The Annals of Statistics
2003, Vol. 31, No. 2, 586–612

VALID ASYMPTOTIC EXPANSIONS FOR THE MAXIMUM LIKELIHOOD ESTIMATOR OF THE PARAMETER OF A STATIONARY, GAUSSIAN, STRONGLY DEPENDENT PROCESS

BY OFFER LIEBERMAN, JUDITH ROUSSEAU AND DAVID M. ZUCKER

Technion and Yale University, CREST-ENSAE and Hebrew University of Jerusalem

We establish the validity of an Edgeworth expansion to the distribution of the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. The result covers ARFIMA-type models, including fractional Gaussian noise. The method of proof consists of three main ingredients: (i) verification of a suitably modified version of Durbin's general conditions for the validity of the Edgeworth expansion to the joint density of the log-likelihood derivatives; (ii) appeal to a simple result of Skovgaard to obtain from this an Edgeworth expansion for the joint distribution of the log-likelihood derivatives; (iii) appeal to and extension of arguments of Bhattacharya and Ghosh to accomplish the passage from the result on the log-likelihood derivatives to the result for the maximum likelihood estimators. We develop and make extensive use of a uniform version of a theorem of Dahlhaus on products of Toeplitz matrices; the extension of Dahlhaus' result is of interest in its own right. A small numerical study of the efficacy of the Edgeworth expansion is presented for the case of fractional Gaussian noise.

1. Introduction. We consider a parametric stationary Gaussian process model for the time series $\{X_t,\ t \in \mathbf{Z}\}$ with autocovariance function $\gamma_\theta(u)$ and spectral density $f_\theta(\lambda)$, where $\theta \in \Theta \subset \mathbb{R}^m$. For simplicity we assume the process is mean zero. The extension of our results to the case in which an unknown mean needs to be replaced by the sample mean, as suggested by Dahlhaus (1989), is done elsewhere. The most popular time series model in practice is the ARMA(j,l) model [Box and Jenkins (1976), Section 3.4], defined as $\Phi(B)X_t = \Psi(B)\varepsilon_t$, where B is the backshift operator $BX_t = X_{t-1}$,

$$\Phi(w) = 1 + \sum_{r=1}^{j} \Phi_r w^r,$$

$$\Psi(w) = 1 + \sum_{r=1}^{l} \Psi_r w^r$$

and $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$. This model is stationary and invertible, provided that the roots of the polynomials Φ and Ψ lie outside the unit circle.

Received February 1999; revised January 2002.
AMS 2000 subject classifications. 62E17, 62M10.
Key words and phrases. Edgeworth expansions, long memory processes, ARFIMA models.

The ARMA(j,l) process is short memory in the sense that the autocovariance $\gamma_\theta(u)$ decays to zero at a geometric rate, that is, is of order $\rho^{-|u|}$ with $\rho \in (0,1)$, as $u \to \infty$. Stemming from Hurst (1951), a literature has developed on models in which the autocovariance function decays more slowly and are therefore more suitable for describing long-range dependence. As discussed by Brockwell and Davis [(1991), Section 13.2], an intermediate memory process is a process for which $\gamma_\theta(u)$ is of order $|u|^{2d-1}$ with $d \in (-\frac{1}{2}, 0)$ as $u \to \infty$ [$\gamma_\theta(u)$ decays hyperbolically but is still absolutely summable], whereas a long memory process is a process for which $\gamma_\theta(u)$ is of order $|u|^{2d-1}$ with $d \in (0, \frac{1}{2})$ as $u \to \infty$ [$\gamma_\theta(u)$ decays hyperbolically and is not absolutely summable]. The most well-known model for intermediate and long memory processes is the fractionally integrated ARMA model ARFIMA(j,d,l) introduced by Hosking (1981) and by Granger and Joyeux (1980) and defined by

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t, \tag{1}$$

where Φ and Ψ are as above, $d \in (-\frac{1}{2}, \frac{1}{2})$, and $(1-B)^d$ is defined by the binomial formula. This process satisfies the condition that $\gamma_\theta(u)$ be of order $|u|^{2d-1}$ as $u \to \infty$.

Very slow decay in the autocovariance function leading to nonsummability corresponds to a pole at the origin in the spectral density function. Accordingly, Robinson (1995) and others have used the terms "strongly dependent process" or "long memory process" to refer to a process with spectral density satisfying

$$f_\theta(\lambda) \sim |\lambda|^{-\alpha(\theta)} A_\theta(\lambda) \qquad \text{as } \lambda \to 0, \tag{2}$$

with $0 < \alpha(\theta) < 1$ and $A_\theta(\lambda)$ slowly varying at 0 in the sense that $\lambda^\delta A_\theta(\lambda)$ is bounded for every $\delta > 0$. The ARFIMA(j,d,l) process with $d > 0$ satisfies this condition with $\alpha(\theta) = 2d$.

This paper concerns asymptotic theory for the maximum likelihood estimator (MLE) $\hat{\theta}_n$ of the parameter θ based on the observations $(X_1, \dots, X_n)$ for stationary Gaussian time series models behaving according to (2). Dahlhaus (1989) established consistency and asymptotic normality of the MLE, drawing on corresponding results by Fox and Taqqu (1986) for Whittle's (1953) approximate MLE. Here we take the theory further by demonstrating the validity of an Edgeworth expansion for the distribution of the MLE. Denote by θ_0 the true value of θ, by $M(\theta)$ the inverse of the limit of n^{-1} times the expected information matrix, and by $\phi_{M(\theta)}$ the multivariate normal density with mean zero and covariance matrix $M(\theta)$. We show, under suitable technical conditions, that

$$\begin{aligned} &\Pr_{\theta_0}\big(\sqrt{n}(\hat{\theta}_n - \theta_0) \le t\big) \\ &\quad = \int_0^t \left[1 + \sum_{r=3}^{s} n^{-(r-2)/2} q_{n,r,\theta_0}(u)\right] \phi_{M(\theta_0)}(u)\,du + o(n^{-(s-2)/2}), \end{aligned} \tag{3}$$

where s is an integer defining the order of the expansion and $q_{n,r,\theta_0}(u)$ are polynomials in u. The form of the expansion and the order of the approximation error are the same as in the i.i.d. case. The coefficients of $q_{n,r,\theta_0}(u)$ are $O(1)$ and are determined by n^{-1} times the cumulants of the log-likelihood derivatives, which in our case, in contrast with the i.i.d. case, depend on n. We arrive at the result (3) by proving a similar Edgeworth expansion for the log-likelihood derivatives (LLDs) and using the argument of Bhattacharya and Ghosh (1978) to proceed from the expansion for the LLDs to that for the MLE.

This result provides the basis for more accurate calculations of p-values and confidence intervals based on the MLE than are obtainable using the normal approximation. In addition, the result provides theoretical support for Bartlett-corrected likelihood ratio tests, for hypotheses of the form $H_0 : \vartheta = \vartheta^{(0)}$, where ϑ is a subvector of θ, as described, for example, in Barndorff-Nielsen and Cox [(1994), Sections 4.4 and 6.5]. According to Barndorff-Nielsen and Hall (1988), if the LLDs and the MLE admit an Edgeworth expansion as above with $s = 6$, the error rate of the Bartlett-corrected likelihood ratio test is $O(n^{-2})$. Bartlett correction of likelihood ratio tests in the context of the ARFIMA model is discussed in detail in Lieberman, Rousseau and Zucker (2000).

There is a vast literature on Edgeworth and related expansions for i.i.d. and weakly dependent processes. Bhattacharya and Rao (1976) give a comprehensive account of Edgeworth expansion in the i.i.d. case. Taniguchi (1984, 1986, 1988, 1991) addresses asymptotic expansions in the ARMA setting. Lahiri (1993) and Götze and Hipp (1994) outline the present state of affairs in regard to Edgeworth expansions in the setting of weakly dependent processes. For strongly dependent processes, however, we are not aware of any work to date, except perhaps for Dahlhaus (1988), on higher order asymptotic theory. Thus, the results presented here constitute a significant advance.

The plan of the paper is as follows. Section 2 presents the basic assumptions and some preliminaries. Section 3 presents a uniform version, needed for our work, of Dahlhaus' (1989) Theorem 5.1 on the limit of the trace of a product of Toeplitz matrices. This result is of interest in its own right. Section 4 provides background on Edgeworth expansions. Section 5 presents the main results. Section 6 presents specific details for the case of ARFIMA processes, including a numerical study. Section 7 presents the proofs of the results.

2. Preliminaries.

2.1. *Background and assumptions.* Basic background and notation have been presented in the Introduction. The data vector is represented by $x = (X_1, \dots, X_n)'$. The basic model assumption is that x is multivariate normal with mean zero and covariance matrix $T_n(f_\theta) = [\gamma_\theta(j-k)]_{1 \le j,k \le n}$, where

$$\gamma_\theta(u) = \int_{-\pi}^{\pi} f_\theta(\lambda) e^{i\lambda u}\, d\lambda. \tag{4}$$

We impose the following assumptions on the parameter space Θ and the spectral density f_θ. The integer s appearing in Assumption II determines the order of the Edgeworth expansion.

ASSUMPTIONS. I. The parameter space Θ is an open subset of $\mathbb{R}^m$.

II. For some positive integer s, $f_\theta(\lambda)$ is $s+1$ times continuously differentiable with respect to θ, and all derivatives are continuous in (λ, θ), $\lambda \neq 0$. In addition, $f_\theta(\lambda)^{-1}$ is continuous in (λ, θ) for all λ and θ.

III. The derivatives $(\partial/\partial\lambda) f_\theta(\lambda)^{-1}$ and $(\partial^2/\partial\lambda^2) f_\theta(\lambda)^{-1}$ are continuous in (λ, θ) for $\lambda \neq 0$. In addition, there exist $\alpha(\theta) \in (0, 1)$ and $c_1(\theta, \delta) < \infty$ such that

$$\left| \left(\frac{\partial}{\partial \lambda} \right)^k f_\theta(\lambda)^{-1} \right| \le c_1(\theta, \delta) |\lambda|^{\alpha(\theta) - k - \delta}$$

for $k = 0, 1, 2$ and all $\delta > 0$.

