The Annals of Statistics
1989, Vol. 17, No. 4, 1749–1766

EFFICIENT PARAMETER ESTIMATION FOR SELF-SIMILAR PROCESSES

BY RAINER DAHLHAUS

Universität Heidelberg

Asymptotic normality of the maximum likelihood estimator for the parameters of a long range dependent Gaussian process is proved. Furthermore, the limit of the Fisher information matrix is derived for such processes which implies efficiency of the estimator and of an approximate maximum likelihood estimator studied by Fox and Taqqu. The results are derived by using asymptotic properties of Toeplitz matrices and an equicontinuity property of quadratic forms.

1. Introduction. Let X_t, $t \in \mathbb{Z}$, be a stationary Gaussian sequence with mean μ and spectral density $f_\theta(x)$, $x \in \Pi := (-\pi, \pi)$, where μ and $\theta \in \Theta \subset \mathbb{R}^p$ are unknown parameters which have to be estimated. We are interested in strongly dependent sequences X_t, that is, in sequences with spectral density $f_\theta(x) \sim |x|^{-\alpha(\theta)} L_\theta(x)$ as $x \to 0$, where $0 < \alpha(\theta) < 1$ and $L_\theta(x)$ varies slowly at 0.

Processes of this type occur in many applied sciences, for example in economics and geophysics [cf. Granger and Joyeux (1980) and Mandelbrot and Wallis (1969)]. Examples of parameteric models of the above type are fractional Gaussian noise, obtained as the increments of self-similar processes [Mandelbrot and Van Ness (1968)] and fractional autoregressive moving average processes [Granger and Joyeux (1980)]. Fox and Taqqu (1986) have considered an approximate maximum likelihood procedure to estimate the parameter θ. They adapted the approach of Whittle (1953), introduced for weakly dependent random variables, and minimized the function

$$\mathscr{L}_N^W(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f_\theta(x) + \frac{I_N(x)}{f_\theta(x)} \right\} dx$$

with respect to $\theta \in \Theta$, where

$$I_N(x) = \frac{1}{2\pi N} \left| \sum_{j=1}^{N} e^{ijx} \left(X_j - \overline{X} \right) \right|^2, \qquad \overline{X}_N = \frac{1}{N} \sum_{j=1}^{N} X_j.$$

We denote the minimizing value by $\hat{\theta}_N$.

Fox and Taqqu (1986), Theorem 2 have proved that $\hat{\theta}_N$ is asymptotically normal with rate of convergence $N^{-1/2}$. In the case $\frac{1}{2} < \alpha(\theta) < 1$ this result is quite surprising, since the basic tool in the proof is a central limit theorem for the quadratic form $\int_{-\pi}^{\pi} \{I_N(x) - EI_N(x)\} \phi(x)\, dx$ with $\phi(x) = \partial/\partial\theta_j\, f_\theta(x)^{-1}$ [proved by Fox and Taqqu (1987), Theorem 4]. We note that the CLT for this

Received October 1986; revised January 1989.

AMS 1980 *subject classifications*. Primary 62F12; secondary 60F99, 62M10.

Key words and phrases. Long range dependence, fractional ARMA, maximum likelihood estimation, efficiency, Toeplitz forms.

quadratic form does not hold in general [it is, e.g., wrong if $\phi(x) = \chi_{[0,\lambda]}(x)$; cf. Ibragimov (1963), Section 6]. It is the special form of $\phi(x)$ that makes the above CLT and therefore also the CLT for $\hat{\theta}_N$ valid.

In this paper we prove that $\hat{\theta}_N$ is not only asymptotically normal but also asymptotically efficient in the sense of Fisher. Furthermore we derive the same for the exact maximum likelihood estimator. In Section 2 we note that certain restrictions of Fox and Taqqu concerning the parametrization can easily be dropped by using essentially the same proof. Section 3 contains consistency and asymptotic normality of the exact maximum likelihood estimator and in Section 4 we derive the efficiency. The proof technique is based on results on the asymptotic behavior of Toeplitz matrices. These results are proved in Section 5. Section 6 contains an equicontinuity property of parametric quadratic forms which is needed in Section 3.

The results are proved under the following assumptions.

(A0) X_t, $t \in \mathbb{Z}$, is a stationary Gaussian sequence with mean μ and spectral density $f_\theta(x)$, $\theta \in \Theta \subset \mathbb{R}^p$, where μ and θ are unknown parameters. Let μ_0 and θ_0 be the true parameters of the process where θ_0 is in the interior of Θ which is assumed to be compact. If $\theta \neq \theta'$ the set $\{x | f_\theta(x) = f_{\theta'}(x)\}$ is supposed to have positive Lebesgue measure.

(A1) $g(\theta) = \int_\Pi \log f_\theta(x)\,dx$ can be differentiated twice under the integral sign.

There exists $\alpha: \Theta \to (0,1)$ such that for each $\delta > 0$:

(A2) $f_\theta(x)$ is continuous at all (x,θ), $x \neq 0$, $f_\theta^{-1}(x)$ is continuous at all (x,θ) and

$$f_\theta(x) = O\big(|x|^{-\alpha(\theta)-\delta}\big).$$

(A3) $\partial/\partial\theta_j\, f_\theta^{-1}(x)$, $\partial^2/\partial\theta_j\,\partial\theta_k\, f_\theta^{-1}(x)$ and $\partial^3/\partial\theta_j\,\partial\theta_k\,\partial\theta_l\, f_\theta^{-1}(x)$ are continuous at all (x,θ):

$$\frac{\partial}{\partial\theta_j} f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-\delta}\big), \qquad 1 \le j \le p,$$

$$\frac{\partial^2}{\partial\theta_j\,\partial\theta_k} f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-\delta}\big), \qquad 1 \le j,k \le p,$$

$$\frac{\partial^3}{\partial\theta_j\,\partial\theta_k\,\partial\theta_l} f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-\delta}\big), \qquad 1 \le j,k,l \le p.$$

(A4) $\partial/\partial x\, f_\theta(x)$ is continuous at all (x,θ), $x \neq 0$, and

$$\frac{\partial}{\partial x} f_\theta(x) = O\big(|x|^{-\alpha(\theta)-1-\delta}\big).$$

(A5) $\partial^2/\partial x\, \partial\theta_j\, f_\theta^{-1}(x)$ is continuous at all (x, θ), $x \neq 0$, and

$$\frac{\partial^2}{\partial x\, \partial\theta_j} f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-1-\delta}\big), \qquad 1 \le j \le p.$$

(A6) $\partial^3/\partial^2 x\, \partial\theta_j\, f_\theta^{-1}(x)$ is continuous at all (x, θ), $x \neq 0$, and

$$\frac{\partial^3}{\partial^2 x\, \partial\theta_j} f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-2-\delta}\big), \qquad 1 \le j \le p.$$

(A7) $\partial f/(\partial x)\, f_\theta^{-1}(x)$ and $\partial^2/(\partial x)^2 f_\theta^{-1}(x)$ are continuous at all (x, θ), $x \neq 0$, and

$$\left(\frac{\partial}{\partial x}\right)^k f_\theta^{-1}(x) = O\big(|x|^{\alpha(\theta)-k-\delta}\big)$$

for $k = 0, 1, 2$.

(A8) The above constants can be chosen independently of θ (not of δ).

(A9) α is assumed to be continuous. Furthermore, there exists a constant C with

$$|f_\theta(x) - f_{\theta'}(x)| \le C|\theta - \theta'| f_{\theta'}(x)$$

uniformly for all x and all θ, θ' with $\alpha(\theta) \le \alpha(\theta')$ where $|\cdot|$ denotes the Euclidean norm.

(A0)–(A6) [apart from the third order differentiability in (A3)] are due to Fox and Taqqu (1986). The third order differentiability in (A3) is used to obtain the equicontinuity of $Z_N^{(2)}$ in Theorem 6.1 (and therefore to obtain asymptotic normality of the exact maximum likelihood estimate). (A7) is needed to derive the limit of the Fisher information matrix. (A8) and (A9) are used to establish the asymptotic properties of the exact maximum likelihood estimate. All conditions are fulfilled for fractional Gaussian noise and fractional ARMA-processes.

