\beta = 2$) it is possible, as in Hurvich, Deo and Brodsky (1997), and in common with many other problems of nonparametric smoothing, to minimize the mean square error $E_f\{\hat{\alpha}_m - \alpha\}^2$ by (cf. Lemma 4.3 below)

$$m = K(\alpha, \beta)n^{2r(\beta)}, \quad (3.5)$$

where $K(\alpha, \beta)$ depends not only on α and β but also, for integer β , on the β -th derivative of $L(\lambda)$ at $\lambda = 0$, and for noninteger β on an analogous quantity. For given β it may be possible to consistently estimate K in (3.5) by some plug-in method or cross-validation. In this paper we wish to adapt to unknown $\beta > 0$ so as to construct an estimate which is (as nearly as possible) adaptive rate-optimal.

The idea of the method we employ is due to Lepskii (1990) (see also Lepskii, et al., 1997) who developed it in a different context. Given that β in $F(\beta)$ is unknown, let $\gamma \leq \beta^*$ be any admissible value, and set

$$m(\gamma) = n^{2r(\gamma)}(\log n)^{\frac{2}{2\gamma+1}} \equiv n^{\frac{2\gamma}{2\gamma+1}}(\log n)^{\frac{2}{2\gamma+1}}. \quad (3.6)$$

Denote $\hat{\alpha}(\gamma) = \hat{\alpha}_{m(\gamma)}$, where $\hat{\alpha}_m$ is defined in (3.3). Let $h = 1/\log n$ and B_h be the h -net of the interval $[0, \beta^*]$

$$B_h = \{\gamma \geq 0 : \gamma = \beta^* - kh, k = 0, 1, 2, \dots\}.$$

Define

$$\hat{\beta} = \sup\{\gamma \in B_h : |\hat{\alpha}(\beta') - \hat{\alpha}(\gamma)| \leq m^{-1/2}(\beta')d(\beta'), \text{ for any } \beta' \leq \gamma, \beta' \in B_h\}, \quad (3.7)$$

where

$$d(\beta') = \frac{4k'}{(2\beta' + 1)^2}, \quad (3.8)$$

and $k' = (\beta^* - \beta')/h$, thereby defining $\hat{\alpha}(\hat{\beta})$.

Our proofs of Theorem 3.1 and Lemmas 3.2 and 4.3 require an assumption that the parameter β is bounded away from 0, i.e. that $\beta \geq \beta_*$ for some $\beta_* > 0$. We also assume that the number of low trimmed frequencies l satisfies the condition

$$l \geq l_*, \quad l = o\left(\frac{m}{(\log m)^3}\right), \quad \text{as } m \rightarrow \infty, \quad (3.9)$$

where the constant $l_* = l_*(\beta_*, \beta^*, C_1, C_2, \delta)$ does not depend on n , but must be chosen sufficiently large. Thus the proportion of trimmed frequencies on $(0, m]$ is negligible. The mildness of (3.9) is due to the particular taper (3.2) used; Theorem 3.1 could be established for tapers which entail less smoothness at the end-points of the sequence $\{h_t\}$ and correspondingly a slower rate of decay of its discrete Fourier transforms at cost of a stronger condition on l . Note that (3.9) is only a sufficient condition. As is common when trimming numbers are introduced for technical reasons, there seems no reasonably precise theoretical guide for the choice of l in practice.

Theorem 3.1 *Under assumption (3.9), uniformly in $\beta \in [\beta_*, \beta^*]$, where $\beta^* > 0$, the sequence $\{\phi_n^*(\beta) = (\log n)^2/n^{r(\beta)}\}$ gives an upper bound to the asymptotic minimax risk for the class $F(\beta) = F(\beta, C_1, C_2, \delta)$, with the estimator $\hat{\alpha}(\hat{\beta})$, defined following (3.6)-(3.8); that is, for some $\bar{C} < \infty$*

$$\limsup_{n \rightarrow \infty} \sup_{\beta \leq \beta^*} \phi_n^{*-2}(\beta) R_n(\hat{\alpha}(\hat{\beta}), F(\beta)) \leq \bar{C}. \quad (3.10)$$

Theorem 3.1 shows that there exists an estimator with the rate of convergence $(\log^2 n/n)^{\beta/(2\beta+1)}$ for all classes $F(\beta)$, $0 < \beta_* \leq \beta \leq \beta^*$. If this rate were equal to the rate $\phi_n(\beta)$ in the lower bound (2.1) for all $\beta_* \leq \beta \leq \beta^*$, then this common sequence could have been called an optimal adaptive rate, and $\hat{\alpha}(\hat{\beta})$ could be referred to as an adaptive rate-optimal estimator. The present results show that the the optimal attainable rate of convergence is between $(\log n/n)^{\beta/(2\beta+1)}$ and $(\log^2 n/n)^{\beta/(2\beta+1)}$, i.e. we have determined it up to logarithmical factor. The optimal attainable adaptive rate remains to be determined.

Notice that Theorem 3.1 would continue to hold if an arbitrary positive factor K were inserted in (3.6) (cf. (3.5)), and arbitrariness in K is then equivalent to arbitrariness in m . Thus $m(\hat{\beta})$ is not an optimal bandwidth with unknown β to the extent that (3.5) can be with known β , and we are concerned here only with showing the existence of an estimator which almost achieves an optimal adaptive rate of convergence, though this aspect is of uppermost importance for sufficiently large n . Note that m in (3.5) increases more slowly than $m(\beta)$ in (3.6), so that $E_f(\hat{\alpha}(\beta) - \alpha)^2 = O(n^{-2r(\beta)})$ decays faster than $\phi_n^{*2}(\beta)$. Since $\phi_n^*(\beta_1) \geq \phi_n^*(\beta_2)$ for $\beta_1 < \beta_2$, it is the largest β such that $f \in F(\beta)$ which determines the rate of convergence of $\hat{\alpha}(\hat{\beta})$ for given f . The grid h is sufficiently fine for our purposes in that $m(\gamma)$ is insensitive, for large n , to $O(1/\log n)$ shifts in γ . Note that $\hat{\beta}$, and thus $\hat{\alpha}(\hat{\beta})$, can be sensitive to the upper bound β^* on the admissible set B_h . In view of our earlier remarks following (1.2), a reasonable choice in many circumstances is $\beta^* = 2$. Of course the outcome $\hat{\beta} = \beta^*$ could indicate that a larger β^* should have been employed.