IV. With $\alpha(\theta)$ as in Assumption III, there exist $c_2(\theta, \delta) < \infty$, $c_3(\theta, \delta) < \infty$ such that, for every $\delta > 0$, the following hold over $\lambda \in (0, \pi)$:

(a) $|f_\theta(\lambda)| \le c_2(\theta, \delta) |\lambda|^{-\alpha(\theta) - \delta}$;

(b) for all $(j_1, \dots, j_k)$, $k \le s + 1$, with duplication among the j_l allowed,

$$\left| \frac{\partial^k f_\theta(\lambda)^{-1}}{\partial \theta_{j_1} \cdots \partial \theta_{j_k}} \right| \le c_3(\theta, \delta) |\lambda|^{\alpha(\theta) - \delta}.$$

V. For any compact subset Θ^* of Θ there exists a constant $C(\Theta^*, \delta) < \infty$ such that the constants $c_1(\theta, \delta), c_2(\theta, \delta)$ and $c_3(\theta, \delta)$ in Assumptions III and IV are bounded by $C(\Theta^*, \delta)$ for all $\theta \in \Theta^*$.

VI. (a) There exists a function $\Omega(\lambda)$ that is integrable over $(0, \pi)$ and a constant $c_4(\theta) < \infty$ such that for all $(j_1, \dots, j_k)$, $k \le s + 1$, with duplication among the j_l allowed,

$$\left| \frac{\partial^k f_\theta(\lambda)}{\partial \theta_{j_1} \cdots \partial \theta_{j_k}} \right| \le c_4(\theta) \Omega(\lambda)$$

for $\lambda \in (0, \pi)$. For any compact subset Θ^* of Θ, there exists a constant $\tilde{C}(\Theta^*) < \infty$ such that $c_4(\theta) \le \tilde{C}(\Theta^*)$ for all $\theta \in \Theta^*$.

(b) When computing derivatives of the form

$$\frac{\partial^k \gamma_\theta(u)}{\partial \theta_{j_1} \cdots \partial \theta_{j_k}}, \qquad k \le s + 1,$$

the derivatives may be taken inside the integral sign of (4).

VII. The function $\alpha(\theta)$ is continuous in θ.

In addition to the foregoing assumptions, we need one further assumption, presented formally in Section 2.3 as Assumption VIII, to the effect that the limiting covariance matrix of the vector of LLDs is positive definite.

Assumption I is a standard background assumption and is made in Bhattacharya and Ghosh's (1978) paper on Edgeworth expansions for MLEs in the i.i.d. setting. Assumptions II–IV are adapted from Dahlhaus (1989). Assumptions III and IV are needed to make use of Dahlhaus' Theorem 5.1 on the limiting behavior of the trace of products of Toeplitz matrices. Assumption III corresponds to Dahlhaus' Assumption (A7). Assumption IV(a) corresponds to the relation (2) presented in the Introduction, while Assumption IV(b) is a extension of Dahlhaus' Assumption (A3) to derivatives of order $k = 3, \dots, s + 1$. Assumption V on uniform bounding constants over compacts in Assumptions III and IV corresponds to Dahlhaus' Assumption (A8) and is needed for the uniform in θ version of Dahlhaus' Theorem 5.1 that we require for our work. Assumption VI is needed to allow interchange of a limit over θ and an integral over λ in integrals involving derivatives of the spectral density and to suitably bound the cumulants of the LLDs. In particular, Assumption VI(b) is needed to allow problem-free calculation of the LLDs. Assumption VII corresponds to Dahlhaus' Assumption (A9) and is used in the proof of the uniform version of Dahlhaus' theorem.

As discussed in Section 6, Assumptions I–VII all hold for ARFIMA processes for all s. Thus, provided Assumption VIII stated in Section 2.3 also holds, for an ARFIMA process the error term in the Edgeworth expansion can be made of any desired order by taking sufficiently many terms in the expansion.

2.2. *General outline of development.* Our ultimate goal is to establish an Edgeworth expansion for the distribution of the MLE. In Section 2.3, we develop expressions for the log-likelihood derivatives (LLDs), their first and second moments and the limiting form of these moments. The expressions for the moments of the LLDs involve traces of products of Toeplitz matrices, the limiting behavior of which has been treated by Dahlhaus [(1989), Theorem 5.1]. Theorem 2 of Section 3 provides a uniform version of Dahlhaus' theorem which is needed for our work. After a review of Edgeworth expansion theory in Section 4, we present our main results, given in Section 5, Theorems 3 and 4. Theorem 3 presents an Edgeworth expansion for the vector of LLDs. The validity of Edgeworth expansion for the density is proved by verifying a suitably modified version of the conditions of Durbin's (1980) Theorem 1. The expansion for the distribution function is then obtained using Skovgaard's (1986) Corollary 3.3. Theorem 4 presents an Edgeworth expansion for the distribution of the MLE. This result is obtained from Theorem 3 on the LLDs using the argument of Bhattacharya and Ghosh (1978).

The following definition will be used throughout:

$$T_n(h) = \left[\int_{-\pi}^{\pi} e^{i(j-k)\xi} h(\xi)\, d\xi \right]_{1 \le j,k \le n}.$$

2.3. *Log-likelihood derivatives.* The log-likelihood for our model is

$$L(\theta) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log\det T_n(f_\theta) - \frac{1}{2}x' T_n^{-1}(f_\theta)x.$$

Let $L_r = \partial L/\partial\theta_r$, $L_{rs} = \partial^2 L/\partial\theta_r\,\partial\theta_s$, $L_{rst} = \partial^3 L/\partial\theta_r\,\partial\theta_s\,\partial\theta_t$ and so on. For a given set of subscripts $\nu = (r_1 \cdots r_q)$, let $\lambda_\nu = E_\theta[L_\nu(\theta)]$ and $l_\nu = L_\nu(\theta) - \lambda_\nu$. When $|\nu| = 1$, $\lambda_\nu = 0$. Explicit expressions for the λ_ν's for derivatives of up to order four are given by Lieberman, Rousseau and Zucker (2000). In general, the L_ν take the form

$$L_\nu = x' B_\nu(\theta)x - F_\nu(\theta). \tag{5}$$

Here

$$B_\nu(\theta) = -\frac{1}{2}\frac{\partial^q T_n^{-1}(f_\theta)}{\partial\theta_{r_1}\cdots\partial\theta_{r_q}} = \sum_{k=1}^{b_\nu} a_k\left[\prod_{j=1}^{p_k} T_n^{-1}(f_\theta)T_n(g_{\theta,j})\right]T_n^{-1}(f_\theta),$$

where the a_k's are constants and the $g_{\theta,j}$'s are derivatives of the spectral density with respect to θ. Since $T_n^{-1}(f_\theta)$ is symmetric, the B_ν's are also symmetric. The $F_\nu(\theta)$ take the following form (with a_k and $g_{\theta,j}$ being of the same nature as, though not identical to, those appearing in the expression for B_ν):

$$F_\nu(\theta) = \sum_{k=1}^{b_\nu} a_k \operatorname{tr}\left[\prod_{j=1}^{p_k} T_n^{-1}(f_\theta)T_n(g_{\theta,j})\right]. \tag{6}$$

The expected LLDs are of the same form as (6). Under Assumptions II–IV and VI(b), it follows from Theorem 5.1 of Dahlhaus (1989) that

$$\lim_{n\to\infty}\frac{1}{n}\operatorname{tr}\left[\prod_{j=1}^{p_k} T_n^{-1}(f_\theta)T_n(g_{\theta,j})\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{j=1}^{p_k}\frac{g_{\theta,j}(\lambda)}{f_\theta(\lambda)}\,d\lambda. \tag{7}$$

This result was proved by Taniguchi (1983) for ARMA processes and extended by Dahlhaus to strongly dependent processes. It will be used extensively in our work. It follows from (7) that $\lambda_\nu = O(n)$ for $|\nu| \geq 2$ as in the i.i.d. case.

We denote by $Z_n(\theta)$ the vector consisting of the LLDs $L_\nu(\theta)$ up to order $s-1$ that are linearly independent as functions of θ and denote the dimension of this vector by d. We let $Y_n(\theta)$ denote the corresponding vector of centered log-likelihood derivatives $l_\nu(\theta)$. We let $\nu(i)$ denote the series of indices ν corresponding to the ith component of $Z_n(\theta)$. We define $W_n(\theta) = Y_n(\theta)/\sqrt{n}$ and $D_n(\theta) = E_\theta[W_n(\theta)W_n(\theta)']$. Because $Y_n(\theta)$ is a vector of central quadratic forms in Gaussian variables plus a vector of nonrandom quantities we have [see Anderson (1984) or Johnson, Kotz and Balakrishnan (1997)] $D_n(\theta)(i,j) = 2\operatorname{tr}(\tilde{B}_i\tilde{B}_j)/n$, where $\tilde{B}_i = B_{\nu(i)}T_n(f_\theta)$. By (7), $D_n(\theta)(i,j)$ converges to the matrix

$$D(\theta)(i,j) = \frac{1}{\pi}\sum_{k=1}^{b_i}\sum_{l=1}^{b_j} a_k^i a_l^j \int_{-\pi}^{\pi}\prod_{r=1}^{p_k}\frac{g_{\theta,r}^{(i)}(\lambda)}{f_\theta(\lambda)}\prod_{s=1}^{p_l}\frac{g_{\theta,s}^{(j)}(\lambda)}{f_\theta(\lambda)}\,d\lambda. \tag{8}$$

Thus W_n is $O_p(1)$, as in the i.i.d. case.

In addition to Assumptions I–VII, we make the following assumption:

VIII. The matrix $D(\theta)$ is positive definite.

Assumption VIII is needed to apply Durbin's (1980) theorem. If $D(\theta)$ is singular, then there exists an index i such that the ith column of $D(\theta)$ is a linear combination of the other columns; that is, there exists $(\alpha_1, \dots, \alpha_d)$ such that, for all $\lambda \neq 0$,

$$\sum_{k=1}^{b_i} a_k^{(i)} \prod_{r=1}^{p_k} \frac{g_{\theta,r}^{(i)}(\lambda)}{f_\theta(\lambda)} = \sum_{j \neq i} \alpha_j \sum_{l=1}^{b_j} a_l^{(j)} \prod_{r=1}^{p_l} \frac{g_{\theta,r}^{(j)}(\lambda)}{f_\theta(\lambda)}. \tag{9}$$

Now observe that

$$-\frac{1}{2} \frac{\partial^q f_\theta^{-1}}{\partial\theta_{r_1} \cdots \partial\theta_{r_q}} = f_\theta^{-1} \sum_{k=1}^{b_\nu} a_k \left[\prod_{j=1}^{p_k} \frac{g_{\theta,j}}{f_\theta} \right]$$

with the same a_k's as in the definition of B_ν. We thus see that (9) is equivalent to the existence of a linear combination between the derivatives of $f_\theta^{-1}(\lambda)$ with the same linear combination coefficients. For any $f_\theta(\lambda)$, therefore, it can be easily verified whether positive definiteness holds or not. Note though that unlike the i.i.d. case, in general singularity of $D(\theta)$ does not imply singularity of $D_n(\theta)$ for finite n. In other words, for finite n, collinearity of derivatives of $f_\theta^{-1}(\lambda)$ does not necessarily imply collinearity of the LLDs. For a weak dependence setting, Taniguchi (1991) proved the validity of an Edgeworth expansion to the distribution of the MLE with an error rate of $o(n^{-1})$. For the class of models he considered, the $D_n(\theta)$ matrix converges to $D(\theta)$ at a n^{-1} rate, and so he concluded that if the LLDs are asymptotically collinear, some of them may be dropped without affecting the order of the error of the expansion [Taniguchi (1991), page 78]. This argument does not carry over in general for higher order expansions. We conclude that under Assumption VIII the Edgeworth expansions in our paper are valid to $o(n^{-(s-2)/2})$ whereas if the assumption is violated, the expansions are generally still valid to at least $o(n^{-1})$. We show in Section 6 that positive definiteness holds for the ARFIMA$(0, d, 0)$ model.

