On Plug-In Estimation of Long Memory Models

By

Offer Lieberman[1]

Technion—Israel Institute of Technology

First version: October 2002

This vesrion: July 29, 2003

[1] Address for correspondence: Faculty of Industrial Engineering and Management, Technion, Haifa 32000, Israel. E-mail: offerl@ie.technion.ac.il

Research on this topic commenced during 2000-2002, while the author was visiting the Cowles Foundation for Research in Economics at Yale University. The author is most grateful to the Cowles Foundation for their generous hospitality and to Donald Andrews and Peter Phillips for numerous helpful comments.

Key words and phrases: ARFIMA, Long memory, Plug-in, Profile likelihood.

JEL Classification Numbers: C12, C13, C15.

Proposed running head: ON PLUG-IN ESTIMATION OF LONG MEMORY

Corresponding author:

Offer Lieberman

Faculty of Industrial Engineering and Management,

Technion-Israel Institute of Technology,

Haifa 32000, Israel.

E-mail: offerl@ie.technion.ac.il

Abstract

We consider the Gaussian ARFIMA(j,d,l) model, with spectral density $f_\theta(\lambda)$, $\theta \in \Theta \subset \mathsf{R}^p$, $\lambda \in (-\pi,\pi)$, $d \in (0,\frac{1}{2})$ and an unknown mean $\mu \in \mathsf{R}$. For this class of models, the information matrix of the full parameter vector, (μ,θ), is asymptotically degenerate. To estimate θ, Dahlhaus (1989) suggested using the maximizer of the plug-in log-likelihood, $L_n(\theta,\tilde{\mu}_n)$, where $\tilde{\mu}_n$ is any $n^{(1-2d)/2}$-consistent estimator of μ. The resulting estimator is a plug-in maximum likelihood estimator (PMLE). This estimator is asymptotically normal, efficient and consistent, but in finite samples it has some serious drawbacks. Primarily, none of the Bartlett identities associated with $L_n(\theta,\tilde{\mu}_n)$ are satisfied for fixed n. Cheung and Diebold (1994) conducted a Monte Carlo simulation study and reported that the bias of the PMLE is about $3-4$ times the bias of the regular MLE. In this paper, we derive asymptotic expansions for the PMLE and show that its second-order bias is contaminated by an additional term, which does not exist in regular cases. This term arises due to the failure of the first Bartlett identity to hold and seems to explain Cheung and Diebold's (1994) simulated results. We derive similar expansions for the Whittle MLE, which is another estimator tacitly using the plug-in principle. An application to the ARFIMA$(0,d,0)$ shows that the additional bias terms are considerable.

1 Introduction

Let $\{X_t, t \in \mathsf{Z}\}$ be a stationary Gaussian ARFIMA(j, d, l) process with an unknown mean $\mu \in \mathsf{R}$ and covariance matrix $T_n(f_\theta)$. The spectral density, $f_\theta(\lambda)$, is parametrized by $\theta \in \Theta \subset \mathsf{R}^p$, $\lambda \in (-\pi, \pi)$ and the memory parameter of the process, d, which is an element of θ, is assumed to lie in $(0, 1/2)$. The main feature of this model is that $f_\theta(\lambda)$ behaves as

$$f_\theta(\lambda) = O\left(|\lambda|^{-2d}\right) \text{ as } |\lambda| \to 0. \tag{1}$$

A process satisfying (1) is said to be long-memory, since it is consistent with a non-summable autocovariance function.

There is a thriving research on long-memory processes, dating back to Hurst (1951). Naturally, much of the econometric thought in this context has been dedicated to the estimation of (μ, θ) and in particular, of the long-memory parameter. A primary candidate for estimation is the Gaussian maximum likelihood estimator (MLE), based on the log-likelihood

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

where 1_n is an n-vector of 1's, n is the sample size and $x_n' = (X_1, \ldots, X_n)$. However, unlike the short-memory case, corresponding to $d = 0$, the estimation of θ is generally not done jointly with μ. The second-order partial derivative of $L_n(\mu, \theta)$ with respect to (wrt) μ equals $-1_n' T_n(f_\theta)^{-1} 1_n$ and Adenstedt (1974, Theorem 5.2) showed that this quantity is of the order $O\left(n^{1-2d}\right)$. This unusual rate is a consequence of the long-memory property of the process. On the other hand, the expected value of the second-order partial derivative of $L_n(\mu, \theta)$ wrt the θ components is $O(n)$. The implication is that in the $d > 0$ case, the information matrix of the full parameter vector, (μ, θ), normalized by n^{-1}, is asymptotically degenerate.

Dahlhaus (1989) adopted the plug-in principle for this class of models. The Gaussian plug-in maximum likelihood estimator (PMLE) of θ is defined as the maximizer of

$$L(\theta, \tilde{\mu}_n) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\det T_n(f_\theta) - \frac{1}{2}(x_n - \tilde{\mu}_n 1_n)' T_n(f_\theta)^{-1}(x_n - \tilde{\mu}_n 1_n), \tag{3}$$

where $\tilde{\mu}_n$ is any $n^{1/2-d}$-consistent estimator of μ. We denote the PMLE by $\tilde{\theta}_n$. The sample mean, $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$, is the most common candidate for $\tilde{\mu}_n$, in both theoretical and empirical research, its variance being of the order $O\left(n^{-1+2d}\right)$. Dahlhaus (1989) showed that the family of PMLE's of θ is consistent, asymptotically normal and efficient. This estimator is often referred to in the literature as the 'exact' MLE, perhaps to distinguish it from estimators such as the Whittle MLE (WMLE), but the PMLE can only be exact if $\tilde{\mu}_n$ is the profile MLE of μ.

We shall refer to the exact MLE (EMLE) as the maximizer of (2) when the mean is known and denote it by $\hat{\theta}_n$. While the PMLE has desirable asymptotic properties, There is evidence that this is not the case in finite samples. Cheung and Diebold (1994) used Monte Carlo simulation to compare the finite sample behavior of various estimators, including the EMLE, the PMLE and the WMLE. Table 2 of their paper reveals that for the simple, zero-mean, ARFIMA$(0, d, 0)$ model, the bias of the PMLE in which $\tilde{\mu}_n = \bar{X}_n$, is about 3–4 times the order of magnitude of the bias of the EMLE. In this paper we provide an analytical explanation for this phenomenon.

The principal cause of the finite sample drawbacks of the PMLE as compared with the EMLE is that the plugged-in likelihood is not a density type of an object and hence, none of the Bartlett identities associated with it is satisfied for fixed n. Among the various implications of this failure are that the stochastic expansion to the PMLE is non-standard, moment expansions associated with the PMLE are generally contaminated by additional terms which are non-negligible to second order, and Bartlett corrections of the likelihood ratio test involve additional terms. In particular, we show that as compared with the EMLE, for each $\delta > 0$ the bias expansion of the PMLE to order $O\left(n^{-3/2+\delta}\right)$ contains the additional term

$$\kappa^{r,s}\tilde{\kappa}_s = O\left(n^{-1+\delta}\right), \tag{4}$$

where $\kappa^{r,s}$ is the n^{-1}-normalized inverse information matrix of the EMLE, $\tilde{\kappa}_s$ is the expected score associated with (3) and $\tilde{\mu}_n = \bar{X}_n$. The additional term (4) is a manifestation of the failure of the first Bartlett identity to hold for (3). It does not appear in the bias expansion of the EMLE because the density associated with (2) is regular.

A competing estimator to the PMLE is the WMLE, denoted by $\hat{\theta}_{W,n}$, which maximizes the function

$$L_{W,n}(\theta) = -n\left[\frac{1}{4\pi}\int_{-\pi}^{\pi}\left\{\log f_\theta(\lambda) + \frac{I_n(\lambda)}{f_\theta(\lambda)}\right\}d\lambda\right], \tag{5}$$

where

$$I_n(\lambda) = \frac{1}{2\pi n}\left|\sum_{j=1}^{n} e^{ij\lambda}\left(X_j - \bar{X}_n\right)\right|^2. \tag{6}$$

The motivation behind the WMLE is mainly computational - the determinant and the matrix inverse in (3) are replaced by simple univariate integral approximations. Fox and Taqqu (1986) proved asymptotic normality of the WMLE at the usual $\sqrt{n}$-rate and Dahlhaus (1989) proved asymptotic efficiency of the estimator. However, since the estimator is based on approximations to $\det(T_n(f_\theta))$ and $T_n^{-1}(f_\theta)$ and since the plug-in principle is tacitly applied in (6), (5) is not a density type of an object and the Bartlett identities associated with it do not hold. We conduct high-order analysis for the WMLE which reveals similar complications as of the PMLE.

Generally speaking, while small sample theory and asymptotic expansions are not as abundant as first-order theory for time series and in particular, for long-memory processes, there is, nevertheless, some progress in the area. Among the various contributions in the area, the reader is referred to Lieberman, Rousseau and Zucker (2000, 2001, 2003), Andrews and Lieberman (2002a, 2002b) and Lieberman and Phillips (2001, 2002). The present paper continues this line of research.

The plan for the remainder of the paper is as follows. In Section 2 we provide Assumptions, basic set up and notation. In Section 3 we outline some well-known results concerning the EMLE, to be used as benchmarks. Asymptotic expansions for the PMLE are given in Section 4. In Section 5 we provide expansions for the WMLE. In Section 6 we discuss the effects of plugging-in on the likelihood ratio test. The ARFIMA$(0, d, 0)$ model is used for illustration in Section 7. Section 8 contains a small simulation study. Section 9 concludes. All proofs are contained in the Appendix.

2 Setup, Assumptions and Notation

The model under consideration is

$$\Phi(B)(1-B)^d(X_t - \mu) = \Psi(B)\varepsilon_t,$$

where B is the backshift operator, $BX_t = X_{t-1}$,

$$\begin{aligned}\Phi(w) &= 1 + \sum_{r=1}^{j}\Phi_r w^r, \\ \Psi(w) &= 1 + \sum_{r=1}^{l}\Psi_r w^r,\end{aligned}$$

and $\varepsilon_t \sim NID(0, \sigma_\varepsilon^2)$. Let $\theta = \left(\sigma_\varepsilon^2, \Phi_1, \ldots, \Phi_j, \Psi_1, \ldots, \Psi_l, d\right)$, $\dim(\theta) = p$. Denote the true covariance parameter of the process by θ_0. Formally, the assumption on the process is:

Assumption 1. $\{X_t, t \in \mathsf{Z}\}$ is a stationary Gaussian process with mean $\mu \in \mathsf{R}$, covariance parameters $\theta \in \Theta \subset \mathsf{R}^p$, spectral density

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

and

$$d \in (0, 1/2).$$

The roots of $\Phi(w)$ and of $\Psi(w)$ lie outside the unit circle.

We remark that all the results of the present paper hold in fact for models more general than ARFIMA(j, d, l). This model is a special case of more general long-memory models considered by Fox and Taqqu (1986), Dahlhaus (1989) and Lieberman, Rousseau and Zucker (2003). In all of these studies, the spectral density of the process is assumed to behave as

$$f_\theta(\lambda) \sim |\lambda|^{-\alpha(\theta)} L_\theta(\lambda) \quad \text{as } |\lambda| \to 0,$$

where $0 < \alpha(\theta) < 1$ and $L_\theta(\lambda) = 0\left(|\lambda|^\delta\right)$ as $|\lambda| \to 0$, $\forall\, \delta > 0$. The spectral density of the ARFIMA(j, d, l) model satisfies all the assumptions in the aforementioned papers. We only work with the ARFIMA(j, d, l) model for the sake of simplicity and clarity of the exposition.

