

**Valid Edgeworth Expansions for the Whittle Maximum Likelihood
Estimator for Stationary Long-memory Gaussian Time Series**

By

Donald W.K. Andrews and Offer Lieberman

April 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1361



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

<http://cowles.econ.yale.edu/>

**Valid Edgeworth Expansions for the Whittle Maximum Likelihood
Estimator for Stationary Long-memory Gaussian Time Series**

DONALD W. K. ANDREWS¹

Cowles Foundation for Research in Economics

Yale University

OFFER LIEBERMAN

Technion and

Cowles Foundation for Research in Economics

Yale University

April 2, 2002

¹Address for correspondence: Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281. Email: donald.andrews@yale.edu

Short title: Expansions for the Whittle MLE

Key words and phrases: ARFIMA, Edgeworth expansion, long memory, Whittle estimator.

AMS 1991 subject classifications: Primary 62E17, 62M10.

Abstract

In this paper, we prove the validity of an Edgeworth expansion to the distribution of the Whittle maximum likelihood estimator for stationary long-memory Gaussian models with unknown parameter $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. The error of the $(s - 2)$ -order expansion is shown to be $o(n^{-(s-2)/2})$ —the usual iid rate—for a wide range of models, including the popular ARFIMA(p, d, q) models. The expansion is valid under mild assumptions on the behavior of spectral density and its derivatives in the neighborhood of the origin. As a by-product, we generalize a Theorem by Fox and Taqqu (1987) concerning the asymptotic behavior of Toeplitz matrices.

Lieberman, Rousseau, and Zucker (2002) (LRZ) establish a valid Edgeworth expansion for the maximum likelihood estimator for stationary long-memory Gaussian models. For a significant class of models, their expansion is shown to have an error of $o(n^{-1})$. The results given here improve upon those of LRZ in that the results provide an Edgeworth expansion for an asymptotically efficient estimator, as LRZ do, but the error of the expansion is shown to be $o(n^{-(s-2)/2})$, not $o(n^{-1})$, for a broad range of models.

1 Introduction

We consider a discrete-time stationary long-memory Gaussian process $\{X_t : t \in \mathbb{Z}\}$ with unknown mean μ and covariance matrix $T_n(f_\theta)$ for $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$. The spectral density of the process satisfies

$$f_\theta(\lambda) \sim A_\theta(\lambda) |\lambda|^{-\alpha(\theta)} \text{ as } |\lambda| \rightarrow 0, \quad (1)$$

where the long-memory parameter $\alpha(\theta)$ is in $(0, 1)$ and $A_\theta(\lambda)$ is slowly varying at the origin. The main feature of (1) is that $f_\theta(\lambda)$ is unbounded at the origin and the autocovariances based on $f_\theta(\lambda)$ are not summable. A popular model that satisfies (1) is the ARFIMA(p, d, q) model for which $d = \alpha(\theta)/2$.

Models that satisfy (1) have been of interest since the early 1950's in a variety of fields, including mathematical statistics, probability, economics, finance, and hydrology. For some key references the reader is referred to Hurst (1951), Mandelbrot and Van Ness (1968), Granger and Joyeux (1980), Hosking (1981), Beran (1994), and Robinson (1995).

A number of estimators of θ are available including the maximum likelihood estimator (MLE) and the Whittle MLE (WMLE). Dahlhaus (1989) establishes consistency, asymptotic efficiency, and asymptotic normality of a plug-in version of the MLE, which we refer to as the PMLE. The PMLE is the maximizer of

$$L_n(\theta, \tilde{\mu}_n) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \det T_n(f_\theta) - \frac{1}{2} (x_n - \tilde{\mu}_n \mathbf{1}_n)' T_n^{-1}(f_\theta) (x_n - \tilde{\mu}_n \mathbf{1}_n), \quad (2)$$

where $\tilde{\mu}_n$ is an $n^{(1-\alpha(\theta))/2}$ -consistent estimator of μ , $x_n = (X_1, \dots, X_n)'$, and $\mathbf{1}_n$ is a column n -vector of ones. The unusual $n^{(1-\alpha(\theta))/2}$ -rate for $\tilde{\mu}_n$ is a consequence of the long-memory property of the process.

Fox and Taquq (1986), hereafter FT86, establish consistency and asymptotic normality of the WMLE. The WMLE is the minimizer of

$$L_n^W(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \{\log f_\theta(\lambda) + f_\theta^{-1}(\lambda) I_n(\lambda)\} d\lambda, \quad (3)$$

where

$$I_n(\lambda) = \frac{1}{2\pi n} \left| \sum_{j=1}^n e^{ij\lambda} (X_j - \bar{X}) \right|^2 \text{ and } \bar{X} = \sum_{j=1}^n X_j.$$

The WMLE and PMLE have the same asymptotic distribution and, hence, the WMLE also is asymptotically efficient. The WMLE, however, has some computational advantages. It does not require the computation of the inverse and determinant of the $n \times n$ covariance matrix $T_n(f_\theta)$.

Recently, Lieberman, Rousseau, and Zucker (2002), henceforth LRZ, proved the validity of the formal Edgeworth expansion to the distribution of the MLE for the parameters of a zero mean, Gaussian long-memory process with spectral density satisfying (1). Andrews and Lieberman (2002) extend their results to the PMLE for the case of unknown mean. For some models, the error of the $(s - 2)$ -order expansion is $o(n^{-(s-2)/2})$. But, for other models, including many ARFIMA(p, d, q) models, the error is shown to be valid only to order $o(n^{-1})$. The reason is that the asymptotic covariance matrix of the log-likelihood derivatives (LLD's) is singular, but the finite sample covariance matrix of the LLD's is not. The Edgeworth expansion for the MLE relies on an Edgeworth expansion for the LLD's and the latter typically requires the asymptotic covariance matrix of the LLD's to be nonsingular. When one discards LLD's to obtain nonsingularity of the asymptotic covariance matrix of the LLD's, it affects the error of the expansion.

In this paper, we prove the validity of an Edgeworth expansion to the distribution of the WMLE for stationary Gaussian processes that satisfy (1). We are able to prove validity of the $(s - 2)$ -order expansion for the WMLE with error $o(n^{-(s-2)/2})$ for a much wider range of models than LRZ do for the MLE. The models covered include the widely-used ARFIMA(p, d, q) models. The generality of the results is possible because the finite sample covariance matrix of the Whittle log-likelihood derivatives (WLD's) is singular whenever its asymptotic covariance matrix is singular. In consequence, when the asymptotic covariance matrix of the WLD's is singular, WLD's that are redundant asymptotically are also redundant in finite samples and one can discard them without affecting the error of the Edgeworth expansion for the WMLE.

The assumptions employed in this paper mainly control the behavior of the spectral density and its derivatives in a neighborhood of the origin. The assumptions are a hybrid of the assumptions of FT86 for the first-order theory for the WMLE and the assumptions of Bhattacharya and Ghosh (1978) for the higher-order theory for the MLE in an iid context. The assumptions differ from those of FT86 primarily in the order of partial derivatives that are assumed to exist. The assumptions are similar to those used in LRZ.

The results of this paper are useful for establishing higher-order improvements of the parametric bootstrap based on the WMLE, see Andrews and Lieberman (2002). The computational advantages of the WMLE over the MLE make the WMLE bootstrap an attractive procedure. In addition, the generality of the results given here allow one to establish more general higher-order improvements for the WMLE-based bootstrap than for the MLE-based bootstrap.

The method of proof used in this paper is outlined briefly as follows. First, we establish validity of an Edgeworth expansion for the WLD's using a general result of Durbin (1980, Thm. 1). A key requirement of Durbin's Theorem concerns the behavior of the cumulants of the WLD's. It is established by generalizing a result of Fox and Taquq (1987, Thm. 1(a)), hereafter FT87, on the properties of the trace of a product of Toeplitz matrices.. Other assumptions of Durbin are verified using the proof of LRZ. Second, we use the argument of Bhattacharya and Ghosh (1978), in which the normalized WMLE is approximated by a function of WLD's, to obtain the desired Edgeworth expansion of the WMLE from that of the WLD's.

Theorem 1(a) of FT87 deals with the asymptotic behavior of $\Pi_n = tr[(T_n(f)T_n(g))^p]$, where $T_n(f)$ and $T_n(g)$ are $n \times n$ Toeplitz matrices and f and g satisfy (1) with exponents $\alpha < 1$ and $\beta < 1$, respectively, in place of $\alpha(\theta)$. We denote the exponent structure of Π_n by $E = \{\alpha, \beta, \dots, \alpha, \beta\}$. In this paper, we need to control the behavior of a more complicated Toeplitz matrix product with a non-homogeneous exponent structure of the form $\tilde{E} = \{b_1, \dots, b_{2p}\}$. Results for this more general case may be of interest in other applications. Interest in algebraic structures of this form originated in Grenander and Szegö's (1956) monograph and has generated considerable interest over the years. For instance, see Dahlhaus (1989) and Taniguchi and Kakizawa (2000).