3. A result on products of Toeplitz matrices. In Theorems 1 and 2 below we present uniform versions of Fox and Taqqu's (1987) Theorem 1.a and Dahlhaus' (1989) Theorem 5.1, which both deal with the limiting behavior of the trace of the product of certain Toeplitz matrices. Theorem 2 is the key result for our purposes, with Theorem 1 being the main building block in its proof. Theorems 1 and 2 are also of independent interest; the behavior of such quantities is a topic of longstanding interest going back to Grenander and Szegö (1958). Theorem 2 is used repeatedly in the proofs of Theorems 3 and 4. Assumptions II–VII ensure that in each application of Theorem 2, the conditions of Theorem 2 are satisfied.

THEOREM 1. *Let* $f_{\theta,1},\dots,f_{\theta,p}$ *and* $g_{\theta,1},\dots,g_{\theta,p}$ *be symmetric real-valued functions defined on* $[-\pi,\pi]$, *continuous on* $\{\lambda : |\lambda| > t\}$, $\forall t > 0$. *Suppose that* $\forall \theta \in \Theta,\ \exists \varepsilon > 0$, *such that* $\forall \delta > 0, \exists M_\theta \ge 0$ *for which*

$$\sup_{|\theta'-\theta|<\varepsilon} |f_{\theta',i}(\lambda)| \le M_\theta |\lambda|^{-\alpha(\theta)-\delta}, \qquad i = 1,\dots,p, \tag{10}$$

$$\sup_{|\theta'-\theta|<\varepsilon} |g_{\theta',i}(\lambda)| \le M_\theta |\lambda|^{-\beta(\theta)-\delta}, \qquad i = 1,\dots,p, \tag{11}$$

for all $\lambda > 0$, *where* $\alpha(\theta) < 1$ *and* $\beta(\theta) < 1$. *Also suppose that* $\forall t > 0,\ \exists M_{t,\theta}$ *such that*

$$\sup_{|\theta'-\theta|<\varepsilon; |\lambda|>t} \left| \frac{d f_{\theta',i}(\lambda)}{d\lambda} \right| \le M_{t,\theta}$$

and

$$\sup_{|\theta'-\theta|<\varepsilon; |\lambda|>t} \left| \frac{d g_{\theta',i}(\lambda)}{d\lambda} \right| \le M_{t,\theta}, \qquad i = 1,\dots,p.$$

Assume further that $p(\alpha(\theta)+\beta(\theta)) < 1$ *for all* θ. *Then for all* θ *and* ε,

$$\begin{aligned} \lim_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \Bigg| \frac{1}{n} \operatorname{tr}\left[\prod_{i=1}^{p} T_n(f_{\theta',i}) T_n(g_{\theta',i}) \right] \\ - (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{i=1}^{p} f_{\theta',i}(\lambda) g_{\theta',i}(\lambda)\, d\lambda \Bigg| = 0. \end{aligned} \tag{12}$$

Consequently, for any compact subset Θ^* *of* Θ,

$$\begin{aligned} \lim_{n\to\infty} \sup_{\theta'\in\Theta^*} \Bigg| \frac{1}{n} \operatorname{tr}\left[\prod_{i=1}^{p} T_n(f_{\theta',i}) T_n(g_{\theta',i}) \right] \\ - (2\pi)^{2p-1} \int_{-\pi}^{\pi} \prod_{i=1}^{p} f_{\theta',i}(\lambda) g_{\theta',i}(\lambda)\, d\lambda \Bigg| = 0. \end{aligned} \tag{13}$$

THEOREM 2. *Let* Θ^* *be a compact subset of* Θ. *Let* p *be a positive integer and let* $\alpha(\theta)$ *and* $\beta(\theta)$ *be continuous functions on* Θ^* *with range in* $(0, 1)$ *satisfying* $\beta(\theta) - \alpha(\theta) < 1/(2p)$. *Suppose that* $f_{\theta,j}(\lambda),\ j \le p$, *are symmetric nonnegative functions and that* $g_{\theta,j}(\lambda),\ j \le p$, *are symmetric real-valued functions satisfying the following conditions*:

A. *The* $f_{\theta,j}(\lambda)$*'s satisfy the conditions stated for* $f_\theta(\lambda)$ (*with* $s = 1$) *in Assumptions* II, III *and* IV(a) *of Section* 2.1 *over* $\lambda \in (0,\pi)$ *and* $\theta \in \Theta^*$, *with bounding constants* c_1, c_2 *and* c_3 *that may depend on* δ *but do not depend on* θ.

B. *The* $g_{\theta,j}$*'s are continuous at all* $\lambda \ne 0$ *and for each* $\delta > 0$ *there exists* $c^*(\delta)$ *such that* $|g_{\theta,j}(\lambda)| \le c^*(\delta)|\lambda|^{-\beta(\theta)-\delta}$ *for all* $\theta \in \Theta^*$.

Then

$$\lim_{n\to\infty}\sup_{\theta\in\Theta^*}\left|\frac{1}{n}\operatorname{tr}\left[\prod_{j=1}^{p}\{T_n(f_{\theta,j})^{-1}T_n(g_{\theta,j})\}\right]\right.$$
$$\left.-\frac{1}{2\pi}\int_{-\pi}^{\pi}\prod_{j=1}^{p}(f_{\theta,j}(\lambda))^{-1}g_{\theta,j}(\lambda)\,d\lambda\right|=0. \tag{14}$$

4. Background on Edgeworth expansions. We now briefly review Edgeworth expansion theory; for further details, see Barndorff-Nielsen and Cox (1989). Let ξ_n be a random p-vector with cumulants $O(n)$, mean $n\mu_n$ and covariance matrix nK_n (assumed positive definite). Let ϕ_{K_n} denote the $N(0, K_n)$ density. Then the $(\tau-2)$-order formal Edgeworth expansion of the density of $U_n = \sqrt{n}(n^{-1}\xi_n - \mu_n)$ is given by

$$a_{\tau-2,n}(u) = \phi_{K_n}(u)\Gamma_{\tau-2,n}(u), \tag{15}$$

with

$$\Gamma_{\tau-2,n}(u) = 1 + \sum_{j=3}^{\tau} n^{-(j-2)/2}P_{nj}(u),$$

where $P_{nj}(u,\theta)$ is the generalized Edgeworth polynomial of order j as defined by Durbin [(1980), Equation (31)]. The putative error rate of the expansion is $o(n^{-(\tau-2)/2})$.

Chambers (1967) discusses computation of $\Gamma_{\tau-2,n}(u)$; Barndorff-Nielsen and Cox [(1989), Equation (6.22)] give the expression for $\tau-2=2$. The expression involves normalized cumulants of ξ_n, which may be replaced by asymptotic approximations without affecting the error rate. Estimating the cumulants does not affect the error rate of the order 1 expansion, but degrades the error rate of the order 2 expansion to $O(n^{-1})$. In (3) we use an equivalent formulation of the the Edgeworth expansion which is built around the limiting values μ and K of μ_n and K_n, respectively.

For maximum likelihood estimators (MLEs), Peers and Iqbal [(1985), page 554] give a expansion of the cumulant generating function with an error of $O(n^{-3/2})$, which yields the first four cumulants of the MLE up to order $O(n^{-3/2})$ in terms of the cumulants of the log-likelihood derivatives (LLDs). In our setting the LLDs are quadratic forms in the multivariate normal variate x [see (5)], whose cumulants are given in Anderson (1984). Alternatively, the cumulants of the MLE may be computed by a bootstrap procedure [cf. Rocke (1989)].

5. Main Edgeworth expansion results. Our main results are that the LLD vector admits a valid Edgeworth expansion of any order and that the MLE admits a valid Edgeworth expansion of order $s-2$, where s is as indicated in the assumptions set forth in Section 2. We begin with a formal statement of the LLD result. Recall that, with Z_n denoting the vector of LLDs, W_n and D_n are defined by $W_n = n^{-1/2}(Z_n - E_\theta[Z_n])$ and $D_n(\theta) = E_\theta[W_n W_n']$.

THEOREM 3. *Let $G_n(u,\theta)$ be the true density of W_n and let $\tilde{G}_n^{(\tau-2)}(u,\theta)$ be its $(\tau-2)$-order formal Edgeworth expansion (as defined in the preceding section), where τ is any integer equal to or greater than* 3. *Then, under Assumptions* I–VIII, *the following results hold*:

(a) *We have*

$$G_n(u,\theta) - \tilde{G}_n^{(\tau-2)}(u,\theta) = o(n^{-(\tau-2)/2})$$

uniformly over u and θ in any compact subset Θ^* of Θ.

(b) *For any $\tau = \tau_0$ taken in part* (a),

$$\Pr_\theta(W_n \in C) = \int_C \tilde{G}_n^{(\tau_0-2)}(u,\theta)\,du + o(n^{-(\tau_0-2)/2+\delta}) \qquad \forall\,\delta > 0,$$

uniformly over all Borel sets C and θ belonging to a compact subset Θ^* of Θ.

