Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence

Liudas Giraitis[1], Heidelberg University

Peter M. Robinson[2], London School of Economics

Alexander Samarov[1,3], University of Massachusetts and MIT.

There exist several estimators of the memory parameter in long-memory time series models with mean μ and the spectrum specified only locally near zero frequency. In this paper we give a lower bound for the rate of convergence of any estimator of the memory parameter as a function of the degree of local smoothness of the spectral density at zero. The lower bound allows one to evaluate and compare different estimators by their asymptotic behavior, and to claim the rate optimality for any estimator attaining the bound. A log-periodogram regression estimator, analysed by Robinson (1992), is then shown to attain the lower bound, and is thus rate optimal.

AMS 1991 *subject classifications.* Primary 62M15; secondary 62G20, 62G05.

1. Research supported by Deutsche Forschungsgemeinschaft.
2. Research supported in part by the ESRC Grant R000233609.
3. Partially supported by NSF Grant DMS-93–06245 and by a grant from IFSRC at MIT.

Key words and phrases. Long range dependence; semiparametric models; optimal rates of convergence; lower bounds.

1. Introduction.

Suppose that we have n observations $X_1, \ldots, X_n$ from a long range dependent, stationary, Gaussian time series $\{X_t\}_{t=-\infty}^{\infty}$ with spectral density

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

and $L(\lambda) \to C$, $C \in (0, \infty)$, as $\lambda \to 0$. We will assume, without loss of generality, that the mean of X_t is zero (note that Theorem 2 below does not use this assumption). The parameter α determines the behavior of the spectrum near zero and is just a re-expression of the self-similarity parameter $H = (\alpha + 1)/2$ and of the fractional differencing parameter $d = \alpha/2$, see, e.g., Beran (1994). When $\alpha = 0$, $f(\lambda)$ tends to a finite positive constant at zero frequency, while if $\alpha \in (0, 1)$ it tends to infinity and if $\alpha \in (-1, 0)$ it tends to zero.

Several estimators of α are now available both for the parametric case when $L(\lambda)$ is specified for all $\lambda \in [-\pi, \pi]$ up to a finite-dimensional parameter (e.g., Fox and Taqqu (1986), Dahlhaus (1989) who required $\alpha \in (0, 1)$) and for the semiparametric case when assumptions are made only about the local, for $\lambda \to 0$, behavior of $L(\lambda)$ (as in (1.1) above), which then acts as an infinite-dimensional nuisance parameter (e.g., Geweke and Porter-Hudak (1983), Künsch (1986), Robinson (1992,1993)).

Perhaps, the best known estimator of α in the semiparametric case is that proposed by Geweke and Porter-Hudak (1983) which is based on the linear least-squares regression of log-periodogram on $\log(2\sin(\lambda_j/2))$, for a certain number of Fourier frequencies $\lambda_j = 2\pi j/n$ close to zero. Robinson (1992) analysed the asymptotic properties of a generalized and modified form of that estimator. It follows from his results that under certain local smoothness conditions, which include the assumption

$$L(\lambda) = C + O(|\lambda|^{\beta}), \ \text{as } \lambda \to 0, \ C \in (0, \infty), \ \beta \in (0, 2], \tag{1.2}$$

the estimator has the convergence rate $n^{-r}M_n$ for all $\alpha \in (-1, 1)$, where $r = r(\beta) = \beta/(2\beta + 1)$ and $M_n \to \infty$ arbitrarily slowly.

In his more recent paper, Robinson (1993) analysed another estimator, originally proposed by Künsch (1987), which was based on the local Whittle approximation of the Gaussian likelihood for the frequencies near zero. Under much weaker assumptions than those for the log-periodogram regression estimator, he established the estimator's asymptotic properties, which imply, under the condition (1.2), the rate $n^{-r}M_n$ with any M_n such that $\log^{-1/(1+2\beta)}(n)/M_n = o(1)$.

Note that Robinson (1992, 1993) proves not just the rate of convergence of the considered estimators but their asymptotic normality with mean zero. By analogy with many other nonparametric estimation problems, one can expect that if the number of Fourier frequencies used is chosen to optimally balance the asymptotic bias and variance, these estimators will attain the rate n^{-r}.

