10-1-2001

Higher-order Improvements of the Parametric Bootstrap for Markov Processes

Donald W.K. Andrews

Higher-order Improvements of the Parametric Bootstrap for Markov Processes

by

Donald W. K. Andrews

October 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1334

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

Higher-order Improvements of the Parametric Bootstrap for Markov Processes

Donald W. K. Andrews[1]

Cowles Foundation for Research in Economics

Yale University

September 2000

Revised: September 2001

Abstract

This paper provides bounds on the errors in coverage probabilities of maximum likelihood-based, percentile-t, parametric bootstrap confidence intervals for Markov time series processes. These bounds show that the parametric bootstrap for Markov time series provides higher-order improvements (over confidence intervals based on first order asymptotics) that are comparable to those obtained by the parametric and nonparametric bootstrap for iid data and are better than those obtained by the block bootstrap for time series. Additional results are given for Wald-based confidence regions.

The paper also shows that k-step parametric bootstrap confidence intervals achieve the same higher-order improvements as the standard parametric bootstrap for Markov processes. The k-step bootstrap confidence intervals are computationally attractive. They circumvent the need to compute a nonlinear optimization for each simulated bootstrap sample. The latter is necessary to implement the standard parametric bootstrap when the maximum likelihood estimator solves a nonlinear optimization problem.

Keywords: Asymptotics, Edgeworth expansion, Gauss-Newton, k-step bootstrap, maximum likelihood estimator, Newton-Raphson, parametric bootstrap, t statistic.

JEL Classification Numbers: C12, C13, C15.

1 Introduction

This paper analyzes the higher-order properties of the parametric bootstrap for maximum-likelihood- (ML) based confidence intervals (CIs) for κ-th order Markov processes possibly with exogenous variables. It is shown that the parametric bootstrap obtains essentially the same higher-order improvements in coverage probabilities relative to standard delta method CIs in the time series context as do the parametric and non-parametric bootstraps for independent and identically distributed (iid) observations. This contrasts with the (nonparametric) block bootstrap for time series, which does not obtain as large improvements, e.g., see Zvingelis (2000), Inoue and Shintani (2000), and Andrews (2001).

In particular, the paper shows that symmetric percentile t CIs constructed using the parametric bootstrap have errors in coverage probability of order $O(N^{-2})$, where N is the sample size. Symmetric percentile t CIs constructed using the delta method, which utilizes the asymptotic normal distribution, have coverage probability errors of magnitude $O(N^{-1})$. Hence, the use of the parametric bootstrap reduces the errors in coverage probability by $O(N^{-1})$. For equal-tailed percentile t CIs, the use of the parametric bootstrap yields errors in coverage probabilities of order $O(N^{-1}\ln N)$, whereas those of the delta method are $O(N^{-1/2}\ln N)$. (The $\ln N$ factors are a product of the method of proof and would not appear in the best possible results.) In contrast, the improvements established in Andrews (2001) for the block bootstrap are only of magnitude $O(N^{-1/4})$ (due to the influence of the independence across blocks, which does not mimic the dependence in the time series of interest).

This paper also analyzes the higher-order properties of a computationally attractive k-step parametric bootstrap procedure for ML estimators. The method was first considered by Davidson and MacKinnon (1999a). For the case of the (nonparametric) block bootstrap, its properties are analyzed in Andrews (2001). The k-step bootstrap is closely related to the one-step and k-step estimators considered by many authors, including Fisher (1925), LeCam (1956), Rothenberg and Leenders (1964), Pfanzagl (1974), Janssen, Jureckova, and Veraverbeke (1985), and Robinson (1988), among others. Let B denote the number of bootstrap repetitions. The standard bootstrap for an ML estimator requires that one solve B nonlinear optimization problems to obtain B bootstrap estimators. These estimators are then used to construct bootstrap CIs, test statistics, etc. In contrast, the k-step bootstrap requires calculation of a closed-form expression for each of the B bootstrap repetitions. Given a bootstrap sample, the k-step bootstrap estimator is obtained by taking k-steps of a Newton-Raphson (NR), default NR, line-search NR, or Gauss-Newton (GN) iterative scheme starting from the estimate based on the original sample.

We show that the distribution function of a k-step bootstrap statistic differs from that of a standard bootstrap statistic by at most N^{-a} with probability $1-o(N^{-a})$ for any $a > 0$, provided k is taken large enough and sufficient smoothness and moment conditions hold. For example, it is often sufficient to take $k \geq 2$ for $a = 1$ and $k \geq 3$ for $a = 2$ for the NR, default NR, and line-search NR k-step bootstraps and $k \geq 3$ for $a = 1$ and $k \geq 5$ for $a = 2$ for the GN k-step bootstrap. These results are used to show that k-step parametric bootstrap CIs yield the same higher-order improvements over delta method CIs as does the standard parametric bootstrap.

The method of proof of the results for the standard parametric bootstrap is as follows. First, we establish an Edgeworth expansion for the ML estimator and the t statistic based on the ML estimator that holds uniformly over a compact set in the parameter space. The method of doing so is similar to that of Bhattacharya and Ghosh (1978). This method is also used by Hall and Horowitz (1996) and Andrews (2001) among others. We utilize an Edgeworth expansion for the normalized sum of strong mixing random variables due to Lahiri (1993), which is an extension of a result of Götze and Hipp (1983), whereas Bhattacharya and Ghosh (1978) consider iid random variables and use a standard Edgeworth expansion for iid random variables. Second, we convert these Edgeworth expansions into Edgeworth expansions for the bootstrap ML estimator and bootstrap t statistic using the fact that the ML estimator lies in a neighborhood of the true value with probability that goes to one at a sufficiently fast rate. Third, we use the argument of Hall (1988) to obtain the error in coverage probability of symmetric percentile t confidence intervals given the Edgeworth expansions for the ML and bootstrap ML t statistics.

To prove the results for the k-step parametric bootstrap, we use the method in Andrews (2001). This method is similar to that used in the numerical analysis literature to establish the quadratic convergence of the Newton-Raphson algorithm. It is also similar to that used in the statistics and econometrics literature to determine the distributional and stochastic differences between statistics, e.g., see Pfanzagl (1974) and Robinson (1988).

This paper provides some Monte Carlo results to illustrate performance of the parametric bootstrap compared to the delta method in the second-order autoregressive (AR(2)) model with Gaussian errors. This model is convenient for Monte Carlo experiments because the ML estimator is the LS estimator, which is available in closed form and, hence, computation is quick. We consider CIs for a nonlinear function of the AR parameters, viz., the cumulative impulse response (CIR), as well as for the AR parameters themselves. We consider sample sizes of 50 and 100 and a variety of different parameter combinations. To see how robust the (Gaussian) parametric bootstrap is to non-normal errors, we also consider errors with t distribution with five degrees of freedom, which exhibits fat tails, and χ^2 distribution with one degree of freedom, which exhibits skewness.

The performances of the delta method and the parametric bootstrap CIs are found to depend on how close the sum of the AR coefficients is from one. When the sum is close to one, both types of CIs perform much more poorly than otherwise. In virtually all parameter combinations, the parametric bootstrap outperforms the delta method in terms of coverage probability. The difference is most pronounced when the sum of AR coefficients is near one. For example, when the AR parameters are .90 and 0.0, the sample size is 100, the errors are normal, and the nominal coverage probabilities of the CIs are .95, the actual coverage probabilities of the delta method, symmetric parametric bootstrap, and equal-tailed parametric bootstrap CIs for the CIR are .714, .876, and .847 respectively. As a second example, when the AR parameters are .50 and 0.0 and everything else is the same as above, the analogous coverage probabilities are .880, .929, and .915. The results change very little when t-5 or χ^2-1 errors are used. Overall, the simulation results indicate that in one Markov model of interest the parametric bootstrap outperforms the delta method.

An alternative bootstrap procedure that can be used in the AR(2) model is the residual-based (RB) bootstrap. We compare the (Gaussian) parametric bootstrap to the RB bootstrap when the errors are normal, t-5, and χ^2-1. For normal and t-5 errors, there is very little difference in the coverage probabilities of the parametric and RB bootstraps. For χ^2-1 errors, the differences are larger. The coverage probabilities of the parametric bootstrap CIs are almost always higher than those of the RB bootstrap CIs. For about half of the parameter combinations considered, the parametric bootstrap coverage probabilities are closer to the nominal value .95 than the RB bootstrap coverage probabilities and vice versa. Hence, the overall performance of the parametric and RB bootstraps are quite similar in the AR(2) model.

No other papers in the literature that we are aware of consider higher-order improvements of the parametric bootstrap for time series processes. In fact, there are few papers that consider higher-order improvements of the parametric bootstrap even for iid observations. One paper that does is Davidson and MacKinnon (1999b). On the other hand, numerous papers in the literature consider different types of bootstrap procedures for time series observations. Rajarishi (1990), Datta and McCormick (1995), and Horowitz (2001) consider a nonparametric bootstrap for Markov processes that utilizes a nonparametric estimator of the transition densities of the process. Bose (1988) and Inoue and Kilian (1999) consider a residual-based bootstrap for AR processes that relies on transforming the data to obtain approximately iid residuals. Paparoditis (1996), Bühlmann (1998), Park (1999), Chang and Park (1999), and Choi and Hall (2000) consider sieve bootstraps for linear time series processes. Many other papers consider the block bootstrap. These include Carlstein (1986), Künsch (1989), Lahiri (1992, 1993, 1996), Hall and Horowitz (1996), Götze and Künsch (1996), Zvingelis (2000), Gonçalves and White (2000), Inoue and Shintani (2000), and Andrews (2001).

The remainder of the paper is organized as follows: Section 2 introduces the parametric Markov model that is considered in the paper and defines the ML estimator and t and Wald statistics. Section 3 defines the parametric bootstrap CIs and CRs. Section 4 states the assumptions. Section 5 provides bounds on the coverage probability errors of the parametric bootstrap CIs and CRs. Section 6 introduces k-step parametric bootstrap CIs and CRs and shows that the same bounds on the coverage probability errors apply as for the standard parametric bootstrap, provided k is taken large enough. Section 7 presents some Monte Carlo simulation results for the parametric bootstrap for an AR(2) model. An Appendix contains proofs of the results.

2 Markov Model and Maximum Likelihood Estimator

In this section, we provide results for likelihood-based methods using the *parametric* bootstrap. The parametric bootstrap utilizes the ML estimator to generate bootstrap samples. It can be used for both bootstrap confidence intervals and tests.

We obtain higher-order improvements of the parametric bootstrap that are the same whether or not the data are dependent.

We consider a correctly specified parametric model for a time series $\{W_i : i = 1, ..., n\}$, where $W_i \in R^{L_w}$. Let $W_i = (Y_i', X_i')'$, where Y_i is a vector of dependent (or response) variables and X_i is a vector of "regressor" variables. The dependent random variables $\{Y_i : i = 1, ..., n\}$ form a κ-th order Markov process. The regressor variables $\{X_i : i = 1, ..., n\}$ are strictly exogenous and, hence, are taken to be fixed (i.e., non-random). All probabilities are based on the randomness in $\{Y_i : i = 1, ..., n\}$ alone.

Assumption 1. (a) The parametric model specifies the density of Y_i given $(X_i, W_{i-1}, W_{i-2}, ..., W_1)$ (with respect to some σ-finite measure μ) to be $d(\cdot|X_i, W_{i-1}, W_{i-2}, ..., W_{i-\kappa}; \theta)$ for $i = \kappa + 1, ..., n$, for some integer $\kappa \geq 0$, where θ is a parameter in the parameter space $\Theta \subset R^{L_\theta}$. (b) For any $\theta_0 \in \Theta$, when $\{Y_i : i \geq 1\}$ is distributed with true parameter θ_0, then $\{Y_i : i \geq 1\}$ is a strong mixing sequence of random variables with strong mixing numbers $\{\alpha(\theta_0, m) : m \geq 1\}$ that satisfy $\sup_{\theta_0 \in \Theta} \alpha(\theta_0, m) \leq C_1 \exp(-C_2 m)$ for some constants $0 < C_1, C_2 < \infty$.

Let E_{θ_0} and P_{θ_0} denote expectation and probability, respectively, when the distribution of the observations is given by the parametric model with true parameter θ_0.