It is emphasized that the LLD expansions are valid for τ as large as we please. We now pass from the LLD result to the MLE result. This type of passage has been discussed in a number of papers, most notably Chibisov (1973), Bhattacharya and Ghosh (1978), Skovgaard (1981) and Taniguchi (1991). We follow the approach of Bhattcharya and Ghosh (BG). The general idea of the argument is as follows. The MLE is the solution to the likelihood equations $L_{r_1}(\hat{\theta}) = 0$, $r_1 = 1,\dots,m$. We expand $L_{r_1}(\theta)$ in a Taylor series as follows [cf. BG's Equation (1.27)]:

$$\begin{aligned}\frac{1}{n}L_{r_1}(\theta) &= \frac{1}{n}L_{r_1}(\theta_0)\\ &\quad+\frac{1}{n}\sum_{q=2}^{s-1}\frac{1}{q!}\sum_{r_2=1}^{m}\cdots\sum_{r_q=1}^{m}L_{r_1,\dots,r_q}(\theta_0)(\theta_{r_1}-\theta_{0,r_1})\cdots(\theta_{r_q}-\theta_{0,r_q})\\ &\quad+R_{n,r_1}(\theta),\end{aligned}\tag{16}$$

with

$$|R_{n,r_1}(\theta)| \le \frac{K}{n}\|\theta-\theta_0\|^s \max_{|\nu|=s+1}\ \sup_{\|\theta'-\theta_0\|\le\|\theta-\theta_0\|}|L_\nu(\theta)|,$$

where K is a constant. The expansion then is inverted to obtain an expansion for the MLE $\hat{\theta}_n$ in terms of the LLDs, from which we develop a formal Edgeworth expansion for the MLE from the Edgeworth expansion for the LLDs. The claim is that the formal expansion for the MLEs thus obtained is a valid expansion. BG provide a rigorous proof of this claim for the i.i.d. setting. We adapt BG's argument to our case. A formal statement of our result follows.

THEOREM 4. *Let* $\tilde{H}_n^{(s-2)}(u,\theta)$ *be the* $(s-2)$*-order formal Edgeworth expansion for the density of the MLE of* θ (*defined as described in the preceding section*). *Let* Θ^* *be a compact subset of* Θ. *Then, under Assumptions* I–VIII,

(a) *there exists a sequence of statistics* $\hat{\theta}_n$ *and a constant* $d_0 = d_0(\Theta^*)$ *such that for any* $\delta > 0$,

$$\inf_{\theta_0 \in \Theta^*} \Pr_{\theta_0}\big(\|\hat{\theta}_n - \theta_0\| < d_0 n^{-1/2+\delta}, \hat{\theta}_n \textit{ solves the likelihood equations}\big)$$
$$= 1 - o(n^{-(s-2)/2});$$

(b) *any estimator* $\hat{\theta}_n$ *satisfying the statement in* (a) *above admits the Edgeworth expansion*

$$\Pr_{\theta_0}\big((\sqrt{n}(\hat{\theta}_n - \theta_0)) \in C\big) = \int_C \tilde{H}_n^{(s-2)}(u,\theta)\,du + o(n^{-(s-2)/2}) \tag{17}$$

uniformly over Θ^* *and every class* $\mathcal{B}$ *of Borel sets satisfying the condition*

$$\sup_{\theta \in \Theta^*} \sup_{C \in \mathcal{B}} \int_{(\partial C)^\varepsilon} \phi_{M(\theta)}(u)\,du = O(\varepsilon) \textit{ as } \varepsilon \downarrow 0,$$

where $M(\theta)$ *is the inverse of the limiting normalized expected information matrix* $\lim_{n\to\infty}[n^{-1}\lambda_{r_1 r_2}(\theta)]$, $\phi_{M(\theta)}$ *is the* $N(0, M(\theta))$ *density*, *and* $(\partial C)^\varepsilon$ *denotes the* ε*-neighborhood of the boundary of* C.

The uniform version of Dahlhaus given in our Theorem 2 leads to the uniform in θ result in Theorem 3. More important, the uniform version is critically necessary to enable us to handle the remainder term in the Taylor expansion (16) of the first-order log-likelihood derivatives.

6. Applications to the ARFIMA model. We illustrate our main results with the ARFIMA(j,d,l) model. For this general model we shall give two examples: one in which all assumptions are satisfied and one in which Assumption VIII does not hold. The spectral density function of $\{X_t\}$ is as follows [Brockwell and Davis (1991), Equation (13.2.18)]:

$$f_\theta(\lambda) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\Psi(e^{i\lambda})|^2}{|\Phi(e^{i\lambda})|^2} |1 - e^{i\lambda}|^{-2d},$$

where $\theta = (\sigma_\varepsilon^2; d; \Phi_1, \dots, \Phi_j; \Psi_1, \dots, \Psi_l)$. Clearly,

$$f_\theta(\lambda) \sim |\lambda|^{-2d} \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\Psi(1)|^2}{|\Phi(1)|^2} \qquad \text{as } \lambda \to 0,$$

and so the condition implied by (2) is satisfied for $0 < d < \frac{1}{2}$. The derivatives of

$f_\theta(\lambda)$ up to order 4 are available on request from the authors. It is very easy to see that Assumptions I–VII hold for this model. For the ARFIMA$(0, d, 0)$ model with $\sigma_\varepsilon^2 = 1$, the inverse spectral density derivatives are $(d^a/dd^a) f_d^{-1}(\lambda) = c^a(\lambda) f_d^{-1}(\lambda)$ with $c(\lambda) = 2\log|1 - e^{i\lambda}|$, $a \geq 1$, and so it is clear that Assumption VIII is also satisfied.

We have conducted a small simulation experiment investigating the performance of the Edgeworth expansions in the Gaussian ARFIMA$(0, d, 0)$ model with a unit variance. The results are summarized in Table 1. We wrote a MATHEMATICA program to calculate the MLE in this model (details are available at http://iew3.technion.ac.il/Home/Users/offerl.phtml).

The likelihood was maximized by a line search on the interval $[-0.49, 0.49]$ with a grid of 0.001. The true d values were taken to be 0.1, 0.2, 0.3, 0.4 and the sample sizes were set to $n = 20, 40$. The simulation consisted of 3,000 replications. The cumulants in the Edgeworth expansions were calculated empirically. We compared the empirical cumulative distribution function of the MLE $\hat{d}$, evaluated at various cutpoints, with the classical asymptotic normal approximation, the normal approximation with centering and scaling based on the finite sample mean and standard deviation of $\hat{d}$ ("corrected normal approximation"), the first order Edgeworth expansion (Edg1), and the second-order Edgeworth expansion (Edg2). The cutpoints chosen were percentiles of the asymptotic normal approximation. A few patterns emerge. Generally speaking, the classical normal approximation is the poorest, the corrected normal approximation somewhat better, and the Edgeworth expansions better still, with Edg2 usually the best. All approximations improve with n, supporting the theoretical asymptotic tendencies established in the paper. The Edgeworth expansions generally behave better at the center of the distribution, and deteriorate towards the far tail. This phenomenon is in line with the behavior of the expansions in the i.i.d. case, see Barndorff-Nielsen and Cox [(1989), Chapter 5]. Overall, the Edgeworth expansions are appealing because of their practical computability and superiority to the normal approximation.

Next, the ARFIMA$(1, d, 1)$ model is very important in applied work. The following argument covers in fact the more general ARFIMA$(1, d, l)$ setup as well, with $l \geq 1$. Denote by $L_{3,d}(\theta)$ the third order LLD with respect to the AR(1) parameter. While $L_{3,d}(\theta)$ is not zero for fixed n, the third-order derivative of the inverse spectral density wrt the AR(1) parameter vanishes. This means that in the $D(\theta)$ matrix, the column corresponding to the asymptotic covariance between any normalized LLD and the normalized $L_{3,d}(\theta)$ is identically zero. The matrix $D_n(\theta)$, however, does not have a zero column. Hence in this case, Assumption VIII does not hold and we cannot simply discard the problematic $L_{3,d}(\theta)$ from the error analysis. Nevertheless, it is still possible to establish an $o(n^{-1})$ error rate for the distribution of the MLE. To do so, we must show that the term involving $L_{3,d}(\theta)$ in the stochastic expansion of the MLE can be absorbed into the remainder without affecting the $o(n^{-1})$ error rate. Following equations

TABLE 1
Edgeworth expansions for the Gaussian ARFIMA$(0, d, 0)$ *model*

	Probabilities, $n = 20$					Probabilities, $n = 40$				
d	asy nrml	empirical	corr nrml	Edg1	Edg2	asy nrml	empirical	corr nrml	Edg1	Edg2
0.1	0.01	0.048	0.033	0.051	0.046	0.01	0.040	0.027	0.042	0.040
	0.05	0.130	0.116	0.125	0.127	0.05	0.100	0.098	0.108	0.102
	0.5	0.589	0.636	0.602	0.594	0.5	0.562	0.593	0.561	0.563
	0.9	0.973	0.939	0.957	0.960	0.9	0.940	0.923	0.937	0.943
	0.95	0.996	0.971	0.988	0.993	0.95	0.978	0.961	0.977	0.980
0.2	0.01	0.066	0.041	0.067	0.057	0.01	0.041	0.028	0.046	0.043
	0.05	0.134	0.130	0.138	0.138	0.05	0.109	0.104	0.115	0.108
	0.5	0.580	0.640	0.593	0.583	0.5	0.580	0.616	0.580	0.580
	0.9	0.984	0.935	0.958	0.967	0.9	0.955	0.934	0.951	0.957
	0.95	1.000	0.967	0.992	1.001	0.95	0.988	0.968	0.986	0.990
0.3	0.01	0.096	0.077	0.106	0.090	0.01	0.054	0.033	0.057	0.049
	0.05	0.179	0.191	0.178	0.183	0.05	0.117	0.112	0.125	0.117
	0.5	0.608	0.674	0.619	0.606	0.5	0.565	0.618	0.569	0.565
	0.9	0.999	0.931	0.962	0.980	0.9	0.967	0.931	0.953	0.964
	0.95	1.000	0.962	0.997	1.014	0.95	1.000	0.965	0.990	0.999
0.4	0.01	0.146	0.171	0.164	0.149	0.01	0.094	0.074	0.105	0.085
	0.05	0.230	0.305	0.231	0.243	0.005	0.171	0.190	0.177	0.177
	0.5	0.640	0.710	0.642	0.639	0.5	0.623	0.689	0.634	0.624
	0.9	1.000	0.917	0.957	0.998	0.9	1.000	0.941	0.975	0.996
	0.95	1.000	0.947	0.997	1.039	0.95	1.000	0.969	1.004	1.020

(3.2.19) and (3.2.30) of Taniguchi (1990), it will be sufficient to show that there is a constant c_1 such that

$$\Pr_{\theta_0}\left(l^2_{1(i)}(\theta)\left|l_{3,d}(\theta)\right| > c_1\rho_n\right) = o(n^{-1}), \tag{18}$$

where $l_{1(i)}(\theta)$ is a normalized score wrt the ith component in the parameter vector, $l_{3,d}(\theta)$ is the normalized $L_{3,d}(\theta)$, both with $\sqrt{n}$-normalization, and ρ_n is a sequence satisfying $\rho_n \to 0$, and $\rho_n\sqrt{n} \to \infty$. It is clear from the proof of Dahlhaus [(1989), Theorem 5.1] and Fox and Taqqu (1987) that the cumulants of $L_{3,d}(\theta)$ are $O(n^{1/2+\delta})$ $\forall\,\delta > 0$. Further, the moments of $l_{1(i)}$ of all orders are $O(1)$. Thus, the condition in (18) follows immediatelly from the Markov and Cauchy–Schwarz inequalities. We remark that the $o(n^{-1})$ error rate is compatible with the rate achieved by Taniguchi (1991) in the short memory, uniparameter context.