Where no ambiguity arises, we shall omit the dependence on n and on θ for brevity. For log-likelihood derivatives and their expectations, we use index notation and write $U_\mu = \partial L/\partial\mu$, $U_{\mu\mu} = \partial^2 L/\partial\mu^2$, $U_r = \partial L/\partial\theta^r$, $U_{rs} = \partial^2 L/\partial\theta^r\partial\theta^s$, $U_{rst} = \partial^3 L/\partial\theta^r\partial\theta^s\partial\theta^t$, etc. The indices r, s, t run from 1 to p. The null cumulants are denoted by $\kappa_\mu = n^{-1}E_{\theta,\mu}U_\mu$, $\kappa_{\mu\mu} = n^{-1}E_{\theta,\mu}U_{\mu\mu}$, $\kappa_r = n^{-1}E_{\theta,\mu}U_r$, $\kappa_{rs} = n^{-1}E_{\theta,\mu}U_{rs}$, $\kappa_{r,s} = n^{-1}E_{\theta,\mu}U_rU_s$, $\kappa_{rs,t} = n^{-1}E_{\theta,\mu}U_{rs}U_t$, etc., and $(\kappa^{r,s})$ in the inverse matrix of $(\kappa_{r,s})_{1\leq r,s\leq p}$. To avoid unnecessary use of additional notation, we shall not distinguish between null cumulants associated with (2) when the mean is either known or not.

The plug-in log-likelihood is denoted by $L(\theta, \tilde{\mu}_n) = \tilde{L}$ and the quantities associated with it are $\tilde{U}_r = \partial\tilde{L}/\partial\theta^r$, $\tilde{U}_{rs} = \partial^2\tilde{L}/\partial\theta^r\partial\theta^s$, $\tilde{\kappa}_r = n^{-1}E_{\theta,\mu}\tilde{U}_r$, $\tilde{\kappa}_{rs} = n^{-1}E_{\theta,\mu}\tilde{U}_{rs}$, etc. Finally, derivatives of the covariance matrix will be denoted by $\dot{T}_r = \partial T(f_\theta)/\partial\theta^r$, $\ddot{T}_{rs} = \partial^2 T(f_\theta)/\partial\theta^r\partial\theta^s$, and $\dddot{T}_{rst} = \partial^3 T(f_\theta)/\partial\theta^r\partial\theta^s\partial\theta^t$.

To give a precise exposition of the problem of joint estimation of (μ, θ), we will need the basic results:

$$\begin{aligned}
A_r &= \frac{\partial}{\partial\theta^r}\left(-\frac{1}{2}\log\det T\right) = -\frac{1}{2}tr\left(T^{-1}\dot{T}_r\right) \qquad (7)\\
A_{rs} &= \frac{\partial^2}{\partial\theta^s\partial\theta^r}\left(-\frac{1}{2}\log\det T\right) = -\frac{1}{2}tr\left(-T^{-1}\dot{T}_sT^{-1}\dot{T}_r + T^{-1}\ddot{T}_{rs}\right)\\
B_r &= \frac{\partial}{\partial\theta^r}\left(-\frac{1}{2}T^{-1}\right) = \frac{1}{2}T^{-1}\dot{T}_rT^{-1}\\
B_{rs} &= \frac{\partial^2}{\partial\theta^s\partial\theta^r}\left(-\frac{1}{2}T^{-1}\right) = \frac{1}{2}\left(-T^{-1}\dot{T}_sT^{-1}\dot{T}_rT^{-1} + T^{-1}\ddot{T}_{rs}T^{-1} - T^{-1}\dot{T}_rT^{-1}\dot{T}_sT^{-1}\right).
\end{aligned}$$

B_{rst} is defined and derived similarly. The first- and second-order partial log-likelihood derivatives wrt the θ components are

$$U_r = A_r + (x - \mu 1)' B_r (x - \mu 1)$$

and

$$U_{rs} = A_{rs} + (x - \mu 1)' B_{rs} (x - \mu 1).$$

The first- and second-order partial log-likelihood derivatives wrt μ are

$$U_\mu = 1' T^{-1} (x - \mu 1)$$

and

$$U_{\mu\mu} = -1' T^{-1} 1.$$

Further,

$$\kappa_r = \frac{1}{n} E_{\theta,\mu} U_r = \kappa_\mu = \frac{1}{n} E_{\theta,\mu} U_\mu = 0,$$

$$\begin{aligned} \kappa_{rs} &= \frac{1}{n} E_{\theta,\mu} U_{rs} \\ &= \frac{1}{n} \left\{ A_{rs} + tr\left(B_{rs} T\right) \right\} \\ &= -\frac{1}{2n} tr\left(T^{-1} \dot{T}_s T^{-1} \dot{T}_r \right) \end{aligned} \tag{8}$$

and

$$\begin{aligned} \kappa_{\mu\mu} &= \frac{1}{n} E_{\theta,\mu} U_{\mu\mu} \\ &= -\frac{1}{n} 1' T^{-1} 1. \end{aligned}$$

The Lemma below follows immediately from Dahlhaus (1989, Theorem 5.1) and Adenstedt (1974, Theorem 5.2).

Lemma 1 Under Assumption 1, $\kappa_{rs} = O(1)$ and $\kappa_{\mu\mu} = O\left(n^{-2d}\right)$.

Since $d > 0$, the lemma implies that $\lim_{n\to\infty} \kappa_{\mu\mu} = 0$ and that the asymptotic information matrix of the full parameter vector (θ, μ) is degenerate. This problem does not occur in the short memory case.

3 Expansions for the EMLE

In this section we outline some well-known results concerning $\hat{\theta}$ and relate them to $\tilde{\theta}$ and $\hat{\theta}_W$ in the following sections. Set $\hat{\delta}^r = \sqrt{n}\left(\hat{\theta} - \theta_0\right)^r$. Formal expansions for regular MLE's are given by, among others, Lawley (1956) and McCullagh (1987). All of these results are based on the assumption that at least one solution to the first-order conditions exists with probability that goes to unity at a fast rate. With this assumption, we can expand $U_r\left(\hat{\theta}\right) = 0$ around θ_0 and then extract $\hat{\delta}$. While most work in the area is done under the iid assumption, the following also holds for long-memory processes.

Proposition 2 Under Assumption 1, the stochastic expansion to $\hat{\delta}^r$ is given by

$$\hat{\delta}^r = \frac{1}{\sqrt{n}} \kappa^{r,s} U_s + \frac{1}{n^{3/2}} \kappa^{r,s} \kappa^{t,u} \left(U_{st} - n\kappa_{st}\right) U_u + \frac{1}{2n^{3/2}} \kappa^{r,s} \kappa^{t,i} \kappa^{u,j} \kappa_{stu} U_i U_j + O_p\left(n^{-1}\right). \tag{9}$$

Equation (9) is the one appearing in McCullagh (1987, p 209) with the exception that his κ's are covariances between log-likelihood derivatives, whereas our κ's are expectations of (product) log-likelihood derivatives, normalized by n^{-1}. Lieberman, Rousseau and Zucker (2003) showed that the κ's are $O(1)$ uniformly in any compact subset of Θ. They also proved the validity of the Edgeworth expansion for the distribution of $\hat{\delta}^r$ under a class of models more general than the ARFIMA(j, d, l) type. We note that the first term in (9) is $O_p(1)$ and the second and third terms are $O_p\left(n^{-1/2}\right)$.

Taking expectations, the bias and covariance of $\hat{\theta}$ are given by

$$E_\theta\left(\hat{\theta}-\theta_0\right)^r = \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(\kappa_{st,u}+\kappa_{stu}/2\right)+O\left(n^{-3/2}\right), \tag{10}$$

and

$$Cov_\theta\left(\hat{\delta}^r,\hat{\delta}^s\right)=\kappa^{r,s}+O(n^{-1}). \tag{11}$$

Formula (10) is given by Cox and Snell (1968). Two remarks are in place. First, the moment expansions (10)–(11) are only formal. There are delicate issues in showing that (10)–(11) are indeed valid asymptotic expansions and we do not attempt to do it in this paper. Secondly, the development of (10) hinges on the fact that the first and second Bartlett identities hold for (2), and that the matrix $(\kappa_{r,s})_{1\le r,s\le p}$ is invertible.

4 Expansions for the PMLE

We consider the plug-in likelihood $L_n(\theta,\tilde{\mu}_n)$. The most popular choice of $\tilde{\mu}_n$, in both theoretical and empirical research, is $\tilde{\mu}_n=\bar{X}_n$. It is stressed that in this case, $L_n(\theta,\bar{x}_n)$ is not even a profile likelihood. Although our developments are more general than this case, we shall concentrate on it in order to illustrate the main points. Here,

$$L_n(\theta,\bar{x}_n)=-\frac{n}{2}\log(2\pi)-\frac{1}{2}\log\det T_n(f_\theta)-\frac{1}{2}x_n' M_n T_n(f_\theta)^{-1}M_n x_n, \tag{12}$$

where

$$M_n=I_n-P_n, P_n=\frac{1}{n}1_n1_n' \tag{13}$$

and I_n is the identity matrix of order n. Since $1_n'M_n=0, x_n'M_n=(x_n-\mu 1_n)'M_n$. Consequently, the object $L_n(\theta,\bar{x}_n)$ is invariant to μ and we may set $\mu=0$ without loss of generality. We obtain:

Lemma 3 Under Assumption 1, for $\tilde{\mu}_n=\bar{X}_n$ and for each $\delta>0$,

$$\tilde{\kappa}_r=O\left(n^{-1+\delta}\right), \text{ and } \tilde{\kappa}_{rs}+\tilde{\kappa}_{r,s}=O\left(n^{-1+\delta}\right), r,s=1,\ldots,p.$$

Thus, the first and second Bartlett identities associated with $L_n(\theta,\bar{x}_n)$ hold only asymptotically. The result plays a critical role in the derivation of the expansions for $\tilde{\theta}$.

Set $\tilde{\delta}=\sqrt{n}\left(\tilde{\theta}-\theta_0\right)$ and let $d_{rs}=\tilde{\kappa}_{rs}-\kappa_{rs}$. The term d_{rs} is then the discrepancy between the plug-in information per observation and the information per observation corresponding to (2) with known μ. We show in Lemma 8 that $d_{rs}=O(n^{-1+\delta}),\forall\delta>0$. As in the regular case, we assume that there exists at least one solution to $\tilde{U}_r=0$, with probability that goes to one at a sufficiently fast rate.