The assumption of a compact parameter space is somewhat unpleasant. However, we do not assume that the estimates are lying in the interior of Θ. They may also lie on its boundary.

For the derivation of the asymptotic properties of the exact maximum likelihood estimator a slightly different set of conditions would be more adequate. One needs mainly conditions on f_θ and its derivations instead of conditions on f_θ^{-1}. However, we do not want to complicate the presentation by a second set of conditions. For example, the property $|\partial/\partial\theta_j\, f_\theta(x)| = O(|x|^{-\alpha(\theta)-\delta})$, $x \neq 0$, can easily be derived from (A2) and (A3).

In our proofs K always denotes a generic constant which may vary from step to step.

2. Asymptotic normality of quasi maximum likelihood estimates. In this section we state a CLT for the estimate $\hat{\theta}_N$. Fox and Taqqu (1986), Theorem

2 considered the special parametrization $f_\theta(x) = \sigma^2 g_\theta(x)$ [there denoted by $\sigma^2 f(x, \theta)$] and assumed $\int_{-\pi}^{\pi} \log g_\theta(x)\, dx = 0$, which is equivalent to the fact that X_t/σ has a one-step prediction standard deviation independent of θ. This assumption on the parametrization is not crucial (it only prescribes a certain choice of the parameter space). However, Fox and Taqqu prove their CLT only for the estimate $\hat{\theta}_N$ of θ and not for the estimate $\hat{\sigma}_N^2$.

In the following theorem this restriction is dropped by considering an arbitrary spectral density $f_\theta(x)$ that fulfills the assumptions (A0)–(A6). This includes the special model $\sigma^2 g_\theta(x)$ since now σ^2 may be, e.g., the first component of θ.

Our proof widely uses the results of Fox and Taqqu. We mainly give a new proof for consistency. That the parameter restriction of Fox and Taqqu can be removed was also noted by Beran (1986), who gave a different proof. Let

$$\nabla g_\theta = \left(\frac{\partial}{\partial \theta_j} g_\theta \right)_{j=1,\dots,p} \quad \text{and} \quad \nabla^2 g_\theta = \left(\frac{\partial^2}{\partial \theta_j \, \partial \theta_k} g_\theta \right)_{j,k=1,\dots,p}.$$

THEOREM 2.1. *Suppose conditions* (A0)–(A6) *hold. Then* $\sqrt{N}(\hat{\theta}_N - \theta_0)$ *tends in distribution to a normal random vector with mean* 0 *and covariance matrix* $\Gamma(\theta_0)^{-1}$ *where*

$$\Gamma(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} (\nabla \log f_\theta(x))(\nabla \log f_\theta(x))' \, dx.$$

PROOF. Let

$$\mathscr{L}(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \left\{ \log f_\theta(x) + \frac{f_{\theta_0}(x)}{f_\theta(x)} \right\} dx.$$

Lemma 1 of Fox and Taqqu (1986) implies

$$\sup_\theta |\mathscr{L}_N^W(\theta) - \mathscr{L}(\theta)| \to 0 \tag{1}$$

with probability 1. Since $\mathscr{L}(\theta)$ is minimized by θ_0 we have

$$\mathscr{L}_N^W(\hat{\theta}_N) \le \mathscr{L}_N^W(\theta_0) \quad \text{and} \quad \mathscr{L}(\theta_0) \le \mathscr{L}(\hat{\theta}_N),$$

which implies $\mathscr{L}(\hat{\theta}_N) \to \mathscr{L}(\theta_0)$ a.s. and therefore also $\hat{\theta}_N \to \theta_0$ with probability 1. Furthermore, we obtain by the mean value theorem

$$\nabla \mathscr{L}_N^W(\hat{\theta}_N)_i - \nabla \mathscr{L}_N^W(\theta_0)_i = \left\{ \nabla^2 \mathscr{L}_N^W(\theta_N^{(i)})(\hat{\theta}_N - \theta_0) \right\}_i,$$

with $|\theta_N^{(i)} - \theta_0| \le |\hat{\theta}_N - \theta_0|$, $i = 1, \dots, p$. If $\hat{\theta}_N$ lies in the interior of Θ, we have $\nabla \mathscr{L}_N^W(\hat{\theta}_N) = 0$. If $\hat{\theta}_N$ lies on the boundary of Θ, then the assumption that θ_0 is in the interior implies $|\hat{\theta}_N - \theta_0| \ge \delta$ for some $\delta > 0$, i.e., we obtain $P(\sqrt{N}\, |\Delta \mathscr{L}_N^W(\hat{\theta}_N)| \ge \varepsilon) \le P(|\hat{\theta}_N - \theta_0| \ge \delta) \to 0$ for all $\varepsilon > \theta$. Thus, the result follows if we prove

(i) $$\nabla^2 \mathscr{L}_N^W(\theta_N^{(i)}) - \nabla^2 \mathscr{L}_N^W(\theta_0) \to_P 0,$$

(ii) $$\nabla^2 \mathscr{L}_N^W(\theta_0) \to_P \Gamma(\theta_0),$$

(iii) $$\sqrt{N}\, \nabla \mathscr{L}_N^W(\theta_0) \to_{\mathscr{D}} \mathscr{N}(0, \Gamma(\theta_0)).$$

(i) and (ii) follow from the smoothness conditions (A0)–(A6), Lemma 1 of Fox and Taqqu (1986) and the consistency of $\hat{\theta}_N$. The proof of (iii) is contained in the proof of Theorem 2 of Fox and Taqqu (1986). □

REMARK 2.2. The assumption of a compact parameter space is only needed for proving $\mathscr{L}(\hat{\theta}_N) \to \mathscr{L}(\theta_0)$ a.s. (if $\hat{\theta}_N \to \theta_0$ a.s. is known, one may derive asymptotic normality in the same way as above by considering $\Theta' = \{\theta \in \Theta: |\theta - \theta_0| \le \delta\}$ with some fixed $\delta > 0$ instead of Θ). In the case considered by Fox and Taqqu (1986), $\theta = (\sigma^2, \tau)$ with $f_\theta(x) = \sigma^2 g_\tau(x)$ where $\int_{-\pi}^{\pi} \log g_\tau(x)\,dx = 0$, $\sigma^2 > 0$ and $\tau \in E$, E compact, it is possible to derive $\mathscr{L}(\hat{\theta}) \to \mathscr{L}(\theta_0)$ a.s. by modifying the above arguments. Alternatively, one may also use Theorem 1 of Fox and Taqqu (1986), where $\hat{\theta}_N \to \theta_0$ a.s. has already been established.

It is normally not possible to solve the estimation equations $\nabla \mathscr{L}_N^W(\theta) = 0$ exactly (they are usually nonlinear). Instead one would determine the estimate, e.g., by a Newton iteration where the integral in $\mathscr{L}_N^W(\theta)$ is replaced by a sum over the Fourier frequencies, i.e., one would minimize

$$\mathscr{L}_N^\dagger(\theta) = \frac{1}{2N} \sum_{s=1}^{N-1} \left\{ \log f_\theta(x_s) + \frac{I_n(x_s)}{f_\theta(x_s)} \right\},$$

with $x_s = 2\pi s/N$. The resulting estimate $\hat{\theta}_N^\dagger$ has the same asymptotic behavior as $\hat{\theta}_N$, which may be proved analogously to above [note that $\sqrt{N}\{\nabla \mathscr{L}_N^W(\theta_0) - \nabla \mathscr{L}_N^\dagger(\theta_0)\} \to_P 0$].

In most cases it is computationally easier to minimize

$$\frac{1}{2}\log \sigma^2(\theta) + \frac{1}{2N} \sum_{s=1}^{N-1} \frac{I_N(x_s)}{f_\theta(x_s)},$$

where

$$\sigma^2(\theta) = \exp\left[\frac{1}{2\pi} \int_\Pi \log\{2\pi f_\theta(\lambda)\}\,d\lambda\right]$$

is the one-step prediction error variance (especially when σ^2 is itself one of the parameters θ_j). This leads also to an equivalent estimate.