Since our goal is to show the existence of an estimator which achieves nearly optimal rate of convergence, we restrict ourselves to the log - periodogram regression estimator (3.3). We expect that Robinson's (1995b) narrow band Gaussian or Whittle estimator, also achieves the nearly optimal rate of convergence; it has the same rate of convergence as the log -periodogram estimator for the same bandwidth sequence. The investigation of this estimator is of interest, bearing in mind its nice statistical properties and its multivariate extension developed by Lobato (1998). An interesting open question is whether using data tapers as in Velasco (1998a, b), the memory parameter range $(-1, 1)$ can be extended to $(-1, 2)$, to cover some nonstationary processes.

The proof of Theorem 3.1 employs two lemmas, proofs of which are left to Section 5. The first describes the covariance properties of the normalised tapered DFT

$$v_h(\lambda) = \frac{w_h(\lambda)}{(c|\lambda|^{-\alpha})^{1/2}}.$$

Lemma 3.1 *For any $j = j_n, k = k_n$, such that $l \leq k \leq j - 3$ and $j \leq n/2$,*

- (a) $E_f v_h(\lambda_j) v_h(\lambda_j) = 1 + O(|\frac{j}{n}|^\beta + \frac{1}{j^2})$;
- (b) $E_f v_h(\lambda_j) v_h(\lambda_j) = O(\frac{1}{j^\beta})$;
- (c) $E_f v_h(\lambda_j) \overline{v_h(\lambda_k)} = O(|\frac{j}{n}|^\beta \frac{1}{|j-k|^2} + \frac{1}{k|j-k|^2} (\frac{j}{k})^{|\alpha|/2})$;
- (d) $E_f v_h(\lambda_j) v_h(\lambda_k) = O(|\frac{j}{n}|^\beta \frac{1}{|j-k|^2} + \frac{1}{k|j-k|^2} (\frac{j}{k})^{|\alpha|/2})$,

uniformly in $f \in F(\beta, C_1, C_2, \delta)$, $0 < \beta \leq \beta^$.*

Remark 3.1 Theorems 2.1-3.1 and Lemmas 3.1-3.2, 4.1-4.3 remain valid after replacing $F(\beta)$ in (1.2) with a class $F^*(\beta)$ with the following 'localized' definition:

$$F^*(\beta) = F^*(\beta, C_0, C_1, C_2, \delta, \lambda_0) = \{f : f(\lambda) = c|\lambda|^{-\alpha}(1 + \Delta(\lambda)), \quad C_0 < c \leq C_1, \\ -1 < \alpha < 1 - \delta, |\Delta(\lambda)| \leq C_2|\lambda|^\beta, \quad \text{for } |\lambda| \leq \lambda_0\},$$

where the constants $0 < C_0, C_1, C_2 < \infty$ and $\delta \in (0, 1)$ are independent of β , and $\lambda_0 > 0$. Class $F^*(\beta)$ does not contain any restriction on spectral densities $f \in F^*(\beta)$ for 'high' frequencies $\lambda \in [\lambda_0, \pi]$.

The only change in this case will be an additional assumption in Lemma 3.1 that the frequencies λ_k, λ_j satisfy the condition $l \leq k \leq j - 3 \leq m$, $m = o(n)$.

Let $A_m = \Sigma - I_{2p}/2$, where I_{2p} is $2p \times 2p$ identity matrix ($p = [(m-l)/3]$) and Σ is the covariance matrix of real and imaginary parts of the $v_h(\lambda_j)$, $j \in I(m)$. Denote by $\|X\|$ the Euclidean norm of the matrix X , $\|X\| = \{\text{tr}(X'X)\}^{1/2}$.

Lemma 3.2 *Under (3.9), there exist $c_1, c_2 \in (0, \infty)$ such that for all sufficiently large n , and $m = o(n)$*

$$E_f \exp\{\sqrt{m}|\hat{\alpha}_m - \alpha(f)|\} \leq c_1 \exp(c_2\|A_m\|^2) \quad (3.11)$$

uniformly in $f \in F(\beta, C_1, C_2, \delta)$, $\beta_ \leq \beta \leq \beta^*$ for any $0 < \beta_* \leq \beta^*$.*

PROOF OF THEOREM 3.1: The proof makes use of ideas of Lepskii (1990), Lepskii and Spokoiny (1995). We decompose the quadratic risk of $\hat{\alpha}(\hat{\beta})$ into two parts corresponding to the events $\{\hat{\beta} \leq \beta\}$ and $\{\hat{\beta} > \beta\}$:

$$\begin{aligned} E_f(\hat{\alpha}(\hat{\beta}) - \alpha(f))^2 &= E_f[(\hat{\alpha}(\hat{\beta}) - \alpha(f))^2 1\{\hat{\beta} \geq \beta\}] + E_f[(\hat{\alpha}(\hat{\beta}) - \alpha(f))^2 1\{\hat{\beta} < \beta\}] \\ &=: R_n^+ + R_n^-, \text{ say.} \end{aligned}$$

Now, (3.10) will follow if we show that uniformly in $f \in F(\beta)$, $\beta_* \leq \beta \leq \beta^*$,

$$R_n^+ = O(\phi_n^*(\beta)^2) \quad (3.12)$$

and

$$R_n^- = O(\phi_n^*(\beta)^2). \quad (3.13)$$

Note further that, since $\phi_n^*(\beta_1) \asymp \phi_n^*(\beta_2)$ when $|\beta_1 - \beta_2| = O(1/\log n)$, it is sufficient to establish (3.12) and (3.13) uniformly over $f \in F(\beta)$, $\beta \in B_h \cap [\beta_*, \beta^*]$.

