CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Research Papers in Economics

Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes

By

Donald W.K. Andrews and Offer Lieberman

August 2002

COWLES FOUNDATION DISCUSSION PAPER NO. 1378

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

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

Higher-order Improvements of the Parametric Bootstrap for Long-memory Gaussian Processes

Donald W. K. Andrews[1]
Cowles Foundation for Research in Economics
Yale University

Offer Lieberman
Technion—Israel Institute of Technology and
Cowles Foundation for Research in Economics
Yale University

October 2001
Revised: August 2002

Abstract

This paper determines coverage probability errors of both delta method and parametric bootstrap confidence intervals (CIs) for the covariance parameters of stationary long-memory Gaussian time series. CIs for the long-memory parameter d_0 are included. The results establish that the bootstrap provides higher-order improvements over the delta method. Analogous results are given for tests. The CIs and tests are based on one or other of two approximate maximum likelihood estimators. The first estimator solves the first-order conditions with respect to the covariance parameters of a "plug-in" log-likelihood function that has the unknown mean replaced by the sample mean. The second estimator does likewise for a plug-in Whittle log-likelihood.

The magnitudes of the coverage probability errors for one-sided bootstrap CIs for covariance parameters for long-memory time series are shown to be essentially the same as they are with iid data. This occurs even though the mean of the time series cannot be estimated at the usual $n^{1/2}$ rate.

Keywords: Asymptotics, confidence intervals, delta method, Edgeworth expansion, Gaussian process, long memory, maximum likelihood estimator, parametric bootstrap, t statistic, Whittle likelihood.

JEL Classification Numbers: C12, C13, C15.

1 Introduction

This paper considers statistical inference for a stationary long-memory Gaussian time series with unknown mean μ_0 and spectral density f_{θ_0} that lies in a parametric family $\{f_\theta : \theta \in \Theta \subset R^{\dim(\theta)}\}$. For this situation, Dahlhaus (1989) establishes the consistency and asymptotic normality of a plug-in maximum likelihood estimator of θ_0, which maximizes the likelihood function with the unknown mean μ_0 replaced by a preliminary consistent estimator, such as the sample mean. Dahlhaus showed that this estimator is asymptotically efficient. His results allow one to construct delta method confidence intervals (CIs) and tests for elements of θ_0, including the long-memory parameter, d_0, using an asymptotic normal approximation. Fox and Taqqu (1986) provide similar results for the Whittle maximum likelihood estimator of θ_0.

In this paper, we establish the asymptotic order of magnitude of coverage probability errors and null rejection rate errors of delta method CIs and tests concerning elements of θ_0. We consider CIs and tests that are based on plug-in maximum likelihood estimators that are defined in terms of the first-order conditions (FOCs) of a plug-in log-likelihood (PLL) function or a plug-in Whittle log-likelihood (PWLL) function. We refer to these estimators as PML and PWML estimators. The PLL and PWLL functions are the Gaussian log-likelihood and Gaussian Whittle log-likelihood, respectively, with the sample mean plugged-in in place of the unknown mean.

In addition, we introduce parametric bootstrap CIs and tests for elements of θ_0 based on PML and PWML estimators and establish bounds on the asymptotic order of magnitude of the coverage probability errors and null rejection rate errors of these procedures. We show that the bootstrap yields higher-order improvements over the delta method in certain cases. To our knowledge there are no other results in the literature, even first-order results, concerning the asymptotic properties of bootstrap methods for long-memory processes.

The results of the paper cover two- and one-sided delta method CIs and t tests. They cover symmetric two-sided and one-sided parametric bootstrap CIs and tests. Both null-restricted and non-null-restricted parametric bootstrap tests are considered. The former are preferred on theoretical grounds.

The coverage probability errors of two- and one-sided delta method CIs for elements of θ_0 are shown to be $O(n^{-1})$ and $O(n^{-1/2})$, respectively, where n is the sample size. These errors are the same as for CIs in models for independent and identically distributed (iid) observations. This occurs even though the mean μ_0 cannot be estimated at the typical $n^{1/2}$ rate. Results for null rejection rate errors of delta method t tests are analogous.

The coverage probability errors of symmetric two-sided and one-sided parametric bootstrap CIs are shown to be $O(n^{-3/2}\ln(n))$, and $O(n^{-1}\ln(n))$, respectively. Apart from the $\ln(n)$ term, latter error is the same as for iid data. The error for symmetric two-sided CIs is not as small as the error $O(n^{-2})$ that has been established for many CIs in iid contexts, see Hall (1988, 1992). This may be because our bound on the error is not sharp.

The results show that symmetric two-sided and one-sided bootstrap CIs exhibit higher-order improvements in terms of coverage probabilities over their delta method counterparts of magnitude at least $n^{-1/2}\ln(n)$.

All of the bootstrap results just stated hold under a certain condition on the variance of the normalized vector of PLL or PWLL derivatives denoted Condition NS_s(iv) below. This condition holds quite generally for PWLL derivatives, but less generally for PLL derivatives. For example, it holds for all stationary Gaussian ARFIMA(p, d, q) processes for PWLL derivatives, but only for ARFIMA$(0, d, q)$ processes for PLL derivatives. If this condition does not hold, then the bounds obtained on the delta method and bootstrap CI coverage probability errors are larger.

We provide some Monte Carlo simulation results for ARFIMA$(0, d, 0)$ processes with unknown variance σ^2. The simulations show that the errors in coverage probabilities of bootstrap CIs tend to be smaller than those of delta method CIs. For example, for nominal 95% CIs, the average over five values of d of the absolute deviations of the true coverage probability from the nominal coverage probability is .016 and .010 for delta method and bootstrap CIs, respectively, based on the PML estimator. For CIs based on the PWML estimator, the corresponding average absolute deviations are .054 and .012. Hence, we conclude that the theoretical asymptotic results of the paper regarding the advantages of the parametric bootstrap over the delta method are reflected in finite samples at least in the limited number of cases considered.

We now outline the method of obtaining the asymptotic results described above. First, we obtain a valid Edgeworth expansion for the normalized vector of PLL or PWLL derivatives. For the PLL case, we do this by extending the results of Lieberman, Rousseau, and Zucker (2003) (LRZ), who consider the long-memory Gaussian case with known mean. In particular, we verify the conditions of Durbin's (1980) Theorem 1, which gives a general result for the validity of an Edgeworth expansion for the density of a normalized random vector that holds uniformly over parameter values in a compact set. Durbin's result is a generalization of an Edgeworth expansion for the density of a normalized sum of iid random variables given by Feller (1971). We convert the Edgeworth expansion for the density of the PLL derivatives into an Edgeworth expansion for their distribution function using a result of Skovgaard (1986, Cor. 3.3). For the PWLL case, we use the Edgeworth expansion derived in Andrews and Lieberman (2002).

Next, we show that t statistics based on the PML and PWML estimators can be approximated arbitrarily closely by a smooth function of a vector of PLL and PWLL derivatives of sufficiently high order. The argument follows that of Bhattacharya and Ghosh (1978, Thm. 3(b)). These results are combined to give Edgeworth expansions for the distributions of the t statistics that hold uniformly over parameter values in a compact set. These Edgeworth expansions are used to determine the coverage probability errors of delta method CIs and analogous results for delta method tests.

We then show that the uniform Edgeworth expansions for the t statistics yield Edgeworth expansions for the bootstrap t statistics because the bootstrap generating (BG) estimator lies in a neighborhood of the true value with probability that goes to one at a sufficiently fast rate as $n \to \infty$. The coefficients of the Edgeworth expansions of the bootstrap t statistics depend on the BG estimator, whereas those of the t statistics depend on the true parameter. We show that the coefficients of the Edgeworth expansions differ by $O(n^{-1/2}\ln(n))$ with probability that goes to one quickly. This implies that the difference between the Edgeworth expansions equals the order of the second term, viz., $n^{-1/2}$, times $O(n^{-1/2}\ln(n))$, which gives a difference of $O(n^{-1}\ln(n))$, on a set with probability that goes to one quickly. This result is used to show that the coverage probability error of the one-sided parametric bootstrap CI is $O(n^{-1}\ln(n))$.

Results for symmetric two-sided bootstrap CIs are obtained by a similar argument. The primary difference is that the second terms in the Edgeworth expansions are order n^{-1} terms, because the $n^{-1/2}$ terms drop out due to symmetry. In consequence, the two Edgeworth expansions differ by $O(n^{-3/2}\ln(n))$, rather than $O(n^{-1}\ln(n))$, and the coverage probability errors are similarly reduced in magnitude.

One drawback of the results of the paper is that the PML and PWML estimators considered are required to satisfy a condition called Condition C_s, which implies consistency of the estimators. We show that there exists a sequence of PML and PWML estimators that satisfies Condition C_s. But, we do not show that all PML and PWML estimators satisfy Condition C_s or that PML and PWML estimators that maximize the PLL or PWLL function necessarily satisfy Condition C_s. The same drawback occurs in the results of Bhattacharya and Ghosh (1978) concerning Edgeworth expansions of PML estimators. If there is a unique solution to the FOCs, then this is not a problem. Or, if one can show that all solutions to the FOCs satisfy Condition C_s for a given parametric specification of the spectral density function, then this is not a problem. Otherwise, one can utilize conservative bootstrap CIs, defined below, that have the property that the probability of under-coverage is less than that of delta method CIs.

Another drawback of the results is their use of the assumption of Gaussianity. If the PML or PWML estimator is $n^{1/2}$-asymptotically normal for non-Gaussian processes, then the delta method and parametric bootstrap CIs considered in the paper are asymptotically correct to first order. For linear non-Gaussian processes, the asymptotic normality of the PWML estimator is established by Giraitis and Surgaillis (1990). However, the asymptotic normality of the PML estimator has not been established for non-Gaussian processes. Since the PWML estimator is an approximate maximum likelihood estimator, the result of Giraitis and Surgaillis (1990) suggests that PML estimators also may be asymptotically normal for linear non-Gaussian processes. But, this may be difficult to prove. On the other hand, Giraitis and Taqqu (1999) show that the PWML estimator is not necessarily $n^{1/2}$-asymptotically normal for nonlinear non-Gaussian long memory processes. In any event, except in special cases, one would not expect the Gaussian parametric bootstrap to yield higher-order improvements over the delta method for non-Gaussian processes.

We conjecture that all of the results of this paper based on the PLL function extend to the case where (i) the mean μ_0 of X_i is replaced by a linear regression function $Z_i'\beta_0$ with regressor vector Z_i, (ii) no conditions are placed on the regressors except that they are non-stochastic and their number is independent of n, and (iii) the PLL function utilizes the least squares estimator of β_0, rather than the sample mean. This extension would allow for deterministic time trends and/or seasonal dummies, among other regressor variables. We have been able to show that all of the PLL-based results of the paper go through in the regression case, except Lemma 5(c). The latter needs to hold with e_n replaced by an arbitrary unit n-vector. We conjecture that it does so, but have not been able to prove it to date.

In addition to the references given above, the results of this paper are related to the parametric bootstrap results of Andrews (2001) for weakly dependent Markov processes. The results also are related to the extensive literature on block bootstraps for weakly dependent time series. For brevity, we do not provide references. Davidson (2001) considers a residual-based bootstrap for testing for cointegration with fractionally integrated processes. He analyzes the properties of this procedure by Monte Carlo, but does not provide any results regarding its first-order or higher-order asymptotic properties.

To our knowledge, the only papers in the literature that consider Edgeworth expansions for statistics based on long-memory processes, other than LRZ and Andrews and Lieberman (2002), are Lieberman, Rousseau, and Zucker (2000) and Giraitis and Robinson (2001). The former provides an Edgeworth expansion for the joint distribution of sample autocorrelations. The latter provides an Edgeworth expansion for the semiparametric local Whittle estimator. Taniguchi (1986, 1991) establishes Edgeworth expansions for (weakly dependent) Gaussian autoregressive-moving average processes.

The remainder of the paper is organized as follows. Section 2 introduces the basic model, the PML and PWML estimators, t statistics based on the PML and PWML estimators, and delta method CIs and tests. Section 3 describes the parametric bootstrap and defines the bootstrap CIs and tests that are considered in the paper. Section 4 states the assumptions. Section 5 establishes sharp bounds on the coverage probability errors of one- and two-sided delta method CIs and analogous results for delta method tests. Section 6 provides bounds on the coverage probability errors of bootstrap CIs and analogous results for bootstrap tests. These results establish the higher-order improvements of the bootstrap. Section 7 provides the Monte Carlo results. An Appendix contains proofs of the results given in Sections 5 and 6.

2 Model

We consider a discrete-time stationary Gaussian long-memory process $\{X_i : i \geq 1\}$ with mean $\mu_0 \in R$ and spectral density $f_{\theta_0}(\lambda)$ for $\lambda \in (-\pi, \pi)$. Both μ_0 and θ_0 are unknown parameters. The true spectral density $f_{\theta_0}(\lambda)$ is assumed to lie in a parametric family $\{f_\theta(\lambda) : \theta \in \Theta\}$, where $\Theta \subset R^{\dim(\theta)}$ is the parameter space for θ. The first element of θ is the long-memory parameter d. That is, $\theta = (\theta_1, \theta_2, ..., \theta_{\dim(\theta)})' = (d, \theta_2, ..., \theta_{\dim(\theta)})'$. The true value of d is denoted d_0. The long-memory feature of the spectral densities in the parametric family is captured by the following basic assumption:[2] for all $\theta \in \Theta$,

$$\begin{aligned} f_\theta(\lambda) &= O(|\lambda|^{-2d-\delta}) \text{ as } |\lambda| \downarrow 0,\ \forall \delta > 0 \text{ and} \\ d &\in (0, 1/2). \end{aligned} \tag{2.1}$$

The spectral density $f_\theta(\lambda)$ is unbounded at the origin, but $f_\theta(\lambda)$ is integrable and the process is covariance stationary, because d is restricted to $(0, 1/2)$. A process whose spectral density satisfies (2.1) exhibits long memory. An example of such a process is the ARFIMA (p, d, q) process. Additional assumptions on the parametric spectral densities $f_\theta(\lambda)$ are given in Section 4 below.

The observed sample of size n is

$$X = (X_1, ..., X_n)'. \tag{2.2}$$

The $n \times n$ (Toeplitz) covariance matrix corresponding to $f_\theta(\lambda)$ is denoted $T_n(f_\theta)$ and has (j, k) element defined by

$$T_n(f_\theta)_{j,k} = \int_{-\pi}^{\pi} e^{i(j-k)\lambda} f_\theta(\lambda) d\lambda. \tag{2.3}$$

The log-likelihood function is

$$L_n(\theta, \mu) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(\det(T_n(f_\theta))) - \frac{1}{2}(X - \mu\mathbf{1}_n)' T_n^{-1}(f_\theta)(X - \mu\mathbf{1}_n), \tag{2.4}$$

where $\mathbf{1}_n$ is an n-vector of ones.

The Whittle log-likelihood function is

$$\begin{aligned} L_{W,n}(\theta, \mu) = & -\frac{n}{2}\ln(2\pi) - \frac{n}{4\pi}\int_{-\pi}^{\pi} \ln(f_\theta(\lambda)) d\lambda \\ & -\frac{1}{2}(X - \mu\mathbf{1}_n)' T_n((2\pi)^{-2} f_\theta^{-1})(X - \mu\mathbf{1}_n). \end{aligned} \tag{2.5}$$

The Whittle log-likelihood is an approximation to the log-likelihood based on the fact that (i) $n^{-1}\ln(\det(T_n(f_\theta))) \rightarrow (2\pi)^{-1}\int_{-\pi}^{\pi}\ln(f_\theta(\lambda))d\lambda$ as $n \rightarrow \infty$ and (ii) $T_n((2\pi)^{-2} f_\theta^{-1})$ approximates the inverse of $T_n(f_\theta)$ for large n in the sense that $T_\infty(f_\theta)T_\infty((2\pi)^{-2} f_\theta^{-1}) = I_\infty$, see Beran (1994, pp. 109–110) for details.

Let $\overline{X}_n = n^{-1}\sum_{i=1}^n X_i$ denote the sample mean. We refer to $L_n(\theta, \overline{X}_n)$ and $L_{W,n}(\theta, \overline{X}_n)$ as the *plug-in log-likelihood* (PLL) and *plug-in Whittle log-likelihood* (PWLL) functions respectively. Like most papers in the literature, we utilize the PLL and PWLL functions rather than log-likelihood functions that depend on both θ and μ. There are three reasons why we do so. First, results of Dahlhaus (1989) and Fox and Taqqu (1986) imply that any consistent solution to the FOCs for the PLL or PWLL function is asymptotically efficient.[3] Second, computation using the PLL or PWLL function is simpler than with the full log-likelihood or Whittle log-likelihood because its argument is of lower dimension. Third, the asymptotic information matrix for the parameter vector $(\theta', \mu)'$ is singular in the long-memory case (when the Hessian is normalized by n^{-1}). This creates a problem when trying to obtain an Edgeworth expansion for the maximum likelihood or Whittle maximum likelihood estimator of $(\theta', \mu)'$. This problem does not arise with estimators based on the PLL or PWLL function, because the asymptotic information matrix for θ alone is nonsingular.

Let I_n denote the n by n identity matrix. Let $\mathbf{1}_n$ denote the n vector of ones.

The PLL and PWLL functions can be written as

$$
\begin{aligned}
L_n(\theta, \overline{X}_n) &= -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln(\det(T_n(f_\theta))) - \frac{1}{2}X' M_n T_n^{-1}(f_\theta) M_n X, \text{ where} \\
M_n &= I_n - P_n, \ P_n = e_n e_n', \ e_n = n^{-1/2}\mathbf{1}_n, \text{ and} \\
L_{W,n}(\theta, \overline{X}_n) &= -\frac{n}{2}\ln(2\pi) - \frac{n}{4\pi}\int_{-\pi}^{\pi} \ln(f_\theta(\lambda)) d\lambda \\
&\quad -\frac{1}{2}X' M_n T_n((2\pi)^{-2} f_\theta^{-1}) M_n X, \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{4\pi}\int_{-\pi}^{\pi} \{\ln(f_\theta(\lambda)) + f_\theta^{-1}(\lambda) I_n(\lambda)\} d\lambda, \text{ where} \\
I_n(\lambda) &= \left| \frac{1}{2\pi n}\sum_{j=1}^{n} e^{ij\lambda}(X_j - \overline{X}_n) \right|^2. \qquad (2.6)
\end{aligned}
$$

Equation (2.6) shows that the PWLL function can be written as a quadratic form in X or as a function of the periodogram $I_n(\lambda)$.