Proposition 4 Under Assumption 1

$$\tilde{\delta}^r=\frac{1}{\sqrt{n}}\kappa^{r,s}\tilde{U}_s+\frac{1}{n^{3/2}}\kappa^{r,s}\kappa^{t,u}\left(\tilde{U}_{st}-n\tilde{\kappa}_{st}\right)\tilde{U}_u+\frac{1}{2n^{3/2}}\kappa^{r,s}\kappa^{t,j}\kappa^{u,k}\kappa_{stu}\tilde{U}_j\tilde{U}_k$$
$$+\frac{1}{\sqrt{n}}\kappa^{r,s}\kappa^{t,u}d_{st}\tilde{U}_u+O_p\left(n^{-1}\right). \tag{14}$$

We show in the Appendix that the first term on the rhs of (14) is $O_p(1)$, the second and third terms are $O_p\left(n^{-1/2}\right)$ and the fourth term is $O_p\left(n^{-1+\delta}\right)$, $\forall\delta>0$. Comparing (9) to (14), we observe that to order $O_p\left(n^{-1}\right)$, (14) contains $\kappa^{r,s}\tilde{U}_r/\sqrt{n}$, which, by Lemma 3, does not have a zero mean, and $\kappa^{r,s}\kappa^{t,u}d_{st}\tilde{U}_u/\sqrt{n} = O_p\left(n^{-1+\delta}\right), \forall\delta > 0$. In addition, the second term on the rhs of (14) involves $\tilde{\kappa}_{st}$, which differs from κ_{st} by an order of $O(n^{-1+\delta})$, $\forall\delta > 0$.

Taking expectations, the formal moment expansions are

$$E_\theta\left(\tilde{\theta} - \theta\right)^r = \kappa^{r,s}\tilde{\kappa}_s + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(\kappa_{st,u} + \kappa_{stu}/2\right) + O\left(n^{-3/2+\delta}\right) \tag{15}$$

and

$$Cov_\theta\left(\tilde{\delta}^r, \tilde{\delta}^s\right) = \kappa^{r,s} + O\left(n^{-1+\delta}\right). \tag{16}$$

Note that on the rhs of (15) all terms apart from $\tilde{\kappa}_s$ are null cumulants associated with $L_n(\theta)$. In contrast with (10), the bias of $\tilde{\theta}$ to order $O\left(n^{-3/2+\delta}\right)$ includes the additional term $\kappa^{r,s}\tilde{\kappa}_r = O\left(n^{-1+\delta}\right)$. This probably explains the huge difference between the biases of the PMLE and the EMLE, see Cheung and Diebold (1994). The error rate in this case is actually $O\left(n^{-3/2+\delta}\right)$, which is slightly different from the $O\left(n^{-3/2}\right)$ rate in the regular case.

The expression (15) can provide a basis for bias correction. The bias corrected estimator is

$$\tilde{\theta}^r_{BC} = \tilde{\theta}^r - \left.\left(\kappa^{r,s}\tilde{\kappa}_s + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(\kappa_{st,u} + \kappa_{stu}/2\right)\right)\right|_{\theta^r = \tilde{\theta}^r}. \tag{17}$$

The correction is generally sufficient for the removal of the first order bias.

5 Expansions for the WMLE

We proceed to investigate the WMLE, which is another estimator using the plug-in principle. We use the notation $\dot{T}_{W,r} = \partial T_W/\partial\theta^r$, $\ddot{T}_{W,r} = \partial^2 T_W/\partial\theta^r\partial\theta^s$, $\dot{f}_r = \partial f_\theta/\partial\theta^r$, $\ddot{f}_{rs} = \partial^2 f_\theta/\partial\theta^r\partial\theta^s$, $U_{W,r} = \partial L_W/\partial\theta^r$, $U_{W,rs} = \partial^2 L_W/\partial\theta^r\partial\theta^s$, $\kappa_{W,r} = n^{-1}E_{\theta,\mu}U_{W,r}$, $\kappa_{W,r,s} = n^{-1}E_{\theta,\mu}U_{W,r}U_{W,s}$, $\kappa_{W,rs} = n^{-1}E_{\theta,\mu}U_{W,rs}$, and so on. The analogue of Lemma 3 is the following.

Lemma 5 Under Assumption 1, for each $\delta > 0$,

$$\kappa_{W,r} = O(n^{-1+\delta}) \text{ and } \kappa_{W,r,s} + \kappa_{W,rs} = O(n^{-1+\delta}), r, s = 1, \ldots, p.$$

Thus, the first and second Bartlett identities associated with $L_{W,n}(\theta)$ hold only asymptotically.

As in the regular case, let $\hat{\delta}^r_W = \sqrt{n}\left(\hat{\theta}_W - \theta_0\right)^r$ and assume the existence of at least one solution to the $U_{W,r} = 0$ with probability that tends to unity fast.

Proposition 6 Under Assumption 1

$$\begin{aligned}\hat{\delta}^r_W &= \frac{1}{\sqrt{n}}\kappa_W^{r,s}U_{W,s} + \frac{1}{n^{3/2}}\kappa_W^{r,s}\kappa_W^{t,u}\left(U_{W,st} - n\kappa_{W,st}\right)U_{W,u} \\ &+ \frac{1}{2n^{3/2}}\kappa_W^{r,s}\kappa_W^{t,i}\kappa_W^{u,j}\kappa_{W,stu}U_{W,i}U_{W,j} + \frac{1}{\sqrt{n}}\kappa_W^{r,s}\kappa_W^{t,u}d_{W,st}U_{W,u} + O_p\left(n^{-1}\right).\end{aligned}$$

We show in the Appendix that

$$\begin{aligned}E_\theta\left(\hat{\theta}_W - \theta\right)^r &= \kappa_W^{r,s}\kappa_{W,s} + \frac{1}{n}\kappa_W^{r,s}\kappa_W^{t,u} \\ &\quad\times\left[\left\{\frac{1}{2n}tr\left(M\ddot{T}_{W,st}MT\right)\left(M\dot{T}_{W,u}MT\right)\right\} + \kappa_{W,stu}/2\right] \\ &\quad + O\left(n^{-3/2}\right).\end{aligned} \tag{18}$$

It follows from Theorems 1 and 3 of Andrews and Lieberman (2002a) and from Lemma 5 of the present paper, that the first term in (18) is $O\left(n^{-1+\delta}\right)$, and the second and third are $O\left(n^{-1}\right)$. Comparing the last expression to (15), we see that the bias expansion is also contaminated by the additional, non-negligible term $\kappa_W^{r,s}\kappa_{W,s}$, due to the failure of the first Bartlett identity to hold. The variance of $\hat{\delta}_W$ is easily shown to be

$$Cov\left(\hat{\delta}_W^r, \hat{\delta}_W^s\right) = \kappa_W^{r,s} + O\left(n^{-1}\right).$$

Similar to (17), we can construct the Whittle based bias corrected estimator of θ as

$$\hat{\theta}_{W,BC}^r = \hat{\theta}_W^r - \left\{\kappa_W^{r,s}\kappa_{W,s} + \frac{1}{n}\kappa_W^{r,s}\kappa_W^{t,u}\left[\left\{\frac{1}{2n}tr\left(M\ddot{T}_{W,st}MT\right)\left(M\dot{T}_{W,u}MT\right)\right\} + \kappa_{W,stu}/2\right]\right\}\Bigg|_{\theta^r = \hat{\theta}_W^r}.$$

6 Testing implications

Following (9), we can write

$$\hat{\delta}^r = a^r + \frac{1}{\sqrt{n}}c^r + O_p\left(n^{-1}\right) \tag{19}$$

with

$$\begin{aligned} a^r &= \frac{1}{\sqrt{n}}\kappa^{r,s}U_s = O_p(1) & (20)\\ c^r &= \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(U_{st} - n\kappa_{st}\right)U_u + \frac{1}{2n}\kappa^{r,s}\kappa^{t,i}\kappa^{u,j}\kappa_{stu}U_iU_j = O_p(1). \end{aligned}$$

Now, half of the LRT is given by

$$\frac{1}{2}W(\theta) = L\left(\hat{\theta}\right) - L(\theta), \tag{21}$$

which can be expanded as

$$\frac{1}{2}W(\theta) = \frac{1}{\sqrt{n}}U_r\hat{\delta}^r + \frac{1}{2n}U_{rs}\hat{\delta}^r\hat{\delta}^s + \frac{1}{6n^{3/2}}U_{rst}\hat{\delta}^r\hat{\delta}^s\hat{\delta}^t + O_p\left(n^{-1}\right).$$

Substituting (19)-(20) into (21) and collecting terms of the same order results in

$$\frac{1}{2}W(\theta) = \frac{1}{2}\kappa^{r,s}z_rz_s + \frac{1}{6\sqrt{n}}\kappa_{rst}z^rz^sz^t + \frac{1}{2\sqrt{n}}z_{rs}z^rz^s + O_p\left(n^{-1}\right), \tag{22}$$

with $z_r = U_r/\sqrt{n}$, $z_{rs} = \left(U_{rs} - n\kappa_{rs}\right)/\sqrt{n}$. Taking expectations,

$$E_\theta\left(W(\theta)\right) = p\left\{1 + \frac{b_n(\theta)}{n} + O\left(n^{-3/2}\right)\right\},$$

where $b_n(\theta)$ is $O(1)$. A full formula for $b_n(\theta)$ is given by Lawley (1956). This result provides the basis for the Bartlett correction of the LRT, viz,

$$W'(\theta) = \frac{W(\theta)}{1 + \frac{b_n(\theta)}{n}}. \tag{23}$$

The mean correction corrects for the distribution as well and it is well known that the distribution of $W'(\theta)$ is $\chi^2(p)$ with an error of the order $O\left(n^{-2}\right)$.

In the plug-in case, it is not difficult to show that

$$\frac{1}{2}\tilde{W}(\theta) = \frac{1}{2}\kappa^{r,s}\tilde{z}_r\tilde{z}_s + \frac{1}{6\sqrt{n}}\kappa_{rst}\tilde{z}^r\tilde{z}^s\tilde{z}^t + \frac{1}{2\sqrt{n}}\tilde{z}_{rs}\tilde{z}^r\tilde{z}^s + O_p\left(n^{-1+\delta}\right), \tag{24}$$

with $\tilde{z}_r = \tilde{U}_r/\sqrt{n}$ and $\tilde{z}_{rs} = (\tilde{U}_{rs} - n\tilde{\kappa}_{rs})/\sqrt{n}$. The expansions (22) and (24) differ wrt the $O_p\left(n^{-1}\right)$ term, which in the plug-in case contains additional terms. As a consequence, the Bartlett correction $(1 + b_n(\theta)/n)$ appearing in (23) is not valid when the plug-in principle is employed. The same argument is true when the the LRT is based on the Whittle estimator.