It is convenient notationally to define overlapping observations $\widetilde{W}_i = (W_i', ..., W_{i+\kappa}')'$ for $i = 1, ..., N$, where $N = n - \kappa$. The sample in terms of the overlapping variables is denoted by χ_N :

$$\chi_N = \{\widetilde{W}_i : i = 1, ..., N\}. \tag{2.1}$$

The normalized negative of the log likelihood function is

$$\begin{aligned} \rho_N(\theta) &= N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i, \theta), \text{ where} \\ \rho(\widetilde{W}_i, \theta) &= -\log d(Y_{i+\kappa}|X_{i+\kappa}, W_{i+\kappa-1}, W_{i+\kappa-2}, ..., W_i; \theta).^2 \end{aligned} \tag{2.2}$$

By definition, the ML estimator, $\widehat{\theta}_N$, solves

$$\min_{\theta \in \Theta} \rho_N(\theta). \tag{2.3}$$

The ML estimator also satisfies the first-order conditions

$$\begin{aligned} N^{-1} \sum_{i=1}^{N} g(\widetilde{W}_i, \widehat{\theta}_N) &= 0, \text{ where} \\ g(\widetilde{W}_i, \theta) &= (\partial/\partial\theta)\rho(\widetilde{W}_i, \theta). \end{aligned} \tag{2.4}$$

The asymptotic covariance matrix, $\Sigma(\theta_0)$, of the ML estimator $\widehat{\theta}_N$ when the true parameter is θ_0 is

$$\Sigma(\theta_0) = D(\theta_0)^{-1} V(\theta_0) D(\theta_0)^{-1}, \text{ where}$$

$$
\begin{aligned}
V(\theta) &= \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E_\theta g(\widetilde{W}_i,\theta)g(\widetilde{W}_i,\theta)' \text{ and} \\
D(\theta) &= \lim_{N\to\infty} N^{-1}\sum_{i=1}^{N} E_\theta \frac{\partial}{\partial\theta'} g(\widetilde{W}_i,\theta).
\end{aligned} \tag{2.5}
$$

A consistent variance matrix estimator Σ_N for $\widehat{\theta}_N$ can be defined in several ways because $D(\theta_0)$ and $V(\theta_0)$ are square matrices and the information matrix equality implies that $D(\theta_0)$ and $V(\theta_0)$ are equal. In particular, one can use

$$
\begin{aligned}
\Sigma_N &= \Sigma_N(\widehat{\theta}_N) \text{ for} \\
\Sigma_N(\theta) &= D_N^{-1}(\theta)V_N(\theta)D_N^{-1}(\theta),\ \Sigma_N(\theta) = D_N^{-1}(\theta),\ \text{or } \Sigma_N(\theta) = V_N^{-1}(\theta),\ \text{where} \\
V_N(\theta) &= N^{-1}\sum_{i=1}^{N} g(\widetilde{W}_i,\theta)g(\widetilde{W}_i,\theta)',\ \text{and} \\
D_N(\theta) &= N^{-1}\sum_{i=1}^{N} \frac{\partial}{\partial\theta'} g(\widetilde{W}_i,\theta).
\end{aligned} \tag{2.6}
$$

Let θ_r, $\theta_{0,r}$, and $\widehat{\theta}_{N,r}$ denote the r–th elements of θ, θ_0, and $\widehat{\theta}_N$ respectively. Let $(\Sigma_N)_{rr}$ denote the (r,r)–th element of Σ_N. The t statistic for testing the null hypothesis $H_0 : \theta_r = \theta_{0,r}$ is

$$
T_N(\theta_{0,r}) = N^{1/2}(\widehat{\theta}_{N,r} - \theta_{0,r})/(\Sigma_N)_{rr}^{1/2}. \tag{2.7}
$$

Suppose $\beta \in R^{L_\beta}$ is a sub-vector of θ, say, $\theta = (\beta', \delta')'$. The Wald statistic for testing $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$ is

$$
\begin{aligned}
\mathcal{W}_N(\beta_0) &= H_N(\widehat{\theta}_N,\beta_0)'H_N(\widehat{\theta}_N,\beta_0),\ \text{where} \\
H_N(\theta,\beta_0) &= ([I_{L_\beta}\vdots 0]\Sigma_N(\theta)[I_{L_\beta}\vdots 0]')^{-1/2}N^{1/2}(\beta-\beta_0).
\end{aligned} \tag{2.8}
$$

3 Parametric Bootstrap

The parametric bootstrap sample $\{W_i^* : i = 1,...,n\}$ is defined as follows. The bootstrap regressors are the same fixed regressors as in the original sample and the bootstrap dependent variables are generated recursively for $i = 1,...,n$ using the parametric density evaluated at the unrestricted ML estimator $\widehat{\theta}_N$. That is, one takes $W_i^* = (Y_i^{*\prime}, X_i')'$, where Y_i^* has density $d(\cdot|X_i, W_{i-1}^*, W_{i-2}^*, ..., W_{i-\kappa_i}^*; \widehat{\theta}_N)$ for $i = 1,...,n$, where $\kappa_i = \min\{\kappa, i+1\}$. The bootstrap observations $\widetilde{W}_i^*$ are defined to be $\widetilde{W}_i^* = (W_i^{*\prime}, ..., W_{i+\kappa}^{*\prime})'$ for $i = 1,...,N$. Under Assumption 1, the conditional distribution of the bootstrap sample given $\widehat{\theta}_N$ is the same as the distribution of the original sample except that the true parameter is $\widehat{\theta}_N$ rather than θ_0.

The bootstrap estimator θ_N^* is defined exactly as the original estimator $\widehat{\theta}_N$ is defined, but with the original sample $\{\widetilde{W}_i : i = 1,...,N\}$ replaced by the bootstrap sample $\{\widetilde{W}_i^* : i = 1, ..., N\}$. That is, θ_N^* solves

$$\min_{\theta \in \Theta} \rho_N^*(\theta), \text{ where } \rho_N^*(\theta) = N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i^*, \theta). \tag{3.1}$$

The bootstrap covariance matrix estimator, Σ_N^*, is defined to be $\Sigma_N^*(\theta_N^*)$ where $\Sigma_N^*(\theta)$ has the same definition as $\Sigma_N(\theta)$ (see (2.6)), but with the bootstrap sample in place of the original sample. (For example, $V_N^*(\theta)$ equals $V_N(\theta)$ with $\widetilde{W}_i$ replaced by $\widetilde{W}_i^*$.)

The bootstrap t and Wald statistics need to be defined such that their distributions mimic the null non-bootstrap distribution even when the sample is generated by a parameter in the alternative hypothesis. This is done by centering the statistics at $\widehat{\theta}_{N,r}$ and $\widehat{\beta}_N$, respectively, rather than at the values specified under the null hypotheses. We define

$$\begin{aligned} T_N^*(\widehat{\theta}_{N,r}) &= N^{1/2}((\theta_N^*)_r - \widehat{\theta}_{N,r})/(\Sigma_N^*)_{rr}^{1/2} \text{ and} \\ \mathcal{W}_N^*(\widehat{\beta}_N) &= H_N^*(\theta_N^*, \widehat{\beta}_N)' H_N^*(\theta_N^*, \widehat{\beta}_N), \text{ where} \\ H_N^*(\theta, \widehat{\beta}_N) &= \left([I_{L_\beta} \vdots 0] \Sigma_N^*(\theta) [I_{L_\beta} \vdots 0]' \right)^{-1/2} N^{1/2}(\beta - \widehat{\beta}_N), \end{aligned} \tag{3.2}$$

$(\theta_N^*)_r$ denotes the r-th element of θ_N^*,[3] and $(\Sigma_N^*)_{rr}$ denotes the (r, r)-th element of Σ_N^*.

Let $z_{|T|,\alpha}^*$, $z_{T,\alpha}^*$, and $z_{\mathcal{W},\alpha}^*$ denote the $1 - \alpha$ quantiles of $|T_N^*(\widehat{\theta}_{N,r})|$, $T_N^*(\widehat{\theta}_{N,r})$, and $\mathcal{W}_N^*(\widehat{\beta}_N)$ respectively. To be precise, we define $z_{|T|,\alpha}^*$ to be a value that minimizes $|P^*(|T_N^*(\widehat{\theta}_{N,r})| \leq z) - (1 - \alpha)|$ over $z \in R$. (This definition allows for discreteness in the distribution of $|T_N^*(\widehat{\theta}_{N,r})|$.) The precise definitions of $z_{T,\alpha}^*$ and $z_{\mathcal{W},\alpha}^*$ are analogous.

The symmetric two-sided bootstrap CI for the r-th element of θ_0, $\theta_{0,r}$, of confidence level $100(1 - \alpha)\%$ is

$$CI_{SYM} = [\widehat{\theta}_{N,r} - z_{|T|,\alpha}^* (\Sigma_N)_{rr}^{1/2} / N^{1/2}, \ \widehat{\theta}_{N,r} + z_{|T|,\alpha}^* (\Sigma_N)_{rr}^{1/2} / N^{1/2}]. \tag{3.3}$$

The equal-tailed two-sided bootstrap CI for $\theta_{0,r}$ of confidence level $100(1 - \alpha)\%$ is

$$CI_{ET} = [\widehat{\theta}_{N,r} - z_{T,\alpha/2}^* (\Sigma_N)_{rr}^{1/2} / N^{1/2}, \ \widehat{\theta}_{N,r} + z_{T,1-\alpha/2}^* (\Sigma_N)_{rr}^{1/2} / N^{1/2}]. \tag{3.4}$$

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

$$CI_{UP} = [\widehat{\theta}_{N,r} - z_{T,\alpha}^* (\Sigma_N)_{rr}^{1/2} / N^{1/2}, \ \infty). \tag{3.5}$$

The bootstrap confidence region for β_0 of confidence level $100(1 - \alpha)\%$ is

$$CR = \{\beta \in R^{L_\beta} : N(\widehat{\beta}_N - \beta)'([I_{L_\beta} \vdots 0] \Sigma_N [I_{L_\beta} \vdots 0]')^{-1} (\widehat{\beta}_N - \beta) \leq z_{\mathcal{W},\alpha}^*\}. \tag{3.6}$$

Correspondingly, the symmetric two-sided bootstrap t test of $H_0 : \theta_r = \theta_{0,r}$ versus $H_1 : \theta_r \neq \theta_{0,r}$ of significance level α rejects H_0 if $|T_N(\theta_{0,r})| > z_{|T|,\alpha}^*$. The equal-tailed two-sided bootstrap t test of significance level α for the same hypotheses rejects H_0 if $T_N(\theta_{0,r}) < z^*_{T,1-\alpha/2}$ or $T_N(\theta_{0,r}) > z^*_{T,\alpha/2}$. The one-sided bootstrap t test of $H_0 : \theta_r \leq \theta_{0,r}$ versus $H_1 : \theta_r > \theta_{0,r}$ of significance level α rejects H_0 if $T_N(\theta_{0,r}) > z^*_{T,\alpha}$.

To carry out tests of the above sort, an alternative parametric bootstrap procedure can be used that employs the restricted ML estimator of θ. Results of Davidson and MacKinnon (1999b) indicate that the error in test rejection probability may be smaller using such a procedure than using a bootstrap based on the unrestricted ML estimator. For this reason, the results of this paper are most useful for CIs and CRs rather than for tests.

4 Assumptions

In this section, we state assumptions that are used in conjunction with Assumption 1 to obtain the results of the paper.

Let a be a non-negative constant such that $2a$ is an integer. The following assumptions depend on a—the larger is a, the stronger are the assumptions. To obtain higher-order improvements of the parametric bootstrap CIs, we require the assumptions to hold with a equal 1, 3/2, or 2 depending upon the CI.

Let $f(\widetilde{W}_i, \theta) \in R^{L_f}$ denote the vector containing the unique components of $g(\widetilde{W}_i, \theta)$ and $g(\widetilde{W}_i, \theta)g(\widetilde{W}_i, \theta)'$ and their partial derivatives with respect to θ through order $d = \max\{2a+2, 3\}$. Let $(\partial^j/\partial\theta^j)g(\widetilde{W}_i, \theta)$ denote the vector of partial derivatives with respect to θ of order j of $g(\widetilde{W}_i, \theta)$. Let $\lambda_{\min}(A)$ denote the smallest eigenvalue of a matrix A. Let $d(\theta, B)$ denote the usual distance between a point θ and a set B (i.e., $d(\theta, B) = \inf\{||\theta - \theta_1|| : \theta_1 \in B\}$).

We establish asymptotic refinements that hold uniformly for the true parameter lying in a subset Θ_0 of Θ. For some $\delta > 0$, let $\Theta_1 = \{\theta \in \Theta : d(\theta, \Theta_0) < \delta/2\}$ be a slightly larger set than Θ_0. To obtain the asymptotic refinements, we need to establish Edgeworth expansions that hold uniformly for the true parameter lying in Θ_1. The reason is that the parametric bootstrap uses $\widehat{\theta}_N$ as the true parameter and Θ_1 contains $\widehat{\theta}_N$ with probability that goes to one (at a sufficiently fast rate) when the true parameter is in Θ_0. In turn, to establish the Edgeworth expansions for all true parameters θ_0 in Θ_1, we need some assumptions to hold uniformly over the slightly larger set $\Theta_2 = \{\theta \in \Theta : d(\theta, \Theta_0) < \delta\}$.

We use the following assumptions.