Let $\widehat{\Theta}_n$ denote the set of solutions to the FOCs of the PLL (or PWLL) function. (For notational simplicity, we do not distinguish between estimators based on the PLL and PWLL functions.) That is,

$$
\frac{\partial}{\partial \theta} L_n(\widehat{\theta}_n, \overline{X}_n) = 0 \text{ (or } \frac{\partial}{\partial \theta} L_{W,n}(\widehat{\theta}_n, \overline{X}_n) = 0) \qquad (2.7)
$$

for all $\widehat{\theta}_n \in \widehat{\Theta}_n$. If no solution to the FOCs exists, then for specificity $\widehat{\Theta}_n$ is defined to contain values that maximize the PLL (or PWLL) function (or maximize it up to some arbitrarily small constant $\varepsilon > 0$). We show below that at least one solution to the FOCs exists with probability that goes to one (at a fast rate) as $n \to \infty$. (In consequence, for the asymptotic results given below, it does not matter how one defines $\widehat{\Theta}_n$ when no solution to the FOCs exists.) Let $\widehat{\theta}_n$ denote an element of $\widehat{\Theta}_n$. We call $\widehat{\theta}_n$ a FOCs plug-in maximum likelihood (PML) estimator (or a PWML estimator).

A complete definition of $\widehat{\Theta}_n$ requires the specification of the set of parameter values θ from which one selects solutions to the FOCs. We allow for two cases. In the first case, this set is the parameter space Θ, which contains the true value θ_0 and which only contains values θ that generate stationary long-memory processes $\{X_i : i \geq 1\}$. Thus, Θ contains parameter values θ for which $d \in (0, 1/2)$. In this case, we refer to $\widehat{\Theta}_n$ as the set of stationarity-restricted PML (SR-PML) (or SR-PWML) estimators.

In the second case, the set of parameter values from which one selects solutions to the FOCs is a set Θ^+ that is larger than the parameter space Θ. The set Θ^+ may be chosen to relax the restriction that $d \in (0, 1/2)$ and allow d to take values in the non-stationary region ($d \geq 1/2$), the weak dependence region ($d = 0$), and/or the intermediate dependence region ($d < 0$). In this case, we refer to $\widehat{\Theta}_n$ as the set of unrestricted PML (UR-PML) (or UR-PWML) estimators. The reason for considering UR-PML and UR-PWML estimators is to cover the case where the researcher does not know a priori that the true value of d lies in $(0, 1/2)$ and hence constructs CIs that do not impose this constraint. (But, the asymptotic results of this paper only apply to the case where the true value of d is in $(0, 1/2)$.)

Dahlhaus (1989) and Fox and Taqqu (1986) show that consistent SR-PML, UR-PML, SR-PWML, and UR-PWML estimators are asymptotically normal and asymptotically efficient provided the true parameter θ_0 lies in the interior of Θ. They also show that the estimator that maximizes the PLL or PWLL function over Θ is consistent and, hence, is an PML or PWML estimator (provided the true parameter lies in the interior of Θ).

The asymptotic covariance matrix of a consistent PML or PWML estimator $\widehat{\theta}_n$ is $\Sigma(\theta_0)$, where

$$\Sigma(\theta) = \left(\frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{\partial}{\partial \theta} \ln(f_\theta(\lambda)) \frac{\partial}{\partial \theta'} \ln(f_\theta(\lambda)) d\lambda \right)^{-1}. \tag{2.8}$$

Provided $f_\theta(\lambda)$ is smooth with respect to θ, a consistent estimator of $\Sigma(\theta_0)$ is $\Sigma(\widehat{\theta}_n)$.

Let θ_H denote some element of Θ. Let $\theta_{0,r}$, $\theta_{H,r}$, and $\widehat{\theta}_{n,r}$ denote the r-th elements of θ_0, θ_H, and $\widehat{\theta}_n$, respectively. Let $\Sigma_{r,r}(\widehat{\theta}_n)$ denote the (r, r)-th element of $\Sigma(\widehat{\theta}_n)$.

The t statistic for testing the null hypothesis $H_0 : \theta_{0,r} = \theta_{H,r}$ is

$$t_n(\theta_{H,r}) = n^{1/2}(\widehat{\theta}_{n,r} - \theta_{H,r})/\Sigma_{r,r}^{1/2}(\widehat{\theta}_n). \tag{2.9}$$

Let z_α denote the $1 - \alpha$ quantile of the standard normal distribution.

The *two-sided delta method* CI for $\theta_{0,r}$ with (approximate) confidence level $100(1-\alpha)\%$ based on the PML estimator $\widehat{\theta}_n$ is

$$\Delta CI_2(\widehat{\theta}_n) = [\widehat{\theta}_{n,r} - z_{\alpha/2}\Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2},\ \widehat{\theta}_{n,r} + z_{\alpha/2}\Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2}]. \tag{2.10}$$

The *upper one-sided delta method* $100(1 - \alpha)\%$ CI for $\theta_{0,r}$ is

$$\Delta CI_{up}(\widehat{\theta}_n) = [\widehat{\theta}_{n,r} - z_\alpha \Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2},\ \infty). \tag{2.11}$$

Correspondingly, the two-sided delta method t test of $H_0 : \theta_{0,r} = \theta_{H,r}$ versus $H_1 : \theta_{0,r} \neq \theta_{H,r}$ with significance level α rejects H_0 if $|t_n(\theta_{H,r})| > z_{\alpha/2}$. The one-sided delta method t test of $H_0 : \theta_{0,r} \leq \theta_{H,r}$ versus $H_1 : \theta_{0,r} > \theta_{H,r}$ with significance level α rejects H_0 if $t_n(\theta_{H,r}) > z_\alpha$.

3 Parametric Bootstrap

Parametric bootstrap samples are generated using an estimator $\widetilde{\theta}_n$ of θ_0 that is referred to as the bootstrap generating (BG) estimator. We allow the BG estimator $\widetilde{\theta}_n$ to differ from the PML or PWML estimator $\widehat{\theta}_n$ that is used to construct CIs and test statistics because for bootstrap tests we want to allow the BG estimator to be a *null-restricted* estimator, as explained below.

By definition, given a BG estimator $\widetilde{\theta}_n$, the parametric bootstrap sample $X^* = (X_1^*, ..., X_n^*)'$ has conditional distribution given X that is the same as the distribution of the original sample except that the true parameters are $(\widetilde{\theta}_n, \overline{X}_n)$ rather than

(θ_0, μ_0). That is, X^* consists of stationary Gaussian random variables with mean $\overline{X}_n$ and spectral density $f_{\widetilde{\theta}_n}(\lambda)$ conditional on the original sample X.

The bootstrap PLL and PWLL functions, $L_n^*(\theta, \overline{X}_n^*)$ and $L_{W,n}^*(\theta, \overline{X}_n^*)$, are defined in the same way as the PLL and PWLL functions are defined, but with X^* and $\overline{X}_n^* = n^{-1}\sum_{i=1}^n X_i^*$ in place of X and $\overline{X}_n$, respectively. Let Θ^* denote the set of solutions in Θ or Θ^+ (depending on whether $\widehat{\Theta}_n$ is defined using solutions in Θ or Θ^+) to the FOCs for the bootstrap PLL or PWLL function. We define the bootstrap estimator θ_n^* to be the element in Θ^* that is closest to $\widehat{\theta}_n$.[4]

The bootstrap t statistic is defined such that its distribution mimics the null distribution of the t statistic even when the sample is generated by a parameter in the alternative hypothesis. This is done by centering the statistic at $\widetilde{\theta}_{n,r}$. We define the bootstrap t statistic to be

$$t_n^*(\widetilde{\theta}_{n,r}) = n^{1/2}(\theta_{n,r}^* - \widetilde{\theta}_{n,r})/\Sigma_{r,r}^{1/2}(\theta_n^*), \tag{3.1}$$

where $\theta_{n,r}^*$ denotes the r-th element of θ_n^*.

Let $z_{|t|,\alpha}^*$ and $z_{t,\alpha}^*$ denote the $1-\alpha$ quantiles of $|t_n^*(\widetilde{\theta}_{n,r})|$ and $t_n^*(\widetilde{\theta}_{n,r})$ respectively. To be precise, we define $z_{|t|,\alpha}^*$ to be a value that minimizes $|P^*(|t_n^*(\widetilde{\theta}_{n,r})| \leq z) - (1-\alpha)|$ over $z \in R$. (This definition allows for discreteness in the distribution of $|t_n^*(\widetilde{\theta}_{n,r})|$. Although the distribution of the absolute value of the parametric bootstrap t statistic undoubtedly is absolutely continuous, it is simpler to allow for discreteness than to prove absolute continuity.) The precise definition of $z_{t,\alpha}^*$ is analogous.

The *symmetric two-sided* bootstrap CI for $\theta_{0,r}$ with (approximate) confidence level $100(1-\alpha)\%$ based on $\widehat{\theta}_n$ is

$$CI_{sym}(\widehat{\theta}_n) = [\widehat{\theta}_{n,r} - z_{|t|,\alpha}^*\Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2},\ \widehat{\theta}_{n,r} + z_{|t|,\alpha}^*\Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2}]. \tag{3.2}$$

The *upper one-sided* bootstrap $100(1-\alpha)\%$ CI for $\theta_{0,r}$ is

$$CI_{up}(\widehat{\theta}_n) = [\widehat{\theta}_{n,r} - z_{t,\alpha}^*\Sigma_{r,r}^{1/2}(\widehat{\theta}_n)/n^{1/2},\ \infty). \tag{3.3}$$

Correspondingly, the symmetric two-sided bootstrap t test of $H_0 : \theta_{0,r} = \theta_{H,r}$ versus $H_1 : \theta_{0,r} \neq \theta_{H,r}$ with significance level α rejects H_0 if $|t_n(\theta_{H,r})| > z_{|t|,\alpha}^*$. The one-sided bootstrap t test of $H_0 : \theta_{0,r} \leq \theta_{H,r}$ versus $H_1 : \theta_{0,r} > \theta_{H,r}$ with significance level α rejects H_0 if $t_n(\theta_{H,r}) > z_{t,\alpha}^*$.

For bootstrap CIs, the BG estimator typically is taken to be the PML or PWML estimator upon which the CI is constructed. When constructing bootstrap tests, the bootstrap is used to generate critical values that reflect the null behavior of the test statistic whether or not the null is true. In consequence, one has two types of estimator that can be used as the BG estimator. First, one can take the BG estimator to be an estimator that does not impose the null hypothesis restriction that the r-th element of θ equals $\theta_{H,r}$. The PML and PWML estimators discussed in the previous section are examples of such estimators. In this case, the centering of the bootstrap t statistic around $\widetilde{\theta}_{n,r}$, rather than $\theta_{H,r}$, ensures that the distribution of the bootstrap test statistic mimics its null distribution.

Alternatively, when considering a bootstrap test, one can take the BG estimator to be an estimator that imposes the null hypothesis restriction that the r-th element of θ equals $\theta_{H,r}$. In this case, the PML or PWML estimator $\widehat{\theta}_n$ that is used to construct the t test statistic is necessarily different from the BG estimator $\widetilde{\theta}_n$, because the former does not impose the null hypothesis. We refer to the resulting bootstrap test as a *null-restricted parametric bootstrap* test. Examples of null-restricted estimators are PML or PWML estimators that solve the FOCs given in (2.7) with the r-th equation deleted and with the r-th element of the estimator equal to $\theta_{H,r}$. We refer to such estimators as NR-PML or NR-PWML estimators. When a null-restricted BG estimator is employed, one centers the bootstrap t statistic at $\theta_{H,r}$, which equals $\widetilde{\theta}_{n,r}$ by definition of the BG estimator.

For carrying out a bootstrap test, it is preferable to use a null-restricted BG estimator. Such an estimator guarantees that the distribution of the bootstrap sample is a null hypothesis distribution. Furthermore, results of Davidson and MacKinnon (1999) indicate that the error in test rejection probability under the null hypothesis for one-sided tests is smaller asymptotically when using a null-restricted BG estimator than when using a BG estimator that is not null-restricted (although their results are not for long-memory cases). The results given below do not demonstrate this, but it may be true. The results given below for null-restricted bootstrap tests are not necessarily sharp.

Finally, we note that for bootstrap CIs the choice between null-restricted and non-null-restricted BG estimators does not arise because there is no null hypothesis upon which to base a null-restricted BG estimator. For CIs one uses a non-null-restricted BG estimator.[5]

4 Assumptions

In this section, we state the assumptions. The assumptions are different, though similar, for the PML and PWML estimators. In consequence, we give the assumptions in two separate sections below. We also specify the parameter values for which the results hold. For example, with an ARFIMA(p, d, q) model, our results do not hold if the parameter value is one for which there are common roots to the autoregressive and moving average components of the model.

4.1 PML Assumptions

The assumptions stated below for the PML estimator are essentially those of LRZ (with their $\alpha(\theta)$ equal to our $2d$). The assumptions are strengthened versions of Dahlhaus' (1989) Assumptions A0, A2, A3, and A7-A9, which Dahlhaus used to establish the asymptotic normality of the PML estimator. Most of these assumptions control the behavior of the spectral density and its derivatives in a neighborhood of the origin. The strengthening of Dahlhaus' assumptions is necessary because Edgeworth expansions require higher-order spectral density derivatives than are necessary for asymptotic normality.

Assumptions II-VI below depend on a positive integer $s \geq 3$ that indexes the order of the PLL derivatives that are used in the Edgeworth expansions employed in the proofs of the CI coverage probability results.

Assumption I. The parameter space Θ is a subset of $R^{\dim(\theta)}$ with non-empty interior.

Assumption II. For some integer $s \geq 3$, $f_\theta(\lambda)$ is $s+1$ times continuously differentiable with respect to θ, and all of its derivatives are continuous in (λ, θ) for $\lambda \neq 0$. In addition, $f_\theta^{-1}(\lambda)$ is continuous in (λ, θ) for all $\lambda \in [0, \pi]$ and $\theta \in \Theta$.

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

$$\left| \frac{\partial^k}{\partial \lambda^k} f_\theta^{-1}(\lambda) \right| \leq c_1(\theta, \delta) |\lambda|^{2d-k-\delta}$$

for $k = 0, 1, 2$ and all $\delta > 0$, where $\theta = (d, \theta_2, ..., \theta_{\dim(\theta)})'$ and $d \in (0, 1/2)$.

Assumption IV. There exist $c_2(\theta, \delta) < \infty$ and $c_3(\theta, \delta) < \infty$ such that for all $\delta > 0$ and $\lambda \in (0, \pi)$:

(a) $|f_\theta(\lambda)| \leq c_2(\theta, \delta) |\lambda|^{-2d-\delta}$ and

(b) for all $(j_1, ..., j_k)$ with $k \leq s+1$, with duplication among the j_i allowed,

$$\left| \frac{\partial^k}{\partial \theta_{j_1} ... \partial \theta_{j_k}} f_\theta^{-1}(\lambda) \right| \leq c_3(\theta, \delta) |\lambda|^{2d-\delta}.$$

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

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

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

for $\lambda \in (0, \pi)$. For any compact subset Θ_c of Θ, there exists a constant $\widetilde{C}(\Theta_c) < \infty$ such that $c_4(\theta) \leq \widetilde{C}(\Theta_c)$ for all $\theta \in \Theta_c$.

(b) When computing derivatives of the form $(\partial^k/\partial\theta_{j_1} ... \partial\theta_{j_k}) \gamma_\theta(u)$ for $k \leq s+1$ and $u = 0, 1, ...$, the derivatives may be taken inside the integral sign of (2.3), where $\gamma_\theta(u) = E_\theta(X_i - \mu_0)(X_{i+u} - \mu_0)$ and E_θ denotes expectation when the true parameter is θ.

See LRZ for a discussion of Assumptions I-VI.

As noted in LRZ, Assumptions I-VI hold for ARFIMA (p, d, q) processes for all $s \geq 3$.

4.2 PWML Assumptions

The assumptions stated below for the PWML estimator are those of Andrews and Lieberman (2002) (with their $\alpha(\theta)$ equal to our $2d$). The assumptions are strengthened versions of Fox and Taqqu's (1986) Assumptions A.1-A.5, which Fox and Taqqu used to establish the asymptotic normality of the PWML estimator. As with the PML assumptions, most of these assumptions control the behavior of the spectral density and its derivatives in a neighborhood of the origin.

The assumptions below depend on a positive integer $s \geq 3$ that indexes the order of the PWLL derivatives that are used in the Edgeworth expansions employed in the proofs of the CI coverage probability results.

Assumption W1. The parameter space Θ is a subset of $R^{\dim(\theta)}$ with non-empty interior.

Assumption 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.