In this paper we consider the semiparametric model (1.1), (1.2) and give a lower bound for the rate of convergence of any estimator of α as a function of the degree of local smoothness of $L(\lambda)$ at zero. More precisely, we show in Section 2 that, the rate n^{-r} cannot be improved on the class of spectral densities defined in (2.1). The lower bound allows one to evaluate and compare different estimators by their asymptotic behavior, and to claim the rate optimality for any estimator attaining the bound. We show, in Section 3, that the lower bound is attained by a modified form of Geweke-Porter-Hudak's (1983) estimator suggested by Robinson (1992), i.e. that this estimator is rate optimal for the class of spectral densities defined in (2.1).

2. The lower bound. For any $\beta > 0$, we define the class $F(\beta, C_0, K_0)$ as in (1.1) and (1.2), but with the fixed bounds C_0 and K_0 for the constants in (1.2):

$$F(\beta, C_0, K_0) = \{f : f(\lambda) = C|\lambda|^{-\alpha}(1 + \Delta(\lambda)),\ 0 < C \le C_0,$$
$$|\alpha| < 1,\ |\Delta(\lambda)| \le K_0|\lambda|^{\beta},\ \lambda \in [-\pi, \pi]\}. \tag{2.1}$$

The proof of the following theorem giving the lower bound for the rate of convergence of arbitrary estimator of α is clearly related to papers of Samarov (1977) and Hall and Welsh (1984). While the theorem states the lower bound for the risk for a $0-1$ loss function $1\{|x| \ge c\}$, it clearly implies similar result for any loss function $l(\cdot)$ such that for some $d > 0$ $l(x) \ge d\, 1\{|x| \ge c\}$ for all x.

The notation $\alpha(f)$ in the theorem is used to emphasize that the parameter α corresponds to the same spectral density f which defines the probability measure P_f.

THEOREM 1. *There exists a positive constant c such that*

$$\liminf_n \inf_{\hat{\alpha}_n} \sup_{f \in F(\beta, C_0, K_0)} P_f\{n^r|\hat{\alpha}_n - \alpha(f)| \ge c\} > 0, \tag{2.2}$$

where inf *is taken over all estimators of α and $r = \beta/(2\beta + 1)$.*

Proof. Let $f_0(\lambda) = 1,\ \lambda \in [-\pi, \pi]$, be the spectral density of the white noise. We assume without the loss of generality that $C_0 > 1$, so that $f_0 \in F(\beta, C_0, K_0)$. Define a sequence of "perturbed" spectral densities as follows:

$$f_n(\lambda) = c_n \lambda^{-h_n}(1 + \Delta_n(\lambda)),\ \lambda \in (0, \pi], \tag{2.3}$$

where

$$\Delta_n(\lambda) = \begin{cases} 0, & 0 < \lambda \le \delta_n := n^{-1/(2\beta+1)} \\ c_n^{-1}\lambda^{h_n} - 1, & \delta_n \le \lambda \le \pi, \end{cases} \tag{2.4}$$

$h_n = \kappa\delta_n^{\beta}$ with some $\kappa > 0$, and

$$c_n = 1 + h_n \log(\delta_n). \tag{2.5}$$

On $[-\pi, 0)$ $f_n(\lambda)$ is defined by symmetry. Note that $f_n(\lambda)$ can be written as

$$f_n(\lambda) = \begin{cases} c_n \lambda^{-h_n}, & 0 < \lambda \leq \delta_n \\ 1, & \delta_n \leq \lambda \leq \pi. \end{cases}$$

LEMMA 1. (cf. Hall and Welsh (1984))
For all sufficiently large n
(i) $f_n \in F(\beta, C_0, K_0)$ *and*
(ii) $\int_{-\pi}^{\pi} (f_n(\lambda) - f_0(\lambda))^2 d\lambda \leq K n^{-1}$ *for some constant* $K > 0$.
The proof of the lemma is given in Section 4.