The results of this paper follow a long tradition on valid asymptotic expansions. The literature started with models for iid data and has gradually expanded to cover models with more complicated dependence structures. In a seminal paper, Bhattacharya and Ghosh (1978) prove validity of the formal Edgeworth expansion to the distribution of the MLE in an iid setting. Taniguchi (1984, 1986, 1988, 1990) establishes a series of validity results applicable mainly to weakly dependent Gaussian ARMA processes. Götze and Hipp (1983, 1994) and Lahiri (1993) establish validity results for the sample mean for non-Gaussian weakly dependent processes. As mentioned above, LRZ and Andrews and Lieberman (2002) provide validity results for the MLE for stationary long-memory Gaussian processes.

The remainder of the paper is organized as follows. Section 2 states the assumptions. Section 3 provides bounds on the cumulants of the WLD's and an Edgeworth expansion for the WLD's. Section 4 gives the Edgeworth expansion for the WMLE. Section 5 discusses an ARFIMA example. Section 6 contains proofs.

2 Assumptions

In this section, we state the assumptions used in the paper and relate them to the assumptions of FT86 and LRZ.

Throughout, $s \geq 3$ denotes a positive integer associated with the order of an expansion. With conditions on derivatives up to order $s + 1$, we prove the validity of the $(s - 2)$ -th order formal Edgeworth expansion to the distribution of the WMLE with an error rate of $o(n^{-(s-2)/2})$.

Assumptions:

W1. Θ has a non-empty interior.

W2. $g(\theta) = \int_{-\pi}^{\pi} \log f_{\theta}(\lambda) d\lambda$ and $h(\theta) = \int_{-\pi}^{\pi} f_{\theta}^{-1}(\lambda) I_n(\lambda) d\lambda$ can be differentiated $s + 1$ times under the integral sign.

There exists $0 < \alpha(\theta) < 1$ such that for each $\delta > 0$:

W3. $f_{\theta}(\lambda)$ is continuous at all (λ, θ) for which $\lambda \neq 0$, $f_{\theta}^{-1}(\lambda)$ is continuous at all (λ, θ) , and $\exists c_1(\theta, \delta) < \infty$ such that

$$|f_{\theta}(\lambda)| \leq c_1(\theta, \delta) |\lambda|^{-\alpha(\theta)-\delta}$$

for all λ in a neighborhood N_{δ} of the origin.

W4. For all (j_1, \dots, j_k) with $k \leq s + 1$ and $j_i \in \{1, \dots, d_{\theta}\}$, $(\partial^k / (\partial \theta_{j_1} \dots \partial \theta_{j_k})) f_{\theta}^{-1}(\lambda)$ is continuous at all (λ, θ) and $\exists c_2(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial^k f_{\theta}^{-1}(\lambda)}{\partial \theta_{j_1} \dots \partial \theta_{j_k}} \right| \leq c_2(\theta, \delta) |\lambda|^{\alpha(\theta)-\delta}, \quad \forall \lambda \in N_{\delta}.$$

W5. $(\partial / \partial \lambda) f_{\theta}(\lambda)$ is continuous at all (λ, θ) for which $\lambda \neq 0$ and $\exists c_3(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial f_{\theta}(\lambda)}{\partial \lambda} \right| \leq c_3(\theta, \delta) |\lambda|^{-\alpha(\theta)-1-\delta}, \quad \forall \lambda \in N_{\delta}.$$

W6. For all (j_1, \dots, j_k) with $k \leq s + 1$ and $j_i \in \{1, \dots, d_{\theta}\}$, $(\partial^{k+1} / (\partial \lambda \partial \theta_{j_1} \dots \partial \theta_{j_k})) f_{\theta}^{-1}(\lambda)$ is continuous at all (λ, θ) for which $\lambda \neq 0$ and $\exists c_4(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial^{k+1} f_{\theta}^{-1}(\lambda)}{\partial \lambda \partial \theta_{j_1} \dots \partial \theta_{j_k}} \right| \leq c_4(\theta, \delta) |\lambda|^{\alpha(\theta)-1-\delta}, \quad \forall \lambda \in N_{\delta}.$$

W7. For any compact subset Θ^* of Θ there exists a constant $C(\Theta^*, \delta) < \infty$ such that the constants $c_i(\theta, \delta)$ for $i = 1, \dots, 4$ are bounded by $C(\Theta^*, \delta)$ for all $\theta \in \Theta^*$.

Assumption W1 is used because the WMLE is asymptotically normal only at points in the interior of Θ . Assumption W1 does not require Θ to be compact, as FT86 does, because we do not prove consistency of the WMLE here. Assumption W2 guarantees the existence of WLD's up to order $s + 1$. Assumption W2 extends Assumption A.1 of FT86 to $s + 1$ derivatives for both parts of equation (3). This assumption is used in place of Assumption VI(b) of LRZ. Assumption W3 characterizes the long-memory property of the process. It corresponds to Assumption A.2 of FT86 and Assumption IV(a) of LRZ. Assumptions W4-W6 restrict the partial derivatives with respect to λ and θ of the spectral density and its inverse for λ in a neighborhood of the origin. Assumptions W4 and W6 extend Assumptions A.3 and A.5 of FT86 to cover $s + 1$ derivatives. Assumption W5 is the same as Assumption A.4 of FT86. Assumption A.6 of FT86 is not used here because we use a different method of analyzing the impact of estimating the mean μ by \bar{X} than does FT86.

Assumption W7 bounds the constants that appear in the preceding assumptions over parameter values θ that lie in compact sets. This assumption is needed to handle the remainder that appears in the approximation of the WMLE by a function of WLD's. It is also needed to deliver Edgeworth expansions that are valid uniformly over certain compact sets Θ^* in Θ . Uniform results of this sort are required to establish the higher-order improvements of the parametric bootstrap based on the WMLE.

Assumptions W1-W7 are satisfied for Gaussian ARFIMA(p, d, q) models.

3 Properties of WLD's

In this section, we define the WLD's, specify the parameter values for which we can obtain Edgeworth expansions for WLD's and the WMLE, determine bounds on the magnitudes of the cumulants of the WLD's, and use these bounds to establish an Edgeworth expansion for the WLD's. This expansion is used in the section below to obtain an Edgeworth expansion for the WMLE.

3.1 Definition of WLD's

Let ν be a set of subscripts (r_1, \dots, r_q) , where r_j is in $\{1, \dots, d_\theta\}$ for all $j \leq q$. Let $D_\nu L_n^W(\theta)$ denote the q -th order WLD with respect to θ specified by ν , viz.,

$$D_\nu L_n^W(\theta) = \frac{\partial^q}{\partial \theta_{r_1} \dots \partial \theta_{r_q}} L_n^W(\theta). \quad (4)$$

In view of (3),

$$D_\nu L_n^W(\theta) = \frac{1}{4\pi} D_\nu \int_{-\pi}^{\pi} \log f_\theta(\lambda) d\lambda + \frac{1}{4\pi} D_\nu \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) I_n(\lambda) d\lambda. \quad (5)$$

The second summand in (5) is

$$\begin{aligned} & \frac{1}{4\pi} D_\nu \int_{-\pi}^{\pi} f_\theta^{-1}(\lambda) \left(\frac{1}{2\pi n} \sum_{j,k=1}^n e^{i(j-k)\lambda} (X_j - \bar{X})(X_k - \bar{X}) \right) d\lambda \\ &= \frac{1}{2n} \sum_{j,k=1}^n (X_j - \bar{X})(X_k - \bar{X}) D_\nu \int_{-\pi}^{\pi} \frac{1}{4\pi^2} f_\theta^{-1}(\lambda) e^{i(j-k)\lambda} d\lambda. \end{aligned} \quad (6)$$

The last integral is the (j, k) element of the Whittle approximation to the inverse of the covariance matrix $T_n(f_\theta)$. More specifically, for an integrable function h on $(-\pi, \pi)$, let $T_n(h)$ denote the $n \times n$ Toeplitz matrix with (j, k) element $\int_{-\pi}^{\pi} e^{i(j-k)\lambda} h(\lambda) d\lambda$. Then, the Whittle approximation to $T_n^{-1}(f_\theta)$ is

$$T_n \left(\frac{1}{4\pi^2} f_\theta^{-1} \right).$$

For simplicity, we write

$$g_{\theta, \nu}(\lambda) = \frac{1}{4\pi^2} D_\nu f_\theta^{-1}(\lambda).$$

Then, the right-hand side (rhs) of (6) is

$$\frac{1}{2n} x_n' M_n T_n(g_{\theta, \nu}) M_n x_n, \quad (7)$$

where $M_n = I_n - P_n$, $P_n = n^{-1} \mathbf{1}_n \mathbf{1}_n'$, and I_n is the identity matrix of order n . Since $M_n \mathbf{1}_n = 0$,

$$x_n' M_n T_n(g_{\theta, \nu}) M_n x_n = (x_n - \mathbf{1}_n \mu)' M_n T_n(g_{\theta, \nu}) M_n (x_n - \mathbf{1}_n \mu).$$

Hence, without loss of generality, we may assume that $\mu = 0$.