Assumption 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 $\forall \delta > 0\ \exists c_1(\theta, \delta) < \infty$ such that

$$|f_\theta(\lambda)| \leq c_1(\theta, \delta) |\lambda|^{-2d-\delta}$$

for all λ in a neighborhood N_δ of the origin, where $\theta = (d, \theta_2, ..., \theta_{\dim(\theta)})'$ and $d \in (0, 1/2)$

Assumption W4. For all $(j_1, \ldots, j_k)$ with $k \leq s+1$ and $j_i \in \{1, ..., d_\theta\}$, $(\partial^k / (\partial\theta_{j_1} \cdots \partial\theta_{j_k})) f_\theta^{-1}(\lambda)$ is continuous at all (λ, θ) and $\forall \delta > 0\ \exists c_2(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial^k f_\theta^{-1}(\lambda)}{\partial\theta_{j_1} \cdots \partial\theta_{j_k}} \right| \leq c_2(\theta, \delta) |\lambda|^{2d-\delta}, \ \forall \lambda \in N_\delta.$$

Assumption W5. $(\partial/\partial\lambda) f_\theta(\lambda)$ is continuous at all (λ, θ) for which $\lambda \neq 0$ and $\forall \delta > 0\ \exists c_4(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial f_\theta(\lambda)}{\partial\lambda} \right| \leq c_4(\theta, \delta) |\lambda|^{-2d-1-\delta}, \ \forall \lambda \in N_\delta.$$

Assumption W6. For all $(j_1, \ldots, j_k)$ with $k \leq s+1$ and $j_i \in \{1, ..., d_\theta\}$, $(\partial^{k+1} / (\partial\lambda \partial\theta_{j_1} \cdots \partial\theta_{j_k})) f_\theta^{-1}(\lambda)$ is continuous at all (λ, θ) for which $\lambda \neq 0$ and $\forall \delta > 0$ $\exists c_5(\theta, \delta) < \infty$ such that

$$\left| \frac{\partial^{k+1} f_\theta^{-1}(\lambda)}{\partial\lambda \partial\theta_{j_1} \cdots \partial\theta_{j_k}} \right| \leq c_5(\theta, \delta) |\lambda|^{2d-1-\delta}, \ \forall \lambda \in N_\delta.$$

Assumption W7. For any compact subset $\overline{\Theta}$ of Θ there exists a constant $C\left(\overline{\Theta},\delta\right)<\infty$ such that the constants $c_i\left(\theta,\delta\right)$ for $i=1,...,6$ and $c_3(\theta)$ are bounded by $C\left(\overline{\Theta},\delta\right)$ $\forall\theta\in\overline{\Theta}$, $\forall\delta>0$.

See Andrews and Lieberman (2002) for a discussion of Assumptions W1-W7.

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

4.3 Parameter Values

We now specify the parameter values θ for which we establish higher-order improvements of the parametric bootstrap.

We only obtain such results for parameter values that are in the interior of Θ and for which the asymptotic covariance matrix, $\Sigma(\theta)$, of the PML or PWML estimator is nonsingular. This is not surprising because, for parameter values that do not satisfy these conditions, the PML or PWML estimator is not asymptotically normal. Thus, 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* parameter values for which we establish higher-order improvements. Rather than excluding such parameter values from the parameter space Θ, which would be unnatural and 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 higher-order improvements.

Next, we introduce some additional notation. Let $\overline{Z}_n(\theta)$ denote $2n$ times the vector of all LLDs, $D_\nu L_n(\theta)$, or WLLDs, $D_\nu L_n^W(\theta)$, up to order $s-1$. (See (8.1) and (8.2) of the Appendix for the form of these partial derivatives for the PLL case. See (8.3) and (8.4) of the Appendix for the PWLL case.) Let $\overline{D}_n(\theta)$ denote the covariance matrix of $n^{-1/2}\overline{Z}_n(\theta)$ when the true parameter is θ. The (j,k) element of $\overline{D}_n(\theta)$ is

$$\begin{aligned}\overline{D}_n\left(\theta\right)_{j,k} &= \frac{2}{n}tr\{M_nB_{n,\nu_j}(\theta)M_nT_n\left(f_\theta\right)M_nB_{n,\nu_k}(\theta)M_nT_n\left(f_\theta\right)\},\text{ where}\\ B_{n,\nu_j}(\theta) &= (-1/2)D_{\nu_j}T_n^{-1}(f_\theta)\text{ for the LLDs,}\\ B_{n,\nu_j}(\theta) &= (-1/2)D_{\nu_j}T_n((2\pi)^{-2}f_\theta^{-1})\text{ for the WLLDs,}\end{aligned} \tag{4.1}$$

and D_{ν_j} denotes partial differentiation with respect to some indices, such as $(\partial^{j_1+\cdots+j_d}/\partial\theta_{j_1}\cdots\partial\theta_{j_d})$. By Lemma 4 in the Appendix for LLDs and by Andrews and Lieberman (2002, eqn. (9)) for WLLDs, the asymptotic covariance matrix $\overline{D}(\theta)$ $(=\lim_{n\to\infty}\overline{D}_n(\theta))$ of $n^{-1/2}\overline{Z}_n(\theta)$ exists and its (j,k) element is

$$\overline{D}\left(\theta\right)_{j,k}=\frac{1}{\pi}\int_{-\pi}^{\pi}\{D_{\nu_j}f_\theta^{-1}\left(\lambda\right)\}\{D_{\nu_k}f_\theta^{-1}(\lambda)\}f_\theta^2\left(\lambda\right)d\lambda. \tag{4.2}$$

Given any sub-vector $Z_n(\theta)$ of $\overline{Z}_n(\theta)$, 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 higher-order improvements for the parametric bootstrap that hold uniformly over compact sets that lie in any set $\widetilde{\Theta}\subset\Theta$ that satisfies the following "nonsingularity" condition:

Condition NS$_s$. (i) $\widetilde{\Theta}$ is an open subset of Θ.

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

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

(iv) For n sufficiently large, the finite-sample covariance matrix of any sub-vector of $n^{-1/2}\overline{Z}_n(\theta)$ that strictly contains $n^{-1/2}Z_n(\theta)$ is singular for all $\theta \in \widetilde{\Theta}$.

Condition NS$_s$ depends on s because $\overline{Z}_n(\theta)$ includes all LLDs or WLLDs up to order $s-1$. Conditions NS$_s$(i) and NS$_s$(ii) restrict consideration to parameter values for which the PML or PWML estimator is asymptotically normal.

Condition NS$_s$(iii) requires that the same LLDs or WLLDs are linearly independent asymptotically for all parameter values in $\widetilde{\Theta}$. This is not very restrictive because our results hold for different choices of the set $\widetilde{\Theta}$ and there is a finite number of different sub-vectors $Z_n(\theta)$ of $\overline{Z}_n(\theta)$ that might be the $Z_n(\theta)$ vector that arises in Condition NS$_s$(iii).

Condition NS$_s$(iv) requires certain finite-sample covariance matrices of LLDs or WLLDs to be singular whenever the corresponding asymptotic covariance matrix is singular. For WLLDs, this always occurs because the finite-sample covariance matrix of the WLLDs closely mirrors the asymptotic covariance matrix, see Andrews and Lieberman (2002) for details. Hence, for WLLDs, Condition NS$_s$(iv) always holds.

On the other hand, for LLDs, Condition NS$_s$(iv) does not always hold and the condition can be restrictive. For example, in an ARFIMA(p,d,q) models with $p \geq 1$, Condition NS$_s$(iv) generally fails whenever $s \geq 4$. This occurs because the third-order derivative of the reciprocal of the spectral density with respect to the lag-one autoregressive parameter is zero, which causes the asymptotic covariance matrix of any set of LLDs that includes this one to be singular, but the finite-sample covariance matrix is not singular. In consequence, the results given below only hold for $s = 3$ in ARFIMA(p,d,q) models with $p \geq 1$. This yields weaker higher-order improvement results for the PML-based parametric bootstrap than are available for models that satisfy Condition NS$_s$(iv), such as ARFIMA$(0,d,q)$ models. The results are also weaker than those that are obtained for the PWML-based parametric bootstrap for ARFIMA(p,d,q) models with any values (p,d,q).

Let
$$W_n(\theta) = n^{-1/2}(Z_n(\theta) - E_\theta Z_n(\theta)), \tag{4.3}$$
where $Z_n(\theta)$ is as in Condition NS$_s$. The higher-order improvement results for the parametric bootstrap are based on an Edgeworth expansion for the vector $W_n(\theta)$ of normalized LLD's or WLD's. Denote the dimension of $Z_n(\theta)$ and $W_n(\theta)$ by d_s.

5 Coverage Probability Errors of Delta Method CIs

In this section, we establish bounds on the coverage probability errors of one- and two-sided delta method CIs based on PML and PWML estimators. These results immediately provide bounds on the errors in the null rejection rates of one- and two-sided delta method t tests.

We say that a sequence of estimators $\{\overline{\theta}_n : n \geq 1\}$ satisfies Condition C_s if for all $\varepsilon > 0$ and all compact subsets Θ_c of Θ,

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(||\overline{\theta}_n - \theta_0|| > n^{-1/2} \ln(n)\varepsilon) = o(n^{-(s-2)/2}) \text{ as } n \to \infty. \tag{5.1}$$

Condition C_s implies that $\overline{\theta}_n$ is consistent. The following Lemma shows that a sequence of PML estimators that satisfies Condition C_s exists.

In the following Lemma and elsewhere below, we make statements like "Suppose Assumptions I-VI or W1-W7 hold. Then, PML or PWML estimators satisfy ... ." By this we mean, if Assumptions I-VI hold, then PML estimators satisfy ... or if Assumptions W1-W7 hold, then PWML estimators satisfy ... .

Lemma 1 *Suppose Assumptions* I-VI *or* W1-W7 *hold for some $s \geq 3$ and the true parameter θ_0 lies in the interior of Θ. Then, there exists a sequence of* PML *or* PWML *estimators $\{\widehat{\theta}_n \in \widehat{\Theta}_n : n \geq 1\}$ that satisfies Condition C_s. (This holds whether $\widehat{\Theta}_n$ is defined to be the set of solutions to the* FOC*s in Θ or Θ^+.)*

The main result of this section is the following.

Theorem 1 *Suppose Assumptions* I-VI *or* W1-W7 *hold, $\{\widehat{\theta}_n \in \widehat{\Theta}_n : n \geq 1\}$ are* PML *or* PWML *estimators that satisfy Condition C_s, and $\widetilde{\Theta}$ is any set that satisfies Condition NS_s with s as specified below. Let Θ_c be any compact subset of $\widetilde{\Theta}$. Then,*
(a) $\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(\theta_0 \in \Delta CI_2(\widehat{\theta}_n)) - (1-\alpha)| = O(n^{-1})$ *for $s = 4$,*
(b) $\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(\theta_0 \in \Delta CI_2(\widehat{\theta}_n)) - (1-\alpha)| = O(n^{-1/2})$ *for $s = 3$, and*
(c) $\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(\theta_0 \in \Delta CI_{up}(\widehat{\theta}_n)) - (1-\alpha)| = O(n^{-1/2})$ *for $s = 3$.*

Comments 1. The errors in coverage probability of delta method CIs in the case of iid data typically are $O(n^{-1})$ and $O(n^{-1/2})$ for two- and one-sided CIs, respectively, e.g., see Hall (1988, 1992). Hence, parts (a) and (c) the Theorem show that delta method CIs in the long memory case have coverage probability errors with the same order of magnitude asymptotically as in the iid case. Note, however, that this is only true for CIs for autocorrelation and variance parameters. It is not true for CIs for the mean parameter in the long-memory case.

2. The error in part (a) of the Theorem is sharp except in the special case where the coefficient on the n^{-1} term of the Edgeworth expansion of $|t_n(\theta_{0,r})|$ is zero. Similarly, the error in part (c) is sharp except when the coefficient on the $n^{-1/2}$ term of the Edgeworth expansion of $t_n(\theta_{0,r})$ is zero. In such cases, sharp errors are determined by the first non-zero terms in the Edgeworth expansions of $|t_n(\theta_{0,r})|$ and

$t_n(\theta_{0,r})$ given in Lemma 9(a) in the Appendix. The error in part (b) may not be sharp.

3. If there exists a unique solution to the FOCs of the PLL or PWLL function with probability $1-o(n^{-(s-2)/2})$, then $\Delta CI_2(\widehat{\theta}_n)$ and $\Delta CI_{up}(\widehat{\theta}_n)$ are uniquely defined and $\widehat{\theta}_n$ satisfies Condition C_s by Lemma 1. If all solutions to the FOCs satisfy Condition C_s, then all CIs $\Delta CI_2(\widehat{\theta}_n)$ and $\Delta CI_{up}(\widehat{\theta}_n)$ for $\widehat{\theta}_n \in \widehat{\Theta}_n$ have coverage probability errors as in Theorem 1. If it is not the case that solutions to the FOCs are unique with probability $1-o(n^{-(s-2)/2})$ and it is not possible to verify Condition C_s for all solutions to the FOCs, then one can consider *conservative delta method* CIs. Let

$$\Delta CI_2 = \bigcup_{\widehat{\theta}_n \in \widehat{\Theta}_n} \Delta CI_2(\widehat{\theta}_n) \text{ and } \Delta CI_{up} = \bigcup_{\widehat{\theta}_n \in \widehat{\Theta}_n} \Delta CI_{up}(\widehat{\theta}_n). \tag{5.2}$$

Combining Lemma 1 and Theorem 1, we have

$$\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\theta_0 \in \Delta CI_2) \geq 1-\alpha+O(n^{-1}), \tag{5.3}$$

provided Assumptions I-VI or W1-W7 and Condition NS_s hold with $s=4$. Hence, the coverage probability of ΔCI_2 is less than $1-\alpha$ by at most $O(n^{-1})$. Similarly, the coverage probability of ΔCI_{up} is less than $1-\alpha$ by at most $O(n^{-1/2})$. If the only solution to the limit as $n \to \infty$ of the FOCs of the PLL or PWLL is θ_0, then all estimators $\widehat{\theta}_n \in \widehat{\Theta}_n$ are consistent and asymptotically normal. In this case, the coverage probabilities of ΔCI_2 and ΔCI_{up} are greater than $1-\alpha$ by $O(1)$ or less.

6 Higher-order Improvements of the Bootstrap

The main result of this paper is the following Theorem. The Theorem establishes bounds on the asymptotic orders of magnitude of coverage probability errors of bootstrap CIs based on PML or PWML estimators.

Theorem 2 *Suppose Assumptions* I-VI *or* W1-W7 *hold, the* PML *or* PWML *estimators* $\{\widehat{\theta}_n \in \widehat{\Theta}_n : n \geq 1\}$ *satisfy Condition* C_s*, the* BG *estimators* $\{\widetilde{\theta}_n : n \geq 1\}$ *satisfy Condition* C_s*, and* $\widetilde{\Theta}$ *is any set that satisfies Condition* NS_s *with* s *as specified below. Let* Θ_c *be any compact subset of* $\widetilde{\Theta}$*. Then,*
(a) $\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(\theta_0 \in CI_{sym}(\widehat{\theta}_n)) - (1-\alpha)| = o(n^{-3/2}\ln(n))$ *for* $s=5$*,*
(b) $\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(\theta_0 \in CI_{up}(\widehat{\theta}_n)) - (1-\alpha)| = o(n^{-1}\ln(n))$ *for* $s=4$*, and*
(c) *the errors in parts* (a) *and* (b) *are* $o(n^{-1/2})$ *for* $s=3$*.*

Comments 1. Theorem 2 provides results for bootstrap tests based on the t statistic $t_n(\theta_{H,r})$ as well as for CIs. For parameter values θ_0 for which $\theta_{0,r}=\theta_{H,r}$, we have $P_{\theta_0}(\theta_{0,r} \in CI_{sym}(\widehat{\theta}_n)) = P_{\theta_0}(|t_n(\theta_{H,r})| \leq z^*_{t,\alpha})$ and likewise for upper CIs and tests. Hence, for parameter values θ_0 in the null hypothesis, parts (a) and (b) of the Theorem give bounds on the error in rejection rates of symmetric two-sided and upper one-sided bootstrap tests respectively. For these results to hold, the estimators $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ only need to satisfy Condition C_s for parameter values θ_0 that satisfy the null hypothesis.

That is, Θ_c can be restricted to parameters that satisfy the null hypothesis. In consequence, the results cover tests based on null-restricted BG estimators, which are consistent only for parameters θ_0 that satisfy the null hypothesis (and hence do not satisfy Condition C_s for all compact sets $\Theta_c \subset \widetilde{\Theta}$), as well as non-null-restricted BG estimators.

The results of the Theorem do not provide information regarding the rejection rates of parametric bootstrap tests when the null hypothesis is false, because $P_{\theta_0}(\theta_{0,r} \in CI_{sym}(\widehat{\theta}_n)) \neq P_{\theta_0}(|t_n(\theta_{H,r})| \leq z^*_{t,\alpha})$ for such parameter values. However, it can be shown that the bootstrap critical values $z^*_{|t|,\alpha}$ and $z^*_{t,\alpha}$ equal $z_{\alpha/2} + o(1)$ and $z_\alpha + o(1)$, respectively, where z_α is the $1-\alpha$ standard normal quantile, whether or not the null hypothesis is true, under fairly general conditions on the BG estimator. In consequence, bootstrap tests have power against non-null parameter values θ_0.

2. Comparison of the results of Theorems 1 and 2 show that the bootstrap CIs $CI_{sym}(\widehat{\theta}_n)$ and $CI_{up}(\widehat{\theta}_n)$ have smaller coverage probability errors than the two- and one-sided delta method CIs, respectively, by the multiplicative factor $o(n^{-1/2}\ln(n))$ (provided the Assumptions and Conditions C_s and NS_s hold for $s = 5$ and 4, respectively).

3. For upper CIs, the bootstrap improvements are almost the same as those that have been established for parametric and non-parametric bootstrap CIs in iid scenarios, which are $O(n^{-1/2})$ typically, e.g., see Hall (1988, 1992). In fact, the slight difference (by a $\ln(n)$ factor) is undoubtedly due to the method of proof. Hence, for upper CIs, the parametric bootstrap for long memory time series performs essentially as well asymptotically as for iid sequences of random variables.

For symmetric two-sided CIs, the higher-order improvements of the Theorem are not as large as those that have been obtained for iid sequences. It may be the case that the errors in Theorem 2(a) are actually $O(n^{-2})$ due to an argument analogous to that of Hall (1988, 1992) for the iid case. It seems difficult to establish such a result rigorously in the long-memory case, however, and we leave such results to future research.

4. If the Assumptions and Conditions C_s and NS_s only hold with $s = 3$, then part (d) of Theorem 2 shows that the parametric bootstrap improves the CI coverage probability error $O(n^{-1/2})$ given in Theorem 1(b) and (c) to $o(n^{-1/2})$. This occurs with CIs based on the PML estimator in ARFIMA(p, d, q) models with $p \geq 1$. Note, however, the results in Theorem 1(b) and Theorem 2(c) may not be sharp.