Denote by P_{f_n} and P_{f_0} the probability measures on R^n generated by n observations $\mathbf{X} = (X_1, \ldots, X_n)$ of the Gaussian stationary sequence with mean zero and spectral densities f_n and f_0 respectively, and let $\Lambda_n = \log(\frac{dP_{f_n}}{dP_{f_0}}(\mathbf{X}))$ denote the log likelihood ratio.

LEMMA 2. (cf. Lemma 1 in Samarov (1977))
There exist positive constants K_1 and K_2 such that for all sufficiently large n
(i) $m_n := E_{f_n} \Lambda_n \leq K_1 < \infty$;
(ii) $\sigma_n^2 := E_{f_n}(\Lambda_n - m_n)^2 \leq K_2 < \infty$.

This lemma, the proof of which is also given in Section 4, guarantees that the measures P_{f_0} and P_{f_n} are close in a certain sense. It is easy to check, for example, that for any event A and any $a > 0$

$$P_{f_n}\{A\} \leq e^a P_{f_0}\{A\} + \frac{M}{a^2}, \tag{2.6}$$

with $M = K_1^2 + K_2$.

Denote now the event

$$U_n(f) = \{n^r |\hat{\alpha}_n - \alpha(f)| \geq c\}$$

and observe that for any μ, $0 \leq \mu \leq 1$,

$$\sup_{f \in F(\beta, C_0, K_0)} P_f\{U_n(f)\} \geq \mu P_{f_0}\{U_n(f_0)\} + (1-\mu) P_{f_n}\{U_n(f_n)\}.$$

Applying here (2.6), we get

$$\sup_{f \in F(\beta, C_0, K_0)} P_f\{U_n(f)\} \geq \mu e^{-a}(P_{f_n}\{U_n(f_0)\} - \frac{M}{a^2}) + (1-\mu) P_{f_n}\{U_n(f_n)\}.$$

Now, since f_n is chosen such that $\alpha(f_0) - \alpha(f_n) = h_n = \kappa n^{-r}$, $U_n(f_0) \cup U_n(f_n) = \Omega$, the certain event, for any $c < \kappa/2$, and we have

$$\sup_{f \in F(\beta, C_0, K_0)} P_f\{U_n(f)\} \geq \mu e^{-a}(1 - P_{f_n}\{U_n(f_n)\} - \frac{M}{a^2}) + (1-\mu) P_{f_n}\{U_n(f_n)\}.$$

Choosing $\mu = 1/(1+e^{-a})$, we get

$$\sup_{f \in F(\beta, C_0, K_0)} P_f\{U_n(f)\} \geq \frac{e^{-a}}{1+e^{-a}}(1 - \frac{M}{a^2}),$$

and the conclusion of the theorem follows if we choose $a > M^{1/2}$. □

Remark 2.1. While the construction of the perturbation $f_n(\lambda)$ in (2.3) is similar to that of Hall and Welsh (1984), it is a bit simpler here since, unlike Hall and Welsh (1984) who work with probability densities, we allow small perturbations in $\int_{-\pi}^{\pi} f(\lambda) d\lambda$. This last difference is also apparently related to the fact that their rate is determined by the ratio β/α while in our case it depends only on β.

Since the rate in (2.2) is independent of α, it suffies for the proof to consider perturbations of a "base" spectral density in $F(\beta, C_0, K_0)$ with any $\alpha \in (-1, 1)$; we have chosen the "base" density with $\alpha = 0$ since this choice simplifies the proof. Note that the lower bound would remain valid for classes of densities with the range of values of α smaller than $(-1, 1)$, e.g. for $\alpha \in [0, 1)$.

Note also that the assumed Gaussianity of the process is not a restriction for the result of Theorem 1, since the lower bound established under the Gaussian assumption will automatically hold for broader classes of processes.

Remark 2.2. Even though we concentrate here on the estimation of the memory parameter α, in practice one also has to estimate the scale parameter C in (2.1). Robinson (1992, 1993) considered several estimates of C and analysed their asymptotic properties. A lower bound, similar to (2.2) but with an additional logarithmic term, for the rate of any such estimate can be easily obtained by only slightly modifying the argument given here.