Assumption 2. (a) Θ is compact and Θ_1 is an open set. (b) $\widehat{\theta}_N$ minimizes $N^{-1}\sum_{i=1}^N \rho(\widetilde{W}_i, \theta)$ over $\theta \in \Theta$. (c) $\rho(\theta, \theta_0) = \lim_{N\to\infty} N^{-1}\sum_{i=1}^N E_{\theta_0}\rho(\widetilde{W}_i, \theta)$ exists and satisfies $\lim_{N\to\infty}\sup_{\theta\in\Theta,\theta_0\in\Theta_1} |N^{-1}\sum_{i=1}^N E_{\theta_0}\rho(\widetilde{W}_i, \theta) - \rho(\theta, \theta_0)| = 0$. (d) For all $\theta_0 \in \Theta_1$, $\rho(\theta, \theta_0)$ is uniquely minimized over $\theta \in \Theta$ by $\theta = \theta_0$. Furthermore, given any $\varepsilon > 0$, these exists $\eta > 0$ such that $||\theta - \theta_0|| > \varepsilon$ implies that $\rho(\theta, \theta_0) - \rho(\theta_0, \theta_0) > \eta$ for all $\theta \in \Theta$ and $\theta_0 \in \Theta_1$. (e) $\sup_{\theta_0\in\Theta_1, i\geq 1} E_{\theta_0}\sup_{\theta\in\Theta} ||g(\widetilde{W}_i, \theta)||^{q_0} < \infty$ and $\sup_{\theta_0\in\Theta_1, i\geq 1} E_{\theta_0}|\rho(\widetilde{W}_i, \theta)|^{q_0} < \infty$ for all $\theta \in \Theta$ for $q_0 = \max\{2a+1, 2\}$.

Assumption 3. (a) $g(\widetilde{w},\theta)$ is $d=\max\{2a+2,3\}$ times partially differentiable with respect to θ on Θ_2 for all $\widetilde{w}$ in the support of $\widetilde{W}_i$ for all $i\geq 1$. (b) $\sup_{\theta_0\in\Theta_1,i\geq 1} E_{\theta_0}||f(\widetilde{W}_i,\theta_0)||^{q_1}<\infty$ for some $q_1>2a+2$. (c) $V(\theta_0)$ and $D(\theta_0)$ satisfy: $\inf_{\theta_0\in\Theta_1}\lambda_{\min}(V(\theta_0))>0, \inf_{\theta_0\in\Theta_1}\lambda_{\min}(D(\theta_0))>0, \lim_{N\to\infty}\sup_{\theta_0\in\Theta_1}|E_{\theta_0}V_N(\theta_0) -V(\theta_0)|=0$ and $\lim_{N\to\infty}\sup_{\theta_0\in\Theta_1}|E_{\theta_0}D_N(\theta_0)-D(\theta_0)|=0$. (d) There is a function $C_f(\widetilde{W}_i)$ such that $||f(\widetilde{W}_i,\theta)-f(\widetilde{W}_i,\theta_0)||\leq C_f(\widetilde{W}_i)||\theta-\theta_0||$ for all $\theta\in\Theta_2$ and $\theta_0\in\Theta_1$ such that $||\theta-\theta_0||<\delta$ and all $i\geq 1$ and $\sup_{\theta_0\in\Theta_1,i\geq 1}E_{\theta_0}C_f^{q_1}(\widetilde{W}_i)<\infty$ for some $q_1>2a+2$.

Assumption 2 imposes some fairly standard conditions used to establish consistency of the ML estimator, as well as some moment conditions. Assumption 3 imposes smoothness and moment conditions on the parametric densities and their derivatives, as well as full rank conditions on the information matrix.

The next assumption comes from Lahiri (1993), which extends results of Götze and Hipp (1983). The assumption guarantees that an Edgeworth expansion holds for $N^{-1/2}\sum_{i=1}^N(f(\widetilde{W}_i,\theta_0)-E_{\theta_0}f(\widetilde{W}_i,\theta_0))$ with remainder $o(N^{-a})$ uniformly over $\theta_0\in\Theta_1$, given the moment condition in Assumption 3(b). The assumption is rather complicated and is not easy to verify in general. Nevertheless, Götze and Hipp (1983, 1994) provide a number of examples in which this condition is verified. For a fixed value θ_0, the assumption is weaker than the corresponding assumptions employed in Hall and Horowitz (1996) and Andrews (1999), which are based on sufficient conditions for the assumption given below.

The following assumption can be replaced by any set of sufficient conditions for an Edgeworth expansion for $N^{-1/2}\sum_{i=1}^N(f(\widetilde{W}_i,\theta_0)-E_{\theta_0}f(\widetilde{W}_i,\theta_0))$ when the true parameter is θ_0 whose remainder is $o(N^{-a})$ uniformly over $\theta_0\in\Theta_1$. For example, there are several Edgeworth expansions in the literature designed specifically for Markov processes. These include Malinovskii (1987, Thm. 1) and Jensen (1989, Thm. 2).[4]

Let $(\Omega,\mathcal{A},P_{\theta_0})$ for $\theta_0\in\Theta$ be the probability space on which the random vectors $\{W_i:i\geq 1\}$ are defined. Let $\mathcal{D}_0,\mathcal{D}_{\pm 1},\mathcal{D}_{\pm 2},...$ be a sequence of sub-σ-fields of $\mathcal{A}$. Let $\mathcal{D}_p^q$ denote the σ-field generated by $\mathcal{D}_j$ for $p\leq j\leq q$.

Assumption 4. (a) There exists a constant $d_1>0$ such that for all $m,i=1,2,...$ with $m>d_1^{-1}$ there exists $\mathcal{D}_{i-m}^{i+m}$-measurable random vectors $Z_{i,m}(\theta_0)$ for which $E_{\theta_0}||f(\widetilde{W}_i,\theta_0)-Z_{i,m}(\theta_0)||<d_1^{-1}\exp(-d_1m)$ for all $\theta_0\in\Theta_1$. (b) There exists a constant $d_2>0$ such that for all $m,i=1,2,...,$ $A\in\mathcal{D}_{-\infty}^i$, and $B\in\mathcal{D}_{i+m}^{\infty}$, $|P_{\theta_0}(A\cap B)-P_{\theta_0}(A)P_{\theta_0}(B)|\leq d_2^{-1}\exp(-dm)$ for all $\theta_0\in\Theta_1$. (c) There exists a constant $d_3>0$ such that for all $m,i=1,2,...$ with $d_3^{-1}<m<i$ and all $t\in R^{L_f}$ with $||t||\geq d$, $E_{\theta_0}|E_{\theta_0}(\exp(\sqrt{-1}t'(\sum_{j=i-m}^{i+m}f(\widetilde{W}_j,\theta_0)))|\mathcal{D}_j:j\neq i)|\leq\exp(-d_3)$ for all $\theta_0\in\Theta_1$. (d) There exists a constant $d_4>0$ such that for all $m,i,p=1,2,...$ and $A\in\mathcal{D}_{i-p}^{i+p}$, $E_{\theta_0}|P_{\theta_0}(A|\mathcal{D}_j:j\neq i)-P_{\theta_0}(A|\mathcal{D}_j:0<|i-j|\leq i+p)|\leq d_4^{-1}\exp(-d_4m)$ for all $\theta_0\in\Theta_1$. (e) There exists matrices $\Omega(\theta_0)\in R^{L_f\times L_f}$ for $\theta_0\in\Theta_1$ such that $\lim_{N\to\infty}\sup_{\theta_0\in\Theta_1}||\mathrm{Var}_{\theta_0}(N^{-1/2}\sum_{i=1}^N f(\widetilde{W}_i,\theta_0))-\Omega(\theta_0)||=0$ and $\Omega(\theta_0)$ has smallest eigenvalue bounded away from 0 over $\theta_0\in\Theta_1$. (f) There exists a constant $d_5>0$ such that for all $i>d_5^{-1}$ and $m>d_5^{-1}$ $\inf\{t'\mathrm{Var}_{\theta_0}(\sum_{j=i}^{i+m}f(\widetilde{W}_j,\theta_0))t:||t||=1,\theta_0\in$

$\Theta_1\} > d_5 m$.

Assumption 4 is a conditional Cramér condition. In the case of an iid sequence of random variables, Assumption 4 reduces to the standard Cramér condition.

5 Higher-order Improvements

One of the main results of this paper is the following Theorem.

Theorem 1 *Suppose Assumptions* 1–4 *hold with* a *in Assumptions 2 and 3 as specified below. Then,*
(a) $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{SYM}) - (1-\alpha)| = O(N^{-2})$ *for* $a = 2$,
(b) $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{ET}) - (1-\alpha)| = o(N^{-1}\ln(N))$ *for* $a = 1$,
(c) $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{UP}) - (1-\alpha)| = o(N^{-1}\ln(N))$ *for* $a = 1$, *and*
(d) $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CR) - (1-\alpha)| = o(N^{-3/2}\ln(N))$ *for* $a = 3/2$.

Comments. 1. The errors in coverage probability of standard delta method CIs and CRs based on asymptotic normal and chi-square approximations are $O(N^{-1})$, $O(N^{-1/2})$, $O(N^{-1/2})$, and $O(N^{-1})$ for symmetric t CIs, equal-tailed t CIs, one-sided t CIs, and elliptical CRs respectively. Hence, the Theorem shows that parametric bootstrap CIs and CRs reduce the coverage errors of standard CIs and CRs by the multiplicative factors $O(N^{-1})$, $o(N^{-1/2}\ln(N))$, $o(N^{-1/2}\ln(N))$, and $o(N^{-1/2}\ln(N))$ respectively. These improvements are almost the same as the improvements that have been established for parametric and non-parametric bootstrap CIs or CRs for a population mean (based on the sample mean) in *iid* scenarios, which are $O(N^{-1})$, $O(N^{-1/2})$, $O(N^{-1/2})$, and $O(N^{-1/2})$, respectively, e.g., see Hall (1988, 1992). Hence, in contrast to the block bootstrap (e.g., see the higher-order improvement results in Andrews (1999)), the parametric bootstrap for time series observations performs essentially as well asymptotically as for independent observations.

2. The result of Theorem 1(a) is sharp and the results of Theorem 1(b) and (c) are very nearly sharp. (Based on results available for population means in iid scenarios, sharp results would be errors of magnitude $O(N^{-1})$ in parts (b) and (c).) But, the result of part (d) for the CR probably is not sharp or nearly sharp. One may be able to obtain an error in part (d) of $O(N^{-2})$ via an argument somewhat similar to that of Hall (1988) for symmetric t CIs. This has not been done in the literature, however, even for the case of a CR for a vector of population means in an iid scenario.

3. The conditions on d, q_0, and q_1 in Assumptions 2 and 3 are as follows. For $a = 1$, the Assumptions require $d \geq 4$, $q_0 \geq 3$, and $q_1 > 4$. For $a = 3/2$, the Assumptions require $d \geq 5$, $q_0 \geq 5$, and $q_1 > 6$. For $a = 2$, the Assumptions require $d \geq 6$, $q_0 \geq 5$, and $q_1 > 6$.

6 k-Step Parametric Bootstrap

In this section, we define the k-step bootstrap estimator, t statistic, and Wald statistic and corresponding CIs and CRs. Then, we establish bounds on the coverage probability errors of these CIs and CRs. Provided k is taken large enough, the bounds are of the same magnitude as those obtained for the standard parametric bootstrap.

The k-step bootstrap estimator is denoted $\theta^*_{N,k}$. The starting value for the k-step estimator is $\widehat{\theta}_N$, the estimator based on the original sample. We define recursively

$$\theta^*_{N,j} = \theta^*_{N,j-1} - (Q^*_{N,j-1})^{-1} N^{-1} \sum_{i=1}^{N} g(\widetilde{W}^*_i, \theta^*_{N,j-1}) \text{ for } 1 \leq j \leq k, \qquad (6.1)$$

where $\theta^*_{N,0} = \widehat{\theta}_N$.

The $L_\theta \times L_\theta$ random matrix $Q^*_{N,j-1}$ depends on $\theta^*_{N,j-1}$. It determines whether the k-step bootstrap estimator is a NR, default NR, line-search NR, GN, or some other k-step bootstrap estimator. The NR, default NR, and line-search NR choices of $Q^*_{N,j-1}$ yield k-step bootstrap estimators that have the same higher-order asymptotic behavior. The results below show that they require fewer steps, k, to approximate the ML bootstrap estimator θ^*_N to a specified accuracy than does the GN k-step estimator. The NR choice of $Q^*_{N,j-1}$ is

$$\begin{aligned} Q^{*,NR}_{N,j-1} &= D^*_N(\theta^*_{N,j-1}), \text{ where} \\ D^*_N(\theta) &= N^{-1} \sum_{i=1}^{N} \frac{\partial}{\partial \theta'} g(\widetilde{W}^*_i, \theta). \end{aligned} \qquad (6.2)$$

The *default* NR choice of $Q^*_{N,j-1}$, denoted $Q^{*,D}_{N,j-1}$, equals $Q^{*,NR}_{N,j-1}$ if $Q^{*,NR}_{N,j-1}$ leads to an estimator $\theta^*_{N,j}$ via (6.1) for which $\rho^*_N(\theta^*_{N,j}) \leq \rho^*_N(\theta^*_{N,j-1})$, but equals some other matrix otherwise. In practice, one wants this other matrix to be such that $\rho^*_N(\theta^*_{N,j}) < \rho^*_N(\theta^*_{N,j-1})$ (but the theoretical results do not require this). For example, one might use the matrix $(1/\varepsilon)I_{L_\theta}$ for some small $\varepsilon > 0$. (See Ortega and Rheinboldt (1970, Theorem 8.2.1) for a result that indicates that such a choice will decrease the criterion function.)