3. Asymptotic normality of exact maximum likelihood estimates. In this section we consider exact maximum likelihood estimates $\tilde{\theta}_N$ obtained by minimizing

$$\mathscr{L}_N(\theta, \tilde{\mu}_N) = \frac{1}{2N} \log\det T_N(f_\theta) + \frac{1}{2N} (\mathbf{X}_N - \tilde{\mu}_N \mathbf{1})' T_N(f_\theta)^{-1} (\mathbf{X}_N - \tilde{\mu}_N \mathbf{1})$$

with respect to Θ, where

$$T_N(f_\theta) = \left\{ \int_\Pi f_\theta(x) \exp(ix(r-s))\,dx \right\}_{r, s = 1, \dots, N}$$

is the Toeplitz matrix of f_θ, $\mathbf{1} = (1, \dots, 1)'$ and $\tilde{\mu}_N$ is a consistent estimate of μ_0 (e.g., the arithmetic mean).

To derive consistency and asymptotic normality we use several asymptotic results for Toeplitz matrices derived in Sections 5 and 6. We define

$$A_\theta^{(1)} = T_N(f_\theta)^{-1}T_N(\nabla f_\theta)T_N(f_\theta)^{-1}T_N(\nabla f_\theta)T_N(f_\theta)^{-1},$$
$$A_\theta^{(2)} = T_N(f_\theta)^{-1}T_N(\nabla^2 f_\theta)T_N(f_\theta)^{-1}$$

and

$$A_\theta^{(3)} = T_N(f_\theta)^{-1}T_N(\nabla f_\theta)T_N(f_\theta)^{-1}.$$

Suppose A is an $n \times n$ matrix. Let

$$\|A\| = \sup_{x\in\mathbb{C}^n}\left(\frac{x^*A^*Ax}{x^*x}\right)^{1/2}$$

be the spectral norm of A (where A^* denotes the conjugate transpose of A) and

$$|A| = [\operatorname{tr}(AA^*)]^{1/2}$$

be the Euclidean norm of A. We use the following well known relations:

$$|\operatorname{tr}(AB)| \le |A|\cdot|B|, \qquad |AB| \le \|A\|\cdot|B|,$$
$$|AB| \le |A|\cdot\|B\|, \qquad \|AB\| \le \|A\|\cdot\|B\|,$$
$$\|A+B\| \le \|A\| + \|B\|, \qquad \|A\| = \|A^*\|, \qquad \|A\| \le |A| \le \sqrt{n}\|A\|,$$
$$x^*Ax \le x^*x\|A\| \qquad \text{for } A > 0$$

[cf. Davies (1973), Appendix II and Graybill (1983), Section 5.6]. Note that as an example $T_N(\nabla f_\theta)$ is a vector of matrices. Thus,

$$|\|T_N(\nabla f_\theta)\|| \quad \text{means} \quad \left(\sum_{i=1}^{p}\left\|T_N\left(\frac{\partial}{\partial\theta_i}f_\theta\right)\right\|^2\right)^{1/2}.$$

We start by proving consistency of $\tilde{\theta}_N$. It would be nice to establish this by proceeding as in Theorem 2.1. However, we were not able to derive (1) with $\mathscr{L}_N$ instead of $\mathscr{L}_N^W$. Therefore, we have adapted the ideas of Walker (1964), Section 2 to our situation.

THEOREM 3.1. *Suppose* (A0), (A2), (A3) *and* (A7)–(A9) *hold and* $\tilde{\mu}_N$ *is a consistent estimate of* μ_0. *Then*

$$\tilde{\theta}_N \to_P \theta_0.$$

PROOF. Since f_θ is uniformly bounded from below, we obtain $\|T_N(f_\theta)^{-1}\| \le K$ and as a consequence

$$\sup_\theta|\mathscr{L}_N(\theta,\tilde{\mu}_N) - \mathscr{L}_N(\theta,\mu_0)| \to_P 0.$$

We now prove that for all $\theta_1 \in \Theta$ there exists a constant $c(\theta_1) > 0$ with

$$\lim_{N\to\infty} E_{\theta_0}\{\mathscr{L}_N(\theta_1,\mu_0) - \mathscr{L}_N(\theta_0,\mu_0)\} \ge c(\theta_1).$$

We obtain with $T_\theta = T_N(f_\theta)$,

$$E_{\theta_0}\{\mathscr{L}_N(\theta_1,\mu_0) - \mathscr{L}_N(\theta_0,\mu_0)\} = \frac{1}{2N}\log\det\{T_{\theta_1}T_{\theta_0}^{-1}\} + \frac{1}{2N}\operatorname{tr}\{T_{\theta_0}T_{\theta_1}^{-1} - I\}.$$

Let $\lambda_{1N}, \dots, \lambda_{NN}$ be the eigenvalues of $T_{\theta_0} T_{\theta_1}^{-1}$. By a Taylor expansion of $\log\det(I + tA)$ around $t = 0$ we obtain that there exists a $\tau \in [0,1]$ such that the first summand is equal to

$$\frac{1}{4N} \sum_{j=1}^{N} \left(\frac{\lambda_{jN} - 1}{1 + \tau(\lambda_{jN} - 1)} \right)^2. \tag{2}$$

Consider the case $\alpha(\theta_1) \geq \alpha(\theta_0)$. (A9) implies $f_{\theta_0}(x) \leq K f_{\theta_1}(x)$ and therefore $T_{\theta_0} \leq K T_{\theta_1}$. Since $T_{\theta_0} T_{\theta_1}^{-1}$ has the same eigenvalues as $T_{\theta_0}^{1/2} T_{\theta_1}^{-1} T_{\theta_0}^{1/2}$, we therefore obtain that there exists a constant $C > 1$ with $0 < \lambda_{jN} \leq C$ for all j and N. This implies that (2) is larger than

$$C^{-2} \frac{1}{4N} \sum_{j=1}^{N} (\lambda_{jN} - 1)^2 = C^{-2} \frac{1}{4N} \operatorname{tr}\left\{ \left(T_{\theta_0} T_{\theta_1}^{-1} - I \right)^2 \right\},$$

which tends to

$$C^{-2} \frac{1}{8\pi} \int_{\Pi} \left(\frac{f_{\theta_0}(x)}{f_{\theta_1}(x)} - 1 \right)^2 dx \geq c(\theta_1)$$

by Theorem 5.1. If $\alpha(\theta_1) < \alpha(\theta_0)$, we obtain $C \leq \lambda_{jN}$ for all j and N with some $C \in (0,1)$, which implies that (2) is larger than

$$C^2 \frac{1}{4N} \sum_{j=1}^{N} \left(1 - \lambda_{jN}^{-1}\right)^2 \to C^2 \frac{1}{8\pi} \int_{\Pi} \left(\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)} - 1 \right)^2 dx \geq c(\theta_1).$$

Since

$$\lim_{N \to \infty} \operatorname{var}_{\theta_0} \{ \mathscr{L}_N(\theta_1, \mu_0) - \mathscr{L}_N(\theta_0, \mu_0) \} = 0,$$

we obtain

$$\lim_{N \to \infty} P_{\theta_0}\left(\mathscr{L}_N(\theta_1, \tilde{\mu}_N) - \mathscr{L}_N(\theta_0, \tilde{\mu}_N) < c(\theta_1)/2 \right) = 0. \tag{3}$$

Furthermore, with a mean value θ,

$$\begin{aligned} \mathscr{L}_N(\theta_2, \mu_0) - \mathscr{L}_N(\theta_1, \mu_0) &= (\theta_2 - \theta_1)' \cdot \frac{1}{2N} \operatorname{tr}\{ T_\theta^{-1} T_N(\nabla f_\theta) \} \\ &\quad + \frac{1}{2N} (\mathbf{X}_N - \mu_0 \mathbf{1})' \left(T_{\theta_2}^{-1} - T_{\theta_1}^{-1} \right) (\mathbf{X}_N - \mu_0 \mathbf{1}). \end{aligned} \tag{4}$$