Remark 2.3. The parameter β defining the class $F(\beta, C_0, K_0)$ determines the local degree of smoothness of the spectral densities in the class. Note that if β grows to infinity, the class $F(\beta, C_0, K_0)$ become closer and closer to the purely parametric family, and rate n^{-r} approaches the parametric rate $n^{-1/2}$.

Remark 2.4. The result of the theorem can be also formulated in terms of modulus of continuity of $\alpha(f)$ defined similarly to Donoho and Liu (1991), but with the L_2 norm which appears to be more natural here than the Hellinger norm.

3. Upper bound. In this section we show that the rate n^{-r}, given in the lower bound in Theorem 1, is attainable and thus optimal. We believe that it can be attained by a number of estimates of α including those discussed in Section 1. We consider here the modified version of Geweke and Porter-Hudak's (1983) estimator which Robinson (1992) showed to be asymptotically normal. It appears to afford the simplest proof in the present circumstances; a slightly more general proof will clearly go through for the other members of the class in Robinson (1992) (which have smaller asymptotic variances), while a rather different and more difficult proof would be needed for the estimates considered by Kűnsch (1987), Robinson (1993).

Define the discrete Fourier transform and periodogram

$$w(\lambda) = \frac{1}{(2\pi n)^{1/2}} \sum_{k=1}^{n} X_k e^{ik\lambda}, \qquad I(\lambda) = |w(\lambda)|^2.$$

Let l and m be integers such that $1 \le l < m \le n/2$, and put $\lambda_j = 2\pi j/n$, $\nu_j = \log j - (m-l)^{-1} \sum_k' \log k$, where $\sum_k' = \sum_{k=l+1}^{m}$. We estimate $\alpha = \alpha(f)$ by

$$\hat{\alpha}_{lmn} = -\frac{\sum_j' \nu_j \log I(\lambda_j)}{\sum_j' \nu_j^2}. \tag{3.1}$$

For any $D_0 > 1$ and r as defined in Theorem 1, the set

$$J_n(r, D_0) = \{l, m:\ 1 \le l < m \le n/2;\ \frac{\log^3(n)}{D_0} \le l \le \frac{D_0 n^{2r}}{\log^3(n)};\ \frac{n^{2r}}{D_0} \le m \le D_0 n^{2r}\}$$

is non-empty for n sufficiently large. In (3.1), m is a bandwidth number which achieves its "optimal rate" in $J_n(r, D_0)$, while l is a limiting number designed to avoid the anomalous behaviour of $I(\lambda_j)$ for finite j as $n \to \infty$, see Kűnsch (1986).

THEOREM 2.

$$\liminf_n \max_{l,m \in J_n(r,D_0)} \sup_{f \in F(\beta, C_0, K_0)} n^{2r} E_f(\hat{\alpha}_{lmn} - \alpha(f))^2 < \infty.$$

Proof. Put $v(\lambda) = w(\lambda)/(C\lambda^{-\alpha})^{1/2}$ and $v_j = (Re\{v(\lambda_j)\}, Im\{v(\lambda_j)\})^T$. Because $\sum_j' \nu_j = 0$,

$$\hat{\alpha}_{lmn} = \alpha(f) - \frac{\sum_j' \nu_j u_j}{\sum_j' \nu_j^2},$$

where $u_j = \log |v(\lambda_j)|^2 + \eta$, where $\eta = 0.5772...$ is Euler's constant, so that the u_j have approximate mean zero, see (3.10) below. From Robinson (1992), for $l, m \in J_n(r, D_0)$

$$\sum_j{}' \nu_j^2 = m(1 + O(\frac{l \log^2(n)}{m})) \sim m, \tag{3.2}$$

as $n \to \infty$. Thus it suffies to show that

$$E(\sum_j{}' \nu_j u_j)^2 = O(n^{2r})$$

uniformly over $l, m \in J_n(r, D_0)$ and $f \in F(\beta, C_0, K_0)$ as $n \to \infty$. (In the rest of the proof, we will use the word "uniformly" in the same sense as here without repeating "over $l, m \in J_n(r, D_0)$ and $f \in F(\beta, C_0, K_0)$".) This will follow if, uniformly as $n \to \infty$,