The *line-search* NR choice of $Q^*_{N,j-1}$, denoted $Q^{*,LS}_{N,j-1}$, uses a scaled version of the NR matrix $Q^{*,NR}_{N,j-1}$ that optimizes the step length. Specifically, let A be a finite subset of $(0, 1]$ of step lengths that includes 1. One computes $\theta^*_{N,j} = \theta^{*,\alpha}_{N,j}$ via (6.1) for $Q^*_{N,j-1}$ $= (1/\alpha)Q^{*,NR}_{N,j-1}$ for each $\alpha \in A$. One takes $Q^{*,LS}_{N,j-1}$ to be the matrix $(1/\alpha)Q^{*,NR}_{N,j-1}$ for the value of α that minimizes $\rho^*_N(\theta^{*\alpha}_{N,j})$ over all $\alpha \in A$. (If the minimizing of value of α is not unique, one takes the largest minimizing value of α in A.)

The GN choice of $Q^*_{N,j-1}$, denoted $Q^{*,GN}_{N,j-1}$, uses a matrix that differs from, but is a close approximation to, the NR matrix $Q^{*,NR}_{N,j-1}$. In particular,

$$Q^{*,GN}_{N,j-1} = D^*_{N,j-1}, \qquad (6.3)$$

where $D^*_{N,j-1}$ is determined by some function $\Delta(\cdot,\cdot)$ as follows:

$$\begin{aligned} D^*_{N,j-1} &= N^{-1} \sum_{i=1}^{N} \Delta(\widetilde{W}^*_i, \theta^*_{N,j-1}) \in R^{L_g \times L_\theta} \text{ and} \\ E^*_{\theta_0} \Delta(\widetilde{W}^*_i, \theta_0) &= E^*_{\theta_0} \frac{\partial}{\partial \theta'} g(\widetilde{W}^*_i, \theta_0) \text{ for all } i \geq 1 \text{ and all } \theta_0 \in \Theta_1. \end{aligned} \qquad (6.4)$$

The latter condition is responsible for $D^*_{N,j-1}$ being a close approximation to $D^*_N(\theta^*_{N,j-1}) = Q^{*,NR}_{N,j-1}$.

An example of a GN matrix $Q^{*,GN}_{N,j-1}$ is the sample outer-product estimator of the bootstrap information matrix. By the information matrix equality,

$$E^*_{\theta_0}\frac{\partial}{\partial\theta'}g(\widetilde{W}^*_i,\theta_0) = E^*_{\theta_0}g(\widetilde{W}^*_i,\theta_0)g(\widetilde{W}^*_i,\theta_0)' \text{ for all } i \geq 1 \text{ and all } \theta_0 \in \Theta_1. \quad (6.5)$$

In this case, the NR matrix $Q^{*,NR}_{N,j-1}$ is the sample analogue of the expectation on the left-hand side of (6.5): $Q^{*,NR}_{N,j-1} = N^{-1}\sum_{i=1}^N(\partial/\partial\theta')g(\widetilde{W}^*_i,\theta^*_{N,j-1})$. The GN matrix $Q^{*,GN}_{N,j-1}$ is the sample analogue of the expectation on the right-hand side of (6.5). Thus, $Q^{*,GN}_{N,j-1}$ is as in (6.3) and (6.4) with

$$\Delta(\widetilde{W}^*_i,\theta) = g(\widetilde{W}^*_i,\theta)g(\widetilde{W}^*_i,\theta)'. \quad (6.6)$$

The GN matrix does not require calculation of the second derivative of the log likelihood function.

Alternatively, one can use a GN matrix $Q^*_{N,j-1}$ based on the *expected* bootstrap information matrix:

$$Q^{*,GN2}_{N,j-1} = N^{-1}\sum_{i=1}^N E^*_\theta \frac{\partial}{\partial\theta'}g(\widetilde{W}^*_i,\theta)\bigg|_{\theta=\theta^*_{N,j-1}}. \quad (6.7)$$

In this case, the function $\Delta(\widetilde{W}^*_i,\theta)$ of (6.4) is $E^*_\theta(\partial/\partial\theta')g(\widetilde{W}^*_i,\theta)$, which is nonrandom. The expected information matrix is often used in the statistical literature on one-step and k-step estimators, e.g., see Pfanzagl (1974).

The bootstrap covariance matrix estimator $\Sigma^*_{N,k}$ is defined as Σ_N is defined in (2.6), but with the bootstrap sample in place of the original sample and $\theta^*_{N,k}$ in place of $\widehat{\theta}_N$.

The k-step bootstrap t and Wald statistics, $T^*_{N,k}(\widehat{\theta}_{N,r})$ and $\mathcal{W}^*_{N,k}(\widehat{\beta}_N)$, are defined as in (3.2), but with θ^*_N and Σ^*_N replaced by $\theta^*_{N,k}$ and $\Sigma^*_{N,k}$ respectively. Let $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, and $z^*_{\mathcal{W},k,\alpha}$ denote the $1-\alpha$ quantiles of $|T^*_{N,k}(\widehat{\theta}_{N,r})|$, $T^*_{N,k}(\widehat{\theta}_{N,r})$, and $\mathcal{W}^*_{N,k}(\widehat{\beta}_N)$ respectively (whose precise definitions are analogous to that of $z^*_{|T|,\alpha}$ given above.)

The k-step bootstrap CIs and confidence regions, denoted $CI_{SYM,k}$, $CI_{ET,k}$, $CI_{UP,k}$, and CR_k, are defined as in (3.3)–(3.6), but with $z^*_{|T|,\alpha}$, $z^*_{T,\alpha}$, and $z^*_{\mathcal{W},\alpha}$ replaced by $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, and $z^*_{\mathcal{W},k,\alpha}$ respectively.

The matrices $\{Q^*_{Nj-1} : j = 1, ..., k\}$ are assumed to satisfy the following assumption.

Assumption 5. The matrices $\{Q^*_{Nj-1} : j = 1, ..., k\}$ satisfy: For some sequence of non-negative constants $\{\psi_N : N \geq 1\}$ with $\lim_{N\to\infty}\psi_N = 0$ and for all $\varepsilon > 0$,

$$\sup_{\theta_0\in\Theta_1} P^*_{\theta_0}(||Q^*_{N,j-1} - D^*_N(\theta^*_{N,j-1})|| > \psi_N) = o(N^{-a}) \text{ for } j = 1, ..., k,$$

where $P^*_{\theta_0}$ denotes the probability when the bootstrap sample is generated using the parameter θ_0 rather than $\widehat{\theta}_N$ and the initial estimator $\theta^*_{N,0}$ is θ_0 rather than $\widehat{\theta}_N$.

We now give sufficient conditions for Assumption 5 for the NR, default NR, line-search NR, and GN choices of $Q^*_{N,j-1}$.

Lemma 1 *Suppose Assumptions* 1–4 *hold for some* $a \geq 0$ *with* $2a$ *an integer. Then, Assumption* 5 *holds with* $\psi_N = 0$ *for all* N *for the NR, default NR, and line-search NR choices of* $Q^*_{N,j-1}$ *for* $j = 1, ..., k$*. In addition, Assumption* 5 *holds with* $\psi_N = N^{-1/2}\ln(N)$ *for the GN choice of* $Q^*_{N,j-1}$ *for* $j = 1, ..., k$ *provided Assumptions* 1 *and* 4 *hold with the elements of* $\Delta(\widetilde{W}_i, \theta)$ *(defined in* (6.4)*) added to* $f(\widetilde{W}_i, \theta)$ *and the function* $\Delta(\cdot, \cdot)$ *satisfies: (i)* $E_{\theta_0}(\Delta(\widetilde{W}_i, \theta_0) - (\partial/\partial\theta')g(\widetilde{W}_i, \theta_0)) = 0$ *for all* $i \geq 1$ *and all* $\theta_0 \in \Theta_1$*, (ii)* $\Delta(\widetilde{W}_i, \theta)$ *is continuously differentiable with respect to* θ *on* Θ_2*, (iii)* $\sup_{\theta_0 \in \Theta_1, i \geq 1} E_{\theta_0}||\Delta(\widetilde{W}_i, \theta_0) - (\partial/\partial\theta')g(\widetilde{W}_i, \theta_0)||^{2a+3} < \infty$*, and (iv)* $\sup_{\theta_0 \in \Theta_1, i \geq 1} E_{\theta_0} \sup_{\theta \in B(\theta_0, \varepsilon)} ||(\partial/\partial\theta_u)(\Delta(\widetilde{W}_i, \theta) - (\partial/\partial\theta')g(\widetilde{W}_i, \theta))||^{q_2} < \infty$ *for all* $u = 1, ..., L_\theta$*, for some* $\varepsilon > 0$*, and for* $q_2 = \max\{2a+1, 2\}$*, where* $B(\theta_0, \varepsilon)$ *denotes an open ball at* θ_0 *of radius* ε*.*

Comment. Conditions (ii)–(iv) of the Lemma hold for the outer-product GN matrix of (6.6) by Assumption 3.

The higher-order asymptotic equivalence of the k-step and standard bootstrap statistics is established in parts (a) and (b) of the following Theorem. Part (b) gives conditions under which the Kolmogorov distances (i.e., the sup norms of the differences between the distribution functions) between $N^{1/2}(\theta^*_{N,k} - \widehat{\theta}_N)$ and $N^{1/2}(\theta^*_N - \widehat{\theta}_N)$, $T^*_{N,k}(\widehat{\theta}_{N,r})$ and $T^*_N(\widehat{\theta}_{N,r})$, and $\mathcal{W}^*_{N,k}(\widehat{\beta}_N)$ and $\mathcal{W}^*_N(\widehat{\beta}_N)$, respectively, are $o(N^{-a})$ for some $a \geq 0$.

In part (a) of the Theorem, the difference between the k-step bootstrap estimator and the standard ML bootstrap estimator is shown to be of greater magnitude than $\mu_{N,k}$ with bootstrap probability $o(N^{-a})$ except on a set with probability $o(N^{-a})$, where

$$\mu_{N,k} = \begin{cases} N^{-2^{k-1}} \ln^{2^k}(N) & \text{for NR, default NR, and line search NR matrices} \\ N^{-(k+1)/2} \ln^{k+1}(N) & \text{for GN matrices.} \end{cases} \tag{6.8}$$

Thus, for the NR procedures, the difference decreases very quickly as k increases and for the GN procedure the difference decreases more slowly as k increases. More generally, for ψ_N as in Assumption 5, $\mu_{N,k}$ is defined by

$$\mu_{N,k} = \max_{j=0,...,k} N^{-2^{k-j-1}} \ln^{2^{k-j}}(N) \psi^j_N. \tag{6.9}$$

The key condition in part (b) of the following Theorem is

$$\mu_{N,k} = o(N^{-(a+1/2)}), \tag{6.10}$$

where $2a$ is a non-negative integer. Given this condition, the Kolmogorov distances between the k-step and bootstrap statistics are $o(N^{-a})$ except on a set with probability $o(N^{-a})$.

If Assumption 5 holds with $\psi_N = 0$, as it does for the NR, default NR, and line-search NR procedures, then (6.10) holds if

$$2^k \geq 2a + 2, \tag{6.11}$$

where $2a$ is an integer. Thus, for $k = 1$, we have $a = 0$; for $k = 2$, we have $a = 1$; for $k = 3$, we have $a = 3$; for $k = 4$, we have $a = 7$; etc.

If Assumption 5 holds with $\psi_N = N^{-1/2}\ln(N)$, as it does for the GN procedure under the conditions in Lemma 1, then (6.10) holds if

$$k \geq 2a + 1, \tag{6.12}$$

where $2a$ is an integer. Thus, for $k = 1$, we have $a = 0$; for $k = 2$, we have $a = 1/2$; for $k = 3$, we have $a = 1$; for $k = 4$, we have $a = 3/2$; etc.