7. Proofs.

7.1. *Proof of Theorem* 1. It suffices to prove (12); the assertion (13) then follows from the fact that any compact set Θ^* can be covered by a finite union of ε-balls. The (j,k)th element of the Toeplitz matrix $T_n(f_\theta)$ is

$$(T_n(f_\theta))_{j,k} = \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f_\theta(\lambda)\,d\lambda.$$

So,

$$\begin{aligned} L_n &= \frac{1}{n}\operatorname{tr}\prod_{i=1}^{p} T_n(f_{\theta,i})T_n(g_{\theta,i}) \\ &= \frac{1}{n}\sum_{j_1=0}^{n-1}\cdots\sum_{j_{2p}=0}^{n-1}\int_{U_\pi} e^{i(j_1-j_2)y_1}e^{i(j_2-j_3)y_2}\,\ldots\,e^{i(j_{2p}-j_1)y_{2p}} \\ &\qquad\qquad \times f_{\theta_1}(y_1)g_{\theta_1}(y_2)\,\cdots\,g_{\theta_p}(y_{2p})\,dy_1\,\cdots\,dy_{2p}, \end{aligned}$$

where $U_\pi = [-\pi,\pi]^{2p}$. We may rewrite the last expression as

$$L_n = \frac{1}{n}\int_{U_\pi} P_n(y)Q_\theta(y)\,dy,$$

with

$$P_n(y) = \sum_{j_1=0}^{n-1}\cdots\sum_{j_{2p}=0}^{n-1} e^{i(j_1-j_2)y_1}e^{i(j_2-j_3)y_2}\,\ldots\,e^{i(j_{2p}-j_1)y_{2p}} \tag{19}$$

and

$$Q_\theta(y) = f_{\theta_1}(y_1)g_{\theta_1}(y_2)\,\cdots\,g_{\theta_p}(y_{2p}).$$

Following Fox and Taqqu [(1987), page 237], we partition U_π into three disjoint sets E_t, F_t and G, satisfying

$$\begin{aligned} E_t &= U_\pi \cap (W \cup U_t)^c, \\ F_t &= U_t \cap W^c, \\ G &= U_\pi \cap W, \end{aligned}$$

where

$$U_t = [-t, t]^{2\pi}, \qquad 0 < t \le \pi,$$

$$W = \bigcup_{i=1}^{2p} W_i, \qquad W_i = \left\{ y \in \mathbb{R}^{2p} : |y_i| \le \frac{|y_{i+1}|}{2} \right\}, \qquad i = 1, \dots, 2p-1,$$

and

$$W_{2p} = \left\{ y \in \mathbb{R}^{2p} : |y_{2p}| \le \frac{|y_1|}{2} \right\}.$$

As $U_\pi = E_t \cup F_t \cup G$ and E_t, F_t, G are disjoint, the sufficient conditions for Theorem 1 to hold are

$$\begin{aligned} \lim_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \Bigg| \int_{E_t} \frac{1}{n} P_n(y) Q_{\theta'}(y)\, dy \\ - (2\pi)^{2p-1} \int_{t \le |z| \le \pi} \prod_{j=1}^{p} f_{\theta',j}(z) g_{\theta',j}(z)\, dz \Bigg| = 0, \end{aligned} \tag{20}$$

$$\lim_{t\to 0} \limsup_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \int_{F_t} \frac{1}{n} P_n(y) Q_{\theta'}(y)\, dy = 0 \tag{21}$$

and

$$\lim_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \int_{G} \frac{1}{n} P_n(y) Q_{\theta'}(y)\, dy = 0. \tag{22}$$

Instead of conditions (21) and (22), it will be sufficient to prove

$$\lim_{t\to 0} \limsup_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \int_{U_t} \frac{1}{n} |P_n(y) Q_{\theta'}(y)|\, dy = 0 \tag{23}$$

and

$$\lim_{n\to\infty} \sup_{|\theta'-\theta|<\varepsilon} \int_{U_\pi \cap W_1} \frac{1}{n} |P_n(y) Q_{\theta'}(y)|\, dy = 0. \tag{24}$$

See Fox and Taqqu [(1987), pages 236–237]. To establish (23) and (24), we recall that under (10) and (11),

$$
\begin{aligned}
(25)\qquad &\sup_{|\theta'-\theta|<\varepsilon;|y_i|\geq t>0} |Q_{\theta'}(y)| \\
&\quad \leq M_\theta^{2p}|y_1|^{-\alpha(\theta)-\delta}|y_2|^{-\beta(\theta)-\delta}\cdots|y_{2p}|^{-\beta(\theta)-\delta} \leq M'_{t,\theta},
\end{aligned}
$$

say. From Fox and Taqqu [(1987), page 237],

$$
|P_n(y)| \leq 4^{2p} h_n(y_1 - y_{2p}) h_n(y_2 - y_1) \cdots h_n(y_{2p} - y_1),
$$

where

$$
h_n(z) = \begin{cases} \min(|z+2\pi|^{-1}, n), & -2\pi \leq z \leq -\pi, \\ \min(|z|^{-1}, n), & -\pi \leq z \leq \pi, \\ \min(|z-2\pi|^{-1}, n), & \pi \leq z \leq 2\pi. \end{cases}
$$

Therefore, the left-hand side of (23) is less than or equal to

$$
\begin{aligned}
&(4M_\theta)^{2p} \lim_{t\to 0} \limsup_{n\to\infty} \int_{U_t} \frac{1}{n} h_n(y_1 - y_{2p}) h_n(y_2 - y_1) \cdots h_n(y_{2p} - y_1) \\
(26)\qquad &\qquad\qquad \times |y_1|^{-\alpha(\theta)-\delta}|y_2|^{-\beta(\theta)-\delta}\cdots|y_{2p}|^{-\beta(\theta)-\delta}\,dy_1\cdots dy_{2p} \\
&\quad = (4M_\theta)^{2p} \lim_{t\to 0} \limsup_{n\to\infty} \int_{U_t} \frac{1}{n} f_n(y)\,dy,
\end{aligned}
$$

say. Note that the last expression does not depend on θ'. Condition (26) is verified by an immediate application of Proposition 6.2.a of Fox and Taqqu [(1987), page 227]. Similarly, condition (24) is verified upon an application of Proposition 6.1.a of Fox and Taqqu [(1987), page 226], and noting that their set V is equal to W_1.

We proceed to show (20). Let $A_t = \{y_1 : t \leq |y_1| \leq \pi\}$, $t > 0$. It is easily seen that

$$
\frac{1}{(2\pi)^{2p-1} n} \int_{[-\pi,\pi]^{2p-1}} P_n(y)\,dy_2\cdots dy_{2p} = 1.
$$

Thus,

$$
\begin{aligned}
&(2\pi)^{2p-1} \int_{A_t} \left[\prod_{j=1}^{p} f_{\theta',j}(y_1) g_{\theta',j}(y_1) \right] dy_1 \\
&\quad = \frac{1}{n} \int_{U_\pi} I_{A_t} \left[\prod_{j=1}^{p} f_{\theta',j}(y_1) g_{\theta',j}(y_1) \right] P_n(y)\,dy \\
&\quad = \frac{1}{n} \int_{U_\pi} I_{A_t} Q_{\theta'}(y_1, \ldots, y_1) P_n(y)\,dy.
\end{aligned}
$$

The main term in (20) may thus be rewritten as

$$\Delta_n(\theta') = \frac{1}{n}\left| \int_{U_\pi} I_{E_t} P_n(y) Q_{\theta'}(y)\, dy - \int_{U_\pi} I_{A_t} P_n(y) Q_{\theta'}(y_1, \ldots, y_1)\, dy \right|. \tag{27}$$

Let $\eta_n = (2p-1)n^{-\beta}$ with $\beta \in (0,1)$, and let

$$V_j = \{y \in U_\pi : |y_j - y_{j-1}| \le n^{-\beta}\}, \qquad j = 2, \ldots, 2p, \qquad V = \bigcap_{j=2}^{2p} V_j. \tag{28}$$

We can re-express (27) as

$$\begin{aligned} \Delta_n(\theta') &= \frac{1}{n}\bigg| \int_{U_\pi} (I_{E_t \cap V \cap A_t} + I_{E_t \cap V \cap A_t^c} + I_{E_t \cap V^c}) P_n(y) Q_{\theta'}(y)\, dy \\ &\qquad - \int_{U_\pi} (I_{E_t \cap V \cap A_t} + I_{(E_t \cap V)^c \cap A_t}) P_n(y) Q_{\theta'}(y_1, \ldots, y_1)\, dy \bigg| \\ &\le \frac{1}{n}\bigg\{ \int_{U_\pi} I_{E_t \cap V \cap A_t} |Q_{\theta'}(y) - Q_{\theta'}(y_1, \ldots, y_1)| |P_n(y)|\, dy \\ &\qquad + \int_{U_\pi} I_{E_t \cap V \cap A_t^c} |Q_{\theta'}(y)| |P_n(y)|\, dy \\ &\qquad + \int_{U_\pi} I_{E_t \cap V^c} |Q_{\theta'}(y)| |P_n(y)|\, dy \\ &\qquad + \int_{U_\pi} I_{A_t \cap W} |Q_{\theta'}(y_1, \ldots, y_1)| |P_n(y)|\, dy \\ &\qquad\qquad + \int_{U_\pi} I_{A_t \cap V^c} |Q_{\theta'}(y_1, \ldots, y_1)| |P_n(y)|\, dy \bigg\}, \end{aligned} \tag{29}$$

where $A_t^c = \{y_1 : |y_1| \le t\}$ and we have used the facts $(E_t \cap V)^c \cap A_t = (A_t \cap E_t^c) \cup (A_t \cap V^c)$ and $(A_t \cap E_t^c) = (A_t \cap W)$. The first term of (29) involves

$$|Q_{\theta'}(y) - Q_{\theta'}(y_1, \ldots, y_1)| = \left| \sum_{j=2}^{2p} \frac{\partial Q_{\theta'}(y)}{\partial y_j}\bigg|_{y=(y_1, c_2, \ldots, c_{2p})} (y_j - y_1) \right|,$$

where $|c_j - y_1| \le |y_1 - y_j| \le \sum_{k=2}^{j} |y_k - y_{k-1}|\ \forall j,\ j = 2, \ldots, 2p$. On V, $|y_j - y_{j-1}| \le n^{-\beta}\ \forall j,\ j = 2, \ldots, 2p$. We thus have

$$\sup_{|\theta' - \theta| < \varepsilon; |y| > t} |Q_{\theta'}(y) - Q_{\theta'}(y_1, \ldots, y_1)| \le (2p-1)\bar{M}_{t,\theta}\eta_n,$$

where $\bar{M}_{t,\theta} \ge \sum_{j=2}^{2p} |\partial Q_{\theta'}(y)/\partial y_j|$. So, the first integral in (29) is bounded by

$$(2p-1)\bar{M}_{t,\theta}\frac{\eta_n}{n} \int_{U_\pi} I_{E_t \cap V \cap A_t} |P_n(y)|\, dy. \tag{30}$$