Using the definition (3.7) of $\hat{\beta}$ and the fact, established in Lemma 4.3, that for $m = m(\beta)$ the estimator $\hat{\alpha}_m$ has mean squared error $O((\frac{m}{n})^{2\beta} + \frac{1}{m})$ in case $f \in F(\beta)$, we have

$$\begin{aligned} R_n^+ &\leq 2E_f[(\hat{\alpha}(\hat{\beta}) - \hat{\alpha}(\beta))^2 1\{\hat{\beta} \geq \beta\}] + 2E_f[(\hat{\alpha}(\beta) - \alpha(f))^2] \\ &\leq \frac{Ck_\beta^2}{(2\beta+1)^4} m(\beta)^{-1} E_f 1\{\hat{\beta} \geq \beta\} + C[(\frac{m(\beta)}{n})^{2\beta} + \frac{1}{m(\beta)}], \end{aligned}$$

where $k_\beta = (\beta^* - \beta)/h = (\beta^* - \beta) \log n$ and C here and below is a generic constant, not always the same. This implies that

$$R_n^+ \leq C[\frac{\log^2 n}{m(\beta)} + (\frac{m(\beta)}{n})^{2\beta} + \frac{1}{m(\beta)}] = O(\phi_n^*(\beta)^2).$$

Turning now to (3.13), we have

$$\begin{aligned} R_n^- &= E_f \sum_{\gamma < \beta, \gamma \in B_h} (\hat{\alpha}(\gamma) - \alpha(f))^2 1\{\hat{\beta} = \gamma\} \\ &= R_{n,1}^- + R_{n,2}^-, \end{aligned}$$

where

$$R_{n,i}^- = \sum_{\gamma < \beta, \gamma \in I_i} E_f (\hat{\alpha}(\gamma) - \alpha(f))^2 1\{\hat{\beta} = \gamma\} \quad i = 1, 2$$

and I_i denote subsets of B_h : $I_1 = \{\gamma \in B_h : (\frac{m(\gamma)}{n})^{2\beta} m(\gamma) \leq 1\}$; $I_2 = \{\gamma \in B_h : (\frac{m(\gamma)}{n})^{2\beta} m(\gamma) > 1\}$. By Cauchy inequality,

$$R_{n,1}^- \leq \sum_{\gamma < \beta, \gamma \in I_1} (E_f (\hat{\alpha}(\gamma) - \alpha(f))^4)^{1/2} P_f^{1/2} \{ \hat{\beta} = \gamma \}. \quad (3.14)$$

From Lemma 3.2, uniformly in $F(\beta)$

$$E(\sqrt{m(\gamma)}(\hat{\alpha}(\gamma) - \alpha(f))^4) \leq c_1 \exp(c_2 \|A_{m(\gamma)}\|^2)$$

and thus

$$\begin{aligned} \left(E_f (\hat{\alpha}(\gamma) - \alpha(f))^4 \right)^{1/2} &\leq C \exp(C \|A_{m(\gamma)}\|^2) n^{-2r(\gamma)} (\log n)^{-\frac{2}{2\gamma+1}} \\ &= C \exp(C \|A_{m(\gamma)}\|^2) n^{-2r(\beta)} (\log n)^{-\frac{2}{2\gamma+1}} \exp\left(\frac{2(k_\gamma - k_\beta)}{(2\gamma+1)(2\beta+1)}\right), \end{aligned}$$

where $k_\gamma = (\beta^* - \gamma)/h = (\beta^* - \gamma) \log n$.

Now we estimate $P_f \{ \hat{\beta} = \gamma \}$ for $\gamma \in I_1$. By definition of $\hat{\beta}$, if $\hat{\beta} = \gamma$, there exists $\beta' \leq \gamma$, $\beta' \in B_h$, such that

$$|\hat{\alpha}(\gamma+h) - \hat{\alpha}(\beta')| > m(\beta')^{-1/2} d(\beta').$$

Using this, we get from (3.8) for $\gamma \in I_1$

$$\begin{aligned} P_f \{ \hat{\beta} = \gamma \} &\leq \sum_{\beta' \leq \gamma, \beta' \in I_1} P_f \left(|\hat{\alpha}(\gamma+h) - \hat{\alpha}(\beta')| \geq m(\beta')^{-1/2} \frac{4k'}{(2\beta'+1)^2} \right) \\ &\leq \sum_{\beta' \leq \gamma, \beta' \in I_1} \exp\left(-\frac{4k'}{(2\beta'+1)^2}\right) E_f \exp(\sqrt{m(\beta')} |\hat{\alpha}(\gamma+h) - \hat{\alpha}(\beta')|), \end{aligned}$$

and since $m(\beta') \leq m(\gamma)$, we have, by Lemma 3.2,

$$\begin{aligned} P_f \{ \hat{\beta} = \gamma \} &\leq C \exp(C \|A_{m(\gamma+h)}\|^2) \sum_{\beta' \leq \gamma, \beta' \in B_h} \exp\left(-\frac{4k'}{(2\beta'+1)^2}\right) \\ &\leq C \exp(C \|A_{m(\gamma+h)}\|^2) \exp\left(-\frac{4k}{(2\gamma+1)^2}\right). \end{aligned} \quad (3.15)$$

Note that Lemma 4.1 and the definition of I_1 imply

$$\|A_{m(\gamma+h)}\|^2 \leq C \left(\left(\frac{m(\gamma+h)}{n} \right)^{2\beta} m(\gamma+h) + 1 \right) \leq C$$

uniformly in $f \in F(\beta)$, $\beta_* \leq \beta \leq \beta^*$. Therefore, combining (3.14)–(3.15), we get

$$\begin{aligned} R_{n,1}^- &\leq C n^{-2r(\beta)} \sum_{\gamma < \beta, \gamma \in I_1} \exp(C \|A_{m(\gamma+h)}\|^2) (\log n)^{-\frac{2}{2\gamma+1}} \exp\left(\frac{2(k_\gamma - k_\beta)}{(2\gamma+1)(2\beta+1)} - \frac{2k_\gamma}{(2\gamma+1)^2}\right) \\ &\leq C n^{-2r(\beta)} \sum_{\gamma < \beta, \gamma \in I_1} (\log n)^{-\frac{2}{2\gamma+1}} \exp\left(\frac{2(k_\gamma - k_\beta)}{(2\gamma+1)(2\beta+1)} - \frac{2k_\gamma}{(2\gamma+1)^2}\right) \end{aligned}$$

and, since $1/(2\beta + 1) \leq 1/(2\gamma + 1)$,

$$\begin{aligned} R_{n,1}^- &\leq C n^{-2r(\beta)} \sum_{\gamma < \beta, \gamma \in I_1} (\log n)^{-\frac{2}{2\beta+1}} \exp\left(\frac{-2k_\beta}{(2\beta+1)^2}\right) \\ &\leq C n^{-2r(\beta)} \log n (\log n)^{-\frac{2}{2\beta+1}} \leq C \phi_n^*(\beta)^2. \end{aligned}$$