3.2 Parameter Values

We now specify the parameter values θ for which we establish Edgeworth expansions for WLD's and the WMLE below. Clearly, only parameter values that are in the interior of Θ and for which the asymptotic covariance matrix of the WMLE is nonsingular are candidates. For example, in an ARFIMA(p, d, q) model, parameter values θ for which there are common roots of the autoregressive and moving average characteristic equations are *not* candidates. Rather than excluding such parameter values from the parameter space Θ , which would be quite artificial, we allow the parameter space Θ

to include such values, but we exclude them from the set of parameter values for which we establish an Edgeworth expansion.

By Theorem 2.1 of Dahlhaus (1989) (also see Theorem 2 of FT86), the asymptotic covariance matrix of the WMLE (suitably normalized) is

$$\Sigma(\theta) = \left(\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \log f_{\theta}(\lambda) \frac{\partial}{\partial \theta'} \log f_{\theta}(\lambda) d\lambda \right)^{-1}.$$

By Dahlhaus (1989, Thm. 2.1 and Sec. 4), $\Sigma(\theta)$ is the inverse of the asymptotic information matrix.

Let $\bar{Z}_n(\theta)$ denote $2n$ times the vector of all WLD's, $D_{\nu} L_n^W(\theta)$, up to order $s-1$. Let $\bar{D}_n(\theta)$ denote the covariance matrix of $n^{-1/2} \bar{Z}_n(\theta)$ when the true parameter is θ . The (j, k) element of $\bar{D}_n(\theta)$ is

$$\bar{D}_n(\theta)_{j,k} = \frac{2}{n} \text{tr} \{ M_n T_n(g_{\theta, \nu_j}) M_n T_n(f_{\theta}) M_n T_n(g_{\theta, \nu_k}) M_n T_n(f_{\theta}) \} \quad (8)$$

(see (11) below). By Proposition 2 and Theorem 3 below, the asymptotic covariance matrix $\bar{D}(\theta)$ ($= \lim_{n \rightarrow \infty} \bar{D}_n(\theta)$) of $n^{-1/2} \bar{Z}_n(\theta)$ exists and its (j, k) element is

$$\bar{D}(\theta)_{j,k} = \frac{1}{\pi} \int_{-\pi}^{\pi} \{ D_{\nu_j} f_{\theta}^{-1}(\lambda) \} \{ D_{\nu_k} f_{\theta}^{-1}(\lambda) \} f_{\theta}^2(\lambda) d\lambda. \quad (9)$$

Given any sub-vector $Z_n(\theta)$ of $\bar{Z}_n(\theta)$, let $D_n(\theta)$ and $D(\theta)$ denote the finite-sample and asymptotic covariance matrices of $n^{-1/2} Z_n(\theta)$, respectively, when the true parameter is θ .

We establish Edgeworth expansions for WLD's and the WMLE that hold uniformly over compact sets that lie in any set $\tilde{\Theta} \subset \Theta$ that satisfies the following ‘‘nonsingularity’’ condition:

Condition NS. (i) $\tilde{\Theta}$ lies in the interior of Θ .

(ii) $\Sigma(\theta)$ is nonsingular for all $\theta \in \tilde{\Theta}$.

(iii) For some sub-vector $Z_n(\theta)$ of $\bar{Z}_n(\theta)$, the asymptotic covariance matrix $D(\theta)$ of $n^{-1/2} Z_n(\theta)$ is nonsingular for all $\theta \in \tilde{\Theta}$ and the asymptotic covariance matrix of any sub-vector of $n^{-1/2} \bar{Z}_n(\theta)$ that strictly contains $n^{-1/2} Z_n(\theta)$ is singular for all $\theta \in \tilde{\Theta}$.

Note that every parameter θ_1 that is in the interior of Θ and for which $\Sigma(\theta_1)$ is nonsingular is in some set $\tilde{\Theta}$ that satisfies Condition NS. This follows because, given θ_1 , there is a sub-vector of $\bar{Z}_n(\theta)$, call it $\bar{Z}_{\theta_1, n}(\theta)$, such that Condition NS(iii) holds for $\theta \in \{\theta_1\}$. Hence, the set $\{\theta_1\}$ is an example of a set $\tilde{\Theta}$ that includes θ_1 and satisfies Condition NS.

The first condition of Condition NS(iii) is utilized because the vector of WLD's $n^{-1/2}Z_n(\theta)$ needs to have a nonsingular covariance matrix to apply Theorem 1 of Durbin (1980), which is used to obtain an Edgeworth expansion for the WLD's. For example, in an ARFIMA(1, d , 1) model, the third derivative with respect to the autoregressive parameter is zero. Hence, $Z_n(\theta)$ does not contain this WLD for any set $\tilde{\Theta}$ that satisfies Condition NS.

In some models, the sub-vector $Z_n(\theta)$ of $\bar{Z}_n(\theta)$ that yields a nonsingular asymptotic covariance matrix $D(\theta)$ in Condition NS(iii) depends on the parameter vector θ . For example, in an ARFIMA(1, d , 1) model, the first partial derivatives with respect to the autoregressive and moving average coefficients are linearly independent for most parameter values. But, at parameter values that yield common roots, these two WLD's are equal. The results given below cover such cases by allowing one to consider different sets $\tilde{\Theta}$, which may have different sub-vectors $Z_n(\theta)$ appearing in Condition NS(iii).

The second condition of Condition NS(iii) guarantees that the finite-sample covariance matrix of any sub-vector $\tilde{Z}_n(\theta)$ of $\bar{Z}_n(\theta)$ that strictly contains $Z_n(\theta)$ is singular (as shown in next paragraph). This is important because we obtain a valid Edgeworth expansion to the WMLE by approximating it by a function of $Z_n(\theta)$ (suitably normalized). If there is a sub-vector $\tilde{Z}_n(\theta)$ of $\bar{Z}_n(\theta)$ that strictly contains $Z_n(\theta)$ and has a nonsingular covariance matrix, then $Z_n(\theta)$ does not contain all of the non-redundant WLD's of order up to $s - 1$ for sample size n . Non-redundant WLD's cannot be omitted from the approximation to the WMLE without affecting the accuracy of the approximation and the remainder of the Edgeworth expansion for the WMLE.

To see why the claim in the first sentence of the previous paragraph is true, let $\tilde{D}_n(\theta)$ and $\tilde{D}(\theta)$ denote the finite-sample and asymptotic covariance matrices of $n^{-1/2}\tilde{Z}_n(\theta)$ respectively. Let $\nu_1, \dots, \nu_{\tilde{d}_s}$ denote the derivative indices corresponding to the elements of $\tilde{Z}_n(\theta)$. By Condition NS(iii), $\tilde{D}(\theta)$ is singular. Hence, by (9), there exists a vector $a = (a_1, \dots, a_{\tilde{d}_s})' \neq 0$ such that

$$\sum_{j=1}^{\tilde{d}_s} a_j D_{\nu_j} f_{\theta}^{-1}(\lambda) = 0 \quad (10)$$

for all λ in a subset of $(-\pi, \pi)$ with Lebesgue measure 2π . From (5), the elements of $\tilde{Z}_n(\theta) - E_{\theta}\tilde{Z}_n(\theta)$ are of the form

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} D_{\nu_j} f_{\theta}^{-1}(\lambda) (I_n(\lambda) - E_{\theta}I_n(\lambda)) d\lambda$$

for $j = 1, \dots, \tilde{d}_s$. Hence, (10) implies that $a'(\tilde{Z}_n(\theta) - E_{\theta}\tilde{Z}_n(\theta)) = 0$. In turn, this implies that $\tilde{D}_n(\theta)$ is singular.