$$\sum\nolimits_j' \nu_j^2 E(u_j)^2 = O(n^{2r}) \tag{3.3}$$

and

$$\sum_{k,j:\ l<k<j\leq m} \nu_j \nu_k E(u_j u_k) = O(n^{2r}). \tag{3.4}$$

By Gaussianity, the moments in (3.3) and (3.4) can be analyzed in terms of the first two moments of the v_j. We have $Ev_j = 0$, $1 \leq j < n$ and the following properties essentially follow from Theorem 1 of Robinson (1992): uniformly over $l, m \in J_n(r, D_0)$ and $f \in F(\beta, C_0, K_0)$ as $n \to \infty$

$$|E_f|v(\lambda_j)|^2 - 1| + |E_f v^2(\lambda_j)| = O(\frac{\log j}{j} + (\frac{j}{n})^\beta) \tag{3.5}$$

uniformly in $j \in [l+1, m]$ and

$$|E_f v(\lambda_j) v(\lambda_k)| + |E_f v(\lambda_j) \bar{v}(\lambda_k)| = O(\frac{\log j}{k} + (\frac{j}{n})^\beta) \tag{3.6}$$

uniformly in $k \in [l+1, j-1]$, $j \in [l+2, m]$.

The proof of (3.3) and (3.4) uses techniques similar to those in the method-of-moments proof of asymptotic normality of $\hat{\alpha}_{lmn}$ in Robinson (1992). Denote by I_d the $d \times d$ identity matrix, by $\phi_d(\cdot)$ the d-variate standard normal density, and for a 2-dimensional vector $x = (x_1, x_2)^T$ put $g(x) = \log(||x||_E^2) + \eta$, where $||A||_E = (Tr(A^T A))^{1/2}$ denotes the Euclidean norm of a matrix A.

To prove (3.3), write

$$Eu_j^2 = |\Sigma_j|^{-1/2} \int g^2(x) \phi_2(\Sigma_j^{-1/2} x) dx, \tag{3.7}$$

where Σ_j is the covariance matrix of v_j and $|\Sigma_j|$ is its determinant. It follows from (3.5) that $\Sigma_j = I_2/2 + o(1)$ uniformly, as $n \to \infty$, so that for n sufficiently large (3.7) is bounded uniformly by

$$\frac{3}{2\pi} \int g^2(x) \exp(-\frac{1}{2}||x||_E^2) dx < \infty$$

from the properties of the normal distribution. Then (3.3) follows from (3.2) and $m \in J_n(r, D_0)$.

To prove (3.4), take $j > k$ and denote by Σ_{jk} the covariance matrix of $(v_j^T, v_k^T)^T$ and, suppressing reference to j and k, $\Phi = \Sigma_{jk}^{-1}$, partitioned into 2×2 submatrices as

$$\Phi = \begin{bmatrix} \Phi_1 & \Phi_{12} \\ \Phi_{21} & \Phi_2 \end{bmatrix}.$$

Put also

$$\tilde{\Phi} = \begin{bmatrix} \Phi_1 & 0 \\ 0 & \Phi_2 \end{bmatrix}, \quad \bar{\Phi} = \begin{bmatrix} 0 & \Phi_{12} \\ \Phi_{21} & 0 \end{bmatrix}.$$

With $z = (x^T, y^T)^T$,

$$E(u_j u_k) = |\Phi|^{1/2} \int g(x)g(y)\phi_4(\Phi^{1/2}z)dz$$

$$(3.8) \qquad = |\Phi|^{1/2} \int g(x)g(y)\phi_4(\tilde{\Phi}^{1/2}z)[\exp(-\frac{1}{2}z^T\bar{\Phi}z) - 1]dz$$

$$(3.9) \qquad +|\Phi|^{1/2}\Pi_{i=1}^2[\int g(x)\phi_2(\Phi_i^{1/2}x)dx].$$

Using (3.5) and (3.6) and arguing as in the proof of Theorems 2 and 3 of Robinson (1992), we have for some $\epsilon > 0$

$$|\exp(-\frac{1}{2}z^T\bar{\Phi}z) - 1 - \frac{1}{2}z^T\bar{\Phi}z| = O([\frac{\log(j)}{k} + (\frac{j}{n})^\beta]^2 \exp(\epsilon||z||_E^2)),$$