Now we estimate $R_{n,2}^-$. By Lemma 4.3,

$$\begin{aligned} R_{n,2}^- &\leq \sum_{\gamma < \beta, \gamma \in I_2} E_f(\hat{\alpha}(m(\gamma)) - \alpha(f))^2 \leq C \sum_{\gamma < \beta, \gamma \in I_2} \left(\left(\frac{m(\gamma)}{n}\right)^{2\beta} + \frac{1}{m(\gamma)} \right) \\ &\leq C \sum_{\gamma < \beta, \gamma \in I_2} \left(\frac{m(\gamma)}{n}\right)^{2\beta}, \end{aligned}$$

by definition of I_2 . Note that $0 < \phi_n^*(\beta) \leq n^{-r(\beta)/2} < 1$ for n large enough. For such n ,

$$\begin{aligned} \sum_{\gamma < \beta, \gamma \in I_2} \left(\frac{m(\gamma)}{n}\right)^{2\beta} &= \sum_{\gamma < \beta} \phi_n^*(\beta)^{2(2\beta+1)/(2\gamma+1)} \leq \phi_n^*(\beta)^2 \sum_{\gamma < \beta} \phi_n^*(\beta)^{4(\beta-\gamma)/(2\gamma+1)} \\ &\leq \phi_n^*(\beta)^2 \sum_{j=0}^{\infty} \phi_n^*(\beta)^{4j/((2\beta+1)\log n)} \leq \phi_n^*(\beta)^2 \sum_{j=0}^{\infty} n^{-2r(\beta)j/((2\beta+1)\log n)} \\ &= \phi_n^*(\beta)^2 \sum_{j=0}^{\infty} e^{-2j\beta(2\beta+1)^{-2}} \leq (1 - e^{-2\beta^*(2\beta^*+1)^{-2}})^{-1} \phi_n^*(\beta)^2. \end{aligned}$$

□

4 Additional lemmas.

The following lemmas are also used, along with Lemmas 3.1 and 3.2, in the proof of Theorem 3.1.

Lemma 4.1 *For any sequence $m = o(n)$,*

$$\|A_m\|^2 \leq C \left[\left(\frac{m}{n}\right)^{2\beta} m + \frac{1}{l} \right] \quad (4.1)$$

uniformly in $f \in F(\beta)$, $0 < \beta \leq \beta^$.*

PROOF OF LEMMA 4.1: Let a_{st} denote the (s, t) -th element of A_m . Consider the contribution to $\|A_m\|^2$ of the a_{st} corresponding to $E_f(r_j r_k)$, $E_f(r_j i_k)$, $E_f(i_j i_k)$, where $r_j = \operatorname{Re} v_h(\lambda_j)$, $i_j = \operatorname{Im} v_h(\lambda_j)$ for $|j - k| \geq 3$. Routine manipulation of Lemma 3.1 (c,d) indicates that these expectations are all $O(p(j, k))$, uniformly in $f \in F(\beta)$, where

$$p(j, k) = (j/n)^\beta (j - k)^{-2} + k^{-1} (j - k)^{-2} (j/k)^{|\alpha|/2}, \quad l \leq k \leq j - 3.$$

Note that for l sufficiently large

$$\begin{aligned} p(j, k) &= p(j, k)1(k \leq j/2) + p(j, k)1(k \geq j/2) \\ &\leq (j/n)^\beta (j - k)^{-2} + k^{-1} (j/2)^{-2} (j/k)^{|\alpha|/2} + 2k^{-1} (j - k)^{-2} \\ &\leq (j/n)^\beta (j - k)^{-2} + 2k^{-1} (j - k)^{-3/2} =: p'(j, k). \end{aligned} \quad (4.2)$$

The contribution of these a_{st} to $\|A_m\|^2 = \sum_{s,t} a_{st}^2$ is

$$\begin{aligned} O\left(\sum_{l < k < j \leq m} p'(j, k)^2\right) &\leq C \sum_{l \leq k < j \leq m} \left[\left(\frac{j}{n}\right)^{2\beta} (j-k)^{-4} + k^{-2} (j-k)^{-3}\right] \\ &\leq C \sum_{l \leq j \leq m} \left[\left(\frac{j}{n}\right)^{2\beta} + j^{-2}\right] \leq C\left(\left(\frac{m}{n}\right)^{2\beta} m + l^{-1}\right) \end{aligned} \quad (4.3)$$

uniformly in $f \in F(\beta)$, $0 < \beta \leq \beta^*$. It is easily seen from Lemma 3.1 (a,b) that the contribution of the $o(m)$ a_{st} corresponding to $E_f(r_j^2)$, $E_f(r_j i_j)$, $E_f(i_j^2)$ is dominated by (4.3). \square

Denote by $\|X\|_{sp}$ the spectral norm of the matrix X , the square root of the largest eigenvalue of $X'X$.

Lemma 4.2 *For any sequence $m = o(n)$,*

$$\|A_m\|_{sp} \leq C\left(\left(\frac{m}{n}\right)^\beta + \frac{1}{l}\right)$$

uniformly in $f \in F(\beta)$, $0 < \beta \leq \beta^$.*

PROOF OF LEMMA 4.2: If x_t is the t -th element of the $2p \times 2p$ vector x , $p = [(m-l)/3]$, then

$$\begin{aligned} \|A_m\|_{sp}^2 &\leq 2 \sup_{\|x\|=1} \sum_{s,t,u} x_s a_{st} a_{tu} x_u \\ &\leq 2 \sup_{\|x\|=1} \sum_{s,t,u} x_s^2 |a_{st} a_{tu}| \leq 2(\max_s \sum_t |a_{st}|)^2, \end{aligned}$$

using Cauchy inequality in the second line. The contribution from the a_{st} corresponding to $E_f(r_j r_k)$, $E_f(r_j i_k)$, $E_f(i_j i_k)$, for $j \neq k$, is from (4.2),

$$\begin{aligned} \max_j \sum_{k:l < k < j} |p'(j, k)| &\leq \max_j \left\{ \left(\frac{j}{n}\right)^\beta \sum_{k < j} |j-k|^{-2} + \frac{1}{l} \sum_{k < j} |j-k|^{-3/2} \right\} \\ &= O\left(\left(\frac{m}{n}\right)^\beta + l^{-1}\right) \end{aligned}$$

uniformly in $f \in F(\beta)$, $0 < \beta \leq \beta^*$, while the contribution of the remaining a_{st} is easily seen to be dominated by this. \square