Let $\partial / \partial \theta_j\, f_\theta = g^+ - g^-$ with $g^+, g^- \geq 0$. With Theorem 12.2.3(3) of Graybill (1983), we obtain

$$\left| \operatorname{tr}\left\{ T_\theta^{-1} T_N\left(\frac{\partial}{\partial \theta_j} f_\theta \right) \right\} \right| \leq \operatorname{tr}\{ T_\theta^{-1} T_N(g^+) \} + \operatorname{tr}\{ T_\theta^{-1} T_N(g^-) \}.$$

Since $T_\theta^{-1} \leq K I_N$ uniformly in θ, we obtain from (A2), (A3) and (A8) for all $\delta > 0$,

$$|\operatorname{tr}\{ T_\theta^{-1} T_N(\nabla f_\theta) \}| \leq K \operatorname{tr}\{ T_N(|\nabla f_\theta|) \} \leq K N \int_{-\pi}^{\pi} |x|^{-\alpha(\theta) - 3\delta}\, dx,$$

which, by using the continuity of α in (A9), is bounded by KN uniformly in N and all θ with $|\theta - \theta_0| \le \eta$ for some $\eta > 0$. Together with Lemma 5.5, the modulus of (4) therefore is bounded by

$$K|\theta_2 - \theta_1|\left\{1 + \frac{1}{N}\sum_{t=1}^{N}(X_t - \mu_0)^2\right\}$$

for all θ_2 with $|\theta_2 - \theta_1| \le \eta$.

Thus, there exists a $\delta > 0$ such that [with $U_\delta(\theta_1) := \{\theta_2 \in \Theta: |\theta_2 - \theta_1| < \delta\}$],

$$\lim_{N\to\infty} P_{\theta_0}\left(\sup_{\theta_2 \in U_\delta(\theta_1)} |\mathscr{L}_N(\theta_2, \tilde{\mu}_N) - \mathscr{L}_N(\theta_1, \tilde{\mu}_N)| \ge c(\theta_1)/4\right) = 0.$$

With (3) we get

$$\lim_{N\to\infty} P_{\theta_0}\left(\inf_{\theta_2 \in U_\delta(\theta_1)} \mathscr{L}_N(\theta_2, \tilde{\mu}_N) - \mathscr{L}_N(\theta_0, \tilde{\mu}_N) \ge c(\theta_1)/4\right) = 1.$$

Now the collection of sets $\{U_{\delta(\theta_1)}(\theta_1)|\theta_1 \ne \theta_0\}$ and the set $U_{\delta_0}(\theta_0)$ where δ_0 is arbitrary, together constitute an open covering of Θ. Since Θ is compact, this contains a finite open covering. This implies

$$\lim_{N\to\infty} P\left(\inf_{\theta\in\Theta} \mathscr{L}_N(\theta, \tilde{\mu}_N) = \inf_{\theta \in U_{\delta_0}(\theta_0)} \mathscr{L}_N(\theta, \tilde{\mu}_N)\right) = 1$$

and therefore

$$\lim_{N\to\infty} P\left(|\hat{\theta}_N - \theta_0| \ge \delta_0\right) = 0. \qquad \square$$

THEOREM 3.2. *Suppose* (A0), (A2), (A3) *and* (A7)–(A9) *hold and* $\tilde{\mu}_N$ *is an* $N^{\{1-\alpha(\theta_0)\}/2}$-*consistent estimate of* μ_0. *Then*

$$\sqrt{N}\left(\tilde{\theta}_N - \theta_0\right) \to_{\mathscr{D}} \mathscr{N}\left(O, \Gamma(\theta_0)^{-1}\right).$$

PROOF. Application of the mean value theorem yields

$$\nabla\mathscr{L}_N\left(\tilde{\theta}_N, \tilde{\mu}_N\right)_i - \nabla\mathscr{L}_N(\theta_0, \tilde{\mu}_N)_i = \sum_{j=1}^{p} \nabla^2\mathscr{L}_N\left(\bar{\theta}_N^{(i)}, \tilde{\mu}_N\right)_{ij}\left(\tilde{\theta}_N - \theta_0\right)_j.$$

Since $\tilde{\theta}_N \to_P \theta_0$ we obtain, as in the proof of Theorem 2.1,

$$\sqrt{N}\,\nabla\mathscr{L}_N\left(\tilde{\theta}_N, \tilde{\mu}_N\right) \to_P 0.$$

For the assertion it is obviously sufficient to prove

(i) $$\sqrt{N}\left(\nabla\mathscr{L}_N(\theta_0, \tilde{\mu}_N) - \nabla\mathscr{L}_N(\theta_0, \mu_0)\right) \to_P 0,$$

(ii) $$\sqrt{N}\,\nabla\mathscr{L}_N(\theta_0, \mu_0) \to_{\mathscr{D}} \mathscr{N}(0, \Gamma(\theta_0)),$$

(iii) $$\sup_\theta |\nabla^2\mathscr{L}_N(\theta, \tilde{\mu}_N) - \nabla^2\mathscr{L}_N(\theta, \mu_0)| \to_P 0,$$

(iv) $\bar{\theta}_N \to_P \theta_0$ implies $|\nabla^2\mathscr{L}_N\left(\bar{\theta}_N, \mu_0\right) - \nabla^2\mathscr{L}_N(\theta_0, \mu_0)| \to_P 0,$

(v) $$\nabla^2\mathscr{L}_N(\theta_0, \mu_0) \to_P \Gamma(\theta_0).$$

(i) We have with $T_\theta = T_N(f_\theta)$ and $T_{\nabla,\theta} = T_N(\nabla f_\theta)$,

$$\nabla \mathscr{L}_N(\theta,\mu) = \frac{1}{2N}\operatorname{tr}\{T_\theta^{-1}T_{\nabla,\theta}\} - \frac{1}{2N}(\mathbf{X}_N - \mu\mathbf{1})' A_\theta^{(3)}(\mathbf{X}_N - \mu\mathbf{1})$$

and therefore

$$\sqrt{N}\{\nabla \mathscr{L}_N(\theta_0,\tilde{\mu}_N) - \nabla \mathscr{L}_N(\theta_0,\mu_0)\}$$
$$= \frac{1}{\sqrt{N}}(\tilde{\mu}_N - \mu_0)\mathbf{1}' A_{\theta_0}^{(3)}(\mathbf{X}_N - \mu_0\mathbf{1}) - \frac{1}{2\sqrt{N}}(\tilde{\mu}_N - \mu_0)^2 \mathbf{1}' A_{\theta_0}^{(3)}\mathbf{1}.$$

Lemma 5.4(d) and Jensen's inequality imply

$$|\mathbf{1}' A_{\theta_0}^{(3)}\mathbf{1}| \le K N^{1-\alpha(\theta_0)+\delta}$$

and

$$E|\mathbf{1}' A_{\theta_0}^{(3)}(\mathbf{X}_N - \mu_0\mathbf{1})| \le |\mathbf{1}' A_{\theta_0}^{(3)} T_{\theta_0} A_{\theta_0}^{(3)}\mathbf{1}|^{1/2} = |\mathbf{1}' A_{\theta_0}^{(1)}\mathbf{1}|^{1/2}$$
$$\le K N^{(1-\alpha(\theta_0)+\delta)/2}.$$

We therefore obtain (i).

(ii) follows with the cumulant method. We have $E\sqrt{N}\,\nabla\mathscr{L}_N(\theta_0,\mu_0) = 0$ and by using the product theorem for cumulants [Brillinger (1981), Theorem 2.3.2],

$$N\operatorname{cov}\{\nabla \mathscr{L}_N(\theta_0,\mu_0), \nabla \mathscr{L}_N(\theta_0,\mu_0)\} = \frac{1}{2N}\operatorname{tr}\{T_{\theta_0}^{-1}T_{\nabla,\theta_0}T_{\theta_0}^{-1}T_{\nabla,\theta_0}\},$$

which tends to $\Gamma(\theta_0)$ by Theorem 5.1. Similarly, we get

$$N^{l/2}\operatorname{cum}\{\nabla \mathscr{L}_N(\theta_0,\mu_0)_{i_1},\dots,\nabla \mathscr{L}_N(\theta_0,\mu_0)_{i_l}\}$$
$$= \frac{1}{2}N^{-l/2}(-1)^l\frac{1}{l}\sum_{\substack{(j_1,\dots,j_l)\\ \text{permutation of}\\ (i_1,\dots,i_l)}} \operatorname{tr}\left[\prod_{k=1}^{l}\left\{T_{\theta_0}^{-1}T_N\left(\frac{\partial}{\partial\theta_{j_k}}f_{\theta_0}\right)\right\}\right],$$

which tends to zero by Theorem 5.1.