7 An application

Consider the ARFIMA$(0,d,0)$ model $(1-B)^d X_t = \varepsilon_t, \varepsilon_t \overset{iid}{\sim} N(0,1)$, and $d \in (0,1/2)$. The true mean of the process in this example is taken to be zero. In this section we develop explicit bias and variance approximations for the EMLE, PMLE and WMLE. For the former, we treat the mean as known, whereas for the PMLE and WMLE, we treat the mean as unknown. The spectral density can be written as

$$f(\lambda) = \frac{1}{2\pi} e^{-dc(\lambda)}, \tag{25}$$

with

$$c(\lambda) = \log(2(1-\cos(\lambda))). \tag{26}$$

First, we deal with the EMLE of d. The first bias approximation to the EMLE, based on (10), is shown in the appendix to be

$$b_{app1}\left(\hat{d};d\right) = -\left\{tr\left(T^{-1}\dot{T}\right)^2\right\}^{-2} tr\left(T^{-1}\dot{T}T^{-1}\ddot{T}\right), \tag{27}$$

so that

$$E_d(\hat{d}-d) = b_{app1}(\hat{d};d) + O(n^{-3/2}).$$

This formula can be simplified by using integral limit approximations to the traces. An approximation to (27), based on an application of Theorem 7 of Lieberman and Phillips (2003), gives

$$b_{app2}\left(\hat{d}\right) = -\frac{108\zeta(3)}{n\pi^4}, \tag{28}$$

so that

$$E_d(\hat{d}-d) = b_{app2}\left(\hat{d}\right) + O(n^{-3/2}).$$

Note that $b_{app2}\left(\hat{d}\right)$ is independent of d, which is a manifestation of Lieberman and Phillips' (2001) result that $\hat{d}$ is second-order pivotal in this model. The bias corrected EMLE in this model is simply

$$\hat{d}_{BC} = \hat{d} + \frac{108\zeta(3)}{n\pi^4}.$$

In view of (11) and (61), the first variance approximation for $\hat{\delta} = \sqrt{n}(\hat{d}-d)$, is

$$V_{app1}\left(\hat{\delta};\delta\right) = \left\{\frac{1}{2n} tr\left(T^{-1}\dot{T}\right)^2\right\}^{-1} \tag{29}$$

so that

$$Var(\hat{\delta};\delta) = V_{app1}\left(\hat{\delta};\delta\right) + O(n^{-1}).$$

Using (65), the simplified variance approximation is

$$V_{app2}\left(\hat{\delta}\right) = \frac{6}{\pi^2}, \tag{30}$$

which is just the well-known asymptotic variance of $\hat{\delta}$.

Next, we investigate the PMLE. The additional bias term in (15) is, using Lemma 7(1) and eqn's (7) and (61),

$$\begin{aligned} \kappa^{r,s}\tilde{\kappa}_s &= \left(\frac{1}{2n} tr\left(T^{-1}\dot{T}\right)^2\right)^{-1} \left\{\frac{1}{n} tr\left(-PB_rT - B_rPT + PB_rPT\right)\right\} \\ &= \left(tr\left(T^{-1}\dot{T}\right)^2\right)^{-1} \left\{tr\left(-2PT^{-1}\dot{T} + PT^{-1}\dot{T}T^{-1}PT\right)\right\}. \end{aligned} \tag{31}$$

The first bias approximation to $\tilde{d}$ is the sum of (31) and $b_{app1}(\hat{d};d)$. The simplified bias approximation will be based on

$$b_{app2}\left(\tilde{d};d\right)=\frac{6}{\pi^2}\left\{\frac{1}{2n}tr\left(-PT^{-1}\dot{T}-\dot{T}T^{-1}P+PT^{-1}\dot{T}T^{-1}PT\right)\right\}-\frac{108\zeta\left(3\right)}{n\pi^4}. \qquad (32)$$

The reason that $\tilde{\kappa}_s$ is not simplified any further is that its limit is zero. The first and second bias approximations to the variance of the PMLE are identical to those of the MLE, as follows from (16). We can construct a bias corrected estimator by subtracting from $\tilde{d}$ the rhs of (32), evaluated at $\tilde{d}$. Note that unlike the corrected EMLE, in this case the correction depends on $\tilde{d}$. Finally, we deal with the WMLE. For the ARFIMA$(0,d,0)$ model,

$$\int_{-\pi}^{\pi}\frac{\dot{f}\left(\lambda\right)}{f\left(\lambda\right)}d\lambda=\int_{-\pi}^{\pi}\left(-c\left(\lambda\right)\right)d\lambda=0\,.$$

Together with (44) and (52), this implies that

$$\kappa_{W,r}=-\frac{1}{2n}tr\left(M\dot{T}_W MT\right) \qquad (33)$$

and

$$\kappa_{W,stu}=-\frac{1}{2n}tr\left(M\ \dddot{T}_W\ MT\right). \qquad (34)$$

The first bias approximation to the WMLE is obtained upon substitution of (55), (33) and (34) into (18). To obtain a simplified approximation to the bias of the WMLE, we apply Proposition 2 and Theorem 3 of Andrews and Lieberman (2002a). This yields

$$b_{app2}\left(\hat{d}_W;d\right)=\kappa_W^{r,s}\kappa_{W,s}-\frac{108\zeta\left(3\right)}{n\pi^4}.$$

We remark that the second summand in $b_{app2}\left(\hat{d}_W;d\right)$ is just $b_{app2}\left(\hat{d};d\right)$. We could also replace $\kappa_W^{r,s}$ by $6/\pi^2$ but empirically the first choice seem to perform better. Bias correction can easily be based on the last formula.

The first variance approximation of $\hat{\delta}_W$ is given by

$$V_{app1}\left(\hat{\delta}_W;\delta\right)=\frac{1}{n}\kappa_W^{r,s}=\frac{1}{n}\left[\frac{1}{2n}tr\left\{M\dot{T}_W MT\right\}^2+n\left\{-\frac{1}{2n}tr\left(M\dot{T}_W MT\right)\right\}^2\right]^{-1}$$

and the integral limit gives the simplified version

$$V_{app2}\left(\hat{\delta}_W;\delta\right)=\frac{6}{n\pi^2}.$$

8 Simulations

We compute the various approximations presented in the previous section and compare them to the simulated results obtained by Cheung and Diebold (1994). Table 1 contains bias of the three estimators for the ARFIMA$(0,d,0)$ model with $n=50,100,300$ and 500 and d in the range $[.05,.45]$ at a .1 grid. The terms 'app1' and 'app2' correspond to the first and second type approximations for each of the estimators, as described in the previous section. The 'adt' terms correspond to the additional bias terms $\kappa^{r,s}\tilde{\kappa}_s$ and $\kappa_W^{r,s}\kappa_{W,s}$, for the PMLE and WMLE, respectively. Cheung and Diebold's (1994) simulated values are taken to be the benchmark.

We start with Table 1. In general, all the approximations improve with n and the approximations of the first type are superior to those of the second type. However, the latter are much simpler to compute, especially for large n. The main feature of the Table is that the additional bias term of the PMLE captures much of the difference between the simulated biases of the EMLE and the PMLE. This shows that the additional term is not only non-negligible, but in fact dominant in terms of real magnitude (as opposed to order of magnitude), for the sample sizes under consideration. The simplified bias approximations of all estimators deteriorate as d increases. The reason is that $b_{app2}\left(\hat{d}\right)$ is the leading term in the approximation of $b_{app1}\left(\hat{d};d\right)$ and while the former is independent of d, the same does not hold for higher order terms in the expansion for $b_{app1}\left(\hat{d};d\right)$. See Lieberman and Phillips (2003) for discussion. Both $b_{app2}\left(\tilde{d};d\right)$ and $b_{app2}\left(\hat{d}_W;d\right)$ are based on $b_{app2}\left(\hat{d}\right)$, so that the deterioration as d increases affects the bias of these estimators as well. As seen from the 'bcor' entries in the Table, the reduction in the bias of the bias corrected estimators can be dramatic.

Mean squared errors are considered in Table 2. It appears that the approximations of the first type are accurate for all estimators and that the approximation errors diminishes with n, as expected. The variance approximation $V_{app2}\left(\hat{\delta}\right)$ is constant but the mse's of the PMLE and the WMLE differ slightly as they involve their non-constant bias approximations. When $n = 500$, all approximations are excellent and are practically indistinguishable from the simulated values.

Figures 1 and 2 contain PP plots of the likelihood ratio tests based on the EMLE (LRT) and the PMLE (PLRT) against the asymptotic $\chi^2(1)$ distribution, with sample sizes of $n = 50$ and 100. While the LRT does not seem to need the Bartlett correction for this model and sample sizes, it appears that the curve associated with the PLRT lies markedly underneath the 45 degree line. It is comforting to know that the first order distortion in the distribution approximation diminishes when we move from the $n = 50$ to the $n = 100$ case. However, the effect of the plugging-in on the $\chi^2(1)$ approximation to the distribution of the PLRT is substantial and cannot be ignored.

9 Conclusions

Gaussian Maximum likelihood estimation of long-memory models with an unknown mean is generally done with an application of the plug-in principle in which the true mean is replaced by an $n^{(1-2d)/2}$-consistent estimator of it, prior to maximization of the plugged-in log-likelihood wrt the covariance parameters. While the principle does not harm the asymptotic properties of the PMLE and the WMLE, the same is not true for the finite sample properties of these estimators.

In this paper we derived asymptotic expansions for the PMLE and the WMLE. It turns out that as compared with regular MLE's, the expansions for the estimators and for their bias are not regular. In particular, the bias expansions for these estimators contain additional terms which do not exist in the regular case. We have found explicit expressions for these terms for general long memory models and calculated them in the simple ARFIMA$(0, d, 0)$ case. For this model, the additional bias term is in fact dominant, causing an inflation in the bias of the PMLE of up to 3-4 times the magnitude of the bias of the EMLE.

While the interest in this paper is in long-memory models, in principle, the PMLE and WMLE expansions can be used for models of short- or intermediate-memory. In the long-memory case the motivation behind using the PMLE is the degeneracy of the asymptotic Fisher information matrix of the full parameter vector (μ, θ). This breakdown does not occur in the short memory case and so it would be difficult to justify the use of the PMLE and with it the need to conduct a high order analysis for it when one can simply use the true, exact MLE. However, this argument clearly does not extend for the WMLE, which is a popular estimator by its own merits. For the WMLE, the results of the paper can be used for processes of short- and intermediate-memory and to our knowledge, similar results are not available yet in the literature.

REFERENCES

Adenstedt, R.K. (1974) On Large-sample estimation for the mean of a stationary random sequence. *Annals of Statistics* 2, 1095–1107.

Andrews, D.W.K. & O. Lieberman (2002a) Valid Edgeworth expansions for the Whittle maximum likelihood estimator for stationary long-memory Gaussian time series. Cowles Foundation Discussion Paper No. 1361, Yale University. Available at http://cowles.econ.yale.edu.

Andrews, D.W.K. & O. Lieberman (2002b) Higher-order improvement of the parametric bootstrap for long-memory Gaussian processes. Cowles Foundation Discussion Paper No. 1378, Yale University. Available at http://cowles.econ.yale.edu.

Cox, D.R. & E.J. Snell (1968) A general definition of residuals (with discussion). *Journal of the Royal Statistical Society B*, 30, 248-275.

Dahlhaus, R. (1989) Efficient parameter estimation for self-similar processes. *Annals of Statistics* 17, 1749–1766.

Cheung, Y.W. & F.X. Diebold (1994) On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean. *Journal of Econometrics* 62, 301-316.

Fox, R. & M.S. Taqqu (1986) Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics* 14, 517–532.

Gradshteyn, I.S. & I.M. Rhyzhik (1980) *Table of Integrals, Series, and Products.* New York: Academic Press.

Hurst, H.E. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770-808.

Lawley, D.N. (1956) A general method for approximating to the distribution of likelihood ratio criteria. *Biometrika* 43, 295-303.

Lieberman, O. (2001) Penalised maximum likelihood estimation for fractional Gaussian processes. *Biometrika* 88, 888-894.

Lieberman, O. & P.C.B. Phillips (2001) Second order expansions for the distribution of the maximum likelihood estimator of the fractional difference parameter. Cowles Foundation Discussion Paper No. 1308, Yale University. Available at http://cowles.econ.yale.edu.