Lemma 4.3 *For $m = o(n)$,*

$$E_f(\hat{\alpha}(m) - \alpha(f))^2 = O\left(\left(\frac{m}{n}\right)^{2\beta} + \frac{1}{m}\right)$$

uniformly in $f \in F(\beta)$, $0 < \beta_ \leq \beta \leq \beta^*$.*

PROOF OF LEMMA 4.3: The proof is similar to that of Theorem 2 of GRS, up to (3.9) of that paper. We deviate from that proof by bounding

$$\left| \exp\left(-\frac{1}{2} z^T \bar{\Phi} z\right) - 1 \right| \leq \|z\|^2 \|\bar{\Phi}\| \exp(\|z\|^2 \|\bar{\Phi}\|) \quad (4.4)$$

where $\bar{\Phi}$ is the 4×4 matrix whose 2×2 blocks on the main diagonal are zero, and whose other elements correspond to those of the inverse of the covariance matrix of (r_j, i_j, r_k, i_k) , for some $j \neq k$. From observations in the proofs of Lemmas 4.1 and (4.2), it follows that $\|\Phi\| \leq Cp'(j, k)$ and thus for any $\epsilon > 0$ $\|\Phi\| \leq \epsilon$ for $l < k < j < m = o(n)$ and l and n large enough, so that (4.4) is

$$O(p'(j, k) \|z\|^2 \exp(\epsilon \|z\|^2)).$$

The remainder of the proof is straightforward, using also (a,b) of Lemma 3.1 and proceeding much as in the proof of Theorem 2 of GRS. \square

5 Proofs of Lemmas 3.1 and 3.2.

PROOF OF LEMMA 3.1: The most important results, so far as the proof of Lemma 3.2 and Theorem 3.1 are concerned, are (c) and (d), and we focus principally on these. Denote:

$$E_{j,k}(\lambda) = \frac{1}{2\pi(3n/8)} D_n^{(h)}(\lambda_j - \lambda) D_n^{(h)}(\lambda - \lambda_k),$$

where for h_t given by (3.2)

$$D_n^{(h)}(\lambda) = \sum_{t=1}^n h_t e^{i\lambda t}$$

and we have used $\sum_{t=1}^n h_t^2 = 3n/8$. From the orthogonality relation

$$\int_{-\pi}^{\pi} E_{j,k}(\lambda) d\lambda = 0, \quad 3 \leq |k| \leq j-3, \quad (5.1)$$

we have for such j, k

$$E_f w_h(\lambda_j) \overline{w_h(\lambda_k)} = \int_{-\pi}^{\pi} f(\lambda) E_{j,k}(\lambda) d\lambda = \int_{-\pi}^{\pi} (f(\lambda) - c\lambda_j^{-\alpha}) E_{j,k}(\lambda) d\lambda =: \sum_{r=1}^3 q_r(j, k), \quad (5.2)$$

where

$$q_r(j, k) := \int_{W_r(j,k)} (f(\lambda) - c\lambda_j^{-\alpha}) E_{j,k}(\lambda) d\lambda, \quad r = 1, 2, 3,$$

and

$$W_1(j, k) = \{|\lambda| \leq \lambda_k/2\}, \quad W_2(j, k) = \{\lambda_k/2 < |\lambda| \leq 3\lambda_j/2\}, \quad W_3(j, k) = \{3\lambda_j/2 \leq |\lambda| \leq \pi\}.$$

It is sufficient to show that

$$|q_r(j, k)| c^{-1} (\lambda_j \lambda_k)^{\alpha/2} \leq C[(j-k)^{-2} (j/n)^\beta + k^{-1} (j-k)^{-2} (j/k)^{|\alpha|/2}], \quad r = 1, 2, 3, \quad (5.3)$$

uniformly in $f \in F(\beta)$, $0 < \beta \leq \beta^*$.

We estimate first $q_1(j, k)$. By definition of $F(\beta)$ we have that

$$|f(\lambda) - c\lambda_j^{-\alpha}| \leq f(\lambda) + c\lambda_j^{-\alpha} \quad (5.4)$$

uniformly over $f \in F(\beta)$, $0 < \beta \leq \beta^*$. Thus,

$$|q_1(j, k)| \leq \int_{|\lambda| \leq |\lambda_k|/2} (f(\lambda) + c\lambda_j^{-\alpha}) |E_{j,k}(\lambda)| d\lambda.$$

Now note that

$$|D_n^{(h)}(\lambda)| \leq \frac{Cn}{(1+n|\lambda|)^3}, \quad 0 < \lambda \leq p_0 < 2\pi, \quad (5.5)$$

for any $0 < p_0 < 2\pi$ as can be established by repeated use of summation by parts (see also Velasco, 1998a, Hannan, 1970, pp.265-7). From (5.5)

$$|E_{j,k}(\lambda)| \leq Ce_{j,k}(\lambda), \quad e_{j,k}(\lambda) := n(1+n|\lambda_j-\lambda|)^{-3}(1+n|\lambda_k-\lambda|)^{-3}. \quad (5.6)$$

Since $|\lambda| \leq \lambda_k/2$ implies $|\lambda - \lambda_k| \geq \lambda_k/2$ and $|\lambda - \lambda_j| \geq \lambda_j - \lambda_k$, we can estimate

$$e_{j,k}(\lambda) \leq n(1+n(\lambda_j-\lambda_k))^{-3}(1+n\lambda_k/2)^{-3} \leq Cn(j-k)^{-3}k^{-3},$$

so that

$$|q_1(j,k)| \leq Cn(j-k)^{-3}k^{-3} \int_{|\lambda| \leq \lambda_k/2} (f(\lambda) + c\lambda_j^{-\alpha})d\lambda.$$

By definition of $F(\beta)$, $f \leq cC|\lambda|^{-\alpha}$ with $\alpha \in (-1, 1-\delta)$, and we get

$$|q_1(j,k)| \leq cCn(j-k)^{-3}k^{-3}(\lambda_k^{-\alpha+1} + \lambda_j^{-\alpha}\lambda_k).$$

Thus,

$$|q_1(j,k)|c^{-1}(\lambda_j\lambda_k)^{\alpha/2} \leq C(j-k)^{-3}k^{-1}(j/k)^{|\alpha|/2}.$$

We have obtained (5.3) for $r = 1$.