The situation is different when establishing an Edgeworth expansion for the MLE, rather than the WMLE, as in LRZ. In this case, one approximates the MLE by a vector of LLD's. Singularity of the asymptotic covariance matrix of a vector of LLD's does not imply singularity of the corresponding finite-sample covariance matrix. For example, with the ARFIMA(1, d , 1) model, one can have a singular asymptotic covariance matrix of LLD's, but a nonsingular finite-sample covariance matrix, see Section 5 for a discussion of this model. In such cases, when one approximates the MLE by a vector of LLD's whose asymptotic covariance matrix is nonsingular, some LLD's that are not redundant in finite samples are omitted. This affects the accuracy of the approximation of the MLE and the remainder of the Edgeworth expansion of the MLE. In consequence, in models with this feature, the Edgeworth expansion for the MLE of LRZ has remainder $o(n^{-1})$, rather than $o(n^{-(s-2)/2})$.

Let

$$W_n(\theta) = n^{-1/2}(Z_n(\theta) - E_\theta Z_n(\theta)).$$

Below we establish an Edgeworth expansion for the vector $W_n(\theta)$ of normalized WLD's. Denote the dimension of $Z_n(\theta)$ and $W_n(\theta)$ by d_s . The elements of $W_n(\theta)$ are of the form

$$n^{-1/2}(x_n' M_n T_n(g_{\theta, \nu}) M_n x_n - E_\theta x_n' M_n T_n(g_{\theta, \nu}) M_n x_n).$$

Note that, although Assumptions W2-W7 concern derivatives up to order $s + 1$, for an Edgeworth expansion to the WMLE with an error rate of order $o(n^{-(s-2)/2})$, we only need an Edgeworth expansion of the joint distribution of a vector of normalized WLD's up to order $s - 1$, viz., $W_n(\theta)$. The reason is that, in the Taylor series approximation of the WMLE by WLD's, the $(s + 1)$ -th order WLD's are in the remainder term and the s -th order WLD's can be replaced in the series expansion by their expectations with the differences between them and their expectations being added to the remainder term. For example, see Taniguchi and Kakizawa (2000, Sec. 4.2).

3.3 WLD Cumulant Bounds

A key step in establishing the validity of an Edgeworth expansion for the distribution of $W_n(\theta)$ that holds uniformly over compact subsets of some set $\tilde{\Theta}$ that satisfies Condition NS is showing that the cumulants of $Z_n(\theta)$ are $O(n)$ uniformly in such sets. This condition is Assumption 4 of Durbin (1980). Durbin's (1980) Thm. 1 is used to obtain the Edgeworth expansion of $W_n(\theta)$.

Let $\kappa_r(\theta)$ denote an r -th order joint cumulant of $Z_n(\theta)$. For simplicity, we drop the subscript n in the following. From the theory of quadratic forms in normal variables, e.g., see Searle (1971, p. 55),

$\kappa_r(\theta)$ can be written as

$$\begin{aligned}\kappa_1(\theta) &= \frac{n}{2\pi} D_{\nu_j} \int_{-\pi}^{\pi} \log f_{\theta}(\lambda) d\lambda + \text{tr}\{MT(g_{\theta, \nu_j})MT(f_{\theta})\} \text{ and} \\ \kappa_r(\theta) &= C_r \text{tr} \left[\prod_{j=1}^r \{MT(g_{\theta, \nu_j})MT(f_{\theta})\} \right] \text{ for } r \geq 2\end{aligned}\quad (11)$$

for some vectors $\{\nu_j : j = 1, \dots, r\}$ of subscripts and some constant $C_r < \infty$. Note that $\kappa_r(\theta)$ involves derivatives of $f_{\theta}^{-1}(\lambda)$, not of $f_{\theta}(\lambda)$.

Theorem 1 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. Then, for all $r \geq 1$, $\kappa_r(\theta) = O(n)$ uniformly over any compact subset Θ^* of $\tilde{\Theta}$.*

To prove Theorem 1, we substitute $M = I - P$ in (11) and rewrite $\kappa_r(\theta)$ for $r \geq 2$ as

$$\kappa_r(\theta) = C_r \sum \text{tr} \left[\prod_{j=1}^r \{(-P)^{\chi_j} T(g_{\theta, \nu_j}) (-P)^{\xi_j} T(f_{\theta})\} \right], \quad (12)$$

where χ_j, ξ_j take on the values zero or one and satisfy $0 \leq \sum_{j=1}^r (\chi_j + \xi_j) \leq 2r$, the summation is over all 2^{2r} possible configurations of $(\chi_1, \xi_1, \dots, \chi_r, \xi_r)$, and $(-P)^0 = I$. The following result establishes that the summand in (12) for which $\chi_j = \xi_j = 0$ for all $j = 1, \dots, r$ is $O(n)$. The result is due to LRZ (see their Theorem 1).

Proposition 2 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. Then, for all $r \geq 1$,*

$$\lim_{n \rightarrow \infty} \sup_{\theta \in \Theta^*} \left| \frac{1}{n} \text{tr} \left[\prod_{j=1}^r \{T(g_{\theta, \nu_j})T(f_{\theta})\} \right] - (2\pi)^{2r-1} \int_{-\pi}^{\pi} \left[\prod_{j=1}^r \{g_{\theta, \nu_j}(\lambda) f_{\theta}(\lambda)\} \right] d\lambda \right| = 0 \quad (13)$$

for any compact subset Θ^* of $\tilde{\Theta}$.

In Proposition 2, Assumption W2 guarantees the existence of the WLD's, Assumptions W3 and W4 specify the exponent structure of the matrix product, Assumptions W5 and W6 are used in the proof of Theorem 1 of LRZ, and Assumption W7 is required for the result to be uniform.

Next, we consider the case where at least one matrix P appears in (12). Because P is of the form $n^{-1}11'$, for any matrices A and B , $\text{tr}[PAPB] = \text{tr}[PA]\text{tr}[PB]$. In consequence, each summand in (12) for which at least one matrix P appears can be written as the product of terms of the following form for different values of p : for $0 \leq p \leq r$,

$$I_{n,p}(\theta) = \text{tr} \left[P \prod_{j=1}^p \{T(g_{\theta, \nu_j})T(f_{\theta})\} \right],$$

$$\begin{aligned}
I_{n,p}^-(\theta) &= \text{tr} \left[P \prod_{j=1}^p \{T(g_{\theta,\nu_j})T(f_\theta)\} T(g_{\theta,\nu_{p+1}}) \right], \text{ and} \\
I_{n,p}^+(\theta) &= \text{tr} \left[P \prod_{j=1}^p \{T(f_\theta)T(g_{\theta,\nu_j})\} T(f_\theta) \right]
\end{aligned} \tag{14}$$

with P replaced by $-P$. In addition, the number of terms of the form $I_{n,p}^-(\theta)$ and $I_{n,p}^+(\theta)$ that appear in each summand must be the same, because each product in (12) must contain the same number r of matrices $T(f_\theta)$ as matrices of the form $T(g_{\theta,\nu_j})$.

The following Theorem makes extensive use of power counting theory as discussed in FT87. The Theorem is analogous to Theorem 1(a) of FT87. Its proof is complicated by the fact that the algebraic structure of the product matrices is not homogenous.

Theorem 3 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. For any $p \geq 0$, any compact set $\Theta^* \subset \tilde{\Theta}$, and any constant $\delta > 0$, there exists a constant $K_p(\Theta^*, \delta) < \infty$ such that*

- (a) $\sup_{\theta \in \Theta^*} |I_{n,p}(\theta)| \leq K_p(\Theta^*, \delta)n^\delta$,
- (b) $\sup_{\theta \in \Theta^*} |I_{n,p}^-(\theta)| \leq K_p(\Theta^*, \delta)n^{-\alpha+\delta}$, and
- (c) $\sup_{\theta \in \Theta^*} |I_{n,p}^+(\theta)| \leq K_p(\Theta^*, \delta)n^{\alpha+\delta}$.

Proposition 2, Theorem 3, and the discussion above combine to establish Theorem 1.

Note that Theorem 3 and Proposition 2 imply that the dominant summand in (12) is the one in which no matrix P appears.

3.4 WLD Edgeworth Expansion

We now state the Edgeworth expansion for the density of $W_n(\theta)$. It is obtained by applying Theorem 1 of Durbin (1980).

Theorem 4 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. For $\theta \in \tilde{\Theta}$, let $G_n(u, \theta)$ be the joint density of $W_n(\theta)$ and let $\tilde{G}_n^{(\tau-2)}(u, \theta)$ be its $(\tau-2)$ -order formal Edgeworth expansion for any integer $\tau \geq 3$. Then,*

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

uniformly over $u \in R^{d_s}$ and θ in any compact subset Θ^ of $\tilde{\Theta}$.*

The Edgeworth expansion for the density of $W_n(\theta)$ can be used to obtain an Edgeworth expansion for the distribution of $W_n(\theta)$ using Corollary 3.3 of Skovgaard (1986).