(iii) We have

$$\nabla^2 \mathscr{L}_N(\theta,\mu) = -\frac{1}{2N}\operatorname{tr}\{T_\theta^{-1}T_{\nabla,\theta}T_\theta^{-1}T_{\nabla,\theta}\} + \frac{1}{2N}\operatorname{tr}\{T_\theta^{-1}T_N(\nabla^2 f_\theta)\}$$
$$+\frac{1}{N}(\mathbf{X}_N - \mu\mathbf{1})' A_\theta^{(1)}(\mathbf{X}_N - \mu\mathbf{1}) - \frac{1}{2N}(\mathbf{X}_N - \mu\mathbf{1})' A_\theta^{(2)}(\mathbf{X}_N - \mu\mathbf{1})$$

and therefore,

$$\sup_\theta|\nabla^2 \mathscr{L}_N(\theta,\tilde{\mu}_N) - \nabla^2 \mathscr{L}_N(\theta,\mu_0)| \le \frac{2}{N}|\tilde{\mu}_N - \mu_0|\sup_\theta|\mathbf{1}' A_\theta^{(1)}(\mathbf{X}_N - \mu_0\mathbf{1})|$$
$$+ \text{three further terms.}$$

Application of the Cauchy–Schwarz inequality and Lemma 5.4 yield for each $\delta > 0$,

$$|\mathbf{1}'A_\theta^{(1)}(\mathbf{X}_N - \mu_0\mathbf{1})| \le \left(\sum_{j,k=1}^{p} |\mathbf{1}'A_\theta^{(1)}\mathbf{1}_{jj}| \, |(\mathbf{X}_N - \mu_0\mathbf{1})'A_\theta^{(1)}(\mathbf{X}_N - \mu_0\mathbf{1})_{kk}| \right)^{1/2}$$
$$\le KN^{(1-\alpha(\theta_0))/2+\delta}\left\{ \sum_{t=1}^{N} (X_t - \mu_0)^2 \right\}^{1/2}$$

and the first term therefore tends to zero. The other terms are treated similarly.

(iv) We have with $R_N(\theta) = (1/2N)\operatorname{tr}\{T_\theta^{-1}T_{\nabla,\theta}T_\theta^{-1}T_{\nabla,\theta}\}$,

$$\nabla^2\mathscr{L}_N(\theta,\mu_0) = Z_N^{(1)}(\theta) - \tfrac{1}{2}Z_N^{(2)}(\theta) + R_N(\theta),$$

with $Z_N^{(1)}(\theta)$ and $Z_N^{(2)}(\theta)$ as in Section 6. Due to the equicontinuity of $Z_N^{(1)}$ and $Z_N^{(2)}$ (Theorem 6.1) it is sufficient to prove that $R_N(\bar{\theta}_N) - R_N(\theta_0) \to_P 0$.

Let η be given with $0 < \eta < \frac{1}{12}$. Choose $\varepsilon > 0$ with $|\alpha(\theta) - \alpha(\theta_0)| < \eta$ for all θ with $|\theta - \theta_0| < \varepsilon$. We have

$$(5)\quad \begin{aligned} |R_N(\bar{\theta}_N) - R_N(\theta_0)| \le |\bar{\theta}_N - \theta_0| \Bigg[\left| \frac{1}{N}\operatorname{tr}\left\{ \left(T_{\theta_1}^{-1}T_{\nabla,\theta_1}\right)^3 \right\} \right| \\ + \left| \frac{1}{N}\operatorname{tr}\left\{ T_{\theta_1}^{-1}T_N\left(\nabla^2 f_{\theta_1}\right)T_{\theta_1}^{-1}T_{\nabla,\theta_1} \right\} \right| \Bigg], \end{aligned}$$

with a mean value θ_1 with $|\theta_1 - \theta_0| < \varepsilon$. We prove that the term in the square brackets is uniformly bounded in θ_1 by a deterministic constant which gives the result. Let $\partial/\partial\theta_i\, f_{\theta_1} = g^+ - g^-$ with $g^+, g^- \ge 0$. We have for all $\delta > 0$,

$$T_N(g^+) \le T_n\left(\left| \frac{\partial}{\partial\theta_i} f_{\theta_1} \right| \right) \le T_N\left(K|\lambda|^{-\alpha(\theta_0)-\eta-\delta}\right).$$

Let $A = T_{\theta_1}^{-1/2}T_N(g^+)T_{\theta_1}^{-1/2}$ and $B = T_{\theta_1}^{-1/2}T_N(K|\lambda|^{-\alpha(\theta_0)-\eta-\delta})T_{\theta_1}^{-1/2}$. Since $0 \le A \le B$ we obtain, with Theorems 9.1.19 and 12.2.3 of Graybill (1983),

$$\frac{1}{N}\operatorname{tr}\{A^3\} \le \frac{1}{N}\operatorname{tr}\{A^2B\} = \frac{1}{N}\operatorname{tr}\{A^{1/2}BA^{1/2}A\} \le \frac{1}{N}\operatorname{tr}\{A^{1/2}BA^{1/2}B\}$$
$$\le \frac{1}{N}\operatorname{tr}\{AB^2\} \le \frac{1}{N}\operatorname{tr}\{B^3\}.$$

Furthermore, we get, with Theorem 12.2.14(2) of Graybill (1983), from (A7), $k = 0$,

$$T_{\theta_1}^{-1} \le T_N\left(K|\lambda|^{-\alpha(\theta_0)+\eta+\delta}\right)^{-1}$$

and the same arguments now give

$$\frac{1}{N}\operatorname{tr}\left[\left\{T_{\theta_1}^{-1}T_N(g^+)\right\}^3\right] \le \frac{1}{N}\operatorname{tr}\left[\left\{T_N\left(K|\lambda|^{-\alpha(\theta_0)+\eta+\delta}\right)^{-1}T_N\left(K|\lambda|^{-\alpha(\theta_0)-\eta-\delta}\right)\right\}^3\right],$$

which tends to $K\int_{-\pi}^{\pi}|\lambda|^{-3\eta-3\delta}\,d\lambda$ by Theorem 5.1. All other terms occurring in (5) can be treated similarly and (iv) therefore is proved.

(v) follows with Theorem 5.1 since $E\,\nabla^2\mathscr{L}_N(\theta_0, \mu_0) \to \Gamma(\theta_0)$ and $\operatorname{var} \nabla^2\mathscr{L}_N(\theta_0, \mu_0) \to 0$. □

The condition that $\tilde{\mu}_N$ is $N^{\{1-\alpha(\theta_0)\}/2}$-consistent is for example fulfilled for the maximum likelihood estimate of μ [Adenstedt (1974)—note that the maximum likelihood estimate of μ is equal to the BLUE-estimate of μ], the arithmetic mean and M-estimates [Beran and Künsch (1985)].

Theorem 3.2 was proved by Yajima (1985) for ARMA(0, d, 0)-processes under the assumption that the mean is known.