Lieberman, O. & P.C.B. Phillips (2003) Error bounds and asymptotic expansions for Toeplitz product functionals of unbounded spectra. Accepted for publication in *Journal of Time Series Analysis.*

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

Lieberman, O., J. Rousseau, and D.M. Zucker (2001) Valid Edgeworth Expansions for the Sample Autocorrelation Function under Long Range Dependence. *Econometric Theory* 17, 257–275.

Lieberman, O., J. Rousseau, and D.M. Zucker (2003) Valid Edgeworth Expansions for the Maximum Likelihood Estimator of the Parameter of a Stationary, Gaussian, Strongly Dependent Process. *Annals of Statistics* 31, 586-612.

McCullagh, P. (1987) *Tensor Methods in Statistics*. New York: Chapman and Hall.

Appendix

Lemmas 7 and 8 below are used throughout. Define

$$\begin{aligned} d_r &= \tilde{\kappa}_r - \kappa_r; d_{rs} = \tilde{\kappa}_{rs} - \kappa_{rs}; d_{r,s} = \tilde{\kappa}_{r,s} - \kappa_{r,s} \\ d_{rst} &= \tilde{\kappa}_{rst} - \kappa_{rst}; d_{rs,t} = \tilde{\kappa}_{rs,t} - \kappa_{rs,t} \end{aligned} \tag{35}$$

and so on. Let

$$a_{r,s} = \frac{2}{n} tr \left\{ (-P)^{\xi_1} B_r (-P)^{\xi_2} T (-P)^{\xi_3} B_s (-P)^{\xi_4} T \right\}, \tag{36}$$

each ξ_i is either 0 or 1, $1 \leq \sum \xi_i \leq 4$, $(-P)^0 = I$ and the summation is over all possible configurations $(\xi_1, \xi_2, \xi_3, \xi_4)$, except for $(0,0,0,0)$. Similarly, we define

$$a_{st,u} = \frac{2}{n} tr \left\{ (-P)^{\xi_1} B_{st} (-P)^{\xi_2} T (-P)^{\xi_3} B_u (-P)^{\xi_4} T \right\}$$

where the ξ_i's satisfy the same conditions as above.

Lemma 7 Under Assumption 1

1. $\tilde{\kappa}_r = d_r = n^{-1} tr \left(-PB_r T - B_r PT + PB_r PT \right).$
2. $d_{rs} = n^{-1} tr \left(-PB_{rs} T - B_{rs} PT + PB_{rs} PT \right).$
3. $\tilde{\kappa}_{r,s} = -\tilde{\kappa}_{rs} + d_{rs} + a_{r,s} + n d_r d_s \neq -\tilde{\kappa}_{rs}.$
4. $d_{r,s} = a_{r,s} + n d_r d_s.$
5. $d_{rst} = n^{-1} tr \left(-PB_{rst} T - B_{rst} PT + PB_{rst} PT \right).$
6. $d_{st,u} = -n \kappa_{s,t} d_u + a_{st,u} + n d_{st} d_u.$

Proof of Lemma 7 We obtain from (12)

$$\tilde{U}_r = A_r + x' M B_r M x$$

and

$$\begin{aligned} \tilde{\kappa}_r &= \frac{1}{n} E_\theta \tilde{U}_r \\ &= \frac{1}{n} E_\theta \left(A_r + x' M B_r M x \right) \\ &= \kappa_r + d_r \\ &= d_r, \end{aligned} \tag{37}$$

where

$$d_r = \frac{1}{n} tr \left(-PB_r T - B_r PT + PB_r PT \right). \tag{38}$$

This establishes part 1 of the lemma. For part 2

$$\tilde{U}_{rs} = A_{rs} + x' M B_{rs} M x,$$

and

$$\begin{aligned} \tilde{\kappa}_{rs} &= \frac{1}{n} E_\theta \tilde{U}_{rs} \\ &= \frac{1}{n} \left\{ A_{rs} + tr \left(M B_{rs} M T \right) \right\} \\ &= \kappa_{rs} + d_{rs} \\ &= -\kappa_{r,s} + d_{rs}, \end{aligned} \tag{39}$$

where

$$d_{rs} = \frac{1}{n} tr\left(-PB_{rs}T - B_{rs}PT + PB_{rs}PT\right). \tag{40}$$

Similarly,

$$\begin{aligned} \tilde{\kappa}_{rst} &= \frac{1}{n} E_\theta \tilde{U}_{rst} \\ &= \kappa_{rst} + d_{rst} \end{aligned}$$

with

$$d_{rst} = \frac{1}{n} tr\left(-PB_{rst}T - B_{rst}PT + PB_{rst}PT\right),$$

yielding part 5. Further,

$$\begin{aligned} \tilde{\kappa}_{r,s} &= \frac{1}{n} E_\theta \left(\tilde{U}_r \tilde{U}_s\right) \\ &= \frac{1}{n} Cov_\theta \left(\tilde{U}_r, \tilde{U}_s\right) + n\tilde{\kappa}_r \tilde{\kappa}_s. \end{aligned} \tag{41}$$

Now, in view of (36),

$$\begin{aligned} Cov_\theta \left(\tilde{U}_r, \tilde{U}_s\right) &= 2tr\left\{\left(MB_rMT\right)\left(MB_sMT\right)\right\} \\ &= Cov_\theta \left(U_r, U_s\right) + na_{r,s} \\ &= E_\theta \left(U_r U_s\right) + na_{r,s} \\ &= n\left(\kappa_{r,s} + a_{r,s}\right), \end{aligned} \tag{42}$$

It follows from (37)-(39) and (41)-(42) that

$$\begin{aligned} \tilde{\kappa}_{r,s} &= \kappa_{r,s} + a_{r,s} + nd_r d_s \\ &= -\kappa_{rs} + a_{r,s} + nd_r d_s \\ &= -\tilde{\kappa}_{rs} + d_{rs} + a_{r,s} + nd_r d_s \\ &\neq -\tilde{\kappa}_{rs}, \end{aligned} \tag{43}$$

which proves part 3. It also follows from (35) and (43) that

$$d_{r,s} = a_{r,s} + nd_r d_s,$$

giving part 4. The remaining null cumulant we require for our analysis is

$$\begin{aligned} n\tilde{\kappa}_{st,u} &= E_\theta \left(\tilde{U}_{st} \tilde{U}_u\right) \\ &= Cov_\theta \left(\tilde{U}_{st}, \tilde{U}_u\right) + E_\theta \left(\tilde{U}_{st}\right) E_\theta \left(\tilde{U}_u\right). \end{aligned} \tag{44}$$

Here

$$\begin{aligned} Cov_\theta \left(\tilde{U}_{st}, \tilde{U}_u\right) &= 2tr\left\{\left(MB_{st}MT\right)\left(MB_uMT\right)\right\} \\ &= Cov_\theta \left(U_{st}, U_u\right) + na_{st,u} \\ &= E_\theta \left(U_{st} U_u\right) + na_{st,u} \\ &= n\left(\kappa_{st,u} + a_{st,u}\right). \end{aligned} \tag{45}$$

We see from (44), (45) and the definition of d_{st} that

$$\begin{aligned} n\tilde{\kappa}_{st,u} &= n\left(\kappa_{st,u} + a_{st,u}\right) + n^2 \tilde{\kappa}_{st} \tilde{\kappa}_u \\ &= n\left(\kappa_{st,u} + a_{st,u}\right) + n^2 \left(\kappa_{st} + d_{st}\right) d_u \end{aligned}$$

or

$$\tilde{\kappa}_{st,u} = \kappa_{st,u} + d_{st,u},$$

with

$$d_{st,u} = a_{st,u} - n\kappa_{s,t} d_u + nd_{st} d_u,$$

showing 6. □

Lemma 8 Under Assumption 1, For each $\delta > 0$,

1. $d_r = O\left(n^{-1+\delta}\right)$.
2. $d_{rs} = O\left(n^{-1+\delta}\right)$.
3. $d_{r,s} = O\left(n^{-1+\delta}\right)$.
4. $d_{rst} = O\left(n^{-1+\delta}\right)$.
5. $a_{r,s} = O\left(n^{-1+\delta}\right)$.
6. $a_{rs,t} = O\left(n^{-1+\delta}\right)$.
7. $d_{rs,t} = O\left(n^{\delta}\right)$.

Note apart from $d_{rs,t}$ which is $O\left(n^{\delta}\right)$, all the other terms in Lemma 8 are $O\left(n^{-1+\delta}\right)$.

Proof of Lemma 8 It will be enough to deal with d_r. Other quantities are established analogously. In view of (7) and (31)

$$|d_r| \leq \frac{1}{n^2}\left|1'B_rT1\right| + \frac{1}{n^2}\left|1'TB_r1\right| + \frac{1}{n^3}\left|1'B_r1\right|\left|1'T1\right|. \tag{46}$$

By the proof of Lemma 3 of Andrews and Lieberman (2002b), which holds under Assumption 1, there exists a $0 \leq K \leq \infty$, such that for each $\delta > 0$, $|1'B_rT1| \leq Kn^{1+\delta}$, $|1'TB_r1| \leq Kn^{1+\delta}$, $|1'B_r1| \leq Kn^{1-2d+\delta}$ and $1'T1 = n^2 Var_\theta\left(\bar{X}_n\right) = O\left(n^{1+2d}\right)$. Thus, the rhs of (41) is at most $Kn^{-1+\delta}$. $\square$

Proof of Lemma 3 The first part follows from Lemma 7 (1) and Lemma 8 (1). The second part follows from Lemma 7 (3) and Lemma 8 (1), 8 (2), 8 (5).

Proof of Proposition 4 We expand $\tilde{U}_r$ around θ_0 to get

$$\begin{aligned} 0 &= \tilde{U}_r + \frac{1}{\sqrt{n}}\tilde{U}_{rs}\tilde{\delta}^s + \frac{1}{2n}\tilde{U}_{rst}\tilde{\delta}^s\tilde{\delta}^t + O_p\left(n^{-1/2}\right) \\ &= \tilde{U}_r + \frac{1}{\sqrt{n}}\left(n\tilde{\kappa}_{rs} + \tilde{U}_{rs} - n\tilde{\kappa}_{rs}\right)\tilde{\delta}^s + \frac{1}{2n}\tilde{U}_{rst}\tilde{\delta}^s\tilde{\delta}^t + O_p\left(n^{-1/2}\right). \end{aligned} \tag{47}$$

Rearranging (47),

$$-\sqrt{n}\tilde{\kappa}_{rs}\tilde{\delta}^s = \tilde{U}_r + \frac{1}{\sqrt{n}}\left(\tilde{U}_{rs} - n\tilde{\kappa}_{rs}\right)\tilde{\delta}^s + \frac{1}{2n}\tilde{U}_{rst}\tilde{\delta}^s\tilde{\delta}^t + O_p\left(n^{-1/2}\right). \tag{48}$$