The aforementioned Theorem is as follows:

Theorem 2 *Suppose Assumptions* 1-5 *hold for some* $a \geq 0$ *with* $2a$ *an integer in parts* (a) *and* (b).
(a) *Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(P^*_{\widehat{\theta}_N}(||\theta^*_{N,k} - \theta^*_N|| > \mu_{N,k}) > N^{-a}\varepsilon) = o(N^{-a}),$$

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(P^*_{\widehat{\theta}_N}(|T^*_{N,k}(\widehat{\theta}_{N,r}) - T^*_N(\widehat{\theta}_{N,r})| > N^{1/2}\mu_{N,k}) > N^{-a}\varepsilon) = o(N^{-a}), \text{ and}$$

$$\sup_{\theta_0 \in \Theta_0} P_{\theta}(P^*_{\widehat{\theta}_N}(|\mathcal{W}^*_{N,k}(\widehat{\beta}_N) - \mathcal{W}^*_N(\widehat{\beta}_N)| > N^{1/2}\mu_{N,k}) > N^{-a}\varepsilon) = o(N^{-a}).$$

(b) *Suppose* $\mu_{N,k} = o(N^{-(a+1/2)})$. *Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}\left(\sup_{z \in R^{L_\theta}} \left| P^*_{\widehat{\theta}_N}(N^{1/2}(\theta^*_{N,k} - \widehat{\theta}_N) \leq z) - P^*_{\widehat{\theta}_N}(N^{1/2}(\theta^*_N - \widehat{\theta}_N) \leq z) \right| > N^{-a}\varepsilon\right) = o(N^{-a}),$$

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\sup_{z \in R} \left| P^*_{\widehat{\theta}_N}(T^*_{N,k}(\widehat{\theta}_{N,r}) \leq z) - P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z) \right| > N^{-a}\varepsilon) = o(N^{-a}), \text{ and}$$

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\sup_{z \in R} \left| P^*_{\widehat{\theta}_N}(\mathcal{W}^*_{N,k}(\widehat{\beta}_N) \leq z) - P^*_{\widehat{\theta}_N}(\mathcal{W}^*_N(\widehat{\beta}_N) \leq z) \right| > N^{-a}\varepsilon) = o(N^{-a}).$$

We use the results of Theorem 2 to show that the errors in coverage probability of the k-step bootstrap CIs are the same as those of the standard bootstrap CIs given in Theorem 1. In consequence, one can obtain higher-order improvements using the bootstrap without doing the nonlinear optimization necessary to compute the standard bootstrap ML estimator.

Theorem 3 (a) *Suppose Assumptions* 1-5 *hold with* $a = 2$ *and* $\mu_{N,k} = o(N^{-5/2})$. *Then,* $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{SYM,k}) - (1-\alpha)| = O(N^{-2})$.
(b) *Suppose Assumptions* 1-5 *hold with* $a = 1$ *and* $\mu_{N,k} = o(N^{-3/2})$. *Then,* $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{ET,k}) - (1-\alpha)| = o(N^{-1}\ln(N))$ *and* $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CI_{UP,k}) - (1-\alpha)| = o(N^{-1}\ln(N))$.
(c) *Suppose Assumptions* 1-5 *hold with* $a = 3/2$ *and* $\mu_{N,k} = o(N^{-2})$. *Then,* $\sup_{\theta_0 \in \Theta_0} |P_{\theta_0}(\theta_0 \in CR_k) - (1-\alpha)| = o(N^{-3/2}\ln(N))$.

Comments. 1. For the NR, default NR, and line-search NR procedures, the condition $\mu_{N,k} = o(N^{-5/2})$ in part (a) is satisfied if $k \geq 3$; the condition $\mu_{N,k} = o(N^{-3/2})$ in part (b) is satisfied if $k \geq 2$; and the condition $\mu_{N,k} = o(N^{-5/2})$ in part (c) is satisfied if $k \geq 3$. For the GN procedure, the condition $\mu_{N,k} = o(N^{-5/2})$ in part (a) is satisfied if $k \geq 5$; the condition $\mu_{N,k} = o(N^{-3/2})$ in part (b) is satisfied if $k \geq 3$; and the condition $\mu_{N,k} = o(N^{-5/2})$ in part (c) is satisfied if $k \geq 4$. Hence, the k-step NR bootstrap procedures require fewer steps than the k-step GN bootstrap procedure to achieve the same higher-order improvements as obtained by the standard parametric bootstrap. But, with NR or GN k-step bootstrap procedures, the number of steps does not need to be very large.

7 Monte Carlo Simulations

In this section, we compare the performance of standard delta method CIs, symmetric percentile t CIs, and equal-tailed percentile t CIs using Monte Carlo simulation. We consider a stationary Gaussian AR(2) model because it is a well-known model, the standard delta method is known to perform poorly when the sum of the AR coefficients is near one, and the parameter estimates are available in closed form, which greatly speeds computation.

7.1 Experimental Design

The model we consider is given by

$$\begin{aligned} Y_i &= \mu + \rho_1 Y_{i-1} + \rho_2 Y_{i-2} + \sigma U_i \text{ for } i = 3, ..., n, \\ Y_1 &= \left(\frac{1}{1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2\rho_2/(1-\rho_2)} \right)^{1/2} U_1, \\ Y_2 &= \frac{\rho_1}{1-\rho_2} Y_1 + \left(\frac{1 - \rho_1^2/(1-\rho_2)^2}{1 - \rho_1^2 - \rho_2^2 - 2\rho_1^2\rho_2/(1-\rho_2)} \right)^{1/2} U_2, \text{ and} \\ U_i &= \text{iid } N(0,1) \text{ for } i = 1, ..., n. \end{aligned} \quad (7.1)$$

As defined, this model is a stationary Gaussian AR(2) model. The model can also be defined in augmented Dickey-Fuller form as

$$\begin{aligned} Y_i &= \mu + \alpha Y_{i-1} - \rho_2 \Delta Y_{i-1} + \sigma U_i \text{ for } i = 3, ..., n, \text{ where} \\ \alpha &= \rho_1 + \rho_2, \\ \Delta Y_{i-1} &= Y_{i-1} - Y_{i-2}, \end{aligned} \quad (7.2)$$

and (Y_1, Y_2, U_i) are as in (7.1).

In terms of the notation of Section 3, $\kappa = 2$, $N = n-2$, $W_i = Y_i$ for $i = 1, ..., n$, $\widetilde{W}_i = (Y_{i+2}, Y_{i+1}, Y_i)'$ for $i = 1, ..., N$, and $\theta = (\mu, \rho_1, \rho_2, \sigma^2)'$. The normalized negative log-likelihood of $\{\widetilde{W}_i : 1 \leq i \leq N\}$ (conditional on Y_1 and Y_2) is

$$\rho_N(\theta) = \frac{1}{2}\log(2\pi) + \frac{1}{2}\log(\sigma^2) + \frac{1}{2}\sum_{i=1}^{N}(Y_{i+2} - \mu - \rho_1 Y_{i+1} - \rho_2 Y_i)^2. \tag{7.3}$$

The parameter space for θ is $R^3 \times R^+$. In consequence, the ML estimators of μ, ρ_1, and ρ_2, denoted $\widehat{\mu}$, $\widehat{\rho}_1$, and $\widehat{\rho}_2$, are the least squares estimators from the regression of Y_{i+2} on $1, Y_{i+1}$, and Y_i for $i = 1, ..., N$. The ML estimator, $\widehat{\alpha}$, of α is $\widehat{\rho}_1 + \widehat{\rho}_2$. The ML estimator of σ^2 is

$$\widehat{\sigma}^2 = (1/N)\sum_{i=1}^{N}(Y_{i+2} - \widehat{\mu} - \widehat{\rho}_1 Y_{i+1} - \widehat{\rho}_2 Y_i)^2. \tag{7.4}$$

Researchers are often interested in the persistence of a time series. This can be measured by the impulse response function (IRF). The IRF traces out the effect of an increase in the innovation σU_i by a unit quantity on the values Y_{i+h}, denoted IRF(h), for $h = 0, 1, ...$ and $i \geq 3$. The cumulative impulse response (CIR), defined by $CIR = \sum_{h=0}^{\infty} IRF(h)$, provides a convenient scalar summary measure of the persistence of the time series. In the model of (7.1), the CIR equals $1/(1-\alpha)$. The ML estimator of CIR is $\widehat{CIR} = 1/(1-\widehat{\alpha})$. (See Andrews and Chen (1994) for further discussion of CIR.)

In the simulation experiment, we consider CIs for the CIR, as well as for the parameters α, ρ_1, and ρ_2. Note that the CIR only depends on the parameter α, so α also is a useful measure of persistence. (The spectrum of $\{Y_i : i \geq 1\}$ at zero equals $\sigma^2/(1-\alpha)^2$ and, hence, is another measure of persistence that depends on the regression coefficients only through α.)

The standard delta method CI for CIR with nominal coverage probability $100(1-\tau)\%$ is given by

$$\begin{aligned} CI_{CIR} &= \left[\widehat{CIR} - \frac{\widehat{\sigma}_{CIR} z_{1-\tau/2}}{\sqrt{N}},\ \widehat{CIR} + \frac{\widehat{\sigma}_{CIR} z_{1-\tau/2}}{\sqrt{N}}\right], \text{ where} \\ \widehat{\sigma}^2_{CIR} &= \widehat{\sigma}^2_{\alpha}/(1-\widehat{\alpha})^4, \end{aligned} \tag{7.5}$$

and $\widehat{\sigma}^2_{\alpha}$ equals $\widehat{\sigma}^2$ times the $(2,2)$ element of the inverse of $N^{-1}\sum_{i=1}^{N}(1, Y_{i-1}, \Delta Y_{i-1}) \times (1, Y_{i-1}, \Delta Y_{i-1})'$. The delta method CIs for α, ρ_1, and ρ_2, denoted CI_{α}, CI_{ρ_1}, and CI_{ρ_2}, respectively, are defined analogously with $\widehat{\sigma}_{CIR}$ replaced by $\widehat{\sigma}_{\alpha}$, $\widehat{\sigma}_{\rho_1}$, and $\widehat{\sigma}_{\rho_2}$, where $\widehat{\sigma}^2_{\rho_1}$ and $\widehat{\sigma}^2_{\rho_2}$ equal $\widehat{\sigma}^2$ times the $(2,2)$ and $(3,3)$ elements, respectively, of the inverse of $N^{-1}\sum_{i=1}^{N}(1, Y_{i-1}, Y_{i-2})(1, Y_{i-1}, Y_{i-2})'$.

The symmetric and equal-tailed parametric bootstrap CIs for CIR, α, ρ_1, and ρ_2 are as defined in (3.3) and (3.4) of Section 3.[5]

Because the ML estimators of CIR, α, ρ_1, and ρ_2 are available in closed form, we do not consider k-step bootstrap CIs.

An alternative to the parametric bootstrap that can be applied in the AR(2) model above is the residual-based (RB) bootstrap. The RB bootstrap is the same as the parametric bootstrap except that the distribution of the bootstrap errors is given by the empirical distribution of the residuals from the original sample, rather than by the normal distribution. Symmetric and equal-tailed RB bootstrap CIs for CIR, α, ρ_1, and ρ_2 are as defined just as with the parametric bootstrap but with the bootstrap errors being iid with distribution given by the empirical distribution of the residuals. We compute RB bootstrap CIs and compare them to the parametric bootstrap CIs.

We report coverage probabilities for 95% CIs for each of the three types of CI, i.e., delta method, symmetric bootstrap, and equal-tailed bootstrap, for each of the four parameters, i.e., CIR, α, ρ_1, and ρ_2. In addition, for the CIs for CIR, we report the probabilities that the CIs miss the true value to the left and to the right and the average length of the CIs. We report results for sample size $N = 100$, as well as some results for $N = 50$.

We consider nine different parameter combinations for ρ_1 and ρ_2, which correspond to four different values of α, viz., .9, .5, $-.5$, and $-.9$, see Table I. These parameter combinations have been chosen because they cover a broad spectrum of different performances of the CIs considered. All results reported are invariant to the values of μ and σ^2, so we set $\mu = 0$ and $\sigma^2 = 1$ without loss of generality.

To assess the robustness of the parametric bootstrap CIs to the distribution of the innovation U_i, we also consider the case where U_i has a t distribution with five degrees of freedom, which has fat tails, and when it has a chi-squared distribution with one degree of freedom (shifted to have mean zero), which has considerable skewness.

All results are based on $R = 10,000$ Monte Carlo repetitions and $B = 5199$ bootstrap repetitions. With this number of Monte Carlo repetitions, the standard deviation of the reported coverage probabilities is .0022.

7.2 Simulation Results

Table I reports results for CIs for CIR for all nine (ρ_1, ρ_2) parameter combinations and $N = 100$. Several features of the results are immediately apparent. First, all three types of CIs perform most poorly when $\alpha = .9$. They perform better when $\alpha = .5$ and best when $\alpha = -.5$ or -1.5.

Second, the error that the CIs make in almost all cases is under-coverage, not over-coverage.

Third, both bootstrap CIs perform better than the delta method CIs in terms of coverage probability whenever $\alpha = .9$, .5, or $-.5$ and are comparable when $\alpha = -1.5$. This is consistent with the asymptotic results of Section 5, which show that the error in coverage probability of the bootstrap CIs converges to zero at a faster rate than for the delta method CIs. When $\alpha = .9$ or .5, the bootstrap CIs perform substantially better than the delta method CIs. For example, when $(\rho_1, \rho_2) = (.9, 0)$, the coverage probabilities of nominal 95% delta, symmetric bootstrap, and equal-tailed bootstrap CIs are .71, .88, and .85, respectively. In this case and others in which the delta method performs quite poorly, the bootstrap CIs perform much better. But, they do not eliminate under-coverage.

Fourth, the symmetric bootstrap CIs perform better in terms of coverage probability than the equal-tailed bootstrap CIs in almost all cases. Especially when $\alpha = .9$, the difference is noticeable. This also is consistent with the asymptotic results of Section 5, which show that the error in coverage probability of the symmetric bootstrap CIs converges to zero at a faster rate than for the equal-tailed bootstrap CIs.