4. Efficiency of the estimates. We now prove that the estimates $\hat{\theta}_N$ and $\tilde{\theta}_N$ are asymptotically efficient in the sense of Fisher [cf. Lehmann (1983), Chapter 6.1]. N^{-1} times the Fisher information matrix $\Gamma_N(\theta_0)$ is equal to

$$\begin{aligned} &NE_{\theta_0}[\{\nabla\mathscr{L}_N(\theta_0, \mu_0)\}\{\nabla\mathscr{L}_N(\theta_0, \mu_0)\}'] \\ &\quad = \frac{1}{4N}\operatorname{cov}\left[(\mathbf{X}_N - \mu_0\mathbf{1})'T_N(f_\theta)^{-1}T_N(\nabla f_\theta)T_N(f_\theta)^{-1}(\mathbf{X}_N - \mu_0\mathbf{1})\right] \\ &\quad = \frac{1}{2N}\operatorname{tr}\left\{T_N(f_{\theta_0})^{-1}T_N(\nabla f_{\theta_0})T_N(f_{\theta_0})^{-1}T_N(\nabla f_{\theta_0})\right\}. \end{aligned}$$

Theorem 5.1 implies that $\Gamma_N(\theta_0)$ tends to $\Gamma(\theta_0)$. We have obtained the following result.

THEOREM 4.1. (a) *Suppose* (A0)–(A7) *hold. Then* $\hat{\theta}_N$ *is an efficient estimate of* θ_0.

(b) *Suppose* (A0), (A2), (A3) *and* (A7)–(A9) *hold. Then* $\tilde{\theta}_N$ *is an efficient estimate of* θ_0.

5. Properties of Toeplitz matrices. In this section we derive some asymptotic results on Toeplitz matrices. The following theorem plays an important role in the proofs of Sections 3 and 4. For classical ARMA-processes it was derived by Taniguchi (1983), Theorem 1 and in a special situation by Davies (1973), Theorem 4.4. In our proof we make use of a result of Fox and Taqqu (1987).

THEOREM 5.1. *Let* $p \in \mathbb{N}$ *and* $\alpha, \beta \in \mathbb{R}$ *with* $0 < \alpha$, $\beta < 1$ *and* $p(\beta - \alpha) < \frac{1}{2}$. *Suppose* $f_j(x)$ *and* $g_j(x)$, $j = 1, \dots, p$, *are symmetric real valued functions, where all* f_j *are nonnegative and satisfy* (A2) *and* (A7) (*for* $k = 0, 1$) (*with the parameter* θ *dropped*), *and all* g_j *are continuous at all* $x \neq 0$ *and satisfy*

$$g_j(x) = O(|x|^{-\beta-\delta})$$

for all $\delta > 0$. *Then*

$$\lim_{N\to\infty} \frac{1}{N}\operatorname{tr}\left[\prod_{j=1}^{p}\left\{T_N(f_j)^{-1}T_N(g_j)\right\}\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\prod_{j=1}^{p}\frac{g_j(x)}{f_j(x)}\right\}dx. \tag{6}$$

Theorem 5.1 is proved at the end of this section. In the proof we approximate $T_N(f_\theta)^{-1}$ by $T_N(\{4\pi^2 f_\theta\}^{-1})$. This approximation is considered in Lemma 5.2. It is the same approximation as for weakly dependent processes [cf. Dzhaparidze and Yaglom (1983), Section 3]. However, due to the long range dependence, the used proof methods are different.

LEMMA 5.2. *Let* $0 < \alpha < 1$. *Suppose* (A2) *and* (A7) (*for* $k = 0, 1$) *hold* (*with the parameter* θ *dropped*). *Then*

$$(7)\qquad \left| I - T_N(f)^{1/2} T_N\left(\{4\pi^2 f\}^{-1}\right) T_N(f)^{1/2} \right| = O(N^{\delta}),$$

for each δ *with* $0 < \delta < \alpha/4$.

PROOF. Let

$$(8)\qquad \Delta_N(x) = \sum_{n=1}^{N} \exp(-ixn).$$

Since $\int_{-\pi}^{\pi} \Delta_N(x-y)\Delta_N(y-z)\,dy = 2\pi\Delta_N(x-z)$ the square of the left-hand side of (7) is equal to

$$(9)\qquad \begin{aligned} N - 2\,\mathrm{tr}\left[T_N(f)T_N\left(\{4\pi^2 f\}^{-1}\right)\right] &+ \mathrm{tr}\left[\left(T_N(f)T_N\left(\{4\pi^2 f\}^{-1}\right)\right)^2\right] \\ = (2\pi)^{-4}\int_{(-\pi,\pi]^4} &\left(\frac{f(x_1)}{f(y_1)} - 1\right)\left(\frac{f(x_2)}{f(y_2)} - 1\right) \\ &\times \Delta_N(x_1 - y_1)\Delta_N(y_1 - x_2)\Delta_N(x_2 - y_2)\Delta_N(y_2 - x_1)\,dx\,dy. \end{aligned}$$

Let $L_N(x)$ be the periodic extension (with period 2π) of

$$L_N(x)' = \begin{cases} N, & |x| \le 1/N, \\ 1/|x|, & \pi \ge |x| > 1/N. \end{cases}$$

Summation by parts gives

$$(10)\qquad |\Delta_N(x)| \le K L_N(x).$$

(A7) implies for each δ with $0 < \delta < \alpha$ and a constant $K \in \mathbb{R}$,

$$(11)\qquad \left| \frac{f(x)}{f(y)} - 1 \right| \le K \frac{|y - x|^{1-3\delta}}{|x|^{1-\delta}}.$$

This may be derived by considering the cases $0 < y \le x/2$, $0 < x/2 < y \le x$, $0 < x < y \le (3x)/2$ and $0 < (3x)/2 < y$ separately. For example in the first case we have $x \le 2(x - y)$ which leads with (A7) to

$$\begin{aligned} \left| \frac{f(x)}{f(y)} - 1 \right| &\le f(x)\left| \frac{1}{f(y)} - \frac{1}{f(x)} \right| \le f(x) \sum_{k=1}^{[x/y]} |f^{-1}((k+1)y) - f^{-1}(ky)| \\ &\le K|x|^{-\alpha-\delta}|y| \sum_k |ky|^{\alpha-1-\delta} \le K|x|^{-2\delta} \le K \frac{|x - y|^{1-3\delta}}{|x|^{1-\delta}}. \end{aligned}$$

(10) and (11) give as an upper bound for (9):

$$K\int_{\Pi^4}|x_1|^{-1+\delta}|x_2|^{-1+\delta}L_N(x_1-y_1)^{3\delta}L_N(y_1-x_2)L_N(x_2-y_2)^{3\delta}L_N(y_2-x_1)\,dx\,dy$$
$$\leq KN^{6\delta}\ln^2 N,$$

which proves the lemma. □

LEMMA 5.3. *Let $f(x)$ and $g(x)$ be positive symmetric functions such that there exist α and β, $0<\alpha,\beta<1$, with*

$$f^{-1}(x)=O(|x|^{\alpha})$$

and

$$g(x)=O(|x|^{-\beta}).$$

Then

$$\|T_N(f)^{-1/2}T_N(g)^{1/2}\|=\|T_N(g)^{1/2}T_N(f)^{-1/2}\|=O(N^{\max\{(\beta-\alpha)/2,0\}}).$$

PROOF. Since $T_N(f)$ has rank N we obtain

$$\begin{aligned}\|T_N(g)^{1/2}T_N(f)^{-1/2}\|^2&=\sup_{|x|=1}\frac{x^*T_N(g)x}{x^*T_N(f)x}\\&\leq K\sup_{|x|=1}\frac{\int_{-\pi}^{\pi}|\gamma|^{-\beta}|\sum_{n=1}^{N}x_n\exp(-i\gamma n)|^2\,d\gamma}{\int_{-\pi}^{\pi}|\gamma|^{-\alpha}|\sum_{n=1}^{N}x_n\exp(-i\gamma n)|^2\,d\gamma}\\&\leq K\sup_{h\in\mathscr{P}_N}\frac{\int_{-\pi}^{\pi}|\gamma|^{-\beta}h(\gamma)\,d\gamma}{\int_{-\pi}^{\pi}|\gamma|^{-\alpha}h(\gamma)\,d\gamma},\end{aligned}$$

where $\mathscr{P}_N=\{h(\gamma)\colon h(\gamma)$ probability density on $[-\pi,\pi]$ with $h(\gamma)\leq N\}$. If $\beta\leq\alpha$, the above expression is bounded; if $\beta>\alpha$, the sup is attained by $h=N\chi_{\{|\gamma|<1/(2N)\}}$ so that the above bound is equal to $O(N^{\beta-\alpha})$. □

The last part of the above proof is due to H. Dehling who improved a former version of this lemma [cf. Taqqu (1986)].