Now, (48) is equivalent to

$$\sqrt{n}\left(\kappa_{r,s} - d_{rs}\right)\tilde{\delta}^s = \tilde{U}_r + \frac{1}{\sqrt{n}}\left(\tilde{U}_{rs} - n\tilde{\kappa}_{rs}\right)\tilde{\delta}^s + \frac{1}{2n}\tilde{U}_{rst}\tilde{\delta}^s\tilde{\delta}^t + O_p\left(n^{-1/2}\right)$$

or

$$\begin{aligned} \tilde{\delta}^r &= \frac{1}{\sqrt{n}}\kappa^{r,s}\tilde{U}_s + \frac{1}{n}\kappa^{r,s}\left(\tilde{U}_{st} - n\tilde{\kappa}_{st}\right)\left(\frac{1}{\sqrt{n}}\kappa^{t,u}\tilde{U}_u + O_p\left(n^{-1/2}\right)\right) \\ &\quad + \frac{1}{2\sqrt{n}}\kappa^{r,s}\tilde{\kappa}_{stu}\left(\frac{1}{\sqrt{n}}\kappa^{t,j}\tilde{U}_j + O_p\left(n^{-1/2}\right)\right)\left(\frac{1}{\sqrt{n}}\kappa^{u,k}\tilde{U}_k + O_p\left(n^{-1/2}\right)\right) \\ &\quad + \kappa^{r,s}d_{st}\left(\frac{1}{\sqrt{n}}\kappa^{t,u}\tilde{U}_u + O_p\left(n^{-1/2}\right)\right) + O_p\left(n^{-1}\right) \\ &= \frac{1}{\sqrt{n}}\kappa^{r,s}\tilde{U}_s + \frac{1}{n^{3/2}}\kappa^{r,s}\kappa^{t,u}\left(\tilde{U}_{st} - n\tilde{\kappa}_{st}\right)\tilde{U}_u + \frac{1}{2n^{3/2}}\kappa^{r,s}\kappa^{t,j}\kappa^{u,k}\tilde{\kappa}_{stu}\tilde{U}_j\tilde{U}_k \\ &\quad + \frac{1}{\sqrt{n}}\kappa^{r,s}\kappa^{t,u}d_{st}\tilde{U}_u + O_p\left(n^{-1}\right). \end{aligned} \tag{49}$$

By the definition of d_{rst} and Lemma 8(4), (49) is equivalent to (14). The term involving d_{st} is $O\left(n^{-1+\delta}\right)$ by Lemma 8(2).□

Proof of eq'n (15) By parts 1,2 of Lemma 8,

$$E_\theta\left(\kappa^{r,s}\kappa^{t,u}d_{st}\tilde{U}_u/\sqrt{n}\right) = \sqrt{n}\kappa^{r,s}\kappa^{t,u}d_{st}\tilde{\kappa}_u = O\left(n^{-3/2+\delta}\right). \tag{50}$$

Taking expectations of (14), using Lemmas 7 and 8, we get

$$\begin{aligned}
E_\theta\left(\tilde{\theta}-\theta_0\right)^r &= \kappa^{r,s}\tilde{\kappa}_r + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\tilde{\kappa}_{st,u} + \frac{1}{n^2}\kappa^{r,s}\kappa^{t,u}\left(-n\tilde{\kappa}_{st}\right)\left(n\tilde{\kappa}_u\right) \\
&\quad + \frac{1}{2n}\kappa^{r,s}\kappa^{t,j}\kappa^{u,k}\kappa_{stu}\tilde{\kappa}_{j,k} + O\left(n^{-3/2+\delta}\right) \\
&= \kappa^{r,s}\tilde{\kappa}_r + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(\kappa_{st,u} - n\kappa_{s,t}d_u + O\left(n^{-1+\delta}\right)\right) \\
&\quad + \kappa^{r,s}\kappa^{t,u}\left(\kappa_{s,t} + O\left(n^{-1+\delta}\right)\right)\tilde{\kappa}_u + \frac{1}{2n}\kappa^{r,s}\kappa^{t,j}\kappa^{u,k}\kappa_{stu}\kappa_{j,k} + O\left(n^{-3/2+\delta}\right) \\
&= \kappa^{r,s}\tilde{\kappa}_r + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\kappa_{st,u} - \kappa^{r,s}d_s \\
&\quad + \kappa^{r,s}\tilde{\kappa}_s + \frac{1}{2n}\kappa^{r,s}\kappa^{t,j}\kappa^{u,k}\kappa_{stu}\kappa_{j,k} + O\left(n^{-3/2+\delta}\right) \\
&= \kappa^{r,s}\tilde{\kappa}_s + \frac{1}{n}\kappa^{r,s}\kappa^{t,u}\left(\kappa_{st,u} + \kappa_{stu}/2\right) + O\left(n^{-3/2+\delta}\right).\square
\end{aligned}$$

Proof of Lemma 5

The score of (5) and its expected value are easily shown to be

$$U_{W,r} = -n\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\dot{f}_r(\lambda)}{f(\lambda)}d\lambda + \frac{1}{2n}x'M\dot{T}_{W,r}Mx\right\} \tag{51}$$

and

$$\kappa_{W,r} = \frac{1}{n}E_\theta\left(U_{W,r}\right) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\dot{f}_r(\lambda)}{f(\lambda)}d\lambda - \frac{1}{2n}trM\dot{T}_{W,r}MT, \tag{52}$$

respectively. It is clear that $\kappa_{W,r}$ is not zero for finite n so that the first-order Bartlett identity does not hold. However,

$$-\frac{1}{2n}tr\left(M\dot{T}_{W,r}MT\right) = -\frac{1}{2n}tr\left(\dot{T}_{W,r}T - 2P\dot{T}_{W,r}T + P\dot{T}_{W,r}PT\right),$$

where P is defined in (13). Theorem 3(a) of Andrews and Lieberman (2002a) imply that for each $\delta > 0$

$$tr\left(P\dot{T}_{W,r}T\right) = O\left(n^\delta\right).$$

In addition,

$$tr\left(P\dot{T}_{W,r}PT\right) = \frac{1}{n^2}\left(1'\dot{T}_{W,r}1\right)\left(1'T1\right).$$

Now, $1'T1/n^2$ is the variance of $\bar{X}_n$ which is $O\left(n^{-1+2d}\right)$ and from the proof of Theorem 3(b) of Andrews and Lieberman (2002a, p 20), $1'\dot{T}_{W,r}1 = O\left(n^{1-2d+\delta}\right)$, $\forall\delta > 0$. Finally, By Theorem 5 of Lieberman and Phillips (2003), $\forall\delta > 0$

$$\begin{aligned}
-\frac{1}{2n}tr\left(\dot{T}_{W,r}T\right) &= -\frac{1}{2}\frac{1}{4\pi^2}2\pi\int_{-\pi}^{\pi}\left(-f^{-2}(\lambda)\dot{f}_r(\lambda)\right)f(\lambda)d\lambda + O\left(n^{-1+\delta}\right) \\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}\frac{\dot{f}_r(\lambda)}{f(\lambda)}d\lambda, +O\left(n^{-1+\delta}\right).
\end{aligned}$$

Thus,

$$\kappa_{W,r} = O\left(n^{-1+\delta}\right), \forall \delta > 0, \tag{53}$$

establishing the first part of the lemma.

Continuing,

$$\begin{aligned} U_{W,rs} &= -n\left\{\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[-f^{-2}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda)+f^{-1}(\lambda)\ddot{f}_{rs}(\lambda)\right]+\frac{1}{2n}x'M\ddot{T}_{W,r}Mx\right\} \\ &= -\frac{n}{4\pi}\int_{-\pi}^{\pi}\left[-f^{-2}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda)+f^{-1}(\lambda)\ddot{f}_{rs}(\lambda)\right]d\lambda \\ &\quad -\frac{1}{2}x'M\ddot{T}_{W,rs}Mx \end{aligned}$$

and

$$\begin{aligned} \kappa_{W,rs} &= \frac{1}{n}E_\theta\left(U_{W,rs}\right) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[-f^{-2}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda)+f^{-1}(\lambda)\ddot{f}_{rs}(\lambda)\right]d\lambda \\ &\quad -\frac{1}{2}tr\left(M\ddot{T}_{W,rs}MT\right). \end{aligned} \tag{54}$$

Also,

$$\begin{aligned} \kappa_{W,r,s} &= \frac{1}{n}E_\theta\left(U_{W,r}U_{W,s}\right) \\ &= \frac{1}{n}Cov_\theta\left(U_{W,r},U_{W,s}\right)+\frac{1}{n}E_\theta\left(U_{W,r}\right)E_\theta\left(U_{W,s}\right) \\ &= \frac{1}{n}\left(-\frac{1}{2}\right)^2 2tr\left\{\left(M\dot{T}_{W,r}MT\right)\left(M\dot{T}_{W,s}MT\right)\right\}+n\kappa_{W,r}\kappa_{W,s} \\ &= \frac{1}{2n}tr\left\{\left(M\dot{T}_{W,r}MT\right)\left(M\dot{T}_{W,s}MT\right)\right\}+n\kappa_{W,r}\kappa_{W,s}. \end{aligned} \tag{55}$$

Eq'ns (54) and (55) show that the second-order Bartlett identity does not hold either. By (53), the second summand in (55) is $O\left(n^{-1+\delta}\right)$, $\forall \delta > 0$. This fact, together with Proposition 2 and Theorem 3 of Andrews and Lieberman (2002a) leads to

$$\begin{aligned} \lim_{n\to\infty}\kappa_{W,rs} &= -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left[-f^{-2}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda)+f^{-1}(\lambda)\ddot{f}_{rs}(\lambda)\right]d\lambda \\ &\quad -\frac{1}{2}\frac{1}{4\pi^2}2\pi\int_{-\pi}^{\pi}\left(2f^{-3}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda)-f^{-2}(\lambda)\ddot{f}_{rs}(\lambda)\right)f(\lambda)\,d\lambda \\ &= -\frac{1}{4\pi}\int_{-\pi}^{\pi}f^{-2}(\lambda)\dot{f}_s(\lambda)\dot{f}_r(\lambda). \end{aligned} \tag{56}$$

and

$$\begin{aligned} \lim_{n\to\infty}\kappa_{W,r,s} &= \frac{1}{2}\left(\frac{1}{4\pi^2}\right)^2(2\pi)^3\int_{-\pi}^{\pi}\left(-f^{-2}(\lambda)\dot{f}_r(\lambda)\right)f(\lambda)\left(-f^{-2}(\lambda)\dot{f}_s(\lambda)\right)f(\lambda)\,d\lambda \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}f^{-2}(\lambda)\dot{f}_r(\lambda)\dot{f}_s(\lambda)\,d\lambda. \end{aligned} \tag{57}$$

Eq'ns (53)-(57) complete the proof. □

Proof of Proposition 6 We expand the Whittle score as

$$\begin{aligned} 0 &= U_{W,r}+\frac{1}{\sqrt{n}}U_{W,rs}\hat{\delta}_W^s+\frac{1}{2n}U_{W,rst}\hat{\delta}_W^s\hat{\delta}_W^t+O_p\left(n^{-1/2}\right) \\ &= U_{W,r}+\frac{1}{\sqrt{n}}\left(n\kappa_{W,rs}+U_{W,rs}-n\kappa_{W,rs}\right)\hat{\delta}^s+\frac{1}{2n}U_{W,rst}\hat{\delta}_W^s\hat{\delta}_W^t+O_p\left(n^{-1/2}\right). \end{aligned}$$

This yields

$$-\sqrt{n}\kappa_{W,rs}\hat{\delta}^s = U_{W,r} + \frac{1}{\sqrt{n}}\left(U_{W,rs} - n\kappa_{W,rs}\right)\hat{\delta}^s_W + \frac{1}{2n}U_{W,rst}\hat{\delta}^s_W\hat{\delta}^t_W + O_p\left(n^{-1/2}\right).$$