$$z^T\tilde{\Phi}z \geq 4\epsilon||z||_E^2$$

uniformly, as $n \to \infty$, while also

$$\int g(x)g(y)\phi_4(\tilde{\Phi}^{1/2}z)z^T\bar{\Phi}zdz = 0$$

for all n large enough, because (3.5) and (3.6) imply that $\tilde{\Phi} = 2I_4 + o(1)$ as $n \to \infty$. Likewise $|\Phi| = O(1)$ and so it is readily deduced that (3.8) is $O((\log(j)/k)^2 + (j/n)^{2\beta})$ uniformly, as $n \to \infty$. To estimate (3.9), note from (3.5) that, for $i = 1, 2$,

$$\phi_2(\Phi_i^{1/2}x) = \phi_2(\sqrt{2}x)\Big(1 + O((\frac{\log(j)}{k} + (\frac{j}{n})^\beta)||x||_E^2)\Big)$$

uniformly, as $n \to \infty$ (cf. (5.26) of Robinson (1992)).

Because

$$(3.10) \qquad \int g(x)\phi_2(\sqrt{2}x)dx = 0, \quad \int |g(x)|||x||_E^2\phi_2(\sqrt{2}x)dx, < \infty$$

it follows that (3.9) is uniformly $O((\log(j)/k)^2 + (j/n)^{2\beta})$ as $n \to \infty$. Then, for $l, m \in J_n(r, D_0)$

$$|\sum_{k,j:\ l<k<j\leq m} \nu_j\nu_k[(\frac{\log(j)}{k})^2 + (\frac{j}{n})^{2\beta}]| \leq \frac{2(\log m)^3}{l}\sum\nolimits_j' |\nu_j| + (\frac{m}{n})^{2\beta}m\sum\nolimits_j' \nu_j^2$$

$$= O(\frac{(\log m)^3}{l}[m \sum_j' \nu_j^2]^{1/2} + \frac{m^{2\beta+2}}{n^{2\beta}}) = O(\frac{m(\log n)^3}{l} + \frac{m^{2\beta+2}}{n^{2\beta}}) = O(n^{2r})$$

to verify (3.4). □

Remark 3.1. Note that (3.5) and (3.6) can be deduced via a somewhat simpler proof than those in Robinson (1992), where f that are not bounded outside a neighborhood of zero are permitted. In Theorem 1 of Robinson (1992), $\beta \in (0, 2]$ is assumed. Observe that if the class F includes $f(\lambda)$ of the form $(2\sin(\lambda/2))^{-\alpha}h(\lambda)$ for a bounded function $h(\lambda)$, as in case of fractionally integrated autoregressive moving average processes, then $\beta \leq 2$ no matter how smooth $h(\lambda)$ is, for example, even when $h(\lambda) \equiv 1$. On the other hand, it seems that if we replace $|\lambda|^{-\alpha}$ by $|2\sin(\lambda/2)|^{-\alpha}$ in the definition of $F(\beta, C_0, K_0)$ then $\beta = \infty$ is permitted in the latter situation, and to take advantage of this we would also need to replace $\log(\lambda_j)$ by $\log(2\sin\lambda_j/2)$ in the definition of $\hat{\alpha}_{lmn}$, as in Geweke-Porter-Hudak's (1983) original form. The point is that λ and $2\sin\lambda/2$ are interchangeable when $\beta \in (0, 2]$, but not when $\beta > 2$.

4. Proof of lemmas.

Proof of Lemma 1. To prove the claim (i), it is enough to show that in (2.3) and (2.4)

$$|\Delta_n(\lambda)| \leq K_0|\lambda|^\beta, \text{ for } \delta_n \leq \lambda \leq \pi. \tag{4.1}$$