As a consequence of Lemma 5.3 we obtain the following results (recall the definition of $A_\theta^{(i)}$ at the beginning of Section 3).

LEMMA 5.4. *Suppose* (A2), (A3), (A7) *and* (A8) *hold. We then obtain for each $\delta>0$ and $i=1,2,3$ with a constant K independent of θ and N*:

(a) $$|\,\|A_\theta^{(i)}\|\,|\leq KN^{\delta},$$

(b) $$|\,\|\nabla A_\theta^{(i)}\|\,|\leq KN^{\delta},$$

(c) $$|x'A_\theta^{(i)}x|\leq Kx'xN^{\delta}\quad\textit{for all } x\in\mathbb{R},$$

(d) $$|\mathbf{1}'A_\theta^{(i)}\mathbf{1}|\leq KN^{1-\alpha(\theta)+\delta}.$$

PROOF. We prove the results for $i=1$. The results for $i=2$ and 3 are obtained similarly.

(a) We obtain with $T_\theta = T_N(f_\theta)$,

$$\left| x' T_\theta^{-1/2} T_N\left(\frac{\partial}{\partial\theta_j} f_\theta\right) T_\theta^{-1/2} x\right| \le x' T_\theta^{-1/2} T_N\left(\left|\frac{\partial}{\partial\theta_j} f_\theta\right|\right) T_\theta^{-1/2} x$$

and therefore with Lemma 5.3 for each $\delta > 0$,

$$\left\| T_\theta^{-1/2} T_N\left(\frac{\partial}{\partial\theta_j} f_\theta\right) T_\theta^{-1/2}\right\| \le \left\| T_\theta^{-1/2} T_N\left(\left|\frac{\partial}{\partial\theta_j} f_\theta\right|\right) T_\theta^{-1/2}\right\| \le K N^{\delta}.$$

Since f_θ is uniformly bounded from below we get $\|T_N(f_\theta)^{-1}\| \le K$ and therefore the result.

(b) follows analogously to (a).

(c) is an immediate consequence of (a).

To prove (d) we note that

$$\begin{aligned}
&\left| \mathbf{1}' T_\theta^{-1} T_N\left(\frac{\partial}{\partial\theta_i} f_\theta\right) T_\theta^{-1} T_N\left(\frac{\partial}{\partial\theta_j} f_\theta\right) T_\theta^{-1}\mathbf{1}\right| \\
&\qquad \le \mathbf{1}' T_\theta^{-1}\mathbf{1} \left\| T_\theta^{-1/2} T_N\left(\frac{\partial}{\partial\theta_i} f_\theta\right) T_\theta^{-1/2}\right\| \left\| T_\theta^{-1/2} T_N\left(\frac{\partial}{\partial\theta_j} f_\theta\right) T_\theta^{-1/2}\right\|.
\end{aligned}$$

Since $\mathbf{1}' T_\theta^{-1}\mathbf{1} \le K N^{1-\alpha(\theta)+\delta}$ for each $\delta > 0$ [cf. Adenstedt (1974), Theorem 5.2] we get the result. □

Proof of Theorem 5.1. Let $k \le 2p$ and $j_1, \dots, j_k \in \{1, \dots, p\}$. By considering all combinations $\sum_{l=1}^k s_l f_{j_l}^{-1}$ and $\sum_{l=1}^k r_l g_{j_l}$ with $s_l, r_l \in \{0, 1\}$ we obtain from Theorem 1.a of Fox and Taqqu (1987),

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{tr}\left\{ \prod_{l=1}^{k} T_N\left(\left\{4\pi^2 f_{j_l}^{-1}\right\}\right) T_N(g_{j_l})\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left\{\prod_{l=1}^{k}\frac{g_{j_l}(x)}{f_{j_l}(x)}\right\} dx. \tag{12}$$

Since $T_N(g_j) = T_N(g_j^+) - T_N(g_j^-)$ with $g_j^+, g_j^- \ge 0$ we now assume without loss of generality that all g_j are nonnegative. We define for $j = 1, \dots, p$,

$$A_j := T_N(g_j)^{1/2} T_N(f_j)^{-1} T_N(g_{j+1})^{1/2}$$

and

$$B_j := T_N(g_j)^{1/2} T_N\left(\left\{4\pi^2 f_j\right\}^{-1}\right) T_N(g_{j+1})^{1/2},$$

where we set $g_{p+1} := g_1$. With this notation the left-hand side of (6) is equal to $(1/N)\operatorname{tr}\{\prod_{j=1}^p A_j\}$. We obtain

$$\begin{aligned}
\left| \frac{1}{N}\operatorname{tr}\left\{\prod_{j=1}^p A_j - \prod_{j=1}^p B_j\right\}\right| &= \left| \sum_{k=1}^p \frac{1}{N}\operatorname{tr}\left\{\left(\prod_{j=1}^{k-1} B_j\right)(A_k - B_k)\left(\prod_{j=k+1}^{p} A_j\right)\right\}\right| \\
&\le \sum_{k=1}^p \frac{1}{N}\left|\prod_{j=1}^{k-1} B_j\right| |A_k - B_k| \prod_{j=k+1}^{p} \|A_j\|.
\end{aligned} \tag{13}$$

(12) implies $\left|\prod_{j=1}^{k-1}B_j\right| = O(N^{1/2})$ and by using Lemma 5.3 we obtain

$$\|A_j\| \le O\left(N^{\max\{\beta-\alpha+2\delta,0\}}\right) \quad \text{for all } j \text{ and all } \delta > 0.$$

Furthermore, we get with Lemmas 5.2 and 5.3,

$$\begin{aligned}|A_k - B_k| &\le \|T_N(g_j)^{1/2}T_N(f_j)^{-1/2}\|\left|I - T_N(f_j)^{1/2}T_N\left(\{4\pi^2 f_j\}^{-1}\right)T_n(f_j)^{1/2}\right|\\ &\quad\times\|T_N(f_j)^{-1/2}T_N(g_{j+1})^{1/2}\|\\ &\le O\left(N^{\beta-\alpha+3\delta}\right)\end{aligned}$$

for all k and all $\delta > 0$. Thus, (13) tends to zero and with (12) we obtain the result. □

The following result is needed in the proof of Theorem 3.1.

LEMMA 5.5. *Suppose* (A2), (A3) *and* (A7)–(A9) *hold. Then there exists a constant K independent of* θ_1, θ_2, *N and y with*

$$\left|y'\left\{T_N(f_{\theta_2})^{-1} - T_N(f_{\theta_1})^{-1}\right\}y\right| \le K|\theta_2 - \theta_1|y'y.$$

PROOF. Without loss of generality we assume $\alpha(\theta_1) \le \alpha(\theta_2)$. We have with $T_\theta = T_N(f_\theta)$,

$$\begin{aligned}\left|y'\left(T_{\theta_2}^{-1} - T_{\theta_1}^{-1}\right)y\right| &\le K\sup_x\left|\frac{x'T_{\theta_2}^{-1}x}{x'T_{\theta_1}^{-1}x} - 1\right|y'y\\ &= K\sup_x\left|\frac{x'T_{\theta_1}^{1/2}T_{\theta_2}^{-1/2}T_{\theta_1}^{1/2}x}{x'x} - 1\right|y'y\\ &= K\sup_x\left|\frac{x'\operatorname{diag}(\lambda_1,\dots,\lambda_N)x}{x'x} - 1\right|y'y,\end{aligned}$$

where $\lambda_1,\dots,\lambda_N$ are the eigenvalues of $T_{\theta_1}^{1/2}T_{\theta_2}^{-1}T_{\theta_1}^{1/2}$. Since $T_{\theta_2}^{-1/2}T_{\theta_1}T_{\theta_2}^{-1/2}$ has the same eigenvalues this is equal to

$$K\sup_x\left|\frac{x'T_{\theta_1}x}{x'T_{\theta_2}x} - 1\right|y'y.$$

(A9) now gives the result. □

6. Equicontinuity of quadratic forms. To prove relation (iv) in the proof of Theorem 3.2 we need an equicontinuity property of certain quadratic forms depending on the parameter θ. This property is established by using a version of the chaining lemma. Let $A_\theta^{(i)}$ be defined as in Section 3 and

$$Z_N^{(i)}(\theta) = \frac{1}{N}(\mathbf{X}_N - \mu_0\mathbf{1})'A_\theta^{(i)}(\mathbf{X}_N - \mu_0\mathbf{1}) - \frac{1}{N}\operatorname{tr}\left\{A_\theta^{(i)}T_N(f_\theta)\right\}.$$