Let

$$d_{W,rs} = \kappa_{W,r,s} + \kappa_{W,rs}. \tag{58}$$

Then

$$\begin{aligned}\sqrt{n}\kappa_{W,r,s}\hat{\delta}^s_W &= \sqrt{n}d_{W,rs}\hat{\delta}^s_W + U_{W,r} + \frac{1}{\sqrt{n}}\left(U_{W,rs} - n\kappa_{W,rs}\right)\hat{\delta}^s_W + \frac{1}{2n}U_{W,rst}\hat{\delta}^s_W\hat{\delta}^t_W + O_p\left(n^{-1/2}\right)\\ &= \sqrt{n}d_{W,rs}\hat{\delta}^s_W + U_{W,r} + \frac{1}{\sqrt{n}}\left(U_{W,rs} - n\kappa_{W,rs}\right)\hat{\delta}^s_W + \frac{1}{2}\kappa_{W,rst}\hat{\delta}^s_W\hat{\delta}^t_W + O_p\left(n^{-1/2}\right).\end{aligned}$$

Now, using the proof of Lemma 5,

$$\begin{aligned}|d_{W,st}| &= |\kappa_{W,st} + \kappa_{W,s,t}| \\ &\le \left|\kappa_{W,st} - \lim_{n\to\infty}\kappa_{W,st}\right| + \left|\kappa_{W,s,t} + \lim_{n\to\infty}\kappa_{W,st}\right| \\ &= \left|\kappa_{W,st} - \lim_{n\to\infty}\kappa_{W,st}\right| + \left|\kappa_{W,s,t} - \lim_{n\to\infty}\kappa_{W,s,t}\right| \\ &\le Kn^{-1+\delta}.\end{aligned} \tag{59}$$

Thus,

$$\begin{aligned}\hat{\delta}^r_W &= \kappa^{r,s}_W d_{W,st}\hat{\delta}^t_W + \frac{1}{\sqrt{n}}\kappa^{r,s}_W U_{W,s} + \frac{1}{n}\kappa^{r,s}_W\left(U_{W,st} - n\kappa_{W,st}\right)\hat{\delta}^t_W + \frac{1}{2\sqrt{n}}\kappa^{r,s}_W\kappa_{W,stu}\hat{\delta}^t_W\hat{\delta}^u_W + O_p\left(n^{-1}\right)\\ &= \kappa^{r,s}_W d_{W,st}\frac{1}{\sqrt{n}}\kappa^{t,u}_W U_{W,u} + \frac{1}{\sqrt{n}}\kappa^{r,s}_W U_{W,s} + \frac{1}{n^{3/2}}\kappa^{r,s}_W\left(U_{W,st} - n\kappa_{W,st}\right)\kappa^{t,u}_W U_{W,u}\\ &\quad + \frac{1}{2\sqrt{n}}\kappa^{r,s}_W\kappa_{W,stu}\frac{1}{\sqrt{n}}\kappa^{t,i}_W U_{W,i}\frac{1}{\sqrt{n}}\kappa^{u,j}_W U_{W,j} + O_p\left(n^{-1}\right).\end{aligned}$$

Hence, the WMLE expansion is

$$\begin{aligned}\hat{\delta}^r_W &= \frac{1}{\sqrt{n}}\kappa^{r,s}_W U_{W,s} + \frac{1}{n^{3/2}}\kappa^{r,s}_W\left(U_{W,st} - n\kappa_{W,st}\right)\kappa^{t,u}_W U_{W,u}\\ &\quad + \frac{1}{2n^{3/2}}\kappa^{r,s}_W\kappa^{t,i}_W\kappa^{u,j}_W\kappa_{W,stu}U_{W,i}U_{W,j} + \frac{1}{\sqrt{n}}\kappa^{r,s}_W\kappa^{t,u}_W d_{W,st}U_{W,u} + O_p\left(n^{-1}\right).\square\end{aligned}$$

Proof of eq'n (18) For the WMLE

$$\begin{aligned}E_\theta\left(\hat{\theta}_W - \theta\right)^r &= \kappa^{r,s}_W\kappa_{W,s} + \frac{1}{n}\kappa^{r,s}_W\kappa^{t,u}_W\kappa_{W,st,u} - \kappa^{r,s}_W\kappa_{W,st}\kappa^{t,u}_W\kappa_{W,u}\\ &\quad + \frac{1}{2n}\kappa^{r,s}_W\kappa^{t,i}_W\kappa^{u,j}_W\kappa_{W,stu}\kappa_{W,i,j} + O\left(n^{-3/2}\right).\end{aligned}$$

In the last line of the calculations we omitted $\kappa^{r,s}_W\kappa^{t,u}_W d_{W,st}\kappa_{W,u} = O\left(n^{-2+\delta}\right)$. Now

$$\begin{aligned}\kappa_{W,st,u} &= \frac{1}{n}E_\theta\left(U_{W,st}U_{W,u}\right)\\ &= \frac{1}{n}Cov_\theta\left(U_{W,st}, U_{W,u}\right) + \frac{1}{n}E_\theta\left(U_{W,st}\right)E_\theta\left(U_{W,u}\right)\\ &= \frac{1}{2n}tr\left(M\ddot{T}_{W,st}MT\right)\left(M\dot{T}_{W,u}MT\right) + n\kappa_{W,st}\kappa_{W,u}.\end{aligned} \tag{60}$$

Because of (58), (59), (60)

$$\begin{aligned}
E_\theta\left(\hat{\theta}_W-\theta\right)^r &= \kappa_W^{r,s}\kappa_{W,s}+\frac{1}{n}\kappa_W^{r,s}\kappa_W^{t,u}\left(\frac{1}{2n}tr\left(M\ddot{T}_W MT\right)\left(M\dot{T}_W MT\right)-n\kappa_{W,s,t}\kappa_{W,u}\right)\\
&\quad+\kappa_W^{r,s}\kappa_{W,s,t}\kappa_W^{t,u}\kappa_{W,u}+\frac{1}{2n}\kappa_W^{r,s}\kappa_W^{t,i}\kappa_W^{u,j}\kappa_{W,stu}\kappa_{W,i,j}+O\left(n^{-3/2}\right)\\
&= \kappa_W^{r,s}\kappa_{W,s}+\frac{1}{n}\kappa_W^{r,s}\kappa_W^{t,u}\\
&\quad\left[\left\{\frac{1}{2n}tr\left(M\ddot{T}_{W,st}MT\right)\left(M\dot{T}_{W,u}MT\right)\right\}+\kappa_{W,stu}/2\right]+O\left(n^{-3/2}\right).
\end{aligned}$$

□

A justification of (27) From (8) we learn that

$$\kappa^{r,s}=\left\{\frac{1}{2n}tr\left(T^{-1}\dot{T}\right)^2\right\}^{-1}. \tag{61}$$

By a manipulation of the third-order Bartlett identity, which holds for (2), we get

$$\kappa_{st,u}+\kappa_{stu}/2=-\left(\kappa_{s,t,u}+\kappa_{s,tu}\right)/2, \tag{62}$$

which avoids the calculation of third-order derivatives. For this model, eq'ns (7)-(8) of Lieberman (2001) show that

$$\left(\kappa_{s,t,u}+\kappa_{s,tu}\right)/2=\frac{1}{4n}tr\left(T^{-1}\dot{T}T^{-1}\ddot{T}\right). \tag{63}$$

Equations (10), (61)-(63) suggest a bias approximation to the MLE based on

$$b_{app1}\left(\hat{\theta};\theta\right)=-\frac{1}{n}\left\{\frac{1}{2n}tr\left(T^{-1}\dot{T}\right)^2\right\}^{-2}\frac{1}{4n}tr\left(T^{-1}\dot{T}T^{-1}\ddot{T}\right), \tag{64}$$

or just (27).□

A justification of (28) By Theorem 7 of Lieberman and Phillips (2003),

$$\begin{aligned}
\frac{1}{2n}tr\left(T^{-1}\dot{T}\right)^2 &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\frac{\dot{f}}{f}\right)^2 d\lambda+O\left(n^{-1/2+\delta}\right),\forall\delta>0\\
&= \frac{1}{4\pi}\int_{-\pi}^{\pi}c^2\left(\lambda\right)d\lambda+O\left(n^{-1/2+\delta}\right),\forall\delta>0\\
&= \frac{1}{4\pi}\frac{2\pi^3}{3}+O\left(n^{-1/2+\delta}\right),\forall\delta>0\\
&= \frac{\pi^2}{6}+O\left(n^{-1/2+\delta}\right),\forall\delta>0.
\end{aligned} \tag{65}$$

The second line above follows from (25) and (26) and the third line follows from a result by Gradshteyn and Rhyzik (1980). The second term needing simplification in (64) is

$$\begin{aligned}
\frac{1}{4n}tr\left(T^{-1}\dot{T}T^{-1}\ddot{T}\right) &= \frac{1}{8\pi}\int_{-\pi}^{\pi}\left(\frac{\dot{f}}{f}\right)^3 d\lambda+O\left(n^{-1/2+\delta}\right)\\
&= -\frac{1}{8\pi}\int_{-\pi}^{\pi}c^3(\lambda)d\lambda+O\left(n^{-1/2+\delta}\right)\\
&= -\frac{1}{8\pi}\left(-24\pi\zeta\left(3\right)\right)+O\left(n^{-1/2+\delta}\right)\\
&= 3\zeta\left(3\right)+O\left(n^{-1/2+\delta}\right),\forall\delta>0,
\end{aligned} \tag{66}$$

the second to last line again follows from Gradshteyn and Rhyzik (1980). Using (65) and (66) in (64), the second bias approximation to the MLE reduces to (28).□

Table 1

Bias of estimators

		n=50					n=100				
	d	.05	.15	.25	.35	.45	.05	.15	.25	.35	.45
	sim	-.0283	-.0213	-.0318	-.0365	-.0433	-.0150	-.0143	-.0137	-.0188	-.0271
EMLE	app1	-.0261	-.0282	0.0317	-.0385	-.0450	-.0131	-.0141	-.0157	-.0198	-.0313
	app2	-.0267	-.0267	-.0267	-.0267	-.0267	-.0133	-.0133	-.0133	-.0133	-.0133
	bcor	-.0016	.0054	-.0051	-.0098	-.0166	-.0017	-.0010	-.0004	-.0055	-.0193
	sim	-.0950	-.0916	-.1068	-.1182	-.1339	-.0494	-.0503	-.0528	-.0594	-.0789
	app1	-.0723	-.0752	-.0789	-.0845	-.0775	-.0401	-.0419	-.0447	-.0500	-.0580
PMLE	app2	-.0729	-.0737	-.0739	-.0727	-.0592	-.0401	-.0411	-.0421	-.0434	-.0398
	adt	-.0462	-.0470	-.0472	-.0460	-.0325	-.0270	-.0280	-.0290	-.0303	-.0267
	bcor	-.0683	-.0649	-.0801	-.0915	-.1072	-.0361	-.0307	-.0395	-.0461	-.0656
	sim	-.1030	-.0972	-.1031	-.0929	-.0886	-.0530	-.0502	-.0471	-.0435	-.0412
	app1	-.0663	-.0675	-.0678	-.0671	-.0656	-.0381	-.0387	-.0384	-.0374	-.0353
WMLE	app2	-.0829	-.0845	-.0846	-.0833	-.0808	-.0441	-.0449	-.0445	-.0429	-.0401
	adt	-.0562	-.0578	-.0579	-.0567	-.0542	-.0308	-.0315	-.0311	-.0296	-.0268
	bcor	-.0763	-.0705	-.0764	-.0662	-.0619	-.0397	-.0369	-.0338	-.0302	-.0279