If the Assumptions and Conditions C_s and NS_s only hold with $s = 4$, not $s = 5$, then the error in part (a) of the Theorem is $o(n^{-1})$. In this case, we see that the parametric bootstrap improves the error $O(n^{-1})$ of the delta method to $o(n^{-1})$.[6]

5. If the $n^{-1/2}$ term in the Edgeworth expansion of $t_n(\theta_{0,r})$ (given in Lemma 9 of the Appendix) does not depend on θ_0 and the Assumptions and Conditions C_s and NS_s hold with $s = 5$, then the error in Theorem 2(b) is reduced by the factor $n^{-1/2}$ to $o(n^{-3/2}\ln(n))$. In this case, the improvement of the parametric bootstrap upper CI over the delta method upper CI is of order $n^{-1}\ln(n)$ (provided the first term in the expansion is not identically zero[7]). (This holds by the same proof as for Theorem 2(b), but with four terms in the Edgeworth expansions instead of three, the second equation of (8.18) for $i = 1$ replaced by $\pi_1(\delta, \kappa_{n,4}(\widetilde{\theta}_n))\Phi(z) = \pi_1(\delta, \kappa_{n,4}(\theta_0))\Phi(z)$ for all z because $\pi_1(\delta, \kappa_{n,4}(\theta_0))$ does not depend on θ_0, and n^{-1} replaced by $n^{-3/2}$ throughout.)

The situation just described occurs in the Gaussian ARFIMA$(0, d, 0)$ model using the PML estimator. Lieberman and Phillips (2001) show that the $n^{-1/2}$ term of the Edgeworth expansion of the PML estimator does not depend on d. Their results are for a zero mean process. The results given in the Appendix show that the $n^{-1/2}$ term of the Edgeworth expansion of the PML estimator is the same whether one uses the sample mean or the true mean in the log-likelihood function. Furthermore, the variance of the PML estimator does not depend on d, so the t statistic $t_n(\theta_{0,r})$ is proportional to the normalized PML estimator, $n^{1/2}(\widehat{\theta}_{n,r} - \theta_{0,r})$, and, hence, has an Edgeworth expansion in which the $n^{-1/2}$ term does not depend on d. Finally, Assumptions I-VI and Conditions C_s and NS_s hold for $s = 4$ in this case. So, we conclude that in the Gaussian ARFIMA$(0, d, 0)$ model with unknown mean $CI_{up}(\widehat{\theta}_n)$ has coverage probability error of magnitude $o(n^{-3/2}\ln(n))$. In contrast, the one-sided delta method CI has error that is $O(n^{-1/2})$.

Similarly, if the n^{-1} term in the Edgeworth expansion of $|t_n(\theta_{0,r})|$ does not depend on θ_0 and the Assumptions and Conditions C_s and NS_s hold with $s = 5$, then the error in Theorem 2(a) is reduced by $n^{-1/2}$ to $o(n^{-2}\ln(n))$.

6. If there exists a unique solution in Θ and/or Θ^+ to the FOCs of the PLL or PWLL function with probability $1 - o(n^{-(s-2)/2})$, then $CI_{sym}(\widehat{\theta}_n)$ is uniquely defined, $\widehat{\theta}_n$ and $\widetilde{\theta}_n$ satisfy Condition C_s by Lemma 1, and $CI_{sym}(\widehat{\theta}_n)$ obtains higher-order improvements over the delta method. If all solutions in Θ and/or Θ^+ to the FOCs satisfy Condition C_s, then all CIs $CI_{sym}(\widehat{\theta}_n)$ obtain higher-order improvements over the delta method. If it is not the case that solutions to the FOCs are unique with probability $1 - o(n^{-(s-2)/2})$ and it is not possible to verify Condition C_s for all solutions to the FOCs, then one can consider *conservative* bootstrap CIs that are higher-order accurate. Let

$$CI_{sym} = \bigcup_{\widehat{\theta}_n \in \widehat{\Theta}_n} CI_{sym}(\widehat{\theta}_n) \text{ and } CI_{up} = \bigcup_{\widehat{\theta}_n \in \widehat{\Theta}_n} CI_{up}(\widehat{\theta}_n). \tag{6.1}$$

Combining Lemma 1 and Theorem 2, we have

$$\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\theta_0 \in CI_{sym}) \geq 1 - \alpha + o(n^{-3/2}\ln(n)), \tag{6.2}$$

provided Assumptions I-VI or W1-W7 and Condition NS_s hold with $s = 5$. Hence, the coverage probability of CI_{sym} is less than $1 - \alpha$ by at most $o(n^{-3/2}\ln(n))$. In contrast, the coverage probability of the symmetric two-sided delta method CI is less than $1 - \alpha$ by at most $O(n^{-1})$. A result analogous to that in (6.2) holds for CI_{up} with $n^{-3/2}$ and $s = 5$ replaced by n^{-1} and $s = 4$, respectively.

7 Monte Carlo Simulations

In this section we compare the coverage probabilities of delta method and parametric bootstrap two-sided CIs for some ARFIMA(p, d, q) processes. We take the number of bootstrap repetitions to be 999 and the number of simulation repetitions to be 1,000. This requires solving roughly one million nonlinear estimation problems for each parameter combination. In consequence, for computational ease, we take $p = q = 0$ and we consider ARFIMA$(0, d, 0)$ processes with unknown long-memory parameter d and unknown variance σ^2.

The values of d considered are 0, .1, .2, .3, and .4. The value of σ^2 is one. The sample size n is 100. Again for computational ease, we approximate the integrals in the definition of the PWML estimator by a finite grid with grid size .0628 $(= 2\pi/n)$. This allows us to use the fast Fourier transform.

Coverage probabilities for nominal 95%, 99%, and 90% confidence levels are reported in Table I. The last column of Table I reports the average absolute deviation of the true coverage probabilities from the nominal coverage probability, where the average is over the five values of d. This column gives a good summary of the relative performances of the different CIs.

Table I shows that the delta method CIs tend to under cover. This is especially true for the PWML-based CI. For example, the latter has an average coverage probability over the five values of d of .896 when the nominal coverage probability is .95. In consequence, the delta method CI based on the PML estimator outperforms that based on the PWML estimator. The bootstrap CIs sometimes under cover and sometimes over cover.

The absolute deviation of the bootstrap CI coverage probability from the nominal confidence level is less than that of the corresponding delta method CI in 24 out of 30 cases. For nominal level 95%, the bootstrap reduces the average absolute deviation from .016 to .010 for the PML-based CIs and from .054 to .012 for the PWML-based CIs. In consequence, the results of Table I indicate that the bootstrap CIs outperform the delta method CIs. This is especially true for the CIs based on the PWML estimator. Table I also shows that the relative performance of the two bootstrap CIs is about equal.

We conclude that the theoretical asymptotic advantages of the bootstrap over the delta method derived above are reflected in the finite sample cases considered here.

8 Appendix of Proofs

8.1 Edgeworth Expansion for the Log-likelihood Derivatives

We begin by establishing an Edgeworth expansion for the d_s-vector of centered and normalized PLL derivatives (LLDs), $W_n(\theta)$, defined in (4.3), that holds uniformly for θ in compact subsets Θ_c of $\widetilde{\Theta}$, where $\widetilde{\Theta}$ satisfies Condition NS_s. The order of the Edgeworth expansion can be arbitrarily large because moments of all orders of $W_n(\theta)$ exist. The Edgeworth expansion of the LLDs is a key ingredient in the proofs of Theorems 1 and 2 given below for the case of the PML estimator.

Let $\nu = (r_1, ..., r_q)'$ denote a q-vector of positive integers each less than or equal to $\dim(\theta)$. We write the real-valued q-th order partial derivative of the PLL objective function specified by ν as

$$\begin{aligned} L_{n,\nu}(\theta) &= D_\nu L_n(\theta, \overline{X}_n) \\ &= \frac{\partial^q}{\partial\theta_{r_1}\cdots\partial\theta_{r_q}} L_n(\theta, \overline{X}_n) \\ &= F_{n,\nu}(\theta) + X' M_n B_{n,\nu}(\theta) M_n X, \end{aligned} \tag{8.1}$$

where

$$\begin{aligned} F_{n,\nu}(\theta) &= -\frac{1}{2} D_\nu \ln(\det(T_n(f_\theta))) \\ &= \sum_{k=1}^{b} a_k \mathrm{tr}\left(\prod_{j=1}^{p_k} T_n^{-1}(f_\theta) T_n(g_{\theta,k,j})\right) \text{ and} \\ B_{n,\nu}(\theta) &= -\frac{1}{2} D_\nu T_n^{-1}(f_\theta) \\ &= \sum_{k=1}^{b} a_k \left(\prod_{j=1}^{p_k} T_n^{-1}(f_\theta) T_n(g_{\theta,k,j})\right) T_n^{-1}(f_\theta) \end{aligned} \tag{8.2}$$

for some fixed constants b, a_k, and p_k that depend on ν and with $g_{\theta,k,j}$ being certain partial derivatives of the spectral density with respect to the components of θ of order q or less. The quantities a_k, p_k, b, and $g_{\theta,k,j}$ are of the same form, but are not identical, in $F_{n,\nu}(\theta)$ and $B_{n,\nu}(\theta)$. For notational simplicity, we do not make a distinction.

In the first equation of (8.2), the second equality holds by application of the facts that if $A = A(\alpha)$ is a nonsingular $n \times n$ matrix that depends on a scalar α, then $(\partial/\partial\alpha)\ln(\det(A)) = \mathrm{tr}(A^{-1}(\partial/\partial\alpha)A)$, e.g., see Dhrymes (2000, Cor. 5.6, p. 159), and $(\partial/\partial\alpha)A^{-1} = -A^{-1}((\partial/\partial\alpha)A)A^{-1}$, e.g., see Dhrymes (2000, Cor. 15.14, p. 167). In the second equation of (8.2), the second inequality holds by repeated application of the second result just stated. See Taniguchi (1986, eqns. (4.4) and (4.5)) for exact expressions for the LLDs $L_{n,\nu}(\theta)$ of orders one, two, and three.

Let $G_n(u, \theta)$ for $u \in R^{d_s}$ be the density of $W_n(\theta)$ when the true parameter is θ. Let $\widetilde{G}_n^{\tau-2}(u, \theta)$ be the formal Edgeworth expansion of $W_n(\theta)$ of order $\tau - 2$. For brevity, we do not specify the precise form of $\widetilde{G}_n^{\tau-2}(u, \theta)$. See LRZ for details.

Theorem 3 *Suppose Assumptions* I-VI *hold and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$. *Then, for all compact sets* $\Theta_c \subset \widetilde{\Theta}$ *and all* $\tau \geq 3$,
(a) $\sup_{\theta_0 \in \Theta_c} \sup_{u \in R^{d_s}} |G_n(u, \theta_0) - \widetilde{G}_n^{\tau-2}(u, \theta_0)| = o(n^{-(\tau-2)/2})$ *and*
(b) $P_{\theta_0}(W_n(\theta_0) \in C) = \int_C \widetilde{G}_n^{\tau-2}(u, \theta_0)du + o(n^{-(\tau-2)/2})$, *uniformly over all Borel sets* C *and all* $\theta_0 \in \Theta_c$.

Comments 1. When Theorem 3 is employed in the proofs of Theorems 1 and 2, we take $\tau = s$.

2. Theorem 3(a) is proved in Section 8.3 by verifying the conditions of Theorem 1 of Durbin (1980), which provides an Edgeworth expansion for the density of $W_n(\theta_0)$. Theorem 3(b) converts the result of part (a) into an Edgeworth expansion for the distribution function of $W_n(\theta_0)$ using Corollary 3.3 of Skovgaard (1986).

8.2 Edgeworth Expansion for the Whittle Log-likelihood Derivatives

In this section, we state an Edgeworth expansion for the vector of centered and normalized PWLL derivatives (WLLDs), $W_n(\theta)$, defined in (4.3), that is established in Andrews and Lieberman (2002). The Edgeworth expansion is used in the proof of Theorems 1 and 2 for the case of the PWML estimator.

We write the real-valued q-th order partial derivative of the PWLL objective function specified by ν as

$$\begin{aligned} L_{W,n,\nu}(\theta) &= D_\nu L_{W,n}(\theta, \overline{X}_n) \\ &= F_{n,\nu}(\theta) + X' M_n B_{n,\nu}(\theta) M_n X, \end{aligned} \tag{8.3}$$

where

$$\begin{aligned} F_{n,\nu}(\theta) &= -\frac{n}{4\pi} \int_{-\pi}^{\pi} D_\nu \ln(f_\theta(\lambda)) d\lambda \text{ and} \\ B_{n,\nu}(\theta) &= -\frac{1}{2} D_\nu T_n((2\pi)^{-2} f_\theta^{-1}). \end{aligned} \tag{8.4}$$

As above, let $G_n(u, \theta)$ for $u \in R^{d_s}$ be the density of $W_n(\theta)$ when the true parameter is θ. Let $\widetilde{G}_n^{\tau-2}(u, \theta)$ be the formal Edgeworth expansion of $W_n(\theta)$ of order $\tau - 2$. We do not specify the precise form of $\widetilde{G}_n^{\tau-2}(u, \theta)$. See Andrews and Lieberman (2002) for details.

Proposition 1 *Suppose Assumptions* W1-W7 *hold and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$. *Then, for all compact sets* $\Theta_c \subset \widetilde{\Theta}$ *and all* $\tau \geq 3$,
(a) $\sup_{\theta_0 \in \Theta_c} \sup_{u \in R^{d_s}} |G_n(u, \theta_0) - \widetilde{G}_n^{\tau-2}(u, \theta_0)| = o(n^{-(\tau-2)/2})$ *and*
(b) $P_{\theta_0}(W_n(\theta_0) \in C) = \int_C \widetilde{G}_n^{\tau-2}(u, \theta_0)du + o(n^{-(\tau-2)/2})$, *uniformly over all Borel sets* C *and all* $\theta_0 \in \Theta_c$.

8.3 Proof of Validity of the Edgeworth Expansion for the Log-likelihood Derivatives

In this section, we prove Theorem 3. The proof uses the following three Lemmas.

As specified in Condition NS$_s$(iii), $Z_n(\theta)$ denotes the d_s-vector of non-redundant LLDs $L_{n,\nu}(\theta)$ or WLLDs $L^W_{n,\nu}(\theta)$ up to order $s-1$. For LLDs, we can write $Z_n(\theta) = (L_{n,\nu(1)}(\theta), ..., L_{n,\nu(d_s)}(\theta))'$, where each vector $\nu(j)$ is of the same form as ν defined above (8.1) for some $q \leq d_s$ for $j = 1, ..., d_s$. For WLLDs, $Z_n(\theta)$ can be written analogously.

Let $\varphi_n(\omega, \theta) = E_\theta \exp(i\omega' Z_n(\theta))$ denote the characteristic function of $Z_n(\theta)$ when θ is the true value, where $\omega \in R^{d_s}$. Let $\eta = (\eta_1, ..., \eta_q)'$ be a q-vector of non-negative integers each of which is less than or equal to d_s. Define $D_{\omega,\eta} = \partial^q/(\partial\omega_{\eta_1} \cdots \partial\omega_{\eta_q})$. Let $\kappa_{n,s}(\theta)_\eta$ denote the η cumulant of $Z_n(\theta)$ when θ is the true value. Note that $\kappa_{n,s}(\theta)_\eta$ is a cumulant of order q. By definition, $\kappa_{n,s}(\theta)_\eta = i^{-q} D_{\omega,\eta} \ln(\varphi_n(\omega, \theta))|_{\omega=0}$, where $i = \sqrt{-1}$. The vector $\kappa_{n,s}(\theta)$ is composed of elements $\kappa_{n,s}(\theta)_\eta$ for vectors η of dimension $q \leq s$.

The following Lemma holds for both the PLL and PWLL cases.

Lemma 2 *Suppose Assumptions* I-VI *or* W1-W7 *hold for some integer* $s \geq 3$. *Then,*
(a) $\varphi_n(\omega, \theta) = \exp(i \sum_{j=1}^{d_s} \omega_j F_{n,\nu(j)}(\theta)) \det^{-1/2}(I_n - 2i \sum_{j=1}^{d_s} \omega_j M_n B_{n,\nu(j)}(\theta) M_n \times T_n(f_\theta))$, *where* $F_{n,\nu(j)}(\theta)$ *and* $B_{n,\nu(j)}(\theta)$ *have the form given in* (8.2) *or* (8.4) *with different constants and spectral density derivatives for each* j,
(b) η *cumulants of order one satisfy:*

$$\kappa_{n,s}(\theta)_\eta = F_{n,\nu(\eta)}(\theta) + tr\left(M_n B_{n,\nu(\eta)}(\theta) M_n T_n(f_\theta)\right),$$

where $F_{n,\nu(\eta)}(\theta)$ *and* $B_{n,\nu(\eta)}(\theta)$ *are of the form given in* (8.2) *or* (8.4) *and* η *is an integer between one and* d_s,
(c) η *cumulants of order* $q \geq 2$ *satisfy:*

$$\kappa_{n,s}(\theta)_\eta = C_q tr\left(\prod_{r=1}^{q} (M_n B_{n,\nu(\eta_r)}(\theta) M_n T_n(f_\theta))\right)$$

for some constant $C_q < \infty$, *where* $B_{n,\nu(\eta_r)}(\theta)$ *is of the form given in* (8.2) *or* (8.4) *with different constants and different spectral density derivatives for each* η_r.

We now establish some results for the case where $Z_n(\theta)$ is based on the PLL function (not the PWLL function).

Let $||A|| = (\text{tr}(A^*A))^{1/2}$ denote the Euclidean norm of A.

Let $Z^0_n(\theta)$ denote the same vector of LLDs as $Z_n(\theta)$, but with $L_{n,\nu}(\theta)$ $(= D_\nu L_n(\theta, \overline{X}_n))$ replaced by $D_\nu L_n(\theta, \mu_0)$, where μ_0 is the true mean. Let $\kappa^0_{n,s}(\theta)$ denote the same vector of cumulants as $\kappa_{n,s}(\theta)$, but for $Z^0_n(\theta)$ rather than $Z_n(\theta)$. Note that $\kappa^0_{n,s}(\theta)_\eta$ equals the same expressions as given for $\kappa_{n,s}(\theta)_\eta$ in Lemma 2(b) and (c), but with M_n deleted. We show that $\sup_{\theta \in \Theta_c} ||\kappa_{n,s}(\theta) - \kappa^0_{n,s}(\theta)|| = O(n^\delta)$ for all $\delta > 0$. LRZ show that $\sup_{\theta \in \Theta_c} ||\kappa^0_{n,s}(\theta)|| = O(n)$. These results combine to give $\sup_{\theta \in \Theta_c} ||\kappa_{n,s}(\theta)|| = O(n)$, which is needed for application of Durbin's (1980) Theorem 1.