THEOREM 6.1. *Suppose* (A2), (A3) *and* (A8) *hold. Then* $Z_N^{(i)}(\theta)$, $i = 1, 2$, *are equicontinuous in probability, i.e., for each* $\eta > 0$ *and* $\varepsilon > 0$ *there exists a* $\delta > 0$ *such that*

$$\limsup_{N\to\infty} P\Big(\sup_{|\theta_1-\theta_2|\le\delta} |Z_N^{(i)}(\theta_1) - Z_N^{(i)}(\theta_2)| > \eta\Big) < \varepsilon.$$

PROOF. Let

$$N(\delta) = \inf\{m | \text{there exist } \theta_1, \dots, \theta_m \in \Theta \text{ with } \inf_i|\theta - \theta_i| \le \delta \text{ for all } \theta \in \Theta\},$$

$$H(\delta) = \log\{4N(\delta)^2/\delta\}$$

and

$$J(\delta) = \int_0^\delta H(u)\,du.$$

Since Θ is compact, $J(\delta) < \infty$. The exponential inequality, proved in Lemma 6.2 and a straightforward modification of Lemma VII.9 of Pollard (1984) imply that

$$P\big(|Z_N^{(i)}(\theta_1) - Z_N^{(i)}(\theta_2)| > 26DJ(|\theta_1 - \theta_2|) \text{ for some } \theta_1, \theta_2 \in \Theta \text{ with } |\theta_1 - \theta_2| \le \varepsilon\big) \le 2\varepsilon, \qquad i = 1, 2.$$

Note for the modification of the chaining lemma that the special form of H was used in Pollard's proof only on page 144 (the second line from the bottom). We need this different H because of our different form of the exponential inequality. If $\eta > 0$ is given we therefore have for δ with $\eta \ge 26DJ(\delta)$,

$$P\Big(\sup_{|\theta_1-\theta_2|\le\delta} |Z_N^{(i)}(\theta_1) - Z_N^{(i)}(\theta_2)| > \eta\Big) \le 2\delta. \qquad \square$$

LEMMA 6.2. *Suppose* (A2), (A3) *and* (A8) *hold. Then there exists a constant* D *such that for* $i = 1, 2$, *all* $\theta_1, \theta_2 \in \Theta$ *and all* $\eta > 0$,

$$P\big(|Z_N^{(i)}(\theta_1) - Z_N^{(i)}(\theta_2)| > \eta|\theta_1 - \theta_2|\big) \le 4p^2\exp(-\eta/D).$$

PROOF. We have to prove an exponential inequality for each component of $Z_N^{(i)}(\theta)$. Let $\theta_1 \ne \theta_2$ and $S = \{Z_N^{(i)}(\theta_1) - Z_N^{(i)}(\theta_2)\}_{jk}/|\theta_1 - \theta_2|$. By using the Markov inequality and Fubini's theorem we obtain for all $t > 0$

$$\begin{aligned} P(|S| > \eta) &\le P(S > \eta) + P(-S > \eta) \\ &\le 2\exp(-t\eta)\sum_{k=0}^{\infty} \frac{t^k}{k!}|ES^k|. \end{aligned}$$

Below we prove that exists a constant C, independent of θ_1, θ_2 and l, with

$$(14) \qquad |\mathrm{cum}_l(S)| \le l!C^l.$$

Application of the product theorem for cumulants [cf. Brillinger (1981), Theorem 2.3.2] yields

$$|ES^k| \le k!(2C)^k$$

with the same constant. Choosing $t = \{4C\}^{-1}$ and $D = 4C$ gives the result.

We now prove (14). We obtain for $l \geq 2$,

$$|\mathrm{cum}_l(S)| = |\theta_1 - \theta_2|^{-l} N^{-l}(l-1)!2^{l-1} \left| \mathrm{tr}\left[\left\{ T_N(f_{\theta_0}) \left(A_{\theta_1}^{(i)} - A_{\theta_2}^{(i)} \right)_{jk} \right\}^l \right] \right|$$

$$\leq |\theta_1 - \theta_2|^{-l} N^{-l}(l-1)!2^{l-1} \left| T_N(f_{\theta_0}) \right|^l \left\| \left(A_{\theta_1}^{(i)} - A_{\theta_2}^{(i)} \right)_{jk} \right\|^l.$$

Application of (x) in Appendix II of Davies (1973) gives with Lemma 5.4(b),

$$\left\| \left(A_{\theta_1}^{(i)} - A_{\theta_2}^{(i)} \right)_{j,k} \right\| \leq |\theta_1 - \theta_2| C(\delta) N^{\delta}.$$

Furthermore we have [with $\Delta_N(x)$ as in (8)]

$$\left| T_N(f_{\theta_0}) \right|^2 = \iint_{-\pi}^{\pi} f_{\theta_0}(x) f_{\theta_0}(y) |\Delta_N(x-y)|^2 \, dx \, dy.$$

Application of Hölder's inequality shows that this can be bounded by $C(\delta)N^{1+\alpha(\theta_0)+2\delta}$ for each $\delta > 0$. If (A8) holds, δ can be chosen independently of θ_0, θ_1 and θ_2, i.e., we obtain (14). □

Acknowledgment. The author is grateful to Professor H. Künsch, ETH Zürich, for bringing the problem to his attention.

REFERENCES

ADENSTEDT, R. K. (1974). On large-sample estimation for the mean of a stationary random sequence. *Ann. Statist.* **2** 1095–1107.

BERAN, J. (1986). Estimation, testing and prediction for self-similar and related processes. Dissertation, ETH Zürich.

BERAN, J. and KÜNSCH, H. (1985). Location estimation for processes with long range dependence. Research Report No. 40, Seminar für Statistik, ETH Zürich.

BRILLINGER, D. R. (1981). *Time Series: Data Analysis and Theory.* Holden-Day, San Francisco.

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

DZHAPARIDZE, K. O. and YAGLOM, A. M. (1983). Spectrum parameter estimation in time series analysis. In *Developments in Statistics* (P. R. Krishnaiah, ed.) **4** 1–181. Academic, New York.

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.

FOX, R. and TAQQU, M. S. (1987). Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields* **74** 213–240.

GRANGER, C. W. J. and JOYEUX, R. (1980). An introduction to long-memory time series and fractional differencing. *J. Time Ser. Anal.* **1** 15–29.

GRAYBILL, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, Calif.

IBRAGIMOV, I. A. (1963). On estimation of the spectral function of a stationary Gaussian process. *Theory Probab. Appl.* **8** 366–401.

LEHMANN, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

MANDELBROT, B. B. and VAN NESS, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10** 422–437.

MANDELBROT, B. B. and WALLIS, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research* **5** 321–340.

POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

TANIGUCHI, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Ann. Statist.* **11** 157–169.

TAQQU, M. S. (1986). Toeplitz matrices and estimation of time series with long-range dependence. Proc. Bernoulli Congress, Tashkent.

WALKER, A. M. (1964). Asymptotic properties of least-squares estimates of parameters of the spectrum of a stationary non-deterministic time-series. *J. Austral. Math. Soc.* **4** 363–384.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2** 423–434.

YAJIMA, Y. (1985). On estimation of long-memory time series models. *Austral. J. Statist.* **27** 303–320.

INSTITUTE OF APPLIED MATHEMATICS
UNIVERSITY OF HEIDELBERG
IM NEUENHEIMER FELD 294
D-6900 HEIDELBERG 1
FEDERAL REPUBLIC OF GERMANY