Corollary 5 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. For any integer $\tau \geq 3$,*

$$Pr_\theta(W_n(\theta) \in C) = \int_C \tilde{G}_n^{(\tau-2)}(u, \theta) du + o(n^{-(\tau-2)/2}) \quad (15)$$

uniformly over all Borel sets C and θ in any compact subset Θ^ of $\tilde{\Theta}$.*

Note that the Edgeworth expansions above are valid to any order under Assumptions W1-W7. (In fact, this feature of Theorem 4 is used to obtain the error $o(n^{-(\tau-2)/2})$ in Corollary 5, rather than $o(n^{-(\tau-2)/2+\delta})$ for arbitrary $\delta > 0$, when applying Skovgaard's result.

4 Edgeworth Expansions for the Whittle MLE

The WMLE $\hat{\theta}_n$ of θ solves

$$\int_{-\pi}^{\pi} \frac{\partial}{\partial \theta_r} \{\log f_\theta(\lambda) - f_\theta^{-1}(\lambda) I_n(\lambda)\} d\lambda = 0 \text{ for } r = 1, \dots, d_\theta. \quad (16)$$

In general, there may be multiple solutions to (16).

Let $\tilde{H}_n^{(s-2)}(u, \theta)$ be the $(s-2)$ -th order formal Edgeworth expansion of the density of $n^{1/2}(\hat{\theta}_n - \theta)$ for $\theta \in \tilde{\Theta}$ given by

$$\tilde{H}_n^{(s-2)}(u, \theta) = \phi_{\Sigma(\theta)}(u) \left\{ 1 + \sum_{r=3}^s n^{-\frac{1}{2}r+1} q_{n,r,\theta}(u) \right\},$$

where $\phi_{\Sigma(\theta)}$ denotes the multivariate normal density with mean zero and covariance matrix $\Sigma(\theta)$ and $\{q_{n,r,\theta}(u) : r = 3, \dots, s\}$ are Edgeworth polynomials whose coefficients are $O(1)$ and depend on the cumulants of the WLD's.

The main result of the paper is the following Theorem. Its form is analogous to that of Theorem 3 of Bhattacharya and Ghosh (1978).

Theorem 6 *Suppose Assumptions W1-W7 hold and $\tilde{\Theta}$ satisfies Condition NS. Let Θ^* denote a compact set in $\tilde{\Theta}$. Then,*

(a) *there exists a sequence of estimators $\{\hat{\theta}_n : n \geq 1\}$ and a constant $d_0 = d_0(\Theta^*)$ such that*

$$\inf_{\theta \in \Theta^*} Pr_\theta(\|\hat{\theta}_n - \theta\| < d_0 n^{-\frac{1}{2}} (\log n)^{\frac{1}{2}}, \hat{\theta}_n \text{ solves (16)}) = 1 - o(n^{-(s-2)/2}), \quad (17)$$

(b) any sequence of estimators $\{\hat{\theta}_n : n \geq 1\}$ that satisfies (17) admits the Edgeworth expansion

$$Pr_{\theta}(\sqrt{n}(\hat{\theta}_n - \theta)) \in C = \int_C \tilde{H}_n^{(s-2)}(u, \theta) du + o(n^{-(s-2)/2}) \quad (18)$$

uniformly over $\theta \in \Theta^*$ and over every class \mathcal{B} of Borel sets that satisfies the condition

$$\sup_{\theta \in \Theta^*} \sup_{C \in \mathcal{B}} \int_{(\partial C)^\varepsilon} \phi_{M(\theta)}(u) du = O(\varepsilon) \text{ as } \varepsilon \downarrow 0, \quad (19)$$

where $(\partial C)^\varepsilon$ denotes the ε -neighborhood of the boundary of C .

Several remarks are in order. First, the error rate in both parts of the Theorem is identical to the iid rate. Second, the error rate can be made arbitrarily small if the assumptions hold for s arbitrarily large. Third, part (a) of the Theorem does not guarantee consistency of the estimator that maximizes the Whittle likelihood. Rather, it shows that a consistent solution to the first-order conditions given in (16) exists. This is analogous to the results of Bhattacharya and Ghosh (1978). Fourth, the regularity condition in (19) is standard. It is the same as in Bhattacharya and Ghosh (1978, eqn. (1.6)).

5 An Example

The ARFIMA(1, d , 1) model is very popular in applied work due to its flexibility. The model is

$$(1 - \phi B)(1 - B)^d X_t = (1 + \psi B)\varepsilon_t,$$

where B is the lag operator, $d \in (0, 1/2)$ is the long-memory parameter, and $|\phi| < 1$ and $|\psi| < 1$ for stationarity and invertibility. Let $\theta = (d, \phi, \psi, \sigma_\varepsilon^2)'$, where σ_ε^2 is the variance of the innovation ε_t .

The spectral density of the ARFIMA(1, d , 1) process is

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

e.g., see Hosking (1981). The spectral density satisfies

$$f_{\theta}(\lambda) \sim C(\theta) |\lambda|^{-\alpha(\theta)} \text{ as } |\lambda| \rightarrow 0,$$

which is a special case of (1).

In this model, the third partial derivative of $f_{\theta}^{-1}(\lambda)$ with respect to (wrt) ϕ is zero. In consequence, by the argument in Section 3.2, the matrices $\bar{D}(\theta)$ and $\bar{D}_n(\theta)$ are both singular. Because the same

degeneracy occurs in $D(\theta)$ and $D_n(\theta)$, the problematic WLD can be deleted from $\bar{Z}_n(\theta)$ without affecting the error in the approximation of the WMLE by the vector of WLD's. That is, for any set $\tilde{\Theta}$ that satisfies Condition NS, the vector $Z_n(\theta)$ does not include the null WLD and this singularity does not cause a problem.

In contrast, when deriving an Edgeworth expansion for the MLE, one considers LLD's rather than WLD's, the asymptotic covariance matrix of an LLD vector that includes the third derivative wrt ϕ is singular, but its finite-sample covariance matrix is nonsingular whenever the sub-matrix without the third derivative wrt ϕ is nonsingular. This occurs because (i) the covariance matrix of all the LLD's up to order $s - 1$ is the same as $\bar{D}_n(\theta)$ defined in (8), but with $T_n(g_{\theta, \nu_j})$ replaced by $D_{\nu_j} T_n^{-1}(f_{\theta})$, (ii) the limit as $n \rightarrow \infty$ of the covariance matrix of all the LLD's up to order $s - 1$ is exactly the same as that of $\bar{D}_n(\theta)$, viz., $\bar{D}(\theta)$, see LRZ, and (iii) the third partial derivative of $T_n^{-1}(f_{\theta})$ wrt ϕ does not equal zero. In consequence, when one drops the third partial derivative wrt ϕ from the vector $Z_n(\theta)$ that is used to approximate the MLE, the approximation of the MLE and the remainder of the Edgeworth expansion are affected.

6 Proofs

Proof of Theorem 3. We prove the results of the Theorem for the case where $p \geq 1$ first. The proof closely follows the work of FT87 using power counting theory. We use their notation. For ease of presentation, we omit the δ from the exponents in the bounds on f and the g_{ν_j} 's. The proof goes through with δ added for some $\delta > 0$ sufficiently small. Then, the results of the Theorem hold for arbitrary $\delta > 0$ because the upper bounds in the Theorem are increasing in δ .

First, we consider $I_{n,p}(\theta)$. We have

$$\begin{aligned} I_{n,p}(\theta) &= \text{tr} \left\{ P \prod_{j=1}^p (T(g_{\nu_j}) T(f)) \right\} \\ &= \sum_{j_1=0}^{n-1} \cdots \sum_{j_{2p+1}=0}^{n-1} P_{j_1, j_2} T(g_{\nu_1})_{j_2, j_3} T(f)_{j_3, j_4} \cdots T(g_{\nu_p})_{j_{2p}, j_{2p+1}} T(f)_{j_{2p+1}, j_1} \\ &= n^{-1} \int_{U_\pi} \tilde{P}_n(y) Q(y) dy, \end{aligned} \tag{20}$$

where

$$\begin{aligned} U_\pi &= [-\pi, \pi]^{2p}, \\ \tilde{P}_n(y) &= \sum_{j_1=0}^{n-1} \cdots \sum_{j_{2p+1}=0}^{n-1} e^{i(j_2-j_3)y_1} e^{i(j_3-j_4)y_2} \cdots e^{i(j_{2p+1}-j_1)y_{2p}}, \\ Q(y) &= g_{\nu_1}(y_1) f(y_2) g_{\nu_2}(y_3) \cdots f(y_{2p}). \end{aligned}$$