		n=300					n=500				
	d	.05	.15	.25	.35	.45	.05	.15	.25	.35	.45
	sim	-.0048	-.0041	-.0041	-.0064	-.0133	-.0019	-.0013	-.0020	-.0018	-.0056
EMLE	app1	-.0044	-.0046	-.0050	-.0061	-.0131	-.0027	-.0028	-.0029	-.0035	-.0076
	app2	-.0044	-.0044	-.0044	-.0044	-.0044	-.0027	-.0027	-.0027	-.0027	-.0027
	bcor	-.0004	.0003	.0003	-.0020	-.0089	.0008	.0014	.0007	.0009	-.0029
	sim	-.0173	-.0174	-.0186	-.0224	-.0349	-.0095	-.0094	-.0119	-.0113	-.0195
	app1	-.0155	-.0163	-.0174	-.0200	-.0296	-.0099	-.0104	-.0111	-.0127	-.0199
PMLE	app2	-.0155	-.0160	-.0168	-.0183	-.0210	-.0099	-.0103	-.0109	-.0120	-.0150
	adt	-.0111	-.0116	-.0124	-.0139	-.0166	-.0072	-.0076	-.0082	-.0093	-.0123
	bcor	-.0129	-.0130	-.0142	-.0180	-.0305	-.0923	-.0067	-.0092	-.0086	-.0168
	sim	-.0184	-.0170	-.0156	-.0147	-.0253	-.0135	-.0116	-.0097	-.0057	-.0105
	app1	-.0152	-.0153	-.0149	-.0137	-.0117	-.0098	-.0099	-.0093	-.0087	-.0059
WMLE	app2	-.0163	-.0165	-.0160	-.0146	-.0124	-.0103	-.0104	-.0098	-.0091	-.0062
	adt	-.0119	-.0121	-.0115	-.0102	-.0079	-.0077	-.0078	-.0071	-.0064	-.0035
	bcor	-.014	-.0126	-.0112	-.0103	-.0209	-.0108	-.0089	-.0070	-.0030	-.0078

EMLE=Exact MLE; PMLE=plug-in MLE; WMLE=Whittle MLE; sim=results from Table 2 of Cheung and Diebold (1944); app1=first order bias term ; app2=an integral approximation to app1; adt=additional bias term due to mean estimation; bcor=bias corrected estimator.

Table 2

Mean-squared error of estimators

	EMLE			PMLE			WMLE		
d	sim	app1	app2	sim	app1	app2	sim	app1	app2
				n=50					
.05	.0156	.0132	.0129	.0280	.0178	.0179	.0335	.0200	.0190
.15	.0127	.0122	.0129	.0274	.0171	.0180	.0342	.0200	.0192
.25	.0128	.0109	.0129	.0296	.0162	.0181	.0343	.0198	.0192
.35	.0102	.0089	.0129	.0297	.0146	.0179	.0311	.0195	.0193
.45	.0079	.0045	.0129	.0302	.0085	.0161	.0307	.0191	.0187
				n=100					
.05	.0076	.0063	.0063	.0110	.0078	.0078	.0121	.0086	.0080
.15	.0062	.0060	.0063	.0103	.0076	.0078	.0111	.0086	.0081
.25	.0060	.0056	.0063	.0105	.0073	.0079	.0112	.0085	.0081
.35	.0049	.0048	.0063	.0103	.0069	.0080	.0109	.0083	.0079
.45	.0035	.0029	.0063	.0114	.0053	.0077	.0113	.0081	.0077
				n=300					
.05	.0022	.0021	.0020	.0027	.0023	.0022	.0027	.0024	.0023
.15	.0021	.0020	.0020	.0026	.0022	.0023	.0027	.0024	.0023
.25	.0020	.0019	.0020	.0026	.0022	.0023	.0027	.0023	.0023
.35	.0020	.0018	.0020	.0027	.0021	.0023	.0028	.0023	.0022
.45	.0014	.0013	.0020	.0028	.0020	.0024	.0026	.0022	.0022
				n=500					
.05	.0014	.0012	.0012	.0016	.0013	.0013	.0016	.0014	.0013
.15	.0013	.0012	.0012	.0014	.0013	.0012	.0015	.0014	.0013
.25	.0012	.0012	.0012	.0015	.0013	.0016	.0015	.0014	.0013
.35	.0011	.0011	.0012	.0014	.0013	.0013	.0015	.0013	.0013
.45	.0008	.0009	.0012	.0013	.0012	.0014	.0012	.0012	.0013

EMLE=Exact MLE; PMLE=plug-in MLE; WMLE=Whittle MLE; sim=results from Table 1 of Cheung and Diebold (1944); app1=first order mse term ; app2=an integral approximation to mse1;

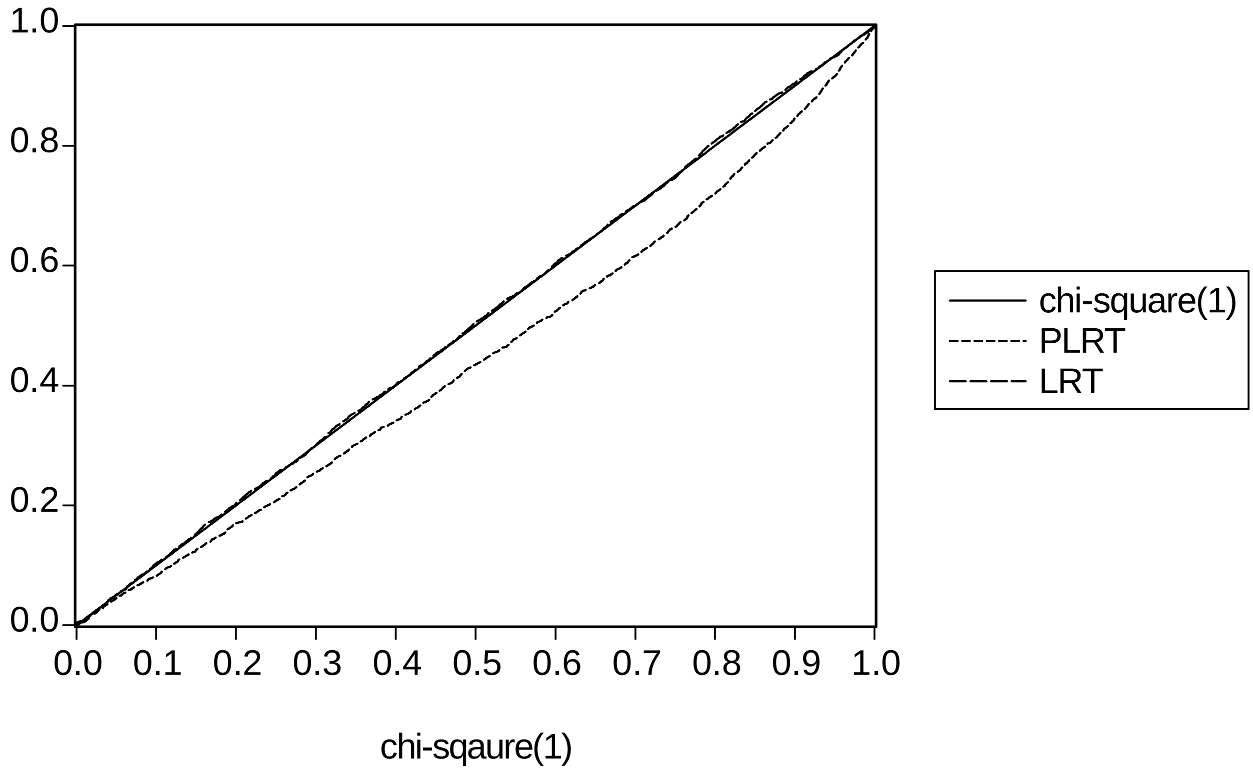

Figure 1: PP plot for the LRT and PLRT in the ARFIMA $(0, d, 0)$ model with $n = 50$ and $\mu = 0$.

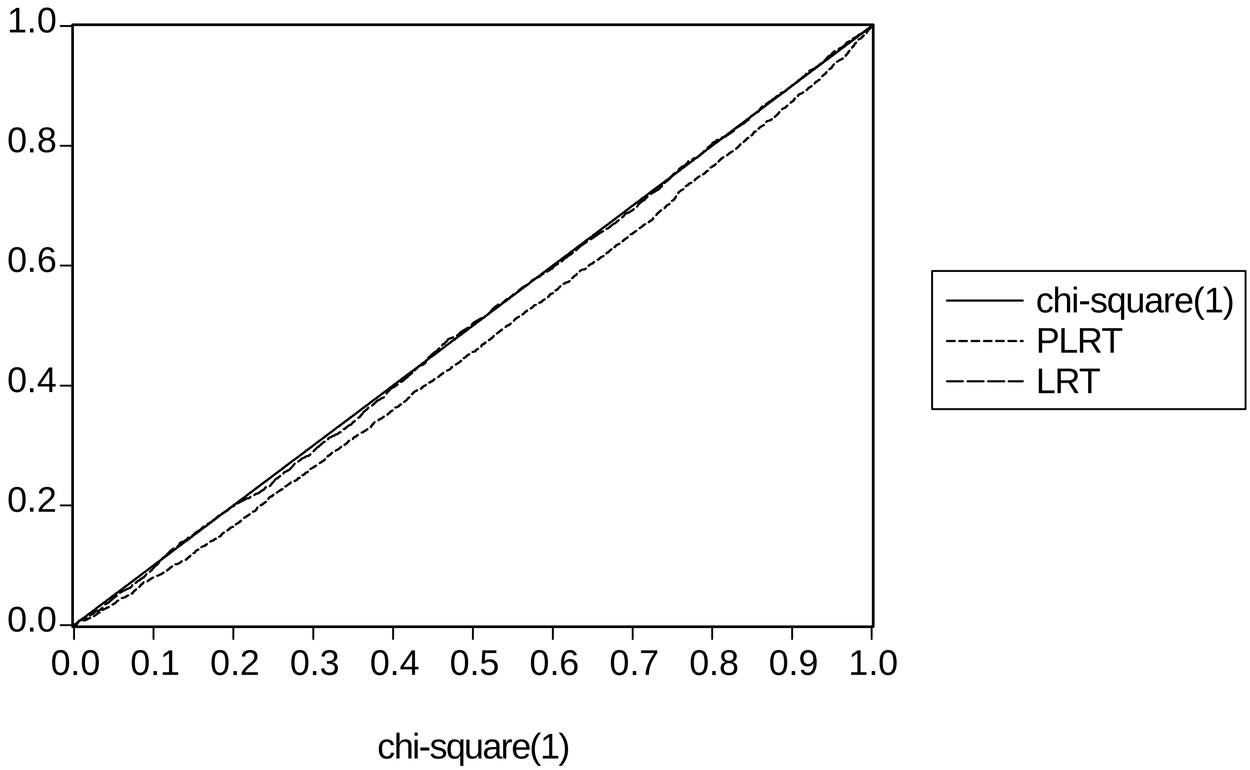

Figure 2: PP plot for the LRT and PLRT in the ARFIMA $(0, d, 0)$ model with $n = 100$ and $\mu = 0$.