Lemma 3 *Suppose Assumptions* I-VI *hold and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$*. Let* Θ_c *be any compact subset of* $\widetilde{\Theta}$*. Then, for the cumulants* $\kappa^0_{n,s}(\theta)$ *and* $\kappa_{n,s}(\theta)$ *based on the log-likelihood and* PLL *functions, respectively, we have*
(a) $\sup_{\theta \in \Theta_c} ||\kappa^0_{n,s}(\theta)|| = O(n)$,
(b) $\sup_{\theta \in \Theta_c} ||\kappa_{n,s}(\theta) - \kappa^0_{n,s}(\theta)|| = O(n^\delta)$ *for all* $\delta > 0$, *and*
(c) $\sup_{\theta \in \Theta_c} ||\kappa_{n,s}(\theta)|| = O(n)$.

Next, we show that the variance of $n^{-1/2} Z_n(\theta)$, denoted $D_n(\theta)$, converges to a matrix $D(\theta)$ as $n \to \infty$ uniformly over $\theta \in \Theta_c$. The proof relies on Theorem 2 of LRZ, which extends Theorem 5.1 of Dahlhaus (1989), on the limiting behavior of traces of products of certain $n \times n$ Toeplitz matrices.

Lemma 4 *Suppose Assumptions* I-VI *hold and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$*. Let* Θ_c *be any compact subset of* $\widetilde{\Theta}$*. Then, for* $D_n(\theta)$ *and* $D(\theta)$ *based on the* PLL *function, we have*

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_c} |D_n(\theta) - D(\theta)| = 0,$$

where the (i, ℓ) *element of* $D(\theta)$ *has the form*

$$[D(\theta)]_{i,\ell} = \sum_{u=1}^{b_i} \sum_{v=1}^{b_\ell} a_u^{(i)} a_v^{(\ell)} \int_{(-\pi,\pi)} f_\theta(\lambda)^{-(p_u+p_v)} \{\prod_{m=1}^{p_u} g^{(i)}_{\theta,m}(\lambda)\} \{\prod_{r=1}^{p_v} g^{(\ell)}_{\theta,r}(\lambda)\} d\lambda,$$

b_i, b_ℓ, $a_u^{(i)}$, $a_v^{(\ell)}$, p_u, p_v, *are constants, and* $g^{(i)}_{\theta,m}(\lambda)$ *and* $g^{(\ell)}_{\theta,r}(\lambda)$ *are partial derivatives with respect to* θ *of* $f_\theta(\lambda)$ *of order* s *or less, as in* (8.2), *for* $i, \ell = 1, ..., d_s$.

Proof of Theorem 3

To prove part (a) we verify the conditions of Theorem 1 of Durbin (1980), which establishes the validity of an Edgeworth expansion for the density of a sequence of random vectors. Durbin's Theorem 1 relies on his Assumptions 2-4 plus the assumptions given in his paragraph containing (28) that (i) $W_n(\theta)$ has a density with respect to Lebesgue measure for n large and (ii) the variance of $W_n(\theta)$, viz., $D_n(\theta)$, converges to a nonsingular matrix $D(\theta_0)$ as $n \to \infty$ and $\theta \to \theta_0$ jointly.

Condition (i) holds because (a) $Z_n(\theta)$ is a vector of partial derivatives of $L_n(\theta, \overline{X}_n)$ up to order $s-1$, (b) by (8.1), each of these partial derivatives is a quadratic form in the multivariate normal random vector X, and (c) the covariance matrix of $n^{-1/2} Z_n(\theta)$ converges uniformly over $\theta \in \Theta_c$ as $n \to \infty$ to a nonsingular matrix $D(\theta)$ by Lemma 4 and Condition NS_s(iii). Condition (ii) holds by Lemma 4.

Durbin's Assumption 4 requires that the cumulants of $Z_n(\theta)$ are $O(n)$ uniformly over $\theta \in \Theta_c$. This holds by Lemma 3(c).

LRZ show that a modified version of Durbin's Assumption 3 can be used in place of his Assumption 3. In verifying Durbin's Assumptions 2 and the modified Assumption 3, the only difference between the present case and the known mean case (which is considered by LRZ) is that terms of the form $\mathrm{tr}(\prod_{\ell=1}^p (M_n B_{n,\nu(\ell)}(\theta) M_n T_n(f_\theta)))$ reduce to $\mathrm{tr}(\prod_{\ell=1}^p (B_{n,\nu(\ell)}(\theta) T_n(f_\theta)))$ in the known mean case. By the proof of Lemma 3(b), the difference between these two expressions is negligible (specifically, it is $O(n^{\delta})$ for all $\delta > 0$) compared to the magnitude of the second expression (which is $O(n)$). In consequence, LRZ's verification of Assumption 2 and the modified Assumption 3 go through in the present case.

Given part (a), part (b) follows by applying Corollary 3.3 of Skovgaard (1986), which converts an Edgeworth expansion of a density to one of a distribution function. □

Proof of Lemma 2

To prove part (a), we use the fact that if A and V are symmetric $n \times n$ matrices, V is positive definite, and $Y \sim N(0, V)$, then $E \exp(iY'AY) = \det^{-1/2}(I_n - 2iAV)$. This holds by straightforward calculations, e.g., see Searle (1971, eqn. (42) with $\mu = 0$, p. 55). Let $U = X - \mu_0 \mathbf{1}_n$, where μ_0 is the true mean. Then, $U \sim N(0, T_n(f_\theta))$, when θ is the true value. Note that $M_n X = M_n U$, because M_n projects onto the space orthogonal to $\mathbf{1}_n$. Using these results and (8.1) gives the result of part (a):

$$\begin{aligned} & E_\theta \exp(i\omega' Z_n(\theta)) \\ &= E_\theta \exp\left(i \sum_{j=1}^{d_s} \omega_j (F_{n,\nu(j)}(\theta) + X' M_n B_{n,\nu(j)}(\theta) M_n X)\right) \\ &= \exp\left(i \sum_{j=1}^{d_s} \omega_j F_{n,\nu(j)}(\theta)\right) E_\theta \exp\left(iU' \left(\sum_{j=1}^{d_s} \omega_j M_n B_{n,\nu(j)}(\theta) M_n\right) U\right) \\ &= \exp\left(i \sum_{j=1}^{d_s} \omega_j F_{n,\nu(j)}(\theta)\right) \det{}^{-1/2}\left(I_n - 2i \sum_{j=1}^{d_s} \omega_j M_n B_{n,\nu(j)}(\theta) M_n T_n(f_\theta)\right). \end{aligned} \tag{8.5}$$

Next, we prove part (b). First order cumulants are first-order moments. Hence, for η equal to an integer, $\kappa_{n,s}(\theta)_\eta = F_{n,\nu(\eta)}(\theta) + E_\theta X' M_n B_{n,\nu(\eta)}(\theta) M_n X$ and $E_\theta X' M_n B_{n,\nu}(\theta) M_n X = E_\theta U' M_n B_{n,\nu}(\theta) M_n U = \mathrm{tr}(M_n B_{n,\nu}(\theta) M_n T_n(f_\theta))$.

To prove part (c), we use the definition of $\kappa_{n,s}(\theta)_\eta$, i.e., $\kappa_{n,s}(\theta)_\eta = i^{-q} D_{\omega,\eta} \ln(\varphi_n(\omega, \theta))|_{\omega=0}$. For η cumulants of order $q \geq 2$,

$$D_{\omega,\eta} \ln(\varphi_n(\omega, \theta)) = D_{\omega,\eta} \ln\left(\det{}^{-1/2}\left(I_n - 2i \sum_{j=1}^{d_s} \omega_j M_n B_{n,\nu(j)}(\theta) M_n T_n(f_\theta)\right)\right). \tag{8.6}$$

The right-hand side (rhs) partial derivative can be calculated using the facts stated in the paragraph following (8.2) that $(\partial/\partial\alpha)\ln(\det(A)) = tr(A^{-1}(\partial/\partial\alpha)A)$ and $(\partial/\partial\alpha)A^{-1} = -A^{-1}((\partial/\partial\alpha)A)A^{-1}$. Since $\det(I_n - 2i\sum_{j=1}^{d_s} \omega_j M_n B_{n,\nu(j)}(\theta) M_n \times T_n(f_\theta))|_{\omega=0} = 1$, the rhs of (8.6) evaluated at $\omega = 0$ is of the form stated in part (c). For more details, see Searle (1971, Thm. 2.5.1, p. 55). □

We use the following notation below. Let $B(\theta, \varepsilon)$ denote an open ball of radius $\varepsilon > 0$ centered at θ. Let $\mathbf{C}$ denote the set of complex numbers. For $z \in \mathbf{C}$, let z^*

denote the conjugate transpose of z and let $||z|| = (z^*z)^{1/2}$ denote the Euclidean norm of z. Let $||A||_{sp} = \sup_{z\in\mathbf{C}^n, ||z||=1}(z^*A^*Az)^{1/2}$ denote the spectral norm of an $n\times n$ matrix A. (Note that the spectral norm of a symmetric matrix equals its largest absolute eigenvalue by diagonalization.) Well-known inequalities are $||A||_{sp} \leq ||A||$, $|z_1^*Az_2| \leq ||z_1||\cdot||z_2||\cdot||A||_{sp}$ for $z_1, z_2 \in \mathbf{C}^n$, and $||AB||_{sp} \leq ||A||_{sp}||B||_{sp}$, for example, see Dahlhaus (1989, p. 1754).

The proof of Lemma 3 uses the following Lemma.

Lemma 5 *Suppose Assumptions* I-VI *hold and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$. *Then, for all compact subsets* Θ_c *of* $\widetilde{\Theta}$,
(a) $\sup_{\theta\in\Theta_c} ||T_n^{-1/2}(f_\theta)T_n^{1/2}(g_\theta)||_{sp} = O(n^\delta)$ *for all* $\delta > 0$, *where* $g_\theta(\lambda)$ *is any partial derivative of* $f_\theta(\lambda)$ *with respect to* θ *of order* s *or less,*
(b) $\sup_{\widetilde{\theta}\in B(\theta,\varepsilon)} e_n'T_n(f_{\widetilde{\theta}})e_n \leq K_\varepsilon n^{2d+3\varepsilon}$ *for all* $\theta \in \Theta_c$ *and all* $\varepsilon > 0$ *sufficiently small, for some constant* $K_\varepsilon < \infty$,
(c) $\sup_{\widetilde{\theta}\in B(\theta,\varepsilon)} e_n'T_n^{-1}(f_{\widetilde{\theta}})e_n \leq K_\varepsilon n^{-2d+\varepsilon}$ *for all* $\theta \in \Theta_c$ *and all* $\varepsilon > 0$ *sufficiently small, for some constant* $K_\varepsilon < \infty$, *and*
(d) $\sup_{\theta\in\Theta_c}(e_n'T_n(f_\theta)e_n \cdot e_n'T_n^{-1}(f_\theta)e_n) = O(n^\delta)$ *for all* $\delta > 0$,
where d *is the first element of* θ *in parts* (b) *and* (c) *and* $e_n = n^{-1/2}\mathbf{1}_n$.

Proof of Lemma 3

Part (a) of the Lemma is proved in Sec. 7.3.2 of LRZ using their Theorem 2.

Part (c) is implied by parts (a) and (b).

Part (b) is proved by showing that $\sup_{\theta\in\Theta_c} |\kappa_{n,s}(\theta)_\eta - \kappa_{n,s}^0(\theta)_\eta| = O(n^\delta)$ for each η of order $q \leq s$. Since $\kappa_{n,s}^0(\theta)_\eta$ is the same as $\kappa_{n,s}(\theta)_\eta$ (given in Lemma 2(b) and (c)), but with M_n deleted, it suffices to show that

$$\sup_{\theta\in\Theta_c}\left|tr\left(\prod_{r=1}^q \left(M_nB_{n,\nu(\eta_r)}(\theta)M_nT_n(f_\theta)\right)\right) - tr\left(\prod_{r=1}^q \left(B_{n,\nu(\eta_r)}(\theta)T_n(f_\theta)\right)\right)\right|$$
$$= O(n^\delta) \tag{8.7}$$

for all $\delta > 0$ and $q \leq s$.

For notational simplicity, let $M = M_n$, $P = P_n$, $I = I_n$, $e = e_n$, $B_r = B_{n,\nu(\eta_r)}$, and $T = T_n(f_\theta)$. The product $\prod_{r=1}^q(MB_rMT) = \prod_{r=1}^q((I-P)B_r(I-P)T)$ can be written as the sum of $\prod_{r=1}^q(B_rT)$ and $2^{2q}-1$ additional terms each of which has the form

$$\pm\prod_{r=1}^q(P^{a(r)}B_rP^{b(r)}T), \tag{8.8}$$

where $a(r)$ and $b(r)$ equal zero or one, $P^0 = I$ by definition, and the number of matrices P that appears in the product, denoted num(P), is at least one and less than or equal to $2q$. Hence, it suffices to show that

$$\sup_{\theta\in\Theta_c}\left|tr\left(\prod_{r=1}^q(P^{a(r)}B_rP^{b(r)}T)\right)\right| = O(n^\delta). \tag{8.9}$$

Because $P = ee'$, the trace in (8.9) is the product of num(P) terms of the form

$$\text{(i) } e' \prod_{\ell=1}^{L} (B_{k_\ell} T) e, \text{ (ii) } e'T \prod_{\ell=1}^{L} (B_{k_\ell} T) e, \text{ and (iii) } e' \prod_{\ell=1}^{L-1} (B_{k_\ell} T) B_{k_L} e \tag{8.10}$$

for $0 \leq L \leq q$. The matrix T appears in the product in (8.9) q times and the number of matrices B_r in the product is q. In consequence, the number of type (ii) and (iii) terms in the product must be equal.

By (8.2), each product $B_r T$ is a finite linear combination of products of the form $\prod_{j=1}^{p_r} (T^{-1} T_{g,j})$, where $T_{g,j} = T_n(g_{\theta,j})$ and $g_{\theta,j}$ is a spectral density partial derivative of order s or less. Hence, the type (i) terms in (8.10) are finite linear combinations of type (i′) terms, the type (ii) terms are finite linear combinations of type (ii′) terms, and the type (iii) terms are finite linear combinations of type (iii′) terms, where terms of type are (i′)-(iii′) defined by

$$\text{(i}'\text{) } e' \prod_{j=1}^{J} (T^{-1} T_{g,j}) e, \text{ (ii}'\text{) } e'T \prod_{j=1}^{J} (T^{-1} T_{g,j}) e, \text{ and (iii}'\text{) } e' \prod_{j=1}^{J} (T^{-1} T_{g,j}) T^{-1} e, \tag{8.11}$$

respectively, for $J \geq 0$.

Now, the absolute value of the type (i′) term can be written as

$$\begin{aligned} &\left| e'T^{-1/2} \prod_{j=1}^{J} \left((T^{-1/2} T_{g,j}^{1/2})(T_{g,j}^{1/2} T^{-1/2}) \right) T^{1/2} e \right| \\ &\leq (e'T^{-1}e)^{1/2} \prod_{j=1}^{J} \left(||T^{-1/2} T_{g,j}^{1/2}||_{sp} \cdot ||T_{g,j}^{1/2} T^{-1/2})||_{sp} \right) (e'Te)^{1/2}, \end{aligned} \tag{8.12}$$

using the inequalities $|x'Ay| \leq ||x|| \cdot ||y|| \cdot ||A||_{sp}$ and $||AB||_{sp} \leq ||A||_{sp} ||B||_{sp}$. Similarly, the absolute values of the type (ii′) and (iii′) terms are bounded by

$$\begin{aligned} &e'Te \prod_{j=1}^{J} \left(||T^{-1/2} T_{g,j}^{1/2}||_{sp} \cdot ||T_{g,j}^{1/2} T^{-1/2})||_{sp} \right) \text{and} \\ &e'T^{-1}e \prod_{j=1}^{J} \left(||T^{-1/2} T_{g,j}^{1/2}||_{sp} \cdot ||T_{g,j}^{1/2} T^{-1/2})||_{sp} \right), \end{aligned} \tag{8.13}$$

respectively.

By Lemma 5(a), $\sup_{\theta \in \Theta_c} ||T^{-1/2} T_{g,j}^{1/2}||_{sp} = O(n^\gamma)$ for all $\gamma > 0$. This result, (8.12), (8.13), the fact that the trace in (8.9) is the product of num(P) $\leq 2q$ terms of types (i)-(iii), and the fact that the number of terms in this product of type (ii) equals that of type (iii) imply that

$$\sup_{\theta \in \Theta_c} \left| tr \left(\prod_{r=1}^{q} (P^{a(r)} B_r P^{b(r)} T) \right) \right| \leq K n^\gamma \sup_{\theta \in \Theta_c} \left((e'T^{-1}e)^{1/2} (e'Te)^{1/2} \right)^{2q} \tag{8.14}$$

for some constant $K < \infty$ and all $\gamma > 0$. The rhs is $O(n^{\varepsilon})$ for all $\varepsilon > 0$ by Lemma 5(d), which establishes the desired result. □

Proof of Lemma 5

Part (a) of the Lemma holds by Lemma 5.3 of Dahlhaus (1989), using Assumptions IV(a) and V to provide the requisite bound on $|f_\theta(\lambda)|$ and using the fact that $|g_{\theta,j}(\lambda)| \leq C_\delta |\lambda|^{-2d+\delta}$ for all $\lambda \in (0,\pi)$, $\theta \in \Theta_c$, and $\delta > 0$ for some constant $C_\delta < \infty$ that does not depend on θ by Assumptions IV and V. Uniformity of the result of part (a) over $\theta \in \Theta_c$ follows easily from Dahlhaus' proof given that C_δ does not depend on θ and $\sup_{\theta \in \Theta_c} c_2(\theta, \delta) < \infty$ for all $\delta > 0$ by Assumption V.