Note that

$$\begin{aligned}\tilde{P}_n(y) &= \left(\sum_{j_1=0}^{n-1} e^{-ij_1 y_{2p}} \right) \left(\sum_{j_2=0}^{n-1} e^{ij_2 y_1} \right) \left(\sum_{j_3=0}^{n-1} e^{ij_3 (y_2 - y_1)} \right) \cdots \left(\sum_{j_{2p+1}=0}^{n-1} e^{ij_{2p+1} (y_{2p} - y_{2p-1})} \right) \\ &= h_n^*(-y_{2p}) h_n^*(y_1) h_n^*(y_2 - y_1) \cdots h_n^*(y_{2p} - y_{2p-1}),\end{aligned}$$

where

$$h_n^*(z) = \sum_{j=0}^{n-1} e^{ijz}.$$

In FT87's case, where the matrix P does not appear in the product, the function $P_n(y)$ is given by

$$P_n(y) = h_n^*(y_1 - y_{2p}) h_n^*(y_2 - y_1) \cdots h_n^*(y_{2p} - y_{2p-1}).$$

Hence, $\tilde{P}_n(y)$ differs from $P_n(y)$ in that the term $h_n^*(-y_{2p}) h_n^*(y_1)$ appears in place of $h_n^*(y_1 - y_{2p})$. In addition, the rhs of (20) contains the n^{-1} multiplicand which is not present in FT87's case. For each $h_n^*(z)$, we use the bound: for all $0 < \eta < 1$,

$$|h_n^*(z)| \leq 4h_{n,\eta}(z) \text{ for } -2\pi \leq z \leq 2\pi,$$

where

$$\begin{aligned}h_{n,\eta}(z) &= n^\eta |z + \xi|^{\eta-1} \\ \xi &= \begin{cases} 0 & \text{for } -\pi \leq z \leq \pi \\ -2\pi & \text{for } \pi \leq z \leq 2\pi \\ 2\pi & \text{for } -2\pi \leq z \leq -\pi, \end{cases}\end{aligned}\tag{21}$$

see p. 227 of FT87. It follows that

$$|I_{n,p}(\theta)| \leq K n^{-1} \int_{U_\pi} \tilde{f}_{n,\eta}(y) dy\tag{22}$$

for some constant $K < \infty$, for all $\theta \in \Theta^*$, where

$$\begin{aligned}\tilde{f}_{n,\eta}(y) &= h_{n,\eta}(y_1) h_{n,\eta}(-y_{2p}) h_{n,\eta}(y_2 - y_1) \cdots h_{n,\eta}(y_{2p} - y_{2p-1}) \\ &|y_1|^\alpha |y_2|^{-\alpha} |y_3|^\alpha \cdots |y_{2p}|^{-\alpha}\end{aligned}$$

and $\alpha = \alpha(\theta)$.

We make the following change of variables:

$$\begin{aligned}x_1 &= y_1 \text{ and} \\ x_k &= y_k - y_{k-1} \text{ for } k = 2, \dots, 2p.\end{aligned}$$

Then, the rhs of (22) is at most

$$\begin{aligned}
& K n^{-1} \int_{U'_\pi} h_{n,\eta}(x_1) h_{n,\eta}(-x_1 - \dots - x_{2p}) h_{n,\eta}(x_2) h_{n,\eta}(x_3) \cdots h_{n,\eta}(x_{2p}) \\
& \times |x_1|^\alpha |x_1 + x_2|^{-\alpha} \cdots |x_1 + \dots + x_{2p-1}|^\alpha |x_1 + \dots + x_{2p}|^{-\alpha} dx,
\end{aligned} \tag{23}$$

where

$$U'_\pi = \left\{ x \in R^{2p} : \left| \sum_{j=1}^k x_j \right| \leq \pi \text{ for } k = 1, \dots, 2p \right\}.$$

Notice that $(-\pi \leq y_1 \leq \pi) \Rightarrow (-\pi \leq x_1 \leq \pi)$ and $(-\pi \leq y_{2p} \leq \pi) \Rightarrow (-\pi \leq x_1 + \dots + x_{2p} \leq \pi)$.

Hence, for all $0 < \gamma_1 < 1$ and $0 < \gamma_2 < 1$, we have

$$h_{n,\gamma_1}(x_1) \leq n^{\gamma_1} |x_1|^{\gamma_1-1} \text{ for } -\pi \leq x_1 \leq \pi$$

and

$$h_{n,\gamma_2}(-x_1 - \dots - x_{2p}) \leq n^{\gamma_2} |x_1 + \dots + x_{2p}|^{\gamma_2-1} \text{ for } -\pi \leq x_1 + \dots + x_{2p} \leq \pi.$$

For all other $h_{n,\eta}$'s appearing in (23), we have $-2\pi \leq x_k \leq 2\pi$ for $k = 2, \dots, 2p$ and so, we need to consider the possibility that some of the ξ 's, defined in (21), are not zero. Thus, the term in (23) is dominated by

$$K n^{-1+(2p-1)\eta+\gamma_1+\gamma_2} \int_{U'_\pi} \tilde{P}_{\eta,\gamma}(x) dx, \tag{24}$$

where

$$\tilde{P}_{\eta,\gamma}(x) = \tilde{P}_{\eta,1}(x) \tilde{P}_{\gamma,2}(x), \tag{25}$$

$$\tilde{P}_{\eta,1}(x) = |x_2 + \xi_2|^{\eta-1} |x_3 + \xi_3|^{\eta-1} \cdots |x_{2p} + \xi_{2p}|^{\eta-1}, \text{ and}$$

$$\tilde{P}_{\gamma,2}(x) = |x_1|^{\gamma_1-1+\alpha} |x_1 + x_2|^{-\alpha} \cdots |x_1 + \dots + x_{2p-1}|^\alpha |x_1 + \dots + x_{2p}|^{\gamma_2-1-\alpha}.$$

The idea is to provide conditions on γ_1 and γ_2 such that the integral in (24) is finite.

It is useful to rewrite $\tilde{P}_{\eta,\gamma}(x)$ as

$$\tilde{P}_{\eta,\gamma}(x) = |L_2|^{b_2} \cdots |L_{4p}|^{b_{4p}}, \tag{26}$$

where

$$L_2 = x_2 + \xi_2, L_3 = x_3 + \xi_3, \dots, L_{2p} = x_{2p} + \xi_{2p}, \tag{27}$$

$$L_{2p+1} = x_1, L_{2p+2} = x_1 + x_2, \dots, L_{4p} = x_1 + \dots + x_{2p},$$

and

$$\begin{aligned}
\{b_2, \dots, b_{4p}\} &= \{b_2, \dots, b_{2p}, \alpha_1, \dots, \alpha_{2p}\} \\
&= \underbrace{\{\eta - 1, \dots, \eta - 1\}}_{2p-1}, \underbrace{\{\gamma_1 - 1 + \alpha, -\alpha, \alpha, -\alpha, \dots, \alpha, \gamma_2 - 1 - \alpha\}}_{2p}
\end{aligned} \tag{28}$$

To proceed, we distinguish between two cases.

Case I: $\xi_2 = \xi_3 = \dots = \xi_{2p} = 0$.

Let

$$P_\eta(x) = P_{\eta,1}(x)P_2(x),$$

where

$$\begin{aligned} P_{\eta,1}(x) &= |x_2 + \dots + x_{2p}|^{\eta-1} |x_2|^{\eta-1} |x_3|^{\eta-1} \dots |x_{2p}|^{\eta-1} \text{ and} \\ P_2(x) &= |x_1|^{-\alpha} |x_1 + x_2|^{-\beta} |x_1 + x_2 + x_3|^{-\alpha} \dots |x_1 + \dots + x_{2p}|^{-\beta}. \end{aligned}$$

Section 5 of FT87 shows that $P_\eta(x)$ is integrable provided

$$0 < \eta < 1, \beta < 1, \alpha < 1, \text{ and } 2\eta > \alpha + \beta. \quad (29)$$

The integrand $\tilde{P}_{\eta,\gamma}(x)$ appearing (24) and defined in (25) differs from $P_\eta(x)$ in two respects:

$$(D1) \ P_{\eta,1}(x) = \tilde{P}_{\eta,1}(x) |x_2 + \dots + x_{2p}|^{\eta-1}.$$

(D2) The exponent structure of $P_2(x)$ is $E_2 = \{-\alpha, -\beta, \dots, -\alpha, -\beta\}$, whereas that of $\tilde{P}_{\gamma,2}(x)$ is

$$\tilde{E}_2 = \{\gamma_1 - 1 + \alpha, \underbrace{-\alpha, \alpha, -\alpha, \dots, \alpha}_{2p-2}, \gamma_2 - 1 - \alpha\}. \quad (30)$$

While the exponent structure of $P_2(x)$ is homogenous, the exponent structure of $\tilde{P}_{\gamma,2}(x)$ is not homogenous. This is important, because the conditions in (29) are not sufficient for integrability in the non-homogeneous case. Moreover, it is clear from the proof of FT87's Proposition 5.5 that the extension of condition (29) to non-homogeneous exponent structures is not trivial.