Fifth, the center of the delta method and symmetric bootstrap CIs is significantly smaller than the true value in all cases. This is reflected in the fact that the probabilities that these CIs miss to the right is essentially zero in all cases. On the other hand, the equal-tailed bootstrap CIs are fairly well centered around the true parameter values. The probabilities that these CIs miss to the left is roughly the same as the probabilities that they miss to the right, in most cases.

Sixth, the average length of the CIs mirrors their coverage probabilities. The delta method CIs are shorter than the bootstrap CIs in all cases except when $\alpha = -1.5$. In these cases, they are too short, which causes their coverage probabilities to be too low. Similarly, the equal-tailed bootstrap CIs are shorter than the symmetric bootstrap CIs in those cases in which the former exhibit under-coverage, which occurs in all cases except when $\alpha = -1.5$.

Overall, it is clear that both bootstrap CIs out perform the delta method CI. The comparison between the two bootstrap CIs is not as clear cut. The symmetric bootstrap CIs outperform the equal-tailed bootstrap CIs in terms of coverage probability. But, the equal-tailed bootstrap CIs are much better centered. Depending upon how one weights these two characteristics of the CIs, one might prefer one bootstrap CI or the other.

Table II reports coverage probabilities for CIs for α, ρ_1, and ρ_2 for the same cases as in Table I. The results for α are quite similar to those for CIR in a qualitative sense. In particular, the delta method CIs under-cover by more than the bootstrap CIs and the equal-tailed bootstrap CIs under-cover by more than the symmetric bootstrap CIs. The main difference is that all three types of CIs perform much better in terms of the amount of under-coverage. For example, the coverage probabilities for $(\rho_1, \rho_2) = (.9, 0)$ are .91, .93, and .92 for the delta, symmetric bootstrap, and equal-tailed bootstrap CIs, respectively. These probabilities are much closer to .95 than the probabilities listed above for the CIR CIs.

Note that one could construct a CI for CIR by transforming the CI for α, because CIR is a monotone transform of α. (That is, the lower endpoint of such a CI for CIR is given by $1/(1 - LE_\alpha)$, where LE_α is the lower endpoint of the CI for α, and the upper endpoint is defined analogously.) The resulting CI for CIR has the same coverage probability as the CI for α.

The results of Table II for ρ_1 and ρ_2 are better than those for α for all three types of CIs. That is, the magnitudes of under-coverage are smaller. In fact, in a few cases there is a small amount of over-coverage. In the cases where the delta method CIs under-cover, the bootstrap CIs under-cover by a smaller amount or by none at all. Hence, the bootstrap CIs for ρ_1 and ρ_2 provide an improvement over those of the delta method.

Tables I and II do not report results for RB bootstrap CIs because they differ very little from the parametric bootstrap results. In most cases, the differences in coverage probabilities are .001 or less. In a few cases, the differences are .002.

Tables III and IV report coverage probability results for the cases of t-5 errors and χ^2-1 errors respectively. These results show that the Gaussian parametric bootstrap CIs still outperform the delta method CIs even when the errors are not Gaussian. In fact, the most salient feature of the results in Tables III and IV is how similar they are to the results when the errors are Gaussian.

Table III does not report results for RB bootstrap CIs because, as in the normal error case, the results are quite similar to those for the parametric bootstrap. The differences between the two for t-5 errors are slightly larger than for N(0, 1) errors, but are still small in most cases. There are a few cases where the differences are as large as .004, but in most cases the differences are .002 or less. The coverage probabilities of the parametric bootstrap CIs are almost always the same as, or closer to, the nominal value .95 than those of the RB bootstrap CIs. This holds because it is almost always the case that the parametric bootstrap CIs have coverage probabilities that are as high or higher than those of the RB bootstrap CIs and both bootstrap CIs usually exhibit under-coverage. These results indicate that the parametric bootstrap CIs are fairly robust to the existence of fat-tailed t-5 errors.

Table IV lists the coverage probabilities of the RB bootstrap CIs for the case of χ^2-1 errors, which are skewed. The differences in coverage probabilities between the parametric and RB bootstrap CIs are noticeably larger than in the N(0, 1) and t-5 error cases. The differences are as large as .021, but usually are smaller. In almost all cases, the coverage probabilities of the parametric bootstrap CIs exceed those of the RB bootstrap CIs. Thus, the parametric bootstrap CIs are more conservative. In roughly half the cases, the parametric bootstrap coverage probabilities are closer to .95 than the RB bootstrap coverage probabilities. Hence, in an overall sense, the parametric bootstrap performs at least as well as the RB bootstrap in the case of (skewed) χ^2-1 errors (at least for sample size 100).

Table V presents results for the case of sample size $N = 50$ and N(0, 1) errors. Comparing the results to those of Tables I and II for $N = 100$, the results are what one would expect. The magnitudes of under-coverage of the CIs are larger and the average lengths of the CIs are larger when $N = 50$ than when $N = 100$. The comparative performances of the delta, symmetric parametric bootstrap, and equal-tailed parametric bootstrap CIs are quite similar for $N = 50$ to those for $N = 100$. The symmetric parametric bootstrap CIs outperform the delta method CIs in terms of coverage probabilities in all cases. The equal-tailed parametric bootstrap CIs outperform the delta method CIs in terms of coverage probabilities in most cases.

8 Appendix of Proofs

In the first subsection of this Appendix, we state Lemmas 2–9 that are used in the proofs of Theorems 1–3 and Lemma 1. In the second subsection, we prove Theorems 1–3. In the third subsection, we prove Lemmas 1–9.

Throughout the Appendix, a denotes a constant that satisfies $a \geq 0$ and $2a$ is an integer, C denotes a generic constant that may change from one equality or inequality to another, and $B(\theta, \varepsilon)$ denotes an open ball of radius $\varepsilon > 0$ centered at θ.

8.1 Lemmas

Lemma 2 *Suppose* $\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\widehat{\theta}_N \notin B(\theta_0, \delta/2)) = o(N^{-a})$ *(for δ as in the definitions of Θ_1 and Θ_2 given in Section 4) and $\{\lambda_N(\theta) : N \geq 1\}$ is a sequence of (non-random) real functions on Θ_1 that satisfies* $\sup_{\theta \in \Theta_1} |\lambda_N(\theta)| = o(N^{-a})$. *Then, for all $\varepsilon > 0$,*

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(|\lambda_N(\widehat{\theta}_N)| > N^{-a}\varepsilon) = o(N^{-a}).$$

Comments. 1. This is a simple, but key, result that is used to obtain bootstrap results from results that hold for statistics based on the original sample uniformly over $\theta_0 \in \Theta_0$. For example, suppose we take $\lambda_N(\theta) = P_\theta^*(||V_N^*(\theta_N^*) - V(\theta)|| > \varepsilon)$ and we show that $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||V_N(\widehat{\theta}_N) - V(\theta_0)|| > \varepsilon) = o(N^{-a})$ and $\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\widehat{\theta}_N \notin B(\theta_0, \delta/2)) = 1 - o(N^{-a})$. Note that $\lambda_N(\theta) = P_\theta(||V_N(\widehat{\theta}_N) - V(\theta)|| > \varepsilon)$ because the bootstrap distribution of $V_N^*(\theta_N^*)$ when the true parameter is θ is the same as the original sample distribution of $V_N(\widehat{\theta}_N)$ when the true parameter is θ. Hence, we know that $\sup_{\theta \in \Theta_1} |\lambda_N(\theta)| = o(N^{-a})$ and, by Lemma 2, we conclude that $\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(P_{\widehat{\theta}_N}^*(||V_N^*(\theta_N^*) - V(\widehat{\theta}_N)|| > \varepsilon) > N^{-a}\varepsilon) = o(N^{-a})$.

2. The condition of Lemma 2 on $\widehat{\theta}_N$ is an implication Lemma 5 below.

Lemma 3 *Suppose Assumption 1 holds.*
(a) *Let $m(\cdot, \theta_0)$ be a matrix-valued function that satisfies $E_{\theta_0} m(\widetilde{W}_i, \theta_0) = 0$ for all $i \geq 1$ and all $\theta_0 \in \Theta_1$ and* $\sup_{\theta_0 \in \Theta_1, i \geq 1} E_{\theta_0} ||m(\widetilde{W}_i, \theta_0)||^p < \infty$ *for $p > 2a$ and $p \geq 2$. Then, for all $\varepsilon > 0$,*

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||N^{-1} \sum_{i=1}^N m(\widetilde{W}_i, \theta_0)|| > \varepsilon) = o(N^{-a}).$$

(b) *Let $m(\cdot, \theta_0)$ be a matrix-valued function that satisfies* $\sup_{\theta_0 \in \Theta_1, i \geq 1} E_{\theta_0} ||m(\widetilde{W}_i, \theta_0)||^p < \infty$ *for $p > 2a$ and $p \geq 2$. Then, there exists $K < \infty$ such that*

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||N^{-1} \sum_{i=1}^N m(\widetilde{W}_i, \theta_0)|| > K) = o(N^{-a}).$$

(c) *Suppose Assumptions* 3(b) *and* 4 *also hold. Then, for all $\varepsilon > 0$,*

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||N^{-1/2} \sum_{i=1}^N (f(\widetilde{W}_i, \theta_0) - E_{\theta_0} f(\widetilde{W}_i, \theta_0))|| > \ln(N)\varepsilon) = o(N^{-a}).$$

Lemma 4 *Suppose Assumptions* 1–3 *hold. Let* $\overline{\theta}_N$ *denote an estimator that satisfies: For all* $\varepsilon > 0$, $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\overline{\theta}_N - \theta_0|| > \varepsilon) = o(N^{-a})$. *Then, for all* $\varepsilon > 0$ *and some* $K < \infty$,

$$\begin{aligned}
\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||V_N(\overline{\theta}_N) - V(\theta_0)|| > \varepsilon) &= o(N^{-a}), \\
\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||D_N(\overline{\theta}_N) - D(\theta_0)|| > \varepsilon) &= o(N^{-a}), \\
\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\frac{\partial^3}{\partial \theta^3}\rho_N(\overline{\theta}_N)|| > K) &= o(N^{-a}), \text{ and} \\
\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||N^{-1}\sum_{i=1}^{N} g(\widetilde{W}_i, \overline{\theta}_N)|| > \varepsilon) &= o(N^{-a}).
\end{aligned}$$

Lemma 5 *Suppose Assumptions* 1–4 *hold. Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(N^{1/2}||\widehat{\theta}_N - \theta_0|| > \ln(N)\varepsilon\right) = o(N^{-a}).$$

Lemma 6 *Suppose Assumption* 1 *holds. Let* $\{A_N(\theta_0) : N \geq 1\}$ *be a sequence of* $L_A \times 1$ *random vectors with Edgeworth expansions for each* $\theta_0 \in \Theta_1$ *with coefficients of order* $O(1)$ *and remainders of order* $o(N^{-a})$ *both uniformly over* $\theta_0 \in \Theta_1$. *(That is, there exist polynomials* $\{\pi_{N,i}(z, \theta_0) : i = 1, ..., 2a\}$ *in* z *whose coefficients are* $O(1)$ *uniformly over* $\theta_0 \in \Theta_1$ *such that* $\sup_{\theta_0 \in \Theta_1} \sup_{B \in \mathcal{B}_{L_A}} |P_{\theta_0}(A_N(\theta_0) \in B) - \int_B (1 + \sum_{i=1}^{2a} N^{-i/2}\pi_{N,i}(z, \theta_0))\phi_{\Omega_N(\theta_0)}(z)dz| = o(N^{-a})$, *where* $\phi_{\Omega_N(\theta_0)}(z)$ *is the density function of a* $N(0, \Omega_N(\theta_0))$ *random variable,* $\Omega_N(\theta_0)$ *has eigenvalues that are bounded away from zero and infinity as* $N \to \infty$ *uniformly over* $\theta \in \Theta_1$, *and* $\mathcal{B}_{L_A}$ *denotes the class of all convex sets in* R^{L_A}.*) Let* $\{\xi_N(\theta_0) : N \geq 1\}$ *be a sequence of random vectors with* $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\xi_N(\theta_0)|| > \omega_N) = o(N^{-a})$ *for some constants* $\omega_N = o(N^{-a})$, *where* $\xi_N(\theta_0) \in R^{L_A}$. *Then,*

$$\sup_{\theta_0 \in \Theta_1} \sup_{B \in \mathcal{B}_{L_A}} |P_{\theta_0}(A_N(\theta_0) + \xi_N(\theta_0) \in B) - P_{\theta_0}(A_N(\theta_0) \in B)| = o(N^{-a}).$$

Let $S_N(\theta) = N^{-1}\sum_{i=1}^{N} f(\widetilde{W}_i, \theta)$ and $S_N^*(\theta) = N^{-1}\sum_{i=1}^{N} f(\widetilde{W}_i^*, \theta)$.