Combining the estimate

$$c_n^{-1} = 1 - h_n \log(\delta_n) + O((h_n \log(\delta_n))^2),$$

which follows from (2.5), with the estimate

$$\lambda^{h_n} - 1 - h_n \log(\lambda) = O((h_n \log(\lambda))^2), \tag{4.2}$$

which holds uniformly for $\lambda \geq \delta_n$, we have

$$\Delta_n(\lambda) = (1 - h_n \log(\delta_n) + O((h_n \log(\delta_n))^2))(1 + h_n \log(\lambda) + O((h_n \log(\lambda))^2)) - 1 \tag{4.3}$$

$$= h_n \log(\frac{\lambda}{\delta_n}) + O((h_n \log(\delta_n))^2).$$

Now, as in Hall and Welsh (1984), it is sufficient to notice that the maximum of $(\delta_n/\lambda)^\beta|\log(\lambda/\delta_n)|$ for $\lambda \geq \delta_n$ is achieved at $\lambda/\delta_n = e^{1/\beta}$. Then, since $h_n = \kappa\delta_n^\beta \leq \kappa\lambda^\beta$, (4.3) implies that for some $K(\beta)$:

$$\Delta_n(\lambda) \leq h_n|\frac{\lambda}{\delta_n}|^\beta \frac{1}{\beta e} + K(\beta)(h_n \log(\lambda))^2 \leq \kappa K(\beta)\lambda^\beta \leq K_0\lambda^\beta,$$

if $\kappa \leq K_0/K(\beta)$, which proves the claim (i).

To prove the second claim, we write

$$\int_{-\pi}^{\pi}(f_n(\lambda) - f_0(\lambda))^2 d\lambda = 2\int_0^{\delta_n}(f_n(\lambda) - f_0(\lambda))^2 d\lambda = 2\int_0^{\delta_n}(c_n\lambda^{-h_n} - 1)^2 d\lambda$$

$$= 2(\int_0^{\delta_n^k}(c_n\lambda^{-h_n} - 1)^2 d\lambda + \int_{\delta_n^k}^{\delta_n}(c_n\lambda^{-h_n} - 1)^2 d\lambda) =: 2(I_1 + I_2), \text{ say},$$

for some $k \geq 2(2\beta + 1)$. It is easy to check that for sufficiently large C_1 and some $n_0(\kappa)$, $I_1 \leq C_1/n$ for all $n \geq n_0(\kappa)$. Arguing similarly to the proof of claim (i), we have, using (2.5) and (4.2),

$$I_2 = \int_{\delta_n^k}^{\delta_n}(c_n\lambda^{-h_n} - 1)^2 d\lambda = \int_{\delta_n^k}^{\delta_n}(h_n \log(\frac{\delta_n}{\lambda}) + O((h_n \log(\delta_n))^2))^2 d\lambda$$

$$\leq 2h_n^2 \int_0^{\delta_n} \log^2(\frac{\delta_n}{\lambda})d\lambda + O(\delta_n h_n^4 \log^4(\delta_n)) \leq K\delta_n^{2\beta+1} = K/n. \qquad \square$$

Proof of Lemma 2 closely follows the proof of Lemma 1 of Samarov (1977). Denote by A_n and B_n the $n \times n$ covariance matrices corresponding to the spectral densities f_0 and f_n respectively. Of course, $A_n = I_n$, the identity matrix. The log likelihood ratio Λ_n has the form

$$\Lambda_n = -\frac{1}{2}(\log|B_n| + (\mathbf{X} - \mu)^T(I_n - B_n^{-1})(\mathbf{X} - \mu)),$$

where $\mathbf{X}^T = (X_1, \ldots, X_n)$, μ is the n-vector of means $(\mu, ..., \mu)$, and $|B_n|$ denotes the determinant of the matrix B_n.

Denoting also $D_n = B_n - I_n$, we have

$$m_n = E_{f_n}\Lambda_n = \frac{1}{2}(Tr(D_n) - \log|B_n|).$$

Set $A_n(\theta) = I_n + \theta D_n$, $0 \leq \theta \leq 1$. Applying mean value theorem to $A_n(0) - A_n(1)$ and using the fact that

$$\frac{d}{d\theta}\log|A_n(\theta)| = Tr(A_n^{-1}(\theta)\frac{d}{d\theta}A_n(\theta)),$$

see, e.g. Davies (1973), we obtain, for some $0 \leq \theta^* \leq 1$,

$$m_n = \frac{1}{2}Tr(D_n - A_n^{-1}(\theta^*)D_n) = \frac{\theta^*}{2}Tr(D_n^2 A_n^{-1}(\theta^*)). \tag{4.4}$$