We prove part (b) using a technique employed in the proof of Lemma 5.3 of Dahlhaus (1989). Let $\mathcal{P}_n$ denote the set of probability densities on $(-\pi, \pi)$ that are bounded by n. Define $h_0(\lambda) = n^{-1}|\sum_{j=1}^n e^{ij\lambda}|^2$. Note that $h_0(\lambda) \in \mathcal{P}_n$. Using (2.3), we have: for all $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}
|e_n' T_n(f_\theta) e_n| &= \left| n^{-1} \int_{-\pi}^{\pi} \sum_{j=1}^n \sum_{k=1}^n e^{i(j-k)\lambda} f_\theta(\lambda) d\lambda \right| \\
&\leq c_2(\theta, \varepsilon) n^{-1} \int_{-\pi}^{\pi} |\sum_{j=1}^n e^{ij\lambda}|^2 |\lambda|^{-2d-\varepsilon} d\lambda \\
&\leq c_2(\theta, \varepsilon) \sup_{h \in \mathcal{P}_n} \int_{-\pi}^{\pi} h(\lambda) |\lambda|^{-2d-\varepsilon} d\lambda \\
&= 2 c_2(\theta, \varepsilon) \int_0^{1/(2n)} n \lambda^{-2d-\varepsilon} d\lambda \\
&= \frac{2^{2d+\varepsilon}}{1 - 2d - \varepsilon} c_2(\theta, \varepsilon) n^{2d+\varepsilon},
\end{aligned} \tag{8.15}$$

where the first inequality uses Assumption IV(a), the second inequality holds because $h_0(\lambda) \in \mathcal{P}_n$, and the second equality holds because the supremum over $h \in \mathcal{P}_n$ is attained by the function $h_n(\lambda) = n1(|\lambda| \leq 1/(2n)) \in \mathcal{P}_n$. Combining (8.15) with $\sup_{\widetilde{\theta} \in B(\theta,\varepsilon), \theta \in \Theta_c} c_2(\theta, \delta) < \infty$ by Assumption V, and $\widetilde{d} \leq d + \varepsilon$ for all $\widetilde{\theta} \in B(\theta, \varepsilon)$, where $\widetilde{\theta} = (\widetilde{d}, \widetilde{\theta}_2, ..., \widetilde{\theta}_{\dim(\theta)})'$, give the result of part (b).

Part (c) holds by the combination of Theorems 4.1 and 5.1 of Adenstedt (1974).

The result of part (d) is proved using parts (b) and (c) of the Lemma. The compact set Θ_c can be covered by a finite number, J_ε, of balls $B(\theta_j, \varepsilon)$ of radius $\varepsilon > 0$ centered at $\theta_j \in \Theta_c$ for $j = 1, ..., J_\varepsilon$. We have

$$\begin{aligned}
\sup_{\theta \in \Theta_c} (e' T_n(f_\theta) e \cdot e' T_n^{-1}(f_\theta) e) &\leq \sum_{j=1}^{J_\varepsilon} \sup_{\theta \in B(\theta_j, \varepsilon)} (e' T_n(f_\theta) e \cdot e' T_n^{-1}(f_\theta) e) \\
&\leq J_\varepsilon K_\varepsilon^2 n^{-2d+3\varepsilon} n^{2d+\varepsilon} = O(n^\delta)
\end{aligned} \tag{8.16}$$

for all $\delta \geq 4\varepsilon > 0$, where the second inequality holds by parts (b) and (c) of the Lemma. □

Proof of Lemma 4

By Lemma 2(c), $[D_n(\theta)]_{i,\ell} = 2n^{-1}\mathrm{tr}(M_n B_{n,\nu(i)}(\theta) M_n T_n(f_\theta) M_n B_{n,\nu(\ell)}(\theta) M_n \times T_n(f_\theta))$ for $i, \ell = 1, ..., d_s$. By Lemma 3(b), $[D_n(\theta)]_{i,\ell} = 2n^{-1}\mathrm{tr}(B_{n,\nu(i)}(\theta)T_n(f_\theta) \times B_{n,\nu(\ell)}(\theta)T_n(f_\theta)) + o(1)$, where $o(1)$ holds uniformly over $\theta \in \Theta_c$. This expression is a sum of traces of products of certain $n \times n$ Toeplitz matrices, because $B_{n,\nu(i)}(\theta)T_n(f_\theta)$ has the form $\sum_{u=1}^{b_i} a_u^{(i)} (\prod_{m=1}^{p_u} T_n^{-1}(f_\theta) T_n(g_{\theta,m}^{(i)}))$ and similarly for $B_{n,\nu(\ell)}(\theta)T_n(f_\theta)$ by (8.2). The stated result now follows from Theorem 2 of LRZ. $\square$

8.4 Lemmas Used in the Proofs of Theorems 1 and 2

Next, we state several lemmas that are used in the proofs of Theorems 1 and 2. Each of the lemmas holds with Θ_c being any compact subset of $\widetilde{\Theta}$, where $\widetilde{\Theta}$ satisfies Condition NS_s for some $s \geq 3$. The first of these lemmas is similar to numerous results that have appeared in the literature.

Lemma 6 *Let $\{A_n(\theta_0) : n \geq 1\}$ be a sequence of $\dim(A) \times 1$ random vectors with Edgeworth expansions for each $\theta_0 \in \Theta_c$ with coefficients of order $O(1)$ and remainders of order $o(n^{-(s-2)/2})$ both uniformly over $\theta_0 \in \Theta_c$. Specifically, there exist polynomials $\{\pi_{n,i}(z, \theta_0) : i = 1, ..., s-2, n \geq 1\}$ in z whose coefficients are $O(1)$ uniformly over $\theta_0 \in \Theta_c$ such that $\sup_{\theta_0 \in \Theta_c} \sup_{B \in \mathcal{B}_{\dim(A)}} |P_{\theta_0}(A_n(\theta_0) \in B) - \int_B (1 + \sum_{i=1}^{s-2} n^{-i/2}\pi_{n,i}(z,\theta_0))\phi_{\Sigma_n(\theta_0)}(z)dz| = o(n^{-(s-2)/2})$, where $\phi_{\Sigma_n(\theta_0)}(z)$ is the density function of a $N(0, \Sigma_n(\theta_0))$ random variable, $\Sigma_n(\theta_0)$ has eigenvalues that are bounded away from zero and infinity as $n \to \infty$ uniformly over $\theta \in \Theta_c$ and $\mathcal{B}_{\dim(A)}$ denotes the class of all convex sets in $R^{\dim(A)}$. Let $\{\xi_n(\theta_0) \in R^{\dim(A)} : n \geq 1\}$ be a sequence of random vectors with $\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(||\xi_n(\theta_0)|| > \omega_n) = o(n^{-(s-2)/2})$ for some constants $\omega_n = o(n^{-(s-2)/2})$. Then,*

$$\sup_{\theta_0 \in \Theta_c} \sup_{B \in \mathcal{B}_{\dim(A)}} |P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B) - P_{\theta_0}(A_n(\theta_0) \in B)| = o(n^{-(s-2)/2}).$$

Let $n^{-1}Z_n^+(\theta_0)$ denote the vector $n^{-1}Z_n(\theta_0)$ of normalized LLDs or WLLDs augmented to include the vector of expected values of all partial derivatives with respect to θ of order s of $n^{-1}L_n(\theta_0, \overline{X}_n)$ or $n^{-1}L_{W,n}(\theta_0, \overline{X}_n)$.

The following lemma is an extension of Theorem 3(b) of Bhattacharya and Ghosh (1978). The lemma shows that the normalized PML or PWML estimator and the t statistic $t_n(\theta_{0,r})$ can be approximated by smooth functions of $n^{-1}Z_n^+(\theta_0)$.

Lemma 7 *Suppose Assumptions* I-VI *or* W1-W7 *hold, the* PML *or* PWML *estimators $\{\widehat{\theta}_n : n \geq 1\}$ satisfy Condition C_s, and $\widetilde{\Theta}$ satisfies Condition NS_s for some integer $s \geq 3$. Let $\Delta_n(\theta_0)$ denote $n^{1/2}(\widehat{\theta}_n - \theta_0)$ or $t_n(\theta_{0,r})$. Let $\dim(\Delta)$ denote the dimension of $\Delta_n(\theta_0)$. For each definition of $\Delta_n(\theta_0)$, there is an infinitely differentiable function $G(\cdot)$ that does not depend on θ_0 that satisfies $G(n^{-1}E_{\theta_0}Z_n^+(\theta)) = 0$ for all n large and all $\theta_0 \in \Theta_c$ and*

$$\sup_{\theta_0 \in \Theta_c} \sup_{B \in \mathcal{B}_{\dim(\Delta)}} |P_{\theta_0}(\Delta_n(\theta_0) \in B) - P_{\theta_0}(n^{1/2}G(n^{-1}Z_n^+(\theta_0)) \in B)| = o(n^{-(s-2)/2}).$$

For some $\delta > 0$, let $\Theta_c^+ = \{\theta \in R^{\dim(\theta)} : \text{dist}(\theta, \Theta_c) \leq \delta\}$ be a compact subset of $\widetilde{\Theta}$ that is slightly larger than Θ_c (where $\text{dist}(\theta, \Theta_c) = \inf\{||\theta - \theta_c|| : \theta_c \in \Theta_c\}$). Let $B(\theta, \varepsilon)$ denote an open ball of radius $\varepsilon > 0$ centered at θ.

The results of Theorem 2 hold uniformly for the true parameter lying in a compact subset Θ_c of $\widetilde{\Theta}$. To obtain the results of Theorem 2, we need to establish Edgeworth expansions (and other results) that hold uniformly for the true parameter lying in the larger set Θ_c^+. The reason is that the parametric bootstrap uses $\widetilde{\theta}_n$ as the true parameter and $\widetilde{\theta}_n \in \Theta_c^+$ with probability that goes to one (at a sufficiently fast rate) when the true parameter is in Θ_c. The next Lemma is a simple, but key, result that allows one to do so. It is used in the proof of Lemma 9(b) below. The condition of the Lemma on $\widetilde{\theta}_n$ is an implication of Condition C_s.

Lemma 8 *Suppose* $\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(\widetilde{\theta}_n \notin B(\theta_0, \delta)) = o(n^{-(s-2)/2})$, *where* Θ_c *is a compact subset of* $\widetilde{\Theta}$ *and* δ *is as in the definition of* Θ_c^+, *and* $\{\lambda_n(\theta) : n \geq 1\}$ *is a sequence of (non-random) real functions on* Θ_c^+ *that satisfies* $\sup_{\theta \in \Theta_c^+} |\lambda_n(\theta)| = o(n^{-(s-2)/2})$. *Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(|\lambda_n(\widetilde{\theta}_n)| > n^{-(s-2)/2}\varepsilon) = o(n^{-(s-2)/2}).$$

The next Lemma provides Edgeworth expansions for the t statistic, $t_n(\theta_{0,r})$, and the bootstrap t statistic, $t_n^*(\widetilde{\theta}_{n,r})$. The Edgeworth expansion for the t statistic is established by utilizing the Edgeworth expansion for the LLDs or WLLDs, given in Theorem 3 or Proposition 1, plus the approximation of the t statistic by a smooth function of the LLDs or WLLDs, given in Lemma 7. The Edgeworth expansion for the bootstrap t statistic is established by utilizing the Edgeworth expansion for the t statistic, given in part (a) of the following Lemma, and Lemma 8.

We now define the components of the Edgeworth expansion of $t_n(\theta_{0,r})$ as well as its bootstrap analogue $t_n^*(\widetilde{\theta}_{n,r})$. Let $\Phi(\cdot)$ denote the distribution function of a standard normal random variable. Let $\overline{\kappa}_{n,s}(\theta) = \kappa_{n,s}(\theta)/n$. By Lemma 3(c), the elements of $\overline{\kappa}_{n,s}(\theta)$ are $O(1)$. Let $\pi_i(\delta, \overline{\kappa}_{n,s}(\theta))$ be a polynomial in $\delta = \partial/\partial z$ whose coefficients are polynomials in the elements of $\overline{\kappa}_{n,s}(\theta)$ and for which $\pi_i(\delta, \overline{\kappa}_{n,s}(\theta))\Phi(z)$ is an even function of z when i is odd and an odd function of z when i is even for $i = 1, ..., s-2$. The Edgeworth expansion of $t_n(\theta_{0,r})$ depends on $\pi_i(\delta, \overline{\kappa}_{n,s}(\theta_0))$. The Edgeworth expansion of $t_n^*(\widetilde{\theta}_{n,r})$ depends on $\pi_i(\delta, \overline{\kappa}_{n,s}(\widetilde{\theta}_n))$.

Lemma 9 *Suppose Assumptions* I-VI *or* W1-W7 *hold, the* PML *or* PWML *estimators* $\{\widehat{\theta}_n \in \widehat{\Theta}_n : n \geq 1\}$ *and the* BG *estimators* $\{\widetilde{\theta}_n : n \geq 1\}$ *satisfy Condition* C_s, *and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some* $s \geq 3$.
(a) *Then,*

$$\sup_{\theta_0 \in \Theta_c} \sup_{z \in R} |P_{\theta_0}(t_n(\theta_{0,r}) \leq z) - [1 + \sum_{i=1}^{s-2} n^{-i/2}\pi_i(\delta, \overline{\kappa}_{n,s}(\theta_0))]\Phi(z)| = o(n^{-(s-2)/2}).$$

(b) *Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0} \left(\sup_{z \in R} |P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \leq z) \right.$$
$$\left. -[1 + \sum_{i=1}^{s-2} n^{-i/2} \pi_i(\delta, \overline{\kappa}_{n,s}(\widetilde{\theta}_n))]\Phi(z)| > n^{-(s-2)/2}\varepsilon \right) = o(n^{-(s-2)/2}).$$

The final lemma shows that the coefficients of the Edgeworth expansions of $t_n(\theta_{0,r})$ and $t^*_n(\widetilde{\theta}_{n,r})$ differ by at most $n^{-1/2}\ln(n)$ except on a set whose probability goes to zero quickly. It is this property that leads to higher-order improvements of bootstrap CIs.

Lemma 10 *Suppose Assumptions* I-VI *or* W1-W7 *hold, the* BG *estimators* $\{\widetilde{\theta}_n : n \geq 1\}$ *satisfy Condition* C_s, *and* $\widetilde{\Theta}$ *satisfies Condition* NS_s *for some integer* $s \geq 3$. *Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(n^{1/2}||\overline{\kappa}_{n,s}(\widetilde{\theta}_n) - \overline{\kappa}_{n,s}(\theta_0)|| > \ln(n)\varepsilon) = o(n^{-(s-2)/2}),$$

where $\overline{\kappa}_{n,s}(\theta)$ *denotes the vector of cumulants of the* PLL *or* PWLL *function.*

8.5 Proofs of Theorems 1 and 2

Proof of Theorem 1

We establish part (c) first. Note that $P_{\theta_0}(\theta_0 \in \Delta CI_{up}(\widehat{\theta}_n)) = P_{\theta_0}(t_n(\theta_{0,r}) \leq z_\alpha)$. In consequence, part (c) follows immediately from Lemma 9(a) with $s = 3$ by replacing z by z_α, since $\Phi(z_\alpha) = 1 - \alpha$.

Next, we prove part (a). We have $P_{\theta_0}(\theta_0 \in \Delta CI_2(\widehat{\theta}_n)) = P_{\theta_0}(|t_n(\theta_{0,r})| \leq z_{\alpha/2})$. By Lemma 9(a) with $s = 4$, we have

$$\sup_{\theta_0 \in \Theta_c} |P_{\theta_0}(|t_n(\theta_{0,r})| \leq z_{\alpha/2}) - [1 + n^{-1}\pi_2(\delta, \overline{\kappa}_{n,4}(\theta_0))](\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}))| = o(n^{-1}) \tag{8.17}$$

because the evenness of $\pi_1(\delta, \overline{\kappa}_{n,4}(\theta_0))\Phi(z)$ in z implies that $\pi_1(\delta, \overline{\kappa}_{n,4}(\theta_0))(\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})) = 0$. Since $\Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2}) = 1 - \alpha$, this establishes part (a).

To prove part (b), we apply Lemma 9(a) with $s = 3$ to obtain (8.17) with $n^{-1}\pi_2(\delta, \overline{\kappa}_{n,4}(\theta_0)$ deleted and $o(n^{-1})$ replaced by $o(n^{-1/2})$. This yields part (b). □

Proof of Theorem 2

We establish part (b) first. Note that $P_{\theta_0}(\theta_{0,r} \in CI_{up}(\widehat{\theta}_n)) = P_{\theta_0}(t_n(\theta_{0,r}) \leq z^*_{t,\alpha})$. We show that the latter equals $1 - \alpha + o(n^{-1}\ln(n))$ uniformly over $\theta_0 \in \Theta_c$. By Lemmas 9(b), 10, and 9(a), respectively, each with $s = 4$, we have: for all $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0} \left(\sup_{z \in R} |P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \leq z) \right.$$
$$\left. -[1 + \sum_{i=1}^{2} n^{-i/2} \pi_i(\delta, \overline{\kappa}_{n,4}(\widetilde{\theta}_n))]\Phi(z)| > n^{-1}\varepsilon \right) = o(n^{-1}),$$

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}\left(\sup_{z\in R}\Big|[\pi_i(\delta,\overline{\kappa}_{n,4}(\widetilde{\theta}_n)) - \pi_i(\delta,\overline{\kappa}_{n,4}(\theta_0))]\Phi(z)| > n^{-1/2}\ln(n)\varepsilon\right) = o(n^{-1}) \text{ for } i=1,2, \text{ and}$$

$$\sup_{\theta_0\in\Theta_c}\sup_{z\in R}|P_{\theta_0}(t_n(\theta_{0,r}) \le z) - [1+\sum_{i=1}^{2} n^{-i/2}\pi_i(\delta,\overline{\kappa}_{n,4}(\theta_0))]\Phi(z)| = o(n^{-1}). \quad (8.18)$$

The three results of (8.18) combine to give

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}(\sup_{z\in R}|P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \le z) - P_{\theta_0}(t_n(\theta_{0,r}) \le z)| > n^{-1}\ln(n)\varepsilon) = o(n^{-1}). \quad (8.19)$$

If $t^*_n(\widetilde{\theta}_{n,r})$ is absolutely continuous, then $P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \le z^*_{t,\alpha}) = 1-\alpha$. Whether or not $t^*_n(\widetilde{\theta}_{n,r})$ is absolutely continuous, the Edgeworth expansion of Lemma 9(b) with $s=4$ implies that

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}(|P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \le z^*_{t,\alpha}) - (1-\alpha)| > n^{-1}\varepsilon) = o(n^{-1}) \quad (8.20)$$

for all $\varepsilon > 0$. This holds because the continuity in z of the Edgeworth expansion in Lemma 9(b) implies that there exists a value $z^{**}_{t,\alpha}$ for which the Edgeworth expansion at $z = z^{**}_{t,\alpha}$ equals $1-\alpha$ and, by definition of $z^*_{t,\alpha}$, $|P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \le z^*_{t,\alpha}) - (1-\alpha)| \le |P^*_{\widetilde{\theta}_n}(t^*_n(\widetilde{\theta}_{n,r}) \le z^{**}_{t,\alpha}) - (1-\alpha)|$.