We estimate now $q_2(j,k)$. Note that for $\lambda \in W_2(j,k)$ we have

$$\begin{aligned} |f(\lambda) - c\lambda_j^{-\alpha}| &\leq |f(\lambda) - c|\lambda|^{-\alpha}| + |c(|\lambda|^{-\alpha} - \lambda_j^{-\alpha})| \\ &\leq cC(\lambda_j^{-\alpha+\beta} + \lambda_k^{-\alpha+\beta} + \lambda_k^{-\alpha-1}|\lambda - \lambda_j|) \end{aligned} \quad (5.7)$$

from $f \in F(\beta)$ and the mean value theorem, which gives for $|\lambda| \geq \lambda_k/2$

$$||\lambda|^{-\alpha} - \lambda_j^{-\alpha}| \leq \sup_{\eta \geq \lambda_k/2} \left| \frac{d\eta^{-\alpha}}{d\eta} \right| |\lambda| - \lambda_j| \leq C|\lambda_k|^{-\alpha-1}|\lambda - \lambda_j|.$$

Note that the greatest distance between λ and λ_j and λ_k equals at least $\frac{1}{2}|\lambda_j - \lambda_k|$. Using

$$\begin{aligned} (1+a)^{-3}(1+b)^{-3} &\leq (1+\max(a,b))^{-3}(1+\min(a,b))^{-3} \\ &\leq \max(a,b)^{-3}[(1+a)^{-3} + (1+b)^{-3}] \end{aligned} \quad (5.8)$$

for $a, b > 0$ we get

$$e_{j,k}(\lambda) \leq C(j-k)^{-3}n[(1+n|\lambda_j-\lambda|)^{-3} + (1+n|\lambda_k-\lambda|)^{-3}]. \quad (5.9)$$

Similarly, by (5.6) and (5.8)

$$\begin{aligned} |\lambda - \lambda_j|e_{j,k}(\lambda) &\leq C(1+n|\lambda_j-\lambda|)^{-2}(1+n|\lambda_k-\lambda|)^{-3} \\ &\leq C(j-k)^{-2}[(1+n|\lambda_k-\lambda|)^{-3} + (1+n|\lambda_j-\lambda|)^{-3}]. \end{aligned}$$

(Note that $(1+a)^{-2}(1+b)^{-3} \leq \max(a,b)^{-2}[(1+a)^{-3} + (1+b)^{-3}]$ for $a, b > 0$.) Hence, by (5.7),

$$\begin{aligned} c^{-1}|q_2(j,k)| &\leq C \int_{\lambda_k/2 \leq |\lambda| \leq 3\lambda_j/2} (\lambda_j^{-\alpha+\beta} + \lambda_k^{-\alpha+\beta} + \lambda_k^{-\alpha-1}|\lambda - \lambda_j|)e_{j,k}(\lambda)d\lambda \\ &\leq C \left\{ (\lambda_j^{-\alpha+\beta} + \lambda_k^{-\alpha+\beta})(j-k)^{-3} + \lambda_k^{-\alpha-1}n^{-1}(j-k)^{-2} \right\} \\ &\quad \int_{-\pi}^{\pi} n \left[(1+n|\lambda_k-\lambda|)^{-3} + (1+n|\lambda_j-\lambda|)^{-3} \right] d\lambda \\ &\leq C(\lambda_j^{-\alpha+\beta} + \lambda_k^{-\alpha+\beta})(j-k)^{-3} + C\lambda_k^{-\alpha}k^{-1}(j-k)^{-2}. \end{aligned}$$

Note that

$$\int_{-\pi}^{\pi} (1 + n|\lambda - \lambda_j|)^{-3} d\lambda \leq 2 \int_0^{2\pi} (1 + n\lambda)^{-3} d\lambda \leq 2n^{-1} \int_0^{\infty} (1 + x)^{-3} dx = n^{-1}.$$

Thus,

$$\begin{aligned} |q_2(j, k)|c^{-1}(\lambda_j \lambda_k)^{\alpha/2} &\leq C(j/k)^{|\alpha|/2} \{(j-k)^{-3} \lambda_j^\beta + k^{-1}(j-k)^{-2}\} \\ &\leq C((j-k)^{-2} \lambda_j^\beta + (j/k)^{|\alpha|/2} k^{-1}(j-k)^{-2}). \end{aligned}$$

Here we used

$$j/(k(j-k)) = 1/k + 1/(j-k) \leq 2.$$

Thus (5.3) holds for $r = 2$. Note that $|\lambda| \geq 3\lambda_j/2$ implies $|\lambda - \lambda_j| \geq |\lambda|/3$ and $|\lambda - \lambda_k| \geq |\lambda|/3$, so we can estimate

$$e_{j,k}(\lambda) \leq Cn^{-5}|\lambda|^{-6}, \quad |\lambda| \geq 3\lambda_j/2. \quad (5.10)$$

Thus, by (5.10), in view of $f(\lambda) \leq cC|\lambda|^{-\alpha}$, $|\lambda| \leq \pi$ uniformly in f , we get

$$\begin{aligned} c^{-1}|q_3(j, k)| &\leq C \int_{3\lambda_j/2 \leq |\lambda| \leq \pi} |\lambda|^{-\alpha} e_{j,k}(\lambda) d\lambda \leq C \int_{3\lambda_j/2 \leq |\lambda| \leq \pi} |\lambda|^{-\alpha} n^{-5} |\lambda|^{-6} d\lambda \\ &= Cn^{-5} \lambda_j^{-\alpha-5} = Cj^{-5} \lambda_j^{-\alpha}. \end{aligned}$$

Therefore

$$|q_3(j, k)|c^{-1}(\lambda_j \lambda_k)^{\alpha/2} \leq C\left(\frac{j}{k}\right)^{|\alpha|/2} j^{-5} \leq C\left(\frac{j}{k}\right)^{|\alpha|/2} k^{-1}(j-k)^{-2},$$

so (5.3) holds for $r = 3$. This completes the proof of (c).

To prove (d), note that for $l < k \leq j - 3 \leq n - 6$,

$$E_f w_h(\lambda_j) w_h(\lambda_k) = \int_{-\pi}^{\pi} f(\lambda) E_{j,-k}(\lambda) d\lambda = \int_{-\pi}^{\pi} (f(\lambda) - c\lambda_j^{-\alpha}) E_{j,-k}(\lambda) d\lambda.$$

Thus, similarly as in (c), from (5.6), it follows that

$$|E_f w_h(\lambda_j) w_h(\lambda_k)| \leq C \int_{-\pi}^{\pi} |f(\lambda) - c\lambda_j^{-\alpha}| \frac{1}{n} \frac{n}{(1+n|\lambda-\lambda_j|)^3} \frac{n}{(1+n|\lambda_k+\lambda|)^3} d\lambda.$$

Since $f(\lambda) = f(-\lambda)$ and

$$(1+n|\lambda-\lambda_j|)(1+n|\lambda+\lambda_k|) \geq (1+n||\lambda-\lambda_j|)(1+n||\lambda-\lambda_k|)$$

we get:

$$|E_f w_h(\lambda_j) w_h(\lambda_k)| \leq C \int_{-\pi}^{\pi} |f(\lambda) - c\lambda_j^{-\alpha}| n(1+n|\lambda-\lambda_j|)^{-3} (1+n|\lambda-\lambda_k|)^{-3} d\lambda.$$

This bound is the same as for the terms in (5.2), and therefore (d) holds by the same argument as in (c).