Further, if we rearrange (19) as

$$P_n(y) = \left(\sum_{j_1=0}^{n-1} e^{ij_1(y_1 - y_{2p})} \right) \left(\sum_{j_2=0}^{n-1} e^{ij_2(y_2 - y_1)} \right) \cdots \left(\sum_{j_{2p}=0}^{n-1} e^{ij_{2p}(y_{2p} - y_{2p-1})} \right),$$

then it follows that

$$|P_n(y)| = \left\{ \left[\prod_{j=2}^{2p} \left| \frac{\sin n(y_j - y_{j-1})/2}{\sin(y_j - y_{j-1})/2} \right| \right] \left| \frac{\sin n(y_1 - y_{2p})/2}{\sin(y_1 - y_{2p})/2} \right| \right\}.$$

Transforming: $z_1 = y_1/2$ and $z_j = (y_j - y_{j-1})/2$, $j = 2, \ldots, 2p$, (30) becomes

$$(31)\quad (2p-1)\bar{M}_{t,\theta} \frac{2^{2p}\eta_n}{n} \int_{\tilde{U}_\pi} I_{E_t \cap V \cap A_t} \left[\prod_{j=2}^{2p} \left| \frac{\sin nz_j}{\sin z_j} \right| \right] \left| \frac{\sin n(z_2 + \cdots + z_{2p})}{\sin(z_2 + \cdots + z_{2p})} \right| dz,$$

where $\tilde{U}_\pi = \{z \in \mathbb{R}^{2p} : |\sum_{j=1}^{i} z_j| \le \pi/2, i = 1, \ldots, 2p\}$. Since $\frac{|\sin nx|}{|\sin x|} \le n$, (31) is less than or equal to

$$(32)\qquad 2^{6p}(2p-1)\bar{M}_{t,\theta}\pi\eta_n \left(\int_0^{\pi/2} \left| \frac{\sin nz}{\sin z} \right| dz \right)^{2p-1} \le K_\theta \eta_n (c + \log n)^{2p-1},$$

where K_θ and c are constants not depending on n. Consequently, (32) tends to 0 as $n \to \infty$.

On the set $E_t \cap V \cap A_t^c$, $|y_1| \in (t - \eta_n, t)$. So, the second term in (29) is

$$\begin{aligned} &\frac{1}{n} \int_{U_\pi} I_{E_t \cap V \cap A_t^c} |Q_{\theta'}(y)||P_n(y)|\, dy \\ &\qquad \le K M'_{t,\theta} \int_{\tilde{U}_\pi} I_{E_t \cap V \cap A_t^c} \prod_{j=2}^{2p} \left| \frac{\sin nz_j}{\sin z_j} \right| dz \\ &\qquad \le 2K M'_{t,\theta} (c + \log n)^{2p-1} \int_{t-\eta_n}^{t} dz_1, \end{aligned}$$

where $M'_{t,\theta}$ is defined in (25), and so this term tends to 0 as $n \to \infty$.

The third and fifth terms in (29) are bounded by

$$(33)\qquad \frac{1}{n} M'_{t,\theta} \int_{V^c} |P_n(y)|\, dy.$$

We shall prove in Lemma 1 that (33) tends to zero. It remains to deal with the fourth term in (29). Using (25) and using the common bound $M'_{t,\theta}$ of $Q_{\theta'}(y_1, \ldots, y_1)$ on A_t and of $Q_\theta(y)$ on E_t,

$$\frac{1}{n} \int_{A_t \cap W} |Q_{\theta'}(y_1, \ldots, y_1)||P_n(y)|\, dy \le \frac{4^{2p} M'_{t,\theta}}{n} \int_{U_\pi} I_{A_t \cap W} f_n(y)\, dy.$$

So the fourth term of (29) is less than or equal to

$$\frac{(2p)4^{2p}M'_{t,\theta}}{n}\int_{U_\pi\cap W_1} f_n(y)\,dy,$$

for which, Proposition 6.1.a of Fox and Taqqu [(1987), page 226] can be readily applied. We have thus completed the proof of Theorem 1.

LEMMA 1. *For all $\beta \in (0,1)$,*

$$\lim_{n\to\infty}\frac{1}{n}\int_{U_\pi} I_{V^c}|P_n(y)|\,dy = 0, \tag{34}$$

where V is defined by (28).

PROOF. Note that

$$\frac{1}{n}P_n(y) = \frac{1}{n}\left\{\prod_{j=2}^{2p}\frac{\sin(n(y_j - y_{j-1})/2)}{\sin((y_j - y_{j-1})/2)}\right\}\frac{\sin(n(y_1 - y_{2p})/2)}{\sin((y_1 - y_{2p})/2)}. \tag{35}$$

For $p = 1$, $(2\pi n)^{-1}P_n(y) = F_n(y_1 - y_2)$, the Fejér kernel of order n, and for $p > 1$, $(2\pi n)^{-1}P_n(y)$ is a generalization of it. The Fejér kernel behaves asymptotically as a delta-Dirac function. The lemma essentially states that the contribution to the integral in (34) outside the neighborhood of the diagonal $\{y \in U_\pi : y_1 = y_2 = \cdots = y_{2p}\}$ is negligible.

Transform

$$x_1 = y_1,$$
$$x_j = y_j - y_{j-1}, \qquad j = 2, \ldots, 2p,$$

and note that the Jacobian of the transformation is unity. Since $V^c = \bigcup_{j=2}^{2p} V_j^c$, (34) is less than or equal to

$$\frac{1}{n}\sum_{j=2}^{2p}\int_{\tilde{U}_\pi} I_{V_j^c(x)}\left\{\prod_{j=2}^{2p}\left|\frac{\sin(nx_j/2)}{\sin(x_j/2)}\right|\right\}\left|\frac{\sin(n\sum_2^{2p}x_j/2)}{\sin(\sum_2^{2p}x_j/2)}\right|dx, \tag{36}$$

where $\tilde{U}_\pi$ is a bounded set satisfying

$$\int_{\tilde{U}_\pi}|P_n(y(x))|\,dx = \int_{U_\pi}|P_n(y)|\,dy$$

and

$$V_j^c(x) = \{x \in \tilde{U}_\pi : |x_j| > n^{-\beta}\}.$$

Tedious albeit straightforward calculations show that (36) is $O(n^{-\beta}\log n)$, which tends to zero $\forall\,\beta \in (0,1)$. $\square$

7.2. *Proof of Theorem* 2. We trace quickly through Dahlhaus' proof of his Theorem 5.1, using his equation numbers. Consider first Dahlhaus' Lemma 5.2 for f_θ, $\theta \in \Theta^*$. The key relations in the proof are (10), which does not involve f_θ, and (11). Under our Assumptions III, IV(a) and V [i.e., the uniform versions of Dahlhaus' Assumptions (A2) and (A7)], (11) holds with a uniform bounding constant K. Next, considering Dahlhaus' Lemma 5.3 for a family of function pairs f, g that obey the order bound conditions of the lemma with uniform bounding constants, it is obvious from Dahlhaus' proof that the conclusion of the lemma holds with a uniform order bound. The remainder of Dahlhaus' Theorem 5.1 uses only these two lemmas, Fox and Taqqu's (1987) Theorem 1.a, and some general linear algebra and analysis. The two lemmas are applied to functions which, under our assumptions, obey the relevant order bound conditions with uniform bounding constants, so that the conclusions of the lemmas hold with uniform bounds as just argued. Our Theorem 1 gives the uniform version of the Fox–Taqqu theorem.

7.3. *Proof of Theorem* 3.

7.3.1. *Preliminaries.* We prove part (a) by verifying the assumptions in Durbin's (1980) general Theorem 1 on validity of the Edgeworth expansion in a multivariate dependent data setting.

Durbin's theorem presupposes the following background assumptions:

(i) $D(\theta) = \lim_{n\to\infty} D_n(\theta)$ exists for all θ.
(ii) $D_n(\theta)$ and $D(\theta)$ are positive definite.
(iii) $D(\theta) \to D(\theta_0)$ as $\theta \to \theta_0$.

Incidentally, condition (iii) is needed to make use of BG's argument in the proof of our Theorem 4; see their Assumption (A_4).

Condition (i) holds for $\theta \in \Theta^*$ by virtue of Dahlhaus' theorem, as indicated in Section 2.3. In regard to condition (ii), positive definiteness of $D(\theta)$ is postulated in our Assumption VIII, and positive definiteness of D_n for sufficiently large n follows from the fact that the eigenvalues of a matrix are continuous functions of the elements of the matrix [Horn and Johnson (1985), Appendix D]. Condition (iii) is a consequence of the continuous differentiability conditions of our Assumption II together with Assumption V, which allows limits over θ to be passed under the integral sign in the expression (8) for D.

The remaining assumptions to be verified are Assumptions 2–4 stated on page 324 of Durbin's paper. We prove Assumption 3 in a modified form that is shown in Section 7.3.3 to be sufficient. Let $\varphi_n(\omega,\theta) = E_\theta[\exp(i\omega' Z_n)]$ be the characteristic function of Z_n. Assumptions 2 and 3 are stated in terms of $\varphi_n(\omega,\theta)$.

7.3.2. *Verification of Durbin's assumptions.*

DURBIN'S ASSUMPTION 2. If n is large enough, $|\varphi_n(\omega, \theta)|$ is integrable over $\mathbb{R}^d$ and

$$\int_{\{\|\omega\| > \delta_1 \sqrt{n}\}} |\varphi_n(\omega/\sqrt{n}, \theta)| d\omega = o(n^{-(\tau-2)/2})$$

for all $\delta_1 > 0$ uniformly for $\theta \in \Theta^*$, for any compact subset Θ^* of Θ.