Lemma 7 *Suppose Assumptions* 1–4 *hold. Let* $\Delta_N(\theta_0)$ *denote* $N^{1/2}(\widehat{\theta}_N - \theta_0)$, $T_N(\theta_{0,r})$, *or* $H_N(\widehat{\theta}_N, \beta_0)$, *where* $\theta_0 = (\beta_0', \delta_0')'$. *Let* L *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* $G(E_{\theta_0}S_N(\theta_0)) = 0$ *for all* N *large and all* $\theta_0 \in \Theta_1$ *and*

$$\sup_{\theta_0 \in \Theta_1} \sup_{B \in \mathcal{B}_L} |P_{\theta_0}(\Delta_N(\theta_0) \in B) - P_{\theta_0}(N^{1/2}G(S_N(\theta_0)) \in B)| = o(N^{-a}).$$

We now define the components of the Edgeworth expansions of $T_N(\theta_{0,r})$ and $\mathcal{W}_N(\beta_0)$, as well as their bootstrap analogues $T_N^*(\widehat{\theta}_{N,r})$ and $\mathcal{W}_N^*(\widehat{\beta}_N)$. Let $\Psi_N(\theta_0) = N^{1/2}(S_N(\theta_0) - E_{\theta_0} S_N(\theta_0))$. Let $\Psi_{N,j}(\theta_0)$ denote the j-th element of $\Psi_N(\theta_0)$. Let $\nu_{N,a}(\theta_0)$ denote a vector of moments of the form $N^{\alpha(m)} E_{\theta_0} \prod_{\mu=1}^m \Psi_{N,j_\mu}(\theta_0)$, where $2 \leq m \leq 2a+2$, $\alpha(m) = 0$ if m is even, and $\alpha(m) = 1/2$ if m is odd. Let $\pi_{Ti}(\delta, \nu_{N,a}(\theta_0))$ be a polynomial in $\delta = \partial/\partial z$ whose coefficients are polynomial functions of the elements of $\nu_{N,a}(\theta_0)$ and for which $\pi_{Ti}(\delta, \nu_{N,a}(\theta_0))\Phi(z)$ is an even function of z when i is odd and is an odd function of z when i is even for $i = 1, ..., 2a$. The Edgeworth expansion of $T_N(\theta_{0,r})$ depends on $\pi_{Ti}(\delta, \nu_{N,a}(\theta_0))$. In contrast, the Edgeworth expansion of $\mathcal{W}_N(\beta_0)$ depends on $\pi_{\mathcal{W}i}(y, \nu_{N,a}(\theta_0))$, where $\pi_{\mathcal{W}i}(y, \nu_{N,a}(\theta_0))$ denotes a polynomial function of y whose coefficients are polynomial functions of the elements of $\nu_{N,a}(\theta_0)$ for $i = 1, ..., [a]$. The Edgeworth expansions of $T_N^*(\widehat{\theta}_{N,r})$ and $\mathcal{W}_N^*(\widehat{\beta}_N)$ depend on $\pi_{Ti}(\delta, \nu_{N,a}(\widehat{\theta}_N))$ and $\pi_{\mathcal{W}i}(y, \nu_{\mathcal{W},N,a}(\widehat{\theta}_N))$ respectively.

Let $\Phi(\cdot)$ denote the distribution function of a standard normal random variable. Let χ^2_λ denote a chi-square random variable with λ degrees of freedom. Let $\theta_{0,r}$ denote the r-th element of θ_0.

Lemma 8 *Suppose Assumptions* 1–4 *hold. Then, for all* $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(N^{1/2} ||\nu_{N,a}(\widehat{\theta}_N) - \nu_{N,a}(\theta_0)|| > \ln(N)\varepsilon) = o(N^{-a}).$$

Lemma 9 *Suppose Assumptions* 1–4 *hold.*
(a) *Then,*

$$\sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(T_N(\theta_{0,r}) \leq z) - [1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{Ti}(\delta, \nu_{N,a}(\theta_0))]\Phi(z)| = o(N^{-a}) \text{ and}$$

$$\sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(\mathcal{W}_N(\beta_0) \leq z) - \int_{-\infty}^{z} d[1 + \sum_{i=1}^{[a]} N^{-i} \pi_{\mathcal{W}i}(y, \nu_{N,a}(\theta_0))] P(\chi^2_{L_H} \leq y)| = o(N^{-a}).$$

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

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\sup_{z \in R} |P^*_{\widehat{\theta}_N}(T_N^*(\widehat{\theta}_{N,r}) \leq z) - [1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{Ti}(\delta, \nu_{N,a}(\widehat{\theta}_N))]\Phi(z)| > N^{-a}\varepsilon\right) = o(N^{-a}) \text{ and}$$

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\sup_{z \in R} |P^*_{\widehat{\theta}_N}(\mathcal{W}_N^*(\widehat{\beta}_N) \leq z) - \int_{-\infty}^{z} d[1 + \sum_{i=1}^{[a]} N^{-i} \pi_{\mathcal{W}i}(y, \nu_{N,a}(\widehat{\theta}_N))] P(\chi^2_{L_H} \leq y)| > N^{-a}\varepsilon\right) = o(N^{-a}).$$

Comments. 1. The terms in the Edgeworth expansions for the Wald statistic only involve integer powers of N^{-1}, not powers $N^{-1/2}$, $N^{-3/2}$, etc. as in the Edgeworth expansions for the t statistic, due to a symmetry property of the expansions.

2. The conditions on q_1 and d in Assumption 3 are not needed in all of the Lemmas above. In particular, Lemmas 4 and 5 only use $q_1 \geq \max\{2a+1, 2\}$ and $d = 3$.

8.2 Proofs of Theorems

8.2.1 Proof of Theorem 1

We establish part (c) first. Note that $P_{\theta_0}(\theta_{0,r} \in CI_{UP}) = P_{\theta_0}(T_N(\theta_{0,r}) \leq z^*_{T,\alpha})$. We show that the latter equals $1 - \alpha + o(N^{-1}\ln(N))$ uniformly over $\theta_0 \in \Theta_0$. By Lemma 9(b), Lemma 8, and Lemma 9(a), respectively, each with $a = 1$, we have: for all $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}\left(\sup_{z \in R} |P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z) - [1 + \sum_{i=1}^{2} N^{-i/2}\pi_{Ti}(\delta, \nu_{N,1}(\widehat{\theta}_N))]\Phi(z)| > N^{-1}\right) = o(N^{-1}),$$

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}\left(\sup_{z \in R} |[\pi_{Ti}(\delta, \nu_{N,1}(\widehat{\theta}_N)) - \pi_{Ti}(\delta, \nu_{N,1}(\theta_0))]\Phi(z)| > N^{-1/2}\ln(N)\varepsilon\right) = o(N^{-1}) \text{ for } i = 1, 2, \text{ and}$$

$$\sup_{\theta_0 \in \Theta_0} \sup_{z \in R} |P_{\theta_0}(T_N(\theta_{0,r}) \leq z) - [1 + \sum_{i=1}^{2} N^{-i/2}\pi_{Ti}(\delta, \nu_{N,1}(\theta_0))]\Phi(z)| = o(N^{-1}). \quad (8.1)$$

The results of (8.1) combine to give

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\sup_{z \in R} |P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z) - P_{\theta_0}(T_N(\theta_{0,r}) \leq z)| > N^{-1}\ln(N)\varepsilon) = o(N^{-1}). \quad (8.2)$$

If $T^*_N(\widehat{\theta}_{N,r})$ is absolutely continuous, then $P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z^*_{T,\alpha}) = 1-\alpha$. Whether or not $T^*_N(\widehat{\theta}_{N,r})$ is absolutely continuous, the Edgeworth expansion of Lemma 9(b) with $a = 1$ implies that

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(|P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z^*_{T,\alpha}) - (1-\alpha)| > N^{-1}\varepsilon) = o(N^{-1}) \quad (8.3)$$

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 $z^*_{T,\alpha}$, $|P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z^*_{T,\alpha}) - (1-\alpha)|$ $\leq |P^*_{\widehat{\theta}_N}(T^*_N(\widehat{\theta}_{N,r}) \leq z^{**}_{T,\alpha}) - (1-\alpha)|$.

Taking $z = z^*_{T,\alpha}$ in (8.2) and combining it with (8.3) gives

$$\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(|1 - \alpha - P_{\theta_0}(T_N(\theta_{0,r}) \leq z^*_{T,\alpha})| > N^{-1}\ln(N)\varepsilon) = o(N^{-1}). \tag{8.4}$$

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

The proof of part (b) is analogous to that for part (c). The proof for part (d) is also analogous to that of part (c), but using the Wald statistic results of Lemmas 8 and 9, rather than the t statistic results, and with these Lemmas applied with $a = 3/2$ rather than $a = 1$. In part (d) the coverage probability error is $o(N^{-3/2}\ln(N))$, rather than $o(N^{-1}\ln(N))$ (which is the error in part (c)), because the first terms in the Edgeworth expansions for the Wald statistic in Lemma 9 are $O(N^{-1})$, whereas those for the t statistic are $O(N^{-1/2})$.

Next, we prove part (a). Note that $P_{\theta_0}(\theta_0 \in CI_{SYM}) = P_{\theta_0}(|T_N(\theta_{0,r})| \leq z^*_{|T|,\alpha})$. We show that the latter is $O(N^{-2})$ uniformly over $\theta_0 \in \Theta_0$.

By Lemma 7 with $a = 2$, it suffices to establish the result with $T_N(\theta_{0,r})$ and $T^*_N(\widehat{\theta}_{N,r})$ replaced by $N^{1/2}G(S_N(\theta_0))$ and $N^{1/2}G(S^*_N(\widehat{\theta}_N))$ respectively. Part (a) now can be established using methods developed for "smooth functions of sample averages," as in Hall (1988, 1992). Define $z_{|G|,\alpha}$ by $P_{\theta_0}(|N^{1/2}G(S_N(\theta_0))| \leq z_{|G|,\alpha}) = 1 - \alpha$ and let $\Delta = z_{|G|,\alpha} - z^*_{|T|,\alpha}$. The idea of the proof is to show that

$$P_{\theta_0}(N^{1/2}G(S_N(\theta_0)) + \Delta \leq z_{|G|,\alpha}) = 1 - \alpha/2 + N^{-3/2}r_1(z_{|G|,\alpha})\phi(z_{|G|,\alpha}) + O(N^{-2})$$

and

$$P_{\theta_0}(N^{1/2}G(S_N(\theta_0)) - \Delta \leq -z_{|G|,\alpha}) = \alpha/2 - N^{-3/2}r_1(-z_{|G|,\alpha})\phi(-z_{|G|,\alpha}) + O(N^{-2}), \tag{8.5}$$

uniformly over $\theta_0 \in \Theta_0$, where $r_1(x)$ is a constant times x and $\phi(\cdot)$ denotes the standard normal density function, as in of Hall (1988). Then,

$$\begin{aligned} P_{\theta_0}(|T_N(\theta_{0,r})| \leq z^*_{|T|,\alpha}) &= P_{\theta_0}(|N^{1/2}G(S_N(\theta_0))| \leq z^*_{|T|,\alpha}) + O(N^{-2}) \\ &= 1 - \alpha + N^{-3/2}r_1(z_{|G|,\alpha})\phi(z_{|G|,\alpha}) \\ &\quad + N^{-3/2}r_1(-z_{|G|,\alpha})\phi(-z_{|G|,\alpha}) + O(N^{-2}) \\ &= 1 - \alpha + O(N^{-2}), \end{aligned} \tag{8.6}$$

uniformly over $\theta_0 \in \Theta_0$, using the fact that $r_1(x)$ is an odd function and $\phi(\cdot)$ is an even function. The results of (8.5) are established by the same argument as used to prove (3.2) of Hall (1988), where his T corresponds to our $N^{1/2}G(S_N(\theta_0))$. (More details of this argument can be found in Hall (1992, Pf. of Thm. 5.3), which considers one-sided confidence intervals, but can be extended to symmetric two-sided confidence intervals.) This argument relies on Edgeworth expansions of $N^{1/2}G(S_N(\theta_0))$ and $N^{1/2}G(S^*_N(\widehat{\theta}_N))$:

$$\sup_{\theta_0 \in \Theta_0} \sup_{z \in R} \Big| P_{\theta_0}(|N^{1/2}G(S_N(\theta_0))| \leq z)$$

$$-[1+N^{-1}\pi_2(\delta,\nu_{N,2}(\theta_0))+N^{-2}\pi_4(\delta,\nu_{N,2}(\theta_0))](\Phi(z)-\Phi(-z))\Big|$$
$$= o(N^{-2}) \text{ and}$$
$$\sup_{\theta_0\in\Theta_0} P_{\theta_0}\left(\sup_{z\in R}\Big|P^*_{\widehat{\theta}_N}(|N^{1/2}G(S^*_N(\widehat{\theta}_N))|\le z)-[1+N^{-1}\pi_2(\delta,\nu_{N,2}(\widehat{\theta}_N))\right.$$
$$\left.+N^{-2}\pi_4(\delta,\nu_{N,2}(\widehat{\theta}_N))](\Phi(z)-\Phi(-z))\Big| > N^{-2}\right)$$
$$= o(N^{-2}), \tag{8.7}$$

which hold by Lemma 9 with $a=2$ and with $T_N(\theta_{0,r})$ and $T^*_N(\widehat{\theta}_{N,r})$ replaced by $N^{1/2}G(S_N(\theta_0))$ and $N^{1/2}G(S^*_N(\widehat{\theta}_N))$, respectively. The former replacements are valid by the proof of Lemma 9. □