To prove (a), we have

$$|E_f w_h(\lambda_j) \overline{w_h(\lambda_j)} - c\lambda_j^{-\alpha}| = \left| \int_{-\pi}^{\pi} (f(\lambda) - c\lambda_j^{-\alpha}) \frac{1}{2\pi(3n/8)} |D_n^{(h)}(\lambda - \lambda_j)|^2 d\lambda \right|$$

$$\begin{aligned}
&\leq \left| \int_{|\lambda| \leq \lambda_j/2} [\dots] d\lambda \right| + \left| \int_{\lambda_j/2 \leq \lambda \leq 3\lambda_j/2} [\dots] d\lambda \right| + \left| \int_{-3\lambda_j/2 \leq \lambda \leq -\lambda_j/2} [\dots] d\lambda \right| + \left| \int_{3\lambda_j/2 \leq |\lambda| \leq \pi} [\dots] d\lambda \right| \\
&=: C(t_1 + t_2 + t_3 + t_4).
\end{aligned}$$

It remains to show that

$$t_i \lambda_j^\alpha \leq cC(\lambda_j^\beta + j^{-2}), \quad j = 1, 2, 3, 4.$$

We start with t_1 . Using (5.5) and $(1 + n|\lambda - \lambda_j|)^{-3} \leq Cj^{-3}$ for $|\lambda| \leq \lambda_j/2$, we have $n^{-1}|D_n^{(h)}(\lambda)|^2 \leq Cnj^{-6}$, and

$$t_1 \leq Cnj^{-6} \int_{|\lambda| \leq \lambda_j/2} (f(\lambda) + c\lambda_j^{-\alpha}) d\lambda.$$

Using the same argument as estimating $q_1(j, k)$ above we get $t_1 \lambda_j^\alpha \leq cCj^{-2}$. Next,

$$\begin{aligned}
c^{-1}t_2 &\leq c^{-1} \left| \int_{\lambda_j/2 \leq \lambda \leq 3\lambda_j/2} [(f(\lambda) - c\lambda^{-\alpha}) + (c|\lambda|^{-\alpha} - c\lambda_j^{-\alpha})] E_{j,j} d\lambda \right| \\
&\leq C \int_{\lambda_j/2 \leq \lambda \leq 3\lambda_j/2} |\lambda|^{-\alpha+\beta} n(1 + n|\lambda - \lambda_j|)^{-6} d\lambda + |D| \\
&\leq C\lambda_j^{-\alpha+\beta} + |D|,
\end{aligned}$$

where

$$D = \int_{\lambda_j/2 \leq \lambda \leq 3\lambda_j/2} (\lambda^{-\alpha} - \lambda_j^{-\alpha}) \frac{1}{2\pi(3n/8)} |D_n^{(h)}(\lambda - \lambda_j)|^2 d\lambda.$$

Using Taylor expansion, (5.5) and $|D_n^{(h)}(\lambda)| = |D_n^{(h)}(-\lambda)|$,

$$\begin{aligned}
D &= \int_{-\lambda_j/2 \leq \lambda \leq \lambda_j/2} ((\lambda_j - \lambda)^{-\alpha} - \lambda_j^{-\alpha}) \frac{1}{2\pi(3n/8)} |D_n^{(h)}(\lambda)|^2 d\lambda \\
&= \int_{-\lambda_j/2 \leq \lambda \leq \lambda_j/2} (-\alpha\lambda_j^{-\alpha-1}\lambda + O(\lambda^2\lambda_j^{-\alpha-2})) \frac{1}{2\pi(3n/8)} |D_n^{(h)}(\lambda)|^2 d\lambda \\
&= O\left(\int_{-\lambda_j/2 \leq \lambda \leq \lambda_j/2} \lambda^2 \lambda_j^{-\alpha-2} n(1 + n|\lambda|)^{-6} d\lambda \right) \\
&= O\left(\lambda_j^{-\alpha-2} n^{-2} \int_0^\infty x^2 (1+x)^{-6} dx \right) = O(\lambda_j^{-\alpha} j^{-2}).
\end{aligned}$$

Thus

$$t_2 c^{-1} \lambda_j^\alpha \leq C(\lambda_j^\beta + j^{-2}).$$

The term t_3 is estimated similarly to t_2 .

We end the proof of item (a) by estimating t_4 . For $\pi \geq |\lambda| \geq 3\lambda_j/2$ we have from (5.5)

$$\frac{1}{2\pi(3n/8)} |D_n^{(h)}(\lambda - \lambda_j)|^2 d\lambda \leq Cn^{-5} |\lambda|^{-6},$$

so

$$t_4 \leq C \int_{3\lambda_j/2 \leq |\lambda| \leq \pi} (f(\lambda) + c\lambda_j^{-\alpha}) n^{-5} |\lambda|^{-6} d\lambda.$$

Therefore, similarly to estimating $q_3(j, k)$ we get $t_4 c^{-1} \lambda_j^\alpha \leq Cj^{-2}$.

To prove (b) for $2 < j < n$, we have:

$$\begin{aligned}
|Ew_h(\lambda_j)w_h(\lambda_j)| &= \left| \int_{-\pi}^{\pi} f(\lambda)E_{j,-j}(\lambda)d\lambda \right| = \left| \int_{-\pi}^{\pi} (f(\lambda) - c|\lambda_j|^{-\alpha})E_{j,-j}(\lambda)d\lambda \right| \\
&\leq C \int_{-\pi}^{\pi} |f(\lambda) - c\lambda_j^{-\alpha}| n(1 + n|\lambda - \lambda_j|)^{-3}(1 + n|\lambda + \lambda_j|)^{-3} d\lambda \\
&= C \left(\int_{|\lambda| \leq \lambda_j/2} [\dots] d\lambda + \int_{\lambda_j/2 \leq |\lambda| \leq 3\lambda_j/2} [\dots] d\lambda + \int_{3\lambda_j/2 \leq |\lambda| \leq \pi} [\dots] d\lambda \right) \\
&=: s_1(j) + s_2(j) + s_3(j).
\end{aligned}$$

Using the argument employed in estimating $q_1(j, k), q_2(j, k), q_3(j, k)$ in (c), we can show that $s_i(j)c^{-1}\lambda_j^\alpha \leq Cj^{-3}$, $i = 1, 2, 3$ uniformly in f and j .