PROOF. From standard theory on quadratic forms in Gaussian variables [Anderson (1984) and Johnson, Kotz and Balakrishnan (1997)] we obtain

$$\begin{aligned} \varphi_n(\omega/\sqrt{n}, \theta) &= E_\theta \exp\left\{ \frac{i}{\sqrt{n}} \sum_{j=1}^{d} \omega_j (A_j + x' B_{\nu(j)} x) \right\} \\ &= \exp\left(\frac{i}{\sqrt{n}} \sum_j \omega_j A_j \right) \det\left[I_n - \frac{2i}{\sqrt{n}} \sum_{j=1}^{d} \omega_j \tilde{B}_j \right]^{-1/2}, \end{aligned} \tag{37}$$

where A_j is a derivative of $-\frac{1}{2} \log\det T_n(f_\theta)$ of order less than or equal to $s-1$.

Let $\rho_1, \dots, \rho_n$ be the eigenvalues of $\sum_{j=1}^{d} \omega_j \tilde{B}_j$. Then

$$\begin{aligned} &|\varphi_n(\omega/\sqrt{n}, \theta)| \\ &\quad = \prod_{r=1}^{n} (1 + 4\rho_r^2/n)^{-1/4} \\ &\quad = \left\{ 1 + \frac{4}{n} \sum_{r=1}^{n} \rho_r^2 + \left(\frac{4}{n}\right)^2 \sum_{r \neq r'} \rho_r^2 \rho_{r'}^2 + \cdots + \left(\frac{4}{n}\right)^n \prod_{r=1}^{n} \rho_r^2 \right\}^{-1/4}. \end{aligned} \tag{38}$$

Recalling that $D_n(\theta)(i, j) = 2\,\mathrm{tr}(\tilde{B}_i \tilde{B}_j)/n$, we see that

$$\frac{4}{n} \sum_{r=1}^{n} \rho_r^2 = \frac{4}{n} \mathrm{tr}\left(\sum_{j=1}^{d} \omega_j \tilde{B}_j \right)^2 = 2\omega' D_n(\theta) \omega. \tag{39}$$

Further, the kth order term in (38) is given by

$$\left(\frac{4}{n}\right)^k \sum_{\mathcal{I}} \rho_{r_1}^2 \cdots \rho_{r_k}^2 = \left(\frac{4}{n}\right)^k \left(\sum_{r=1}^{n} \rho_r^2 \right)^k - \left(\frac{4}{n}\right)^k \sum_{\mathcal{J}} \rho_{r_1}^2 \cdots \rho_{r_k}^2,$$

where $\mathcal{I}$ denotes the set of indices such that no two indices are equal and $\mathcal{J}$ denotes the set of indices with at least two indices equal. The first term on the right-hand side, from (39), equals $(2\omega' D_n(\theta)\omega)^k$. The second term is bounded by

$$\left(\frac{4}{n}\right)^k \frac{k(k-1)}{2} \left(\sum_{r=1}^{n} \rho_r^4 \right) \left(\sum_{r=1}^{n} \rho_r^2 \right)^{k-2}.$$

Now $\sum \rho_r^4 = \operatorname{tr}(\sum_{j=1}^d \omega_j \tilde{B}_j)^4$ which, by Theorem 2, is no more than $O(n\|\omega\|^4)$, with the bound uniform in θ. We thus find that the kth order term in (38) equals $(2\omega' D_n(\theta)\omega)^k$ minus a term of magnitude no more than $O(n^{-1}\|\omega\|^{2k})$.

Now let η be half the smallest eigenvalue of $D(\theta)$ over $\theta \in \Theta^*$, and let $k \geq 2(d+\tau-2)$ be a fixed integer. Then for large enough n,

$$\begin{aligned}\int_{\{\|\omega\|>\delta_1\sqrt{n}\}} |\varphi_n(\omega/\sqrt{n},\theta)|\,d\omega &\leq \left(\frac{1}{2n\eta}\right)^{k/4} \int_{\{\|\omega\|>\delta_1\sqrt{n}\}} (\|\omega\|^2)^{-k/4}\,d\omega \\ &= n^{-(k/4)+(d/2)}(2\eta)^{-k/4} \int_{\{\|\bar{\omega}\|>\delta_1\}} \|\bar{\omega}\|^{-k/2}\,d\bar{\omega},\end{aligned}$$

and the integral on the right-hand side is finite. The desired conclusion follows. □

DURBIN'S ASSUMPTION 3 (Modified version). (a) For any $r = (r_1,\ldots,r_d)$, with $|r| \leq \tau$, the derivatives $\partial^{|r|}\log\varphi_n(\omega;\theta)/\partial\omega^r$ exist for ω in a neighborhood of the origin.

(b) For r as above, the quantity $n^{-1}\partial^{|r|}\log\varphi_n(0;\theta)/\partial\omega^r$ has a limit as $n\to\infty$.

(c) For any vector ξ with $\|\xi\|=1$, the quantity $n^{-1}d^\tau\varphi_n(\mu\xi;\theta)/d\mu^\tau$ has a limit as $n\to\infty$ and $\mu\to 0$, with convergence uniform over ξ.

PROOF. (a) From (37),

$$\log\varphi_n(\omega;\theta) = i\sum_{j=1}^d \omega_j A_j - \tfrac{1}{2}\log\det\left(I_n - 2i\sum_{j=1}^d \omega_j \tilde{B}_j\right). \tag{40}$$

The second term is the log of a polynomial function of ω and is therefore infinitely differentiable in ω provided the determinant does not vanish. The proof of Assumption 2 shows that this proviso holds.

(b) We focus on the second term since the first term is easily dealt with. Derivatives of the second term times $2/n$ take the form (up to sign)

$$n^{-1}\operatorname{tr}\left(\Omega^{-1}\tilde{B}_{j_1}\Omega^{-1}\tilde{B}_{j_2}\cdots\Omega^{-1}\tilde{B}_{j_m}\right)$$

with $\Omega = I - 2i\sum_{j=1}^d \omega_j\tilde{B}_j$. Evaluation at $\omega = 0$ yields $n^{-1}\operatorname{tr}(\tilde{B}_{j_1}\tilde{B}_{j_2}\cdots\tilde{B}_{j_m})$, which has a limit as $n\to\infty$ by virtue of Theorem 2.

(c) We again focus on the second term. For given ξ, the second term evaluated at $\omega = \mu\xi$ may be written as $\frac{1}{2}\log\det(I-\mu G)$ where $G = 2i\sum_{j=1}^d \xi_j\tilde{B}_j$. For any given integer m, the mth derivative of this quantity with respect to μ, multiplied by $-2/n$, is given by $n^{-1}\operatorname{tr}(\Omega^{-1}G\cdots\Omega^{-1}G)$, where $\Omega = I-\mu G$. But Ω, and hence Ω^{-1}, commutes with G and so this quantity may be expressed as $n^{-1}\operatorname{tr}(G^m\Omega^{-m})$. Now

$$\frac{1}{n}\operatorname{tr}(G^m) = (2i)^m \sum_{j_1,\ldots,j_m=1}^d \xi_{j_1}\cdots\xi_{j_m}\left[\frac{1}{n}\operatorname{tr}\left(\tilde{B}_{j_1}\cdots\tilde{B}_{j_m}\right)\right].$$

The quantity in brackets has a limit as $n \to \infty$ by Theorem 2, and thus so does $n^{-1}\,\mathrm{tr}(G^m)$. Convergence is clearly uniform in ξ since there are only finitely many terms of the type in brackets and $|\xi_j| \le 1$ for all j (since $\|\xi\| = 1$).

Next, using the matrix norm relations given on page 1754 of Dahlhaus (1989), we obtain

$$\begin{aligned}
&\left|\frac{1}{n}\,\mathrm{tr}(G^m\Omega^{-m}) - \frac{1}{n}\,\mathrm{tr}(G^m)\right| \\
&\qquad = \left|\frac{1}{n}\,\mathrm{tr}\big(G^m(I-\Omega^{-m})\big)\right| \le \frac{1}{n}\,|G^m|\,|I-\Omega^{-m}| \\
&\qquad = \frac{1}{n}\,|G^m|\,|(\Omega^m - I)\Omega^{-m}| \le \frac{1}{n}\,|G^m|\,|\Omega^m - I|\,\|\Omega^{-1}\|^m,
\end{aligned}$$

where $|Q|$ denotes the Euclidean norm of Q ($|Q| = \mathrm{tr}(QQ^*)^{1/2}$) and $\|Q\|$ denotes the spectral norm of Q (the square root of the largest eigenvalue of Q^*Q), with Q^* denoting the conjugate transpose of Q.

By an argument similar to that already given for $\mathrm{tr}(G^m)$, we find that $|G^m|^2 = O(n)$ uniformly in ξ. As for $|\Omega^m - I|$, we have

$$|\Omega^m - I| \le \sum_{c=1}^{m} \binom{m}{c} |\mu|^c\,|G^c| \le M n^{1/2}\{(1+|\mu|)^m - 1\}$$

for some constant M, since, by previous arguments, $|G^c|^2 = O(n)$ uniformly in ξ.

Turning to $\|\Omega^{-1}\|$, recall that $\Omega = I - 2i\mu H$ with

$$H = \sum_{j=1}^{d} \xi_j \tilde{B}_j,$$

where the ξ_j are real numbers and the $\tilde{B}_j$ are real symmetric matrices. Let $\zeta_1, \dots, \zeta_n$ denote the eigenvalues of H, all of which are real. We have $\Omega^*\Omega = I + 4\mu^2 H^2$, which has eigenvalues $1 + 4\mu^2\zeta_c^2,\ c = 1, \dots, n$. It follows that

$$\|\Omega^{-1}\| = \max_{1\le c\le n} (1 + 4\mu^2\zeta_c^2)^{-1/2} \le 1.$$

We thus find that

$$\left|\frac{1}{n}\,\mathrm{tr}(G^m\Omega^{-m}) - \frac{1}{n}\,\mathrm{tr}(G^m)\right| \le K\{(1+|\mu|)^m - 1\}$$

for a suitable constant K independent of ξ. Clearly, the right-hand side converges to zero as $\mu \to 0$. The desired conclusion follows. The result holds uniformly in θ by virtue of the uniformity in Theorem 2. $\square$

DURBIN'S ASSUMPTION 4. The cumulants of L_ν are $O(n)$ uniformly in θ over Θ^*.

PROOF. In view of (40), the Jth cumulant of L_ν, $J > 1$, is proportional to $\operatorname{tr}(\prod_{j=1}^J \tilde{B}_j)$ and by Theorem 2 this term is $O(n)$ uniformly in θ. □

We have thus completed the proof of part (a) of the theorem. Part (b) is an immediate consequence of Skovgaard's (1986) Corollary 3.3.

7.3.3. *Justification that the modified Assumption* 3 *is sufficient.* We trace through Durbin's proof, using his notation, and with (Dxx) denoting equation number xx in Durbin's paper.