Denote by $||A||_{sp} = \sup_{||x||_E=1}||Ax||_E$ the spectral norm of a matrix A. The following four inequalities are well known, see, e.g. Davies (1973):

(4.5) $\qquad (i) \qquad Tr(CD) \leq ||C||_E||D||_E.$

$$(4.6) \qquad (ii) \qquad ||CD||_E \leq ||C||_E ||D||_{sp}.$$

Let $C = \{c_{j-k}\}_{1\leq j,k\leq n}$ be a $n \times n$ Toeplitz matrix generated by a function $\gamma(\lambda) = \sum_{j=-\infty}^{\infty} c_j e^{ij\lambda}$, $c_{-j} = c_j$, and $\sum_{j=-\infty}^{\infty} c_j^2 < \infty$. Then

$$(4.7) \qquad (iii) \qquad ||C||_E^2 \leq \frac{n}{2\pi}\int_{-\pi}^{\pi} \gamma^2(\lambda)d\lambda$$

and, if C is positive definite,

$$(4.8) \qquad (iv) \qquad ||C^{-1}||_{sp} \leq \sup_{\lambda\in[-\pi,\pi]} |1/\gamma(\lambda)|.$$

Applying (4.5) and (4.6) in (4.4), we get

$$(4.9) \qquad m_n \leq \frac{1}{2}||D_n||_E^2 ||A_n^{-1}(\theta^*)||_{sp}.$$

But by (4.7) and Lemma 1 (ii)

$$||D_n||_E^2 \leq \frac{n}{2\pi}\int_{-\pi}^{\pi}(f_n(\lambda) - f_0(\lambda))^2 d\lambda \leq \tilde{K},$$

as $n \to \infty$.

It is easy to see that the function $1+\theta^*(f_n(\lambda)-1)$ generating the matrix $A_n(\theta^*)$ is bounded away from 0 for $\lambda \in [-\pi,\pi]$ and $0 \leq \theta^* \leq 1$. Therefore, by (4.8) $||A_n^{-1}(\theta^*)||_{sp}$ is bounded, which together with (4.9) gives the first claim of Lemma 2.

To prove the second claim, we use the well-known expression for the variance of the Gaussian log likelihood (see, e.g., Davies (1973)), (4.7), and Lemma 1:

$$\sigma_n^2 = E_{f_n}(\Lambda_n - m_n)^2 = \frac{1}{2}||B_n(I_n - B_n^{-1})||_E^2 = \frac{1}{2}||B_n - I_n||_E^2 \leq K_2$$

as $n \to \infty$. □

Acknowledgement. A part of this paper was written while the first and third author (LG and AS) were visiting the Institut für Angewandte Mathematik, University of Heidelberg. They would like to thank the institute and in particular Professor Rainer Dahlhaus for providing the ideal working environment.

References.

Beran, J. (1994). *Statistics for long-memory processes.* Chapman & Hall, New York.

Davies, R. (1973). Asymptotic inference in stationary Gaussian time-series. *Adv. in Appl. Probab.* **5** 469-497.

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

Donoho, D. and Liu, R. (1991). Geometrizing rates of convergence, II. *Ann. Statist.* **19** 633-667.

Fox, R. and Taqqu, M. (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.

Hall, P. and Welsh, A. (1984). Best attainable rates of convergence for estimates of parameters of regular variation. *Ann. Statist.* **12** 1079-1084.

Kűnsch, H. (1986). Discrimination between monotonic trends and long-range dependence. *J. Appl. Probab.* **23** 1025-1030.

Kűnsch, H. (1987). Statistical properties of self-similar processes. *Proc. 1st World Congress of the Bernoulli Society* **4** 67-74, VNU Science Press.

Robinson, P. (1992). Log-periodogram regression of time series with long range dependence. Preprint.

Robinson, P. (1993). Gaussian semiparametric estimation of long range dependence. Preprint.

Samarov, A. (1977). Lower bound for the risk of spectral density estimates. *Problems of Info. Transm.* **13** 67-72.