Our goal is to show that the function $\tilde{P}_{\eta,\gamma}(x)$ is integrable by accommodating for the differences (D1)-(D2). The first difference leads to a simplification, whereas the latter leads to a complication. We deal first with (D1). It is clear from the discussion on p. 222 of FT87 that it is enough to consider sets $W \subset T$ that do not contain $x_2 + \dots + x_{2p}$, where FT87's set of functions T is given by

$$T = \{x_2 + \dots + x_{2p}, x_2, x_3, \dots, x_{2p}, x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + \dots + x_{2p}\}.$$

Note that T is the set of multiplicands of $P_\eta(x)$ without the exponents or absolute values. The analogous set in our case is

$$\begin{aligned} \tilde{T} &= \{x_2, x_3, \dots, x_{2p}, x_1, x_1 + x_2, x_1 + x_2 + x_3, \dots, x_1 + \dots + x_{2p}\} \\ &= T \setminus \{x_2 + \dots + x_{2p}\}. \end{aligned}$$

For any $W \subset T$, let $s(W) = T \cap \text{span}(W)$. Although it is enough to consider the integrability of $P_\eta(x)$ with the restriction $x_2 + \dots + x_{2p} \notin W$, FT87 still need to consider the case $x_2 + \dots + x_{2p} \in s(W)$. For any $W \subset \tilde{T}$, let $\tilde{s}(W) = \tilde{T} \cap \text{span}(W)$. Because $x_2 + \dots + x_{2p} \notin \tilde{T}$, it is also true that $x_2 + \dots + x_{2p} \notin \tilde{s}(W)$. Hence, unlike the situation in FT87, we do not need to consider the case $x_2 + \dots + x_{2p} \in \tilde{s}(W)$.

As in Section 3 of FT87, we define the dimension of $P_{\eta,\gamma}$ with respect to a set $W \subset \tilde{T}$ to be

$$d(\tilde{P}_{\eta,\gamma}, W) = |W| + \sum_{\{j: L_j \in \tilde{s}(W)\}} b_j,$$

where $|W|$ denotes the cardinality of W and the b_j 's and L_j 's are defined in (27)-(28). One can think of the elements of \tilde{T} arranged in columns as follows:

$$\left| \begin{array}{c} x_2 \\ x_1 + x_2 \end{array} \right| \left| \begin{array}{c} x_3 \\ x_1 + x_2 + x_3 \end{array} \right| \cdots \left| \begin{array}{c} x_{2p} \\ x_1 + \dots + x_{2p} \end{array} \right|.$$

The top element in each column arises from the bound on $P_n(x)$ and the bottom element in each column, apart from the first and last columns, arises from the bound on $Q(x)$. As in FT87, p. 223, we consider a set W that contains at most one element from each column. We partition W into contiguous "blocks" such that $W = \bigcup_{i=1}^n B_i$, where a set $B \subset W$ is a "block" if there exist $\ell_B < r_B$ such that (i) W contains neither column $\ell_B - 1$ nor column $r_B + 1$ and (ii) B contains column ℓ_B through r_B and no other columns. As in FT87, the integral in (24) is finite if $d(\tilde{P}_{\eta,\gamma}, B_i) > 0$ for all B_i .

Let B denote one of the blocks B_i . It contains a block of columns l through to r , so $|B| = r - l + 1$. Let m denote the smallest k satisfying $x_1 + \dots + x_k \in B$. We can write $\tilde{s}(B) = W_1 \cup W_2$, where

$$\begin{aligned} W_1 &= \{x_l, x_{l+1}, \dots, x_{m-1}, x_{m+1}, \dots, x_r\} \text{ and} \\ W_2 &= \{x_1 + \dots + x_m, x_1 + \dots + x_{m+1}, \dots, x_1 + \dots + x_r\}. \end{aligned}$$

We count the powers associated with W_1 and obtain $(r - l)(\eta - 1)$. Similarly, the powers associated with W_2 contribute $\sum_{j=m}^r \alpha_j$, where the α_j 's are defined in (28). Thus,

$$\begin{aligned} d(\tilde{P}_{\eta,\gamma}, B) &= (r - l + 1) + (r - l)(\eta - 1) + \sum_{j=m}^r \alpha_j \\ &= (r - l)\eta + (1 + \sum_{j=m}^r \alpha_j). \end{aligned}$$

Sufficient conditions for $d(\tilde{P}_{\eta,\gamma}, B)$ to be positive are

$$\eta > 0 \text{ and } \inf_{m \leq r} \sum_{j=m}^r \alpha_j > -1. \quad (31)$$

Recall that

$$\{\alpha_1, \dots, \alpha_{2p}\} = \{\gamma_1 - 1 + \alpha, \underbrace{-\alpha, \alpha, -\alpha, \dots, \alpha}_{2p-2}, \gamma_2 - 1 - \alpha\}.$$

Hence, the second condition in (31) is satisfied if

$$\gamma_2 > \alpha, \quad \gamma_1 > 0, \quad \text{and} \quad \gamma_1 + \gamma_2 > 1, \quad (32)$$

because $\inf_{\theta \in \Theta^*} \alpha = \inf_{\theta \in \Theta^*} \alpha(\theta) > 0$. Returning to (24), we see that the entire expression is at most $K_p(\Theta^*, \delta)n^\delta$ for some constant $K_p(\Theta^*, \delta) < \infty$ for all $\theta \in \Theta^*$, $\forall \delta > 0$, because $\sup_{\theta \in \Theta^*} \alpha = \sup_{\theta \in \Theta^*} \alpha(\theta) < 1$.

As noted above, we do not need to consider the case $x_2 + \dots + x_{2p} \in \tilde{s}(W)$, so the proof is complete for the $\xi_2 = \dots = \xi_{2p} = 0$ case.

Case II: At least one $\xi_j \neq 0$ for $j = 2, \dots, 2p$.

This case is dealt with in Section 6 of FT87. It is clear from (26)-(27) that the only L 's affected in this case are L_2, \dots, L_{2p} . In FT87's case, the L 's affected are L_1, \dots, L_{2p} each of which has exponent $\eta - 1$. In their case, FT87 fix a permutation $\{\sigma_1, \dots, \sigma_{4p}\}$ of $\{1, \dots, 4p\}$ and define

$$E_\sigma^t = \{x \in U_t' : |L_{\sigma_1}| \leq |L_{\sigma_2}| \leq \dots \leq |L_{\sigma_{4p}}|\}.$$

A basis is constructed for T satisfying

$$|L_{\tau_1}| \leq |L_{\tau_2}| \leq \dots \leq |L_{\tau_r}|, \quad \text{where } r = \text{rank}(T).$$

By the proofs of Propositions 6.1 and 6.2 of FT87, it follows that

$$\int_{E_\sigma^\pi} \tilde{P}_{\eta, \gamma}(x) dx < \infty,$$

provided the conditions in (32) hold. Because $U_\pi = \bigcup_\sigma E_\sigma^\pi$, we are done in this case as well.

Next, we consider $I_{n,p}^-(\theta)$. We have

$$\begin{aligned} I_{n,p}^-(\theta) &= \text{tr} \left[P \prod_{j=1}^p \{T(g_{\nu_j})T(f)\} T(g_{\nu_{p+1}}) \right] \\ &= \sum_{j_1=0}^{n-1} \dots \sum_{j_{2p+2}=0}^{n-1} P_{j_1, j_2} T(g_{\nu_1})_{j_2, j_3} T(f)_{j_3, j_4} \dots T(g_{\nu_p})_{j_{2p}, j_{2p+1}} T(f)_{j_{2p+1}, j_{2p+2}} T(g_{\nu_{p+1}})_{j_{2p+2}, j_1} \\ &= n^{-1} \int_{U_\pi} \tilde{P}_n(y) Q(y) dy, \end{aligned}$$

where in this case $U_\pi = [-\pi, \pi]^{2p+1}$,

$$\begin{aligned}\tilde{P}_n(y) &= \sum_{j_1=0}^{n-1} \dots \sum_{j_{2p+2}=0}^{n-1} e^{i(j_2-j_3)y_1} e^{i(j_3-j_4)y_2} \dots e^{i(j_{2p+2}-j_1)y_{2p+1}} \text{ and} \\ Q(y) &= g_{\nu_1}(y_1)f(y_2)g_{\nu_2}(y_3) \dots f(y_{2p})g_{\nu_{p+1}}(y_{2p+1}).\end{aligned}$$

The exponent structure in this case is

$$\tilde{E}_2 = \{\gamma - 1 + \alpha, \underbrace{-\alpha, \alpha, -\alpha, \dots, -\alpha}_{2p-1}, \gamma - 1 + \alpha\}.$$

Note that here we choose $\gamma_1 = \gamma_2 = \gamma$ in the first and last terms in \tilde{E}_2 . The second condition in (31) is satisfied if we take $2\gamma > 1 - \alpha$. Together with the condition that $0 < \gamma < 1$, it follows that $\sup_{\theta \in \Theta^*} |I_{n,p}^-(\theta)| \leq K_p(\Theta^*, \delta)n^{-\alpha+\delta}$ for some constant $K_p(\Theta^*, \delta) < \infty, \forall \delta > 0$.