1. Modified Assumption 3a makes Durbin's $\psi_{nr}(z,\theta)$ well defined.
2. Up to (D33), no use is made of Assumption 3.
3. From the paragraph after (D33) up to (D35), Durbin is involved with bounding $|\alpha - \beta|$. Observe that the quantity $\chi_n^{(r)}(p,\theta)$ defined in the line below (D34) is precisely the quantity referred to in our Assumption 3c. Thus, the conclusion that there exist δ_1 and N_1 independent of z_0 such that for $0 < u \le p < \delta_1$ and $n > N_1$ we have $|\chi_n^{(r)}(u,\theta) - \chi_n^{(r)}(0,\theta)| < r!\,\varepsilon$ follows from our Assumption 3c. This takes us to (D36).
4. Durbin now proceeds to bound $|\beta|$. To get the conclusion $|\psi_{nr}(z;\theta)| < \delta_3 p^3$ (for $p < \delta_1, n > N_1$) what we need is for $\chi_n^{(j)}(0,\theta)$ to be bounded. This follows from our Assumption 3b. This takes us to (D37).
5. From (D37) to (D39) we have a straightforward progression building on previous results, with no further use of Assumption 3. The same is true of the remainder of the argument dealing with the first integral on the right-hand side of (D32).
6. The second integral on the right-hand side of (D32) is handled by Assumption 2 only.
7. The last displayed expression on page 327 is obtained as follows. The factor $\exp(-\frac{1}{2}z' D_n(\theta)z)$ is replaced by $\exp(-\frac{1}{2}p^2\{\lambda(\theta_0) - \frac{1}{2}\delta_4\})$ by virtue of (D39). In regard to $n\psi_{nr}(z/\sqrt{n},\theta)$, the argument is as follows: the line just below (D36) gives
$$\psi_{nr}(z,\theta) = \sum_{j=3}^{r} \chi_n^{(j)}(0,\theta)\frac{p^j}{j!}.$$
Replacing z by $z/\sqrt{n}$ and p by $p/\sqrt{n}$, we get
$$n\psi_{nr}(z/\sqrt{n},\theta) = \sum_{j=3}^{r} n^{-(j-2)/2}\chi_n^{(j)}(0,\theta)\frac{p^j}{j!}.$$
But we saw before that $\chi_n^{(j)}(0,\theta)$ are bounded, so the result follows.

8. We are left to deal with the last integral on the right-hand side of (D32). To handle this, Durbin needs the quantities $n^{-1}\partial^j \log\phi_n(0,\theta)/\partial z_j$ for $j = 3,\dots,r$ to be uniformly bounded in n. This follows from our Assumption 3b.

7.4. *Proof of Theorem* 4 (*Sketch*). BG's proof goes through largely without change; we focus on the points calling for special attention. Below Θ^{**} denotes a compact subset of Θ whose interior contains Θ^*. It follows from our Theorems 1 and 2 that all the error bounds mentioned below are uniform in θ.

In BG's proof of their Theorem 2, their Q_n corresponds to the joint distribution of our W_n. Application as in BG of the result of James (1955, 1958) and James and Mayne (1962) requires only the condition, identical to Durbin's Assumption 4 verified above, that the jth cumulant of our Y_n/n is $O(n^{-j+1})$. A new argument is needed to obtain an analogue of BG's (2.21), to the effect that the expected value of a polynomial function of W_n can be approximated adequately by the integral of the function in question with respect to the approximate density $\tilde{G}_n^{(\tau-2)}$ (in our notation). Let δ be an arbitrarily small positive number, and write

$$E_{\theta_0}[W_{ni}^k] = E_{\theta_0}[W_{ni}^k I(\|W_n\| \le n^{\delta/k})] + E_{\theta_0}[W_{ni}^k I(\|W_n\| > n^{\delta/k})]. \tag{41}$$

By a variation on Markov's inequality, the second term is bounded in absolute value by $E_{\theta_0}[\|W_n\|^{k+m}]n^{-\delta m/k}$. Because $Y_{ni} = \sqrt{n}\, W_{ni}$ has cumulants $O(n)$, the moment $E_{\theta_0}[\|W_n\|^{k+m}]$ is $O(1)$ and so the second term in (41) is $O(n^{-\delta m/k})$. Since m may be taken arbitrarily large, this term may be neglected. Our Theorem 3 implies that the first term in (41) is equal to

$$\int_{\|u\|\le n^{\delta/k}} u_i^k \tilde{G}_n^{(\tau-2)}(u,\theta_0)\,du + o(n^{-(\tau-2)/2+\delta}) \qquad \forall\,\delta > 0. \tag{42}$$

Through simple manipulations using the elementary inequality

$$\int_\zeta^\infty v^\ell \phi(v)\,dv \le M_\ell \max\{\zeta^{\ell-1},1\}e^{-\zeta^2/2},$$

where ϕ denotes the standard normal density, ℓ a nonnegative integer, and M_ℓ a positive constant, it may be seen that the restriction on the integral in (42) may be removed without affecting the error rate. This gives the desired result for a power of a component of W_n. A similar argument, with the aid of the Cauchy–Schwarz inequality, can be made to deal with a product of powers of several components. Since $\tilde{G}_n^{(\tau)}(u,\theta_0)$ is a valid asymptotic expansion to any order with steps of $O(n^{-1/2})$, we have obtained BG's (2.21) with the required $o(n^{-(s-2)/2})$ error rate.

BG's proof of their Theorem 3 is built around the Taylor expansion (16). We need an analogue of BG's formula (2.32), or Taniguchi's (1990) Equations (3.2.24)–(3.2.28), from which the rest of BG's proof carries through easily.

Let δ be a small positive constant. It will be sufficient to show that there is a constant d_1 such that

$$P_{\theta_0}\big(|n^{-1/2}l_\nu(\theta_0)| > d_1 n^\delta\big) = o(n^{-(s-2)/2}), \tag{43}$$

where $l_\nu(\theta_0)$ is a centered LLD, for any ν satisfying $1 \le |\nu| \le s+1$. The left-hand side of (43) is less than or equal to $(d_1 n^\delta)^{-2\eta} E_{\theta_0}(n^{-1/2} l_\nu(\theta_0))^{2\eta}$, where η is a positive integer. Since the LLDs have cumulants $O(n)$, $n^{-1/2} l_\nu(\theta_0)$ have moments $O(1)$. Setting η to satisfy $2\eta\delta > (s-2)/2$, we obtain the desired result.

REFERENCES

ANDERSON, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. Wiley, New York.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1989). *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London.

BARNDORFF-NIELSEN, O. E. and COX, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

BARNDORFF-NIELSEN, O. E. and HALL, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* **75** 374–378.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis*: *Forecasting and Control*, rev. ed. Holden-Day, San Francisco.

BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series*: *Theory and Methods*, 2nd ed. Springer, New York.

CHAMBERS, J. M. (1967). On methods of asymptotic approximation for multivariate distributions. *Biometrika* **54** 367–383.

CHIBISOV, D. M. (1973). An asymptotic expansion for a class of estimators containing maximum likelihood estimators. *Theory Probab. Appl.* **18** 295–303.

DAHLHAUS, R. (1988). Small sample effects in time series analysis: A new asymptotic theory and a new estimate. *Ann. Statist.* **16** 808–841.

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

DURBIN, J. (1980). Approximations for densities of sufficient estimators. *Biometrika* **67** 311–333.

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.

GÖTZE, F. and HIPP, C. (1994). Asymptotic distribution of statistics in time series. *Ann. Statist.* **22** 2062–2088.

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.

GRENANDER, U. and SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Univ. California Press, Berkeley. (2nd ed., Chelsea, New York, 1984.)

HORN, R. A. and JOHNSON, C. R. (1985). *Matrix Analysis.* Cambridge Univ. Press.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.

HURST, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engin.* **116** 770–808.

JAMES, G. S. (1955). Cumulants of a transformed variate. *Biometrika* **42** 529–531.

JAMES, G. S. (1958). On moments and cumulants of systems of statistics. *Sankyā* **20** 1–30.

JAMES, G. S. and MAYNE, A. J. (1962). Cumulants of functions of a random variable. *Sankyā Ser. A* **24** 47–54.

JOHNSON N. L, KOTZ, S. and BALAKRISHNAN, N. (1997). *Continuous Univariate Distributions* **2**, 2nd ed. Wiley, New York.

LAHIRI, S. N. (1993). Refinements in asymptotic expansions for sums of weakly dependent random vectors. *Ann. Probab.* **21** 791–799.

LIEBERMAN, O., ROUSSEAU, J. and ZUCKER, D. M. (2000). Small-sample likelihood-based inference in the ARFIMA model. *Econometric Theory* **16** 231–248.

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

PEERS, H. W. and IQBAL, M. (1985). Asymptotic expansions for confidence limits in the presence of nuisance parameters, with applications. *J. Roy. Statist. Soc. Ser. B* **47** 547–554.

ROBINSON, P. M. (1995). Time series with strong dependence. In *Advances in Econometrics*. Sixth World Congress (C. A. Sims, ed.) **1** 47–95. Cambridge Univ. Press.

ROCKE, D. M. (1989). Bootstrap Bartlett adjustment in seemingly unrelated regression. *J. Amer. Statist. Assoc.* **84** 598–601.

SKOVGAARD, I. M. (1981). Edgeworth expansions of the distributions of maximum likelihood estimators in the general (non i.i.d.) case. *Scand. J. Statist.* **8** 227–236.

SKOVGAARD, I. M. (1986). On multivariate Edgeworth expansions. *Internat. Statist. Rev.* **54** 169–186.

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

TANIGUCHI, M. (1984). Validity of Edgeworth expansions for statistics of time series. *J. Time Ser. Anal.* **5** 37–51.

TANIGUCHI, M. (1986). Third order asymptotic properties of maximum likelihood estimators for Gaussian ARMA processes. *J. Multivariate Anal.* **18** 1–31.

TANIGUCHI, M. (1988). Asymptotic expansions of the distributions of some test statistics for Gaussian ARMA processes. *J. Multivariate Anal.* **27** 494–511.

TANIGUCHI, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis. Lecture Notes in Statist.* **68**. Springer, Berlin.

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

O. LIEBERMAN
FACULTY OF INDUSTRIAL
ENGINEERING MANAGEMENT
TECHNION
HAIFA 32000
ISRAEL
E-MAIL: offerl@ie.technion.ac.il

J. ROUSSEAU
LABORATOIRE DE STATISTIQUES
UFR BIOMEDICALE
UNIVERSITÉ DE PARIS 5
75006 PARIS CEDEX
FRANCE

D. M. ZUCKER
DEPARTMENT OF STATISTICS
HEBREW UNIVERSITY
OF JERUSALEM
JERUSALEM 91905
ISRAEL