Taking $z = z^*_{t,\alpha}$ in (8.19) and combining it with (8.20) gives

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}(|1-\alpha - P_{\theta_0}(t_n(\theta_{0,r}) \le z^*_{t,\alpha})| > n^{-1}\ln(n)\varepsilon) = o(n^{-1}). \quad (8.21)$$

The expression inside the absolute value sign is non-random. Hence, for n large, $|1-\alpha-P_{\theta_0}(t_n(\theta_{0,r}) \le z^*_{t,\alpha})| \le n^{-1}\ln(n)\varepsilon$, which establishes part (b) of the Theorem.

Next, we prove part (a). We have $P_{\theta_0}(\theta_{0,r} \in CI_{sym}(\widehat{\theta}_n)) = P_{\theta_0}(|t_n(\theta_{0,r})| \le z^*_{t,\alpha})$. We show that the latter equals $1-\alpha + o(n^{-3/2}\ln(n))$ uniformly over $\theta_0 \in \Theta_c$. By Lemmas 9(b), 10, and 9(a), respectively, each with $s=5$, we have: for all $\varepsilon > 0$,

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}\left(\sup_{z\in R}|P^*_{\widetilde{\theta}_n}(|t^*_n(\widetilde{\theta}_{n,r})| \le z) - [1+n^{-1}\pi_2(\delta,\overline{\kappa}_{n,5}(\widetilde{\theta}_n))](\Phi(z)-\Phi(-z))| > n^{-3/2}\varepsilon\right) = o(n^{-3/2}),$$

$$\sup_{\theta_0\in\Theta_c} P_{\theta_0}\left(\sup_{z\in R}\Big|[\pi_2(\delta,\overline{\kappa}_{n,5}(\widetilde{\theta}_n)) - \pi_2(\delta,\overline{\kappa}_{n,5}(\theta_0))](\Phi(z)-\Phi(-z))| > n^{-1/2}\ln(n)\varepsilon\right) = o(n^{-3/2}), \text{ and}$$

$$\sup_{\theta_0\in\Theta_c}\sup_{z\in R}|P_{\theta_0}(|t_n(\theta_{0,r})| \le z) - [1+n^{-1}\pi_2(\delta,\overline{\kappa}_{n,5}(\theta_0))](\Phi(z)-\Phi(-z))| = o(n^{-3/2}), \quad (8.22)$$

using the evenness of $\pi_j(\delta, \overline{\kappa}_{n,5}(\widetilde{\theta}_n))\Phi(z)$ and $\pi_j(\delta, \overline{\kappa}_{n,5}(\theta_0))\Phi(z)$ in z for $j = 1, 3$ in the first and third results respectively.

The three results of (8.22) combine to give

$$\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(\sup_{z \in R} |P^*_{\widetilde{\theta}_n}(|t^*_n(\widetilde{\theta}_{n,r})| \leq z) - P_{\theta_0}(|t_n(\theta_{0,r})| \leq z)| > n^{-3/2}\ln(n)\varepsilon) = o(n^{-3/2}). \tag{8.23}$$

Given (8.23), the remainder of the proof of part (a) is analogous to that of part (b) above.

Part (c) holds by the same proofs as for parts (a) and (b) but in the proof of part (a) $n^{-3/2}$ is replaced by $n^{-1/2}$ throughout, the terms $n^{-1}\pi_2(\delta, \overline{\kappa}_{n,5}(\widetilde{\theta}_n))$ and $n^{-1}\pi_2(\delta, \overline{\kappa}_{n,5}(\theta_0))$ are deleted in (8.22), and $\ln(n)$ is deleted in (8.23) and in the proof of part (b) n^{-1} is replaced by $n^{-1/2}$ throughout, the sum over i from 1 to 2 is replaced by the summand for $i = 1$ alone in (8.18), and $\ln(n)$ is deleted in (8.19), (8.21), and the sentence following (8.21). □

8.6 Proofs of Lemmas 1 and 6-10

Proof of Lemma 1 The proof for PML estimators is analogous to that of Theorem 4(a) of LRZ using Lemma 3(b). For PWML estimators, the result holds by Theorem 6(a) of Andrews and Lieberman (2002). □

Proof of Lemma 6

For any convex set $B \subset R^{\dim(A)}$ and any $\tau > 0$, let $B^+_\tau = \{x \in R^{\dim(A)} : ||x - y|| \leq \tau \text{ for some } y \in B\}$. We have

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_c, B \in \mathcal{B}_{\dim(A)}} (P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B) - P_{\theta_0}(A_n(\theta_0) \in B)) \\
= &\sup_{\theta_0 \in \Theta_c, B \in \mathcal{B}_{\dim(A)}} (P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B, ||\xi_n(\theta_0)|| \leq \omega_n) - P_{\theta_0}(A_n(\theta_0) \in B) \\
&\quad + P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B, ||\xi_n(\theta_0)|| > \omega_n)) \\
\leq &\sup_{\theta_0 \in \Theta_c, B \in \mathcal{B}_{\dim(A)}} \big(P_{\theta_0}(A_n(\theta_0) \in B^+_{\omega_n}) - P_{\theta_0}(A_n(\theta_0) \in B)\big) \\
&\quad + \sup_{\theta_0 \in \Theta_c} P_{\theta_0}(||\xi_n(\theta_0)|| > \omega_n).
\end{aligned} \tag{8.24}$$

The second term on the right-hand side is $o(n^{-(s-2)/2})$ by assumption.

Given that $A_n(\theta_0)$ has an Edgeworth expansion with remainder $o(n^{-(s-2)/2})$ uniformly over $\theta_0 \in \Theta_c$, the first term on the rhs of (8.24) is bounded by

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_c, B \in \mathcal{B}_{\dim(A)}} \left(\int_{B^+_{\omega_n}} (1 + \sum_{i=1}^{[s-2]} n^{-i/2} \pi_{n,i}(z, \theta_0)) \phi_{\Sigma_n(\theta_0)}(z) dz \right. \\
&\qquad \left. - \int_B (1 + \sum_{i=1}^{[s-2]} n^{-i/2} \pi_{n,i}(z, \theta_0)) \phi_{\Sigma_n \theta_0)}(z) dz \right) + o(n^{-(s-2)/2}).
\end{aligned} \tag{8.25}$$

The expression in (8.25) is $O(\omega_n) = o(n^{-(s-2)/2})$ because $\phi_{\Sigma_n \theta_0)}(z)$ and its derivatives of all orders are bounded over $z \in R^{\dim(A)}$ given the assumptions on $\Sigma_n(\theta_0)$ and the polynomials $\{\pi_{n,i}(z, \theta_0) : i = 1, ..., s-2\}$ have coefficients that are $O(1)$ uniformly over $\theta_0 \in \Theta_c$. Hence, the left-hand side of (8.24) is less than or equal to $o(n^{-(s-2)/2})$.

Let $B_\tau^- = \{x \in B : ||x - y|| \geq \tau \text{ for all } y \in B^c\}$, where B^c denotes the complement of B. We have

$$P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B) \geq P_{\theta_0}(A_n(\theta_0) \in B_{\omega_n}^-, ||\xi_n(\theta_0)|| \leq \omega_n). \tag{8.26}$$

Using this, an analogous argument to that of (8.24) and (8.25) shows that

$$\sup_{\theta_0 \in \Theta_c, B \in \mathcal{B}_{\dim(A)}} (P_{\theta_0}(A_n(\theta_0) \in B) - P_{\theta_0}(A_n(\theta_0) + \xi_n(\theta_0) \in B)) \leq o(n^{-(s-2)/2}), \tag{8.27}$$

which completes the proof. □

Proof of Lemma 7

Let $\rho_n(\theta) = n^{-1} L_n(\theta, \overline{X}_n)$ or $n^{-1} L_{W,n}(\theta, \overline{X}_n)$. Suppose $\Delta_n(\theta_0) = n^{1/2}(\widehat{\theta}_n - \theta_0)$. By Conditions C_s and NS_s(i), which implies that θ_0 lies in the interior of Θ, we have $\inf_{\theta_0 \in \Theta_c} P_{\theta_0}(\widehat{\theta}_n$ is in the interior of $\Theta) = 1 - o(n^{-(s-2)/2})$ and $\inf_{\theta_0 \in \Theta_c}$ $P_{\theta_0}((\partial/\partial\theta)\rho_n(\widehat{\theta}_n)) = 0) = 1 - o(n^{-(s-2)/2})$. Element by element Taylor expansions of $(\partial/\partial\theta)\rho_n(\widehat{\theta}_n)$ about θ_0 of order $s-1$ give

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\rho_n(\widehat{\theta}_n) = \frac{\partial}{\partial\theta}\rho_n(\theta_0) + \sum_{j=1}^{s-2} \frac{1}{j!} D^j \frac{\partial}{\partial\theta}\rho_n(\theta_0)(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0) \\
&+ \frac{1}{(s-1)!} ED^{s-1} \frac{\partial}{\partial\theta}\rho_n(\theta_0)(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0) + \zeta_{1n}(\theta_0) + \zeta_{2n}(\theta_0), \text{ where} \\
\zeta_{1n}(\theta_0) &= \frac{1}{(s-1)!}(D^{s-1}\frac{\partial}{\partial\theta}\rho_n(\theta_n^+) - D^{s-1}\frac{\partial}{\partial\theta}\rho_n(\theta_0))(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0), \\
\zeta_{2n}(\theta_0) &= \frac{1}{(s-1)!}(D^{s-1}\frac{\partial}{\partial\theta}\rho_n(\theta_0) - ED^{s-1}\frac{\partial}{\partial\theta}\rho_n(\theta_0))(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0),
\end{aligned} \tag{8.28}$$

θ_n^+ lies between $\widehat{\theta}_n$ and θ_0, and $D^j(\partial/\partial\theta)\rho_n(\theta_0)(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0)$ denotes $D^j(\partial/\partial\theta)\rho_n(\theta_0)$ as a j-linear map, whose coefficients are partial derivatives of $(\partial/\partial\theta)\rho_n(\theta_0)$ of order j, applied to the j-tuple $(\widehat{\theta}_n - \theta_0, ..., \widehat{\theta}_n - \theta_0)$. Note that $n^{-1}Z_n^+(\theta_0)$ is the column vector whose elements are the non-redundant components of $(\partial/\partial\theta)\rho_n(\theta_0)$, $D^1(\partial/\partial\theta)\rho_n(\theta_0)$, ..., $D^{s-2}(\partial/\partial\theta)\rho_n(\theta_0)$ plus the components of $ED^{s-1}(\partial/\partial\theta)\rho_n(\theta_0)$. Let $e_n(\theta_0) = ((\zeta_{1n}(\theta_0) + \zeta_{2n}(\theta_0))', 0, ..., 0)'$ be conformable with $Z_n^+(\theta_0)$. The first equation in (8.28) can be written as $\nu(n^{-1}Z_n^+(\theta_0) + e_n(\theta_0), \widehat{\theta}_n - \theta_0) = 0$, where $\nu(\cdot, \cdot)$ is an infinitely differentiable function that satisfies $\nu(n^{-1}E_{\theta_0}Z_n^+(\theta_0), 0) = 0$ for all $n \geq 1$, and $(\partial/\partial x)\nu(n^{-1}E_{\theta_0}Z_n^+(\theta_0), x)|_{x=0} = n^{-1}E_{\theta_0}(\partial^2/\partial\theta\partial\theta')L_n(\theta_0, \overline{X}_n)$ is negative definite for n large because it converges to $-\Sigma^{-1}(\theta_0)$ as $n \to \infty$, using the information matrix equality, and the latter is negative definite by Condition NS_s(ii).

The same is true in the PWLL case. Hence, the implicit function theorem can be applied to $\nu(\cdot,\cdot)$ at the point $(n^{-1}E_{\theta_0}Z_n^+(\theta_0),0)$ to obtain

$$\inf_{\theta_0\in\Theta_c} P_{\theta_0}(\widehat{\theta}_n-\theta_0=\Lambda(n^{-1}Z_n^+(\theta_0)+e_n(\theta_0)))=1-o(n^{-(s-2)/2}), \tag{8.29}$$

where Λ is a function that does not depend on n or θ_0, is infinitely differentiable in a neighborhood of $n^{-1}E_{\theta_0}Z_n^+(\theta_0)$ for all n large, and satisfies $\Lambda(n^{-1}E_{\theta_0}Z_n^+(\theta_0))=0$.

We apply Lemma 6 with $A_n(\theta_0) = n^{1/2}\Lambda(n^{-1}Z_n^+(\theta_0))$ and $\xi_n(\theta_0) = n^{1/2}(\Lambda(n^{-1}Z_n^+(\theta_0) + e_n(\theta_0))-\Lambda(n^{-1}Z_n^+(\theta_0)))$ to obtain

$$\sup_{\theta_0\in\Theta_c,B\in\mathcal{B}_{\dim(\theta)}} |P_{\theta_0}(n^{1/2}\Lambda(n^{-1}Z_n^+(\theta_0)+e_n(\theta_0))\in B)$$
$$-P_{\theta_0}(n^{1/2}\Lambda(n^{-1}Z_n^+(\theta_0))\in B)|=o(n^{-(s-2)/2}). \tag{8.30}$$

Lemma 6 applies because (i) $P_{\theta_0}(||\xi_n(\theta_0)||>\omega_n)\leq P_{\theta_0}(Cn^{1/2}||e_n(\theta_0)||>\omega_n)$ by a mean value expansion, (ii) $||e_n(\theta_0)||=||\zeta_{1n}(\theta_0)||+||\zeta_{2n}(\theta_0)||$, (iii) $\zeta_{jn}(\theta_0)$ satisfies $\inf_{\theta_0\in\Theta_c}P_{\theta_0}(||\zeta_{jn}(\theta_0)||\leq C||\widehat{\theta}_n-\theta_0||^s)=1-o(n^{-(s-2)/2})$ for $j=1,2$, (iv) ω_n, which is defined to equal $n^{1/2-s/2}\ln^s(n)$, is $o(n^{-(s-2)/2})$, (v) $\sup_{\theta_0\in\Theta_c}P_{\theta_0}(n^{1/2}||e_n(\theta_0)||>\omega_n)\leq\sup_{\theta_0\in\Theta_c}P_{\theta_0}(Cn^{1/2}||\widehat{\theta}_n-\theta_0||^s>\omega_n)+o(n^{-(s-2)/2})=o(n^{-(s-2)/2})$ by Condition C_s, and (vi) $A_n(\theta_0)=n^{1/2}\Lambda(n^{-1}Z_n^+(\theta_0))$ has an Edgeworth expansion (with remainder $o(n^{-(s-2)/2})$ uniformly over $\theta_0\in\Theta_c$) by the proof of Lemma 9(a) below. Property (iii) for $j=1$ uses the assumption that $f_\theta(\lambda)$ is $s+1$ times partially differentiable with respect to θ. Property (iii) holds for $j=2$ because it can be shown that $D^{s-1}(\partial/\partial\theta)\rho_n(\theta_0)-ED^{s-1}(\partial/\partial\theta)\rho_n(\theta_0)=O_p(n^{-1/2})$ using a large deviation inequality.

Equations (8.29) and (8.30) yield part (a) of the Lemma with $G(\cdot)=\Lambda(\cdot)$.