Finally, the proof for the bound on $I_{n,p}^+(\theta)$ uses the same ideas as above. In this case, the exponent structure is

$$\tilde{E}_2 = \{\gamma - 1 - \alpha, \underbrace{\alpha, -\alpha, \alpha, \dots, \alpha}_{2p-1}, \gamma - 1 - \alpha\}.$$

Sufficient conditions for integrability are $0 < \gamma < 1$ and $2\gamma > 1 + \alpha$, from which it follows that $\sup_{\theta \in \Theta^*} |I_{n,p}^+(\theta)| \leq K_p(\Theta^*, \delta)n^{\alpha+\delta}, \forall \delta > 0$. The proof of Theorem 3 is now complete for the case where $p \geq 1$.

To finish the proof, we consider the case where $p = 0$. We have $I_{n,0}(\theta) = \text{tr}[P] = 1$, so part (a) of the Theorem holds trivially. Next, we have

$$\begin{aligned}I_{n,0}^-(\theta) &= \text{tr}[PT(g_{\theta, \nu_1})] = n^{-1}1'T(g_{\theta, \nu_1})1 = n^{-1} \int_{-\pi}^{\pi} \sum_{j,k=0}^{n-1} e^{i(j-k)\lambda} g_{\theta, \nu_1}(\lambda) d\lambda, \text{ and} \\ |I_{n,0}^-(\theta)| &\leq n^{-1} \int_{-\pi}^{\pi} \left| \sum_{j,k=0}^{n-1} e^{i(j-k)\lambda} \right| \cdot |g_{\theta, \nu_1}(\lambda)| d\lambda,\end{aligned}$$

where 1 denotes an n -vector of ones. We use the inequality

$$\left| \sum_{j=0}^{n-1} e^{ij\lambda} \right| \leq 4n^\eta |\lambda|^{\eta-1}, \quad \forall \lambda \in [-\pi, \pi], \quad \forall \eta \in (0, 1),$$

see FT87 pp. 227 and 237. Because $|g_{\theta, \nu_1}(\lambda)| \leq c_2(\theta, \delta) |\lambda|^\alpha, \forall \lambda \in N_\delta, \forall \delta > 0$, where $\alpha = \alpha(\theta) - \delta$, by Assumption W4, and $\int_{-\pi}^{\pi} |g_{\theta, \nu_1}(\lambda)| d\lambda < \infty$ by Assumption W2, there exists a constant $K < \infty$, such that

$$|I_{n,0}^-(\theta)| \leq Kn^{2\eta-1} \int_{-\pi}^{\pi} |\lambda|^{2(\eta-1)+\alpha} d\lambda.$$

The integral is finite provided $2(\eta - 1) + \alpha > -1$ or $2\eta - 1 > -\alpha$. Take η such that $2\eta - 1$ is arbitrarily close to $-\alpha$ to obtain the desired result.

Similarly, $I_{n,0}^+(\theta) = \text{tr}[PT(f_\theta)] = n^{-1}1'T(f_\theta)1$. By Assumption W3, $|f_\theta(\lambda)| \leq c_2(\theta, \delta) |\lambda|^\alpha$, $\forall \lambda \in N_\delta$, $\forall \delta > 0$, where $\alpha = -\alpha(\theta) - \delta$. The argument above for $I_{n,0}^-(\theta)$ holds for both positive and negative α , as long as $|\alpha| < 1$. Hence, the desired result holds for $I_{n,0}^+(\theta)$. \square

Proof of Theorem 4. Theorem 4 is proved by verifying Assumptions 1-4 of Durbin's (1980) Theorem 1. Durbin's Assumption 1 corresponds to our Assumption W8. Durbin's Assumption 4 requires that the joint cumulants of the WLD's are $O(n)$ uniformly on Θ^* . This is satisfied by our Theorem 1. Durbin's Assumptions 2 and 3 concern the characteristic function of $Z_n(\theta)$, which we denote by $\varphi_n(\omega, \theta) = E_\theta[\exp(i\omega'Z_n(\theta))]$. From standard theory on quadratic forms in Gaussian variables (e.g., see Searle (1971, p. 55)), we have

$$\begin{aligned} \varphi_n(\omega/\sqrt{n}, \theta) &= E_\theta \exp \left\{ \frac{i}{\sqrt{n}} \sum_{j=1}^{d_s} \omega_j (A_j(\theta) + x_n' B_j(\theta) x_n) \right\} \\ &= \exp \left(\frac{i}{\sqrt{n}} \sum_{j=1}^{d_s} \omega_j A_j(\theta) \right) \det \left(I_n - \frac{2i}{\sqrt{n}} \sum_{j=1}^{d_s} \omega_j \tilde{B}_j(\theta) \right)^{-1/2}, \end{aligned} \quad (33)$$

where $A_j(\theta)$ is a partial derivative of $g(\theta)$ (defined in Assumption W2) of order less than or equal to $s-1$, $B_j(\theta) = MT(g_{\theta, \nu_j})M$, g_{θ, ν_j} is a partial derivative of $(4\pi^2)^{-1}f_\theta^{-1}(\lambda)$ of order less than or equal to $s-1$, and $\tilde{B}_j(\theta) = B_j(\theta)T(f_\theta)$. Because Proposition 2 and Theorem 3 ensure that $\text{tr}(\prod_{j=1}^p \tilde{B}_j(\theta)) = O(n)$ uniformly on Θ^* for any finite p and because $D(\theta) > 0$ by Assumption W8, LRZ's verification of Durbin's (1980) Assumptions 2 and 3 goes through without change. \square

Proof of Theorem 6. The proof of the Theorem relies on a passage from the result of Corollary 5 to the result of the Theorem. The proof of Theorem 4 of LRZ shows that Bhattacharya and Ghosh's (1978, Thms. 2 and 3) argument extends to the long-memory case. The main step in the proof concerns tail probability behavior of the centered WLD's, as in the equations in (2.32) of Bhattacharya and Ghosh (1978). As shown in Taniguchi and Kakizawa (2000, Pf. of Thm. 4.2.7), slightly weaker conditions than those of (2.32) are sufficient for Bhattacharya and Ghosh's (1978) proof of Theorem 3 to go through. These weaker conditions can be verified straightforwardly using Markov's inequality in Taniguchi and Kakizawa's case and in our case, rather than using von Bahr's inequality for iid random variables as Bhattacharya and Ghosh (1978) do. This completes the proof. \square

REFERENCES

- Andrews, D. and O. Lieberman (2002). Higher-order Improvement of the Parametric Bootstrap for Long-memory Gaussian Processes. Working paper. Cowles Foundation, Yale University.
- Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.
- Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434-451.
- Dahlhaus, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766.
- Durbin, J. (1980). Approximations for densities of sufficient statistics. *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. (1983). Asymptotic expansions for sums of weakly dependent random vectors. *Z. Wahrsch. Verw. Gebiete* **64** 211–239.
- 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** 15–30.
- Grenander, U. and Szegö, G. (1956). *Toeplitz Forms and Their Applications*. University of California Press, Berkeley CA (second edition, Chelsea, New York, 1984).
- Hosking, J. R. M. (1981). Fractional differencing. *Biometrika* **68** 165–176.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civ. Engin.* **116** 770-808.
- 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. (2002). Valid Asymptotic expansions for the maximum likelihood estimator of the parameter of a stationary, Gaussian, strongly dependent process. Cowles Foundation Discussion Paper No. 1351, Yale University. Tentatively accepted for publication in *Ann. Statist.*
- Mandelbrot, B. B. and Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* **10** 422–437.
- Robinson, P. M. (1995). Log periodogram regression of time series with long range dependence. *Ann. Statist.* **23** 1048–1072.
- Searle, S. R. (1971). *Linear Models*. Wiley, New York.
- Skovgaard, I. M. (1986). On multivariate Edgeworth expansions. *Internat. Statist. Rev.* **54** 169–186.
- 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. (1990). Higher Order Asymptotic Theory for Time Series Analysis (J. Berger, S. Fienberg, J. Gani, K. Krickeberg, I. Olkin and B. Singer, eds.), Springer Lecture Notes in Statistics **68**.
- Taniguchi, M, and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer-Verlag, New York.