This completes the proof of (b) and Lemma 3.1. \square

PROOF OF LEMMA 3.2: Put $\mu_j = \nu_j\sqrt{p}/\sum_j' \nu_k^2$, where \sum_j' denotes the sum $\sum_{j \in I(m)}$. Because $\sum_j' \nu_j = 0$ and $\sum_j' \nu_j^2 \sim p$ as $n \rightarrow \infty$ (cf Robinson (1995b)), it follows that

$$\sum_j' \mu_j = 0, \quad \sum_j' \mu_j^2 \rightarrow 1 \quad m \rightarrow \infty. \quad (5.11)$$

From (3.3)

$$\sqrt{m}(\hat{\alpha}_m - \alpha(f)) = -\sum_j' \mu_j u_j,$$

where $u_j = \log|v_h(\lambda_j)|^2 + \eta$, with $\eta = 0.5772\dots$ Euler's constant. To prove (3.11) we have to show that for all sufficiently large n and $m = o(n)$

$$J := E_f \exp(\pm \sum_j' \mu_j u_j) = E_f [\prod_j' |v_h(\lambda_j)|^{\pm 2\mu_j}] < c_1 \exp(c_2 \|A_m\|^2), \quad (5.12)$$

uniformly in $f \in F(\beta)$, $0 < \beta_* \leq \beta \leq \beta^*$, where $\prod_j' = \prod_{j \in I(m)}$. The expectation in (5.12) is with respect to the $2p$ -dimensional Gaussian distribution with covariance matrix $\Sigma = I_{2p}/2 + A_m$ which is nonsingular because $\|A_m\|_{sp} < 1/2$, as follows for large enough n from Lemma 4.2 and (3.9). Denoting by $x_j, j \in I(m)$ the two-dimensional components of x ,

$$\begin{aligned}
J &= \frac{|\Sigma|^{-1/2}}{(2\pi)^p} \int \prod_j' (x_j^T x_j)^{\pm \mu_j} \exp(-2^{-1} x^T \Sigma^{-1} x) dx \\
&= \frac{|\Sigma|^{-1/2}}{(2\pi)^p} \int \prod_j' (x_j^T x_j)^{\pm \mu_j} \exp\left(-\frac{x^T x}{2}\right) \exp\left(-\frac{1}{2} x^T (\Sigma^{-1} - I_{2p}) x\right) dx \\
&\leq J_0^{1/2} \prod_j' J_j^{1/2}, \quad (5.13)
\end{aligned}$$

by Cauchy-Schwarz inequality, where

$$J_0 = \frac{|\Sigma|^{-1}}{(2\pi)^p} \int \exp(-x^T (\Sigma^{-1} - I_{2p}) x) dx, \quad J_j = \frac{1}{2\pi} \int (x_j^T x_j)^{\pm 2\mu_j} \exp(-x_j^T x_j) dx_j, \quad j > 0.$$

Now

$$J_0 = |2\Sigma - 2\Sigma^2|^{-1/2} = 2^p \prod_j' (1 - 4\eta_j^2)^{-1/2}$$

where η_j is the j -th eigenvalue of A_m . From the inequality $1 - x \geq e^{-2x}$, $0 < x < 1/2$, and Lemma 4.2, for n large enough

$$J_0 \leq 2^p \exp(4 \sum_j' \eta_j^2) = 2^p \exp(4 \|A_m\|^2). \quad (5.14)$$

On the other hand, after transformation to the polar coordinates, as $\mu_j \rightarrow 0$ (which follows from $\max_{j \in I(m)} |\mu_j| \rightarrow 0$ as $n \rightarrow \infty$)

$$J_j = \int_0^\infty r^{\pm 4\mu_j + 1} \exp(-r^2) dr = \frac{1}{2} \Gamma(1 \pm 2\mu_j) = \frac{1}{2} \exp\{\mp 2\mu_j \eta + O(\mu_j^2)\}$$

from the two-term mean value expansion for $\log \Gamma(1 + z)$. From (5.11), $\prod_j' J_j \leq C 2^{-p}$. Then (5.12) follows from (5.13) and (5.14). \square

REFERENCES

- GEWEKE, J. AND PORTER-HUDAK, S. (1983). The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221-238.
- 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**, 49-60.
- HANNAN, E.J. (1970). *Multiple Time Series*. Wiley, New York.
- HALL, P. AND WELSH, A. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.*, **12**, 1079-84.
- HURVICH, C.M., DEO, R.S. AND BRODSKY, J. (1997). The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long memory time series . *J. Time Ser. Anal.*, *forthcoming*
- HURVICH, C.M. AND RAY, B (1995). Estimation of the memory parameter for nonstationary or noninvertable fractionally integrated processes. *J. Time Ser. Anal.*, **16**, 17-42.
- KÜNSCH, H. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* , **23**, 1025-30.
- KÜNSCH, H. (1987). Statistical properties of self-similar processes. *Proc. 1st World Congress of the Bernoulli Society*, **4**, 67-74, VNU Science Press.
- LEPSKII, O.J. (1990). On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Apl.*, **35** 459-470.
- LEPSKII, O.J., MAMMEN, E., AND SPOKOINY, V.G. (1997). Ideal spatial adaptation to inhomogeneous smoothness: approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* , **25**, 929-947.
- LOBATO, I. (1998). A semiparametric two - step estimator in a multivariate long memory model. *J. Econometrics*, *forthcoming*.
- ROBINSON, P.M. (1995a). Log-periodogram regression of time series with long range dependence *Ann. Statist.* , **23**, 1048-1072.
- ROBINSON, P.M. (1995b). Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, **23**, 1630-1661.
- VELASCO, C. (1998a). Non-stationary log-periodogram regression. *J. Econometrics*, *forthcoming*.
- VELASCO, C. (1998b). Gaussian semiparametric estimation of non - stationary time series. *J. Time Ser. Anal.*, *forthcoming*.

Liudas Giraitis, London School of Economics, Department of Economics, Houghton Street, London WC2A 2AE. *email*: L.Giraitis@lse.ac.uk

Peter M. Robinson, London School of Economics, Department of Economics, Houghton Street, London WC2A 2AE. *email*: P.M.Robinson@lse.ac.uk

Alexander Samarov, Sloan School of Management, E40-131, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge, MA, 02139, USA. *email*: samarov@mit.edu