Next, suppose $\Delta_n(\theta_0)$ equals $t_n(\theta_{0,r})$. The t statistic $\Delta_n(\theta_0)$ is a function of $\widehat{\theta}_n$. We take a Taylor expansion of $\Delta_n(\theta_0)/n^{1/2}$ about $\widehat{\theta}_n=\theta_0$ to order $s-1$, where the highest-order term involves the expectation of the partial derivatives rather than the partial derivatives themselves (as in (8.28)), to obtain

$$\Delta_n(\theta_0)=n^{1/2}(\widetilde{\Lambda}(n^{-1}Z_n^+(\theta_0),\widehat{\theta}_n-\theta_0)+\widetilde{\zeta}_n(\theta_0)), \tag{8.31}$$

where $\widetilde{\Lambda}$ is an infinitely differentiable function that does not depend on θ_0, $\widetilde{\Lambda}(n^{-1}E_{\theta_0}Z_n^+(\theta_0),0)=0$ for n large, $\widetilde{\zeta}_n(\theta_0)$ is the remainder term in the Taylor expansion, and $||\widetilde{\zeta}_n(\theta_0)||=O(||\widehat{\theta}_n-\theta_0||^s)$. Combining (8.29) with (8.31) gives $\Delta_n(\theta_0)=n^{1/2}(\widetilde{\Lambda}(n^{-1}Z_n^+(\theta_0),\ \Lambda(n^{-1}Z_n^+(\theta_0)+e_n(\theta_0)))+\widetilde{\zeta}_n(\theta_0))$. We apply Lemma 6 again, using the result above for $||\widetilde{\zeta}_n(\theta_0)||$, to obtain an analogue of (8.30) with $A_n(\theta_0) = n^{1/2}\widetilde{\Lambda}(n^{-1}Z_n^+(\theta_0),\Lambda(n^{-1}Z_n^+(\theta_0)))$. We can write $G(n^{-1}Z_n^+(\theta_0)) = \widetilde{\Lambda}(n^{-1}Z_n^+(\theta_0),\ \Lambda(n^{-1}Z_n^+(\theta_0)))$, where $G(\cdot)$ is infinitely differentiable and $G(n^{-1}E_{\theta_0}Z_n^+(\theta_0))=\widetilde{\Lambda}(n^{-1}E_{\theta_0}Z_n^+(\theta_0),\ \Lambda(n^{-1}E_{\theta_0}Z_n^+(\theta_0)))=\widetilde{\Lambda}(n^{-1}E_{\theta_0}Z_n^+(\theta_0),0)=0$ for all n large. Combining this, the analogue of (8.30), and (8.31) gives the result of the Lemma for $\Delta_n(\theta_0)$ equal to $t_n(\theta_{0,r})$. □

Proof of Lemma 8

We have

$$
\begin{aligned}
&\sup_{\theta_0\in\Theta_c} P_{\theta_0}(|\lambda_n(\widetilde{\theta}_n)| > n^{-(s-2)/2}\varepsilon)\\
&\le \sup_{\theta_0\in\Theta_c} P_{\theta_0}(|\lambda_n(\widetilde{\theta}_n)| > n^{-(s-2)/2}\varepsilon, \widetilde{\theta}_n \in B(\theta_0,\delta)) + \sup_{\theta_0\in\Theta_c} P_{\theta_0}(\widetilde{\theta}_n \notin B(\theta_0,\delta))\\
&\le \sup_{\theta_0\in\Theta_c} P_{\theta_0}(\sup_{\theta\in\Theta_c^+}|\lambda_n(\theta)| > n^{-(s-2)/2}\varepsilon) + o(n^{-(s-2)/2})\\
&= 1(o(n^{-(s-2)/2}) > n^{-(s-2)/2}\varepsilon) + o(n^{-(s-2)/2})\\
&= o(n^{-(s-2)/2}), \qquad (8.32)
\end{aligned}
$$

where the second inequality uses the fact that when $\widetilde{\theta}_n \in B(\theta_0,\delta)$ and $\theta_0 \in \Theta_c$ one has $\widetilde{\theta}_n \in \Theta_c^+$. □

Proof of Lemma 9

We establish part (a) first. By Lemma 7, it suffices to show that the result of part (a) holds with $t_n(\theta_{0,r})$ replaced by $n^{1/2}G(n^{-1}Z_n^+(\theta_0))$. That is, it suffices to show that $n^{1/2}G(n^{-1}Z_n^+(\theta_0))$ possesses the Edgeworth expansion given in part (a) of the present Lemma with remainder $o(n^{-(s-2)/2})$ uniformly over $\theta_0 \in \Theta_c$. Theorem 3 for the PLL case and Proposition 1 for the PWLL case establish Edgeworth expansions for $W_n(\theta_0) = n^{1/2}(n^{-1}Z_n(\theta_0) - n^{-1}E_{\theta_0}Z_n(\theta_0))$ for each $\theta_0 \in \Theta_c$. An Edgeworth expansion for $n^{1/2}G(n^{-1}Z_n^+(\theta_0))$ is obtained from that of $n^{1/2}(n^{-1}Z_n(\theta_0) - n^{-1}E_{\theta_0}Z_n(\theta_0))$ by the argument in Bhattacharya (1985, Pf. of Thm. 1) or Bhattacharya and Ghosh (1978, Pf. of Thm. 2) using the smoothness of $G(\cdot)$, $G(n^{-1}E_{\theta_0}Z_n^+(\theta_0)) = 0$ for all $n \ge 1$ and all $\theta_0 \in \Theta_c$, and Condition NS_s(ii).

Part (b) follows from Lemma 8 with

$$
\begin{aligned}
\lambda_n(\theta_0) &= \sup_{z\in R}|P^*_{\theta_0}(t^*_n(\theta_{0,r}) \le z) - [1 + \sum_{i=1}^{s-2} n^{-i/2}\pi_i(\delta, \overline{\kappa}_{n,s}(\theta_0))]\Phi(z)|\\
&= \sup_{z\in R}|P_{\theta_0}(t_n(\theta_{0,r}) \le z) - [1 + \sum_{i=1}^{s-2} n^{-i/2}\pi_i(\delta, \overline{\kappa}_{n,s}(\theta_0))]\Phi(z)|. \qquad (8.33)
\end{aligned}
$$

The first condition of Lemma 8 holds by Condition C_s and the second condition of Lemma 8 holds by part (a) of the present Lemma with Θ_c replaced by the compact set Θ_c^+. The second equality of (8.33) holds because the distribution of the bootstrap t statistic, $t^*_n(\theta_{0,r})$, when the bootstrap sample is generated by θ_0 is the same as the distribution of the original sample t statistic, $t_n(\theta_{0,r})$, when the original sample is generated by θ_0, because $t^*_n(\theta_{0,r})$ and $t_n(\theta_{0,r})$ are invariant with respect to $\overline{X}_n$ and the true mean μ_0 respectively. □

Proof of Lemma 10

Let $\overline{\kappa}_{n,s}(\theta)_\eta$ denote an element of $\overline{\kappa}_{n,s}(\theta)$. By a mean-value expansion, for all $\theta_0 \in \Theta_c$ and all $\theta \in \Theta_c^+$ such that $||\theta - \theta_0|| < \delta$ (where δ is as in the definition of Θ_c^+),

$$|\overline{\kappa}_{n,s}(\theta)_\eta - \overline{\kappa}_{n,s}(\theta_0)_\eta| \le K_n||\theta - \theta_0||, \text{ where}$$

$$K_n = \sup_{\theta \in \Theta_c^+, i=1,\ldots,\dim(\theta)} |(\partial/\partial\theta_i)\overline{\kappa}_{n,s}(\theta)_\eta|. \tag{8.34}$$

We show below that K_n is a constant that satisfies $\limsup_{n\to\infty} K_n < \infty$.

Let $\gamma > 0$ satisfy $\gamma < \varepsilon/(\dim^{1/2}(\overline{\kappa}) \limsup_{n\to\infty} K_n)$, where $\dim(\overline{\kappa})$ denotes the dimension of $\overline{\kappa}_{n,s}(\theta)$. Then, we have

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_c} P_{\theta_0}(n^{1/2}||\overline{\kappa}_{n,s}(\widetilde{\theta}_n) - \overline{\kappa}_{n,s}(\theta_0)|| > \ln(n)\varepsilon) \\
&\le \sup_{\theta_0 \in \Theta_c} P_{\theta_0}(n^{1/2}||\overline{\kappa}_{n,s}(\widetilde{\theta}_n) - \overline{\kappa}_{n,s}(\theta_0)|| > \ln(n)\varepsilon,\ n^{1/2}||\widetilde{\theta}_n - \theta_0|| \le \ln(n)\gamma) \\
&\quad + \sup_{\theta_0 \in \Theta_c} P_{\theta_0}(n^{1/2}||\widetilde{\theta}_n - \theta_0|| > \ln(n)\gamma) \\
&\le \sup_{\theta_0 \in \Theta_c} P_{\theta_0}(\dim^{1/2}(\overline{\kappa}) K_n n^{1/2}||\widetilde{\theta}_n - \theta_0|| > \ln(n)\varepsilon,\ n^{1/2}||\widetilde{\theta}_n - \theta_0|| \le \ln(n)\gamma) \\
&\quad + o(n^{-(s-2)/2}) \\
&= o(n^{-(s-2)/2}),
\end{aligned} \tag{8.35}$$

where the second inequality uses (8.34) and Condition C_s.

We have $\limsup_{n\to\infty} K_n < \infty$ provided

$$\sup_{\theta \in \Theta_c} |(\partial/\partial\theta_i)\kappa_{n,s}(\theta)_\eta| = O(n) \text{ for all } i \le \dim(\theta), \tag{8.36}$$

for any compact set $\Theta_c \subset \widetilde{\Theta}$ (since Θ_c^+ is a compact subset in $\widetilde{\Theta}$).

First, consider the case of cumulants of the PLL function. Suppose $\kappa_{n,s}(\theta)_\eta$ is a cumulant of order two or greater. By Lemma 2(c) and the chain rule, $(\partial/\partial\theta_i)\kappa_{n,s}(\theta)_\eta$ is a finite sum of terms of the form

$$C_q \mathrm{tr}\left(\prod_{r=1}^{q} (M_n \overline{B}_r M_n \overline{T}_r\right),$$

where $\overline{B}_r$ equals either $B_{n,\nu(\eta_r)}(\theta)$ or $(\partial/\partial\theta_i)B_{n,\nu(\eta_r)}(\theta)$ and $\overline{T}_r$ equals either $T_n(f_\theta)$ or $(\partial/\partial\theta_i)T_n(f_\theta) = T_n(g_{\theta,i})$.

Note that $(\partial/\partial\theta_i)B_{n,\nu(\eta_r)}(\theta)$ has the same form as $B_{n,\nu(\eta_r)}(\theta)$ because it is a partial derivative of $-(1/2)T_n^{-1}(f_\theta)$, just as $B_{n,\nu(\eta_r)}(\theta)$ is, see (8.2). Hence, $\overline{B}_r$ has the same form as $B_{n,\nu(\eta_r)}(\theta)$. Also, $\overline{B}_r\overline{T}_r$ has the same form as $B_{n,\nu(\eta_r)}(\theta)T_n(f_\theta)$. If $\overline{T}_r$ equals $T_n(f_\theta)$, this follows from the previous result. If $\overline{T}_r$ equals $T_n(g_{\theta,i})$, then $\overline{B}_r\overline{T}_r$ is of the same form as

$$B_{n,\nu(\eta_r)}(\theta)T_n(g_{\theta,i}) = \sum_{k=1}^{b} a_k \left(\prod_{j=1}^{p_k} T_n^{-1}(f_\theta)T_n(g_{\theta,k,j})\right) T_n^{-1}(f_\theta)T_n(g_{\theta,i}), \tag{8.37}$$

using (8.2), and the rhs is of the same form as $B_{n,\nu(\eta_r)}(\theta)T_n(f_\theta)$.

Now, because $\overline{B}_r\overline{T}_r$ has the same form as $B_{n,\nu(\eta_r)}(\theta)T_n(f_\theta)$, $(\partial/\partial\theta_i)\kappa_{n,s}(\theta)_\eta$ has the same form as $\kappa_{n,s}(\theta)_\eta$ itself given in Lemma 2(c). In consequence, the proof of Lemma 3(a) and (b) shows that (8.36) holds.

Next, suppose $\kappa_{n,s}(\theta)_\eta$ is a cumulant of order one. Then, the second summand in its expression given in Lemma 2(b) is of the same form as cumulants of order two or greater and, hence, is dealt with by the argument above. The first summand in the expression in Lemma 2(b), viz., $F_{n,\nu(\eta)}(\theta)$, has partial derivative $(\partial/\partial\theta_i)F_{n,\nu(\eta)}(\theta)$ that is the same form as $F_{n,\nu(\eta)}(\theta)$ itself given in (8.2) by inspection. Hence, again the proof of Lemma 3(a) and (b) shows that (8.36) holds.

Lastly, consider cumulants of the PWLL function. By (8.1)-(8.4), the cumulants of the PWLL function are the same as those of the PLL function except that $F_{n,\nu}(\theta)$ and $B_{n,\nu}(\theta)$ are defined differently. Nevertheless, for both the PLL and PWLL functions, $(\partial/\partial\theta_i)F_{n,\nu}(\theta)$ and $(\partial/\partial\theta_i)B_{n,\nu}(\theta)$ have the same form as $F_{n,\nu}(\theta)$ and $B_{n,\nu}(\theta)$. In consequence, the argument given above for the PLL function also holds for the PWLL function. $\square$

Footnotes

[1] The authors thank Vadim Marmer for carrying out the Monte Carlo simulations. The first author gratefully acknowledges the research support of the National Science Foundation via grant numbers SBR-9730277 and SES-0001706.

[2] The condition (2.1) on $f_\theta(\lambda)$ is satisfied if $f_\theta(\lambda) = O(|\lambda|^{-2d})$ as $|\lambda| \downarrow 0$. The latter condition often appears in the literature. It is slightly stronger than (2.1), but is simpler. On the other hand, some of this simplicity is lost when one considers derivatives of $f_\theta(\lambda)$ with respect to d, because a $\log(|\lambda|)$ term arises. Condition (2.1) has the advantage of including cases where $f_\theta(\lambda) = |\lambda|^{-2d} g_\theta(\lambda)$ and $g_\theta(\lambda)$ is slowly varying at $\lambda = 0$.

[3] In fact, this is true with $\overline{X}_n$ replaced by any estimator $\widehat{\mu}_n$ of μ_0 for which $n^{1/2-d_0}(\widehat{\mu}_n - \mu_0) = O_p(1)$.

[4] If the closest element is not unique, then any of the closest elements can be used.

[5] Strictly speaking, this is not true. One could use a null-restricted BG estimator to construct t tests for $H_0 : \theta_{0,r} = \theta_{H,r}$ for a range of values of $\theta_{H,r}$ and invert the tests to obtain a CI. That is, take the CI for $\theta_{0,r}$ to be the set of all values $\theta_{H,r}$ for which the null-restricted parametric bootstrap t test of $H_0 : \theta_{0,r} = \theta_{H,r}$ fails to reject the null hypothesis. Such a CI has the same higher-order properties as the null-restricted tests upon which it is based. This bootstrap CI has the disadvantage, however, that it may be difficult to compute. To compute the CI, one has to compute null-restricted t tests for a range of values of $\theta_{H,r}$.

[6] The stated result holds by the same proof as for part (a) with $n^{-3/2}$ replaced by n^{-1} throughout and with $\ln(n)$ deleted in (8.23).

[7] If the $n^{-1/2}$ term in the Edgeworth expansion of $t_n(\theta_{0,r})$ is identically zero, not just independent of θ_0, then the error for the delta method upper CI is given in Comment 3 following Theorem 1, rather than in Theorem 1(c).

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. (2001): "Higher-order Improvements of the Parametric Bootstrap for Markov Processes," Cowles Foundation Discussion Paper No. 1334, Yale University. Available at http://cowles.econ.yale.edu.

Andrews, D. W. K. and O. Lieberman (2002): "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.

Beran, J. (1994): *Statistics for Long-memory Processes.* New York: Chapman and Hall.

Bhattacharya, R. N. (1985): "Some Recent Results on Cramér-Edgeworth Expansions with Applications," *Multivariate Analysis—VI.* Ed. by P. R. Krishnaiah. New York: Elsevier, 57–75.

Bhattacharya, R. N. and J. K. Ghosh (1978): "On the Validity of the Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

Chambers, J. (1967): "On Methods of Asymptotic Approximation for Multivariate Distributions," *Biometrika,* 54, 367–383.

Dahlhaus, R. (1989): "Efficient Parameter Estimation for Self-similar Processes," *Annals of Statistics,* 17, 1749–1766.

Davidson, J. (2001): "Alternative Bootstrap Procedures for Testing Cointegration in Fractionally Integrated Processes," unpublished manuscript, Cardiff University.

Davidson, R. and J. G. MacKinnon (1999): "The Size Distortion of Bootstrap Tests," *Econometric Theory,* 15, 361–376.

Durbin, J. (1980): "Approximation for Densities of Sufficient Estimators," *Biometrika,* 67, 311–333.

Dhrymes, P. (2000): *Mathematics for Econometrics.* Third edition. New York: Springer.

Feller, W. (1971): *An Introduction to Probability Theory: Volume II.* New York: Wiley.

Fox, R. and M. S. Taqqu (1986): "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series," *Annals of Statistics,* 14, 517–532.

Giraitis, L. and P. M. Robinson (2001): "Edgeworth Expansions for the Semiparametric Whittle Estimation of Long Memory," unpublished manuscript, London School of Economics.

Giraitis, L. and D. Surgailis (1990): "A Central Limit Theorem for Quadratic Forms in Strongly Dependent Linear Variables and Applications to Asymptotic Normality of Whittle's Estimate," *Probability Theory and Related Fields,* 86, 87-104.

Giraitis, L. and M. S. Taqqu (1999): "Whittle Estimator for Finite-variance non-Gaussian Time Series with Long Memory," *Annals of Statistics,* 27, 178–203.

Hall, P. (1988): "On Symmetric Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society,* Series B, 50, 35–45.

——— (1992): *The Bootstrap and Edgeworth Expansion.* New York: Springer.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moment Estimators," *Econometrica,* 64, 891–916.

Lieberman, O. and P. C. P. 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., J. Rousseau, and D. M. Zucker (2000): "Valid Edgeworth Expansions for the Sample Autocorrelation Function under Long Range Dependence," *Econometric Theory,* 16, 231–248.

——— (2003): "Valid Edgeworth Expansions for the Maximum Likelihood Estimator of the Parameter of a Stationary, Gaussian, Strongly Dependent Process," *Annals of Statistics,* 31, forthcoming.

Searle, S. R. (1971): *Linear Models.* New York: Wiley.

Skovgaard, I. M. (1986): "On Multivariate Edgeworth Expansions," *International Statistical Review,* 54, 169–186.

Taniguchi, M. (1986): "Third Order Asymptotic Properties of Maximum Likelihood Estimators for Gaussian ARMA Processes," *Journal of Multivariate Analysis,* 18, 1–31.

——— (1991): *Higher Order Asymptotic Theory for Time Series Analysis.* Lecture Notes in Statistics, Vol. 68. New York: Springer.

Table I

Coverage Probabilities of Delta Method and Parametric Bootstrap Confidence Intervals for the Long-memory Parameter d Based on Plug-in ML and Plug-in Whittle ML Estimators for ARFIMA$(0, d, 0)$ Processes

	d					
	0	.1	.2	.3	.4	Avg Abs Dev
(a) Nominal 95% Confidence Interval						
Delta Method PML	.928	.930	.935	.954	.970	.016
Bootstrap PML	.944	.960	.948	.958	.972	.010
Delta Method PWML	.886	.890	.883	.895	.928	.054
Bootstrap PWML	.950	.942	.924	.940	.964	.012
(b) Nominal 99% Confidence Interval						
Delta Method PML	.978	.988	.977	.988	.996	.007
Bootstrap PML	.992	.990	.992	.995	.999	.004
Delta Method PWML	.966	.954	.956	.971	.970	.027
Bootstrap PWML	.991	.989	.990	.990	.990	.000
(c) Nominal 90% Confidence Interval						
Delta Method PML	.869	.872	.882	.913	.940	.026
Bootstrap PML	.890	.886	.876	.880	.937	.021
Delta method PWML	.816	.821	.810	.831	.891	.066
Bootstrap PWML	.894	.889	.888	.873	.938	.019