8.2.2 Proof of Theorem 2

Define $\widehat{\theta}_{N,k}$, $Q_{N,j-1}$, $T_{N,k}(\theta_{0,r})$, and $\mathcal{W}_{N,k}(\beta_0)$ just as $\theta^*_{N,k}$, $Q^*_{N,j-1}$, $T^*_{N,k}(\widehat{\theta}_{N,r})$, and $\mathcal{W}^*_{N,k}(\widehat{\beta}_N)$ are defined but with the bootstrap sample $\{\widetilde{W}^*_i : i=1,2,...,N\}$ replaced by the original sample $\{\widetilde{W}_i : i=1,2,...,N\}$ and with the initial estimator $\widehat{\theta}_{N,0}$ used to generate $\widehat{\theta}_{N,k}$ given by the true parameter θ_0. To establish part (a) of the Theorem, we apply Lemma 2 three times with

$$\begin{aligned}
\lambda_N(\theta_0) &= P^*_{\theta_0}(||\theta^*_{N,k}-\theta^*_N||>\mu_{N,k}) = P_{\theta_0}(||\widehat{\theta}_{N,k}-\widehat{\theta}_N||>\mu_{N,k}),\\
\lambda_N(\theta_0) &= P^*_{\theta_0}(|T^*_{N,k}(\theta_{0,r})-T^*_N(\theta_{0,r})|>N^{1/2}\mu_{N,k})\\
&= P_{\theta_0}(|T_{N,k}(\theta_{0,r})-T_N(\theta_{0,r})|>N^{1/2}\mu_{N,k}), \text{ and}\\
\lambda_N(\theta_0) &= P^*_{\theta_0}(|\mathcal{W}^*_{N,k}(\beta_0)-\mathcal{W}^*_N(\beta_0)|>N^{1/2}\mu_{N,k})\\
&= P_{\theta_0}(|\mathcal{W}_{N,k}(\beta_0)-\mathcal{W}_N(\beta_0)|>N^{1/2}\mu_{N,k}).
\end{aligned} \tag{8.8}$$

The condition of Lemma 2 on $\widehat{\theta}_N$ is established in Lemma 5. In consequence, to establish part (a) of the Theorem, it suffices to show that

$$\begin{aligned}
\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||\widehat{\theta}_{N,k}-\widehat{\theta}_N||>\mu_{N,k}) &= o(N^{-a}),\\
\sup_{\theta_0\in\Theta_1} P_{\theta_0}(|T_{N,k}(\theta_{0,r})-T_N(\theta_{0,r})|>N^{1/2}\mu_{N,k}) &= o(N^{-a}), \text{ and}\\
\sup_{\theta_0\in\Theta_1} P_{\theta_0}(|\mathcal{W}_{N,k}(\beta_0)-\mathcal{W}_N(\beta_0)|>N^{1/2}\mu_{N,k}) &= o(N^{-a}).
\end{aligned} \tag{8.9}$$

We establish the first result of (8.9) first. A Taylor expansion about $\widehat{\theta}_{N,k-1}$ gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_N)\\
&= \frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,k-1})+\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,k-1})(\widehat{\theta}_N-\widehat{\theta}_{N,k-1})+R_{N,k}\\
&= \frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,k-1})+Q_{N,k-1}(\widehat{\theta}_{N,k}-\widehat{\theta}_{N,k-1})+Q_{N,k-1}(\widehat{\theta}_N-\widehat{\theta}_{N,k})
\end{aligned}$$

$$+(\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,k-1}) - Q_{N,k-1})(\widehat{\theta}_N - \widehat{\theta}_{N,k-1}) + R_{N,k}$$

$$= Q_{N,k-1}(\widehat{\theta}_N - \widehat{\theta}_{N,k}) + (\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,k-1}) - Q_{N,k-1})(\widehat{\theta}_N - \widehat{\theta}_{N,k-1}) + R_{N,k},$$

where

$$R_{N,k} = \left[(\widehat{\theta}_N - \widehat{\theta}_{N,k-1})'\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}\rho_N(\theta^+_{N,k-1,u})(\widehat{\theta}_N - \widehat{\theta}_{N,k-1})/2\right]_{L_\theta}, \qquad (8.10)$$

$[\xi_u]_{L_\theta}$ denotes an L_θ vector whose u-th element is ξ_r, $\theta^+_{N,k-1,u}$ lies between $\widehat{\theta}_N$ and $\widehat{\theta}_{N,k-1}$, the first equality holds except with supremum P_{θ_0}-probability over $\theta_0 \in \Theta_1$ equal to $o(N^{-a})$ by Lemma 5, and the fourth equality holds because $(\partial/\partial\theta)\rho_N(\widehat{\theta}_{N,k-1})$ + $Q_{N,k-1}$ $(\widehat{\theta}_{N,k} - \widehat{\theta}_{N,k-1}) = 0$ by the definition of $\widehat{\theta}_{N,k}$. Rearranging (8.10) yields

$$\begin{aligned}
&||\widehat{\theta}_{N,k} - \widehat{\theta}_N|| \\
&\leq ||(Q_{N,k-1})^{-1}R_{N,k}|| + ||(Q_{N,k-1})^{-1}(\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,k-1}) - Q_{N,k-1})(\widehat{\theta}_{N,k-1} - \widehat{\theta}_N)|| \\
&\leq \zeta_N(||\widehat{\theta}_{N,k-1} - \widehat{\theta}_N||^2 + \psi_N||\widehat{\theta}_{N,k-1} - \widehat{\theta}_N||), \text{ where}
\end{aligned}$$

$$\zeta_N = \max_{j=1,\ldots,k}\{||(Q_{N,j-1})^{-1})|| \cdot \sum_{u=1}^{L_\theta}||\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}\rho_N(\theta^+_{N,j-1,u})/2|| + ||(Q_{N,j-1})^{-1})|| \cdot \widetilde{\psi}_N||\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,j-1}) - Q_{N,j-1}|| + 1\}, \qquad (8.11)$$

$\widetilde{\psi}_N = \psi_N^{-1}$ if $\psi_N > 0$ and $\widetilde{\psi}_N = 0$ if $\psi_N = 0$. Repeated substitution into the right-hand side of the inequality gives an upper bound that is a finite sum of terms with dominant terms of the form:

$$C\zeta_N^\phi||\widehat{\theta}_{N,0} - \widehat{\theta}_N||^{2^{k-j}}\psi_N^j \text{ for } j = 0, ..., k, \qquad (8.12)$$

where ϕ is a positive integer and $\widehat{\theta}_{N,0} = \theta_0$ when the true parameter is θ_0. To see this, consider the solution in terms of x_0 of the equation $x_k = x_{k-1}^2 + \lambda x_{k-1}$. Collect all terms in powers of λ that are multiplied by the smallest number of x_0 terms.

An upper bound on the right–hand side of the inequality in (8.11) is

$$C\zeta_N^\phi \max_{j=0,\ldots,k}(\gamma_N)^{2^{k-j}}N^{-2^{k-j-1}}\ln^{2^{k-j}}(N)\psi_N^j, \text{ where } \gamma_N = N^{1/2}||\widehat{\theta}_{N,0} - \widehat{\theta}_N||\ln^{-1}(N). \qquad (8.13)$$

For all $\varepsilon > 0$, $\sup_{\theta_0\in\Theta_1} P_{\theta_0}(\gamma_N > \varepsilon) = o(N^{-a})$ by Lemma 5 because $\widehat{\theta}_{N,0} = \theta_0$. In addition, by Lemma 4 and Assumptions 3(a) and 5, there exists a finite constant K such that $\sup_{\theta_0\in\Theta_1} P_{\theta_0}(\zeta_N > K) = o(N^{-a})$. Assumption 5 applies here because $P^*_{\theta_0}(||Q^*_{N,j-1} - D^*_N(\theta^*_{N,j-1})|| > \psi_N) = P_{\theta_0}(||Q_{N,j-1} - (\partial^2/\partial\theta\partial\theta')\rho_N(\widehat{\theta}_{N,j-1})|| > \psi_N)$. Combining these results with (8.11) and (8.13) gives:

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}\left(||\widehat{\theta}_{N,k} - \widehat{\theta}_N|| > \max_{j=0,\ldots,k} N^{-2^{k-j-1}}\ln^{2^{k-j}}(N)\psi_N^j\right)$$

$$\le \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(C\zeta_N^{\phi}\lambda_N > 1)$$
$$= \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(CK^{\phi}\varepsilon > 1) + o(N^{-a})$$
$$= o(N^{-a}), \tag{8.14}$$

where the last equality holds for $\varepsilon > 0$ sufficiently small. Hence, the first result of part (a) of the Theorem holds.

Next, we establish the second result of part (a) of the Theorem. Let Σ_r denote $(\Sigma_N)_{rr}$. Let $\Sigma_{k,r}$ denote Σ_r with $\widehat{\theta}_N$ replaced by $\widehat{\theta}_{N,k}$ in all parts of its definition in (2.6). We use the following:

$$|T_{N,k}(\theta_{0,r}) - T_N(\theta_{0,r})| \le N^{1/2}||\widehat{\theta}_{N,k} - \widehat{\theta}_N||/\Sigma_{k,r}^{1/2}$$
$$+N^{1/2}||\widehat{\theta}_N - \theta_0|| \cdot |\Sigma_{k,r}^{1/2} - \Sigma_r^{1/2}|/(\Sigma_{k,r}\Sigma_r)^{1/2}. \tag{8.15}$$

By (8.13), the second result of part (a) is implied by the first result plus the following: There exists a $K < \infty$ and a $\delta > 0$ such that

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(|\Sigma_{k,r}^{1/2} - \Sigma_r^{1/2}| > \mu_{N,k}) = o(N^{-a}), \tag{8.16}$$
$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\widehat{\theta}_N - \theta_0|| > K) = o(N^{-a}), \tag{8.17}$$
$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\Sigma_{k,r} < \delta) = o(N^{-a}), \text{ and} \tag{8.18}$$
$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\Sigma_r < \delta) = o(N^{-a}). \tag{8.19}$$

Equation (8.17) holds by Lemma 5. Equations (8.18) and (8.19) hold by Lemma 5, the first result (8.9), and the first and/or second results of Lemma 4.

Equation (8.16) is implied by (8.18), (8.19), and

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(|\Sigma_{k,r} - \Sigma_r| > \mu_{N,k}) = o(N^{-a}) \tag{8.20}$$

by a mean value expansion. Equation (8.20) is implied by

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||D_N(\widehat{\theta}_{N,k}) - D_N(\widehat{\theta}_N)|| > \mu_{n,k}) = o(N^{-a}) \text{ and/or}$$
$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||V_N(\widehat{\theta}_{N,k}) - V_N(\widehat{\theta}_N)|| > \mu_{n,k}) = o(N^{-a}). \tag{8.21}$$

These results hold by mean value expansions, Lemma 3(b) with $m(\widetilde{W}_i, \theta_0) = \sup_{\theta \in \Theta_2}$ $||(\partial^2/\partial\theta_u\partial\theta')g(\widetilde{W}_i, \theta)||$ and $m(\widetilde{W}_i, \theta_0) = \sup_{\theta \in \Theta_2} ||(\partial/\partial\theta_u)(g(\widetilde{W}_i, \theta)g(\widetilde{W}_i, \theta)')||$ for $u = 1, ..., L_\theta$, Lemma 5, the first result of (8.9), and Assumption 3.

We now prove the third result of part (a). Let $H_N = H_N(\widehat{\theta}_N)$ and $H_{N,k} = H_N(\widehat{\theta}_{N,k})$. We have

$$|\mathcal{W}_{N,k}(\beta_0) - \mathcal{W}_N(\beta_0)| = |(H_{N,k} - H_N)'H_{N,k} + H_N'(H_{N,k} - H_N)|$$
$$\le ||H_{N,k} - H_N||(||H_{N,k}|| + ||H_N||). \tag{8.22}$$

Hence, it suffices to show that

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||H_{N,k} - H_N|| > N^{1/2}\mu_{N,k}) = o(N^{-a}) \text{ and}$$
$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||H_N|| > M) = o(N^{-a}) \text{ for some } M < \infty. \quad (8.23)$$

The second result of (8.23) holds by Lemma 9(a) because $||H_N||^2 = \mathcal{W}_N(\beta_0)$. The first result of (8.23) is implied by the matrix version of (8.20) and the first result of (8.9).

To establish part (b) of the Theorem, we apply Lemma 2 three times with

$$\begin{aligned} \lambda_N(\theta_0) &= \sup_{z \in R^{L_\theta}} \left| P^*_{\theta_0}(N^{1/2}(\theta^*_{N,k} - \theta_0) \leq z) - P^*_{\theta_0}(N^{1/2}(\theta^*_N - \theta_0) \leq z) \right| \\ &= \sup_{z \in R^{L_\theta}} \left| P_{\theta_0}(N^{1/2}(\widehat{\theta}_{N,k} - \theta_0) \leq z) - P_{\theta_0}(N^{1/2}(\widehat{\theta}_N - \theta_0) \leq z) \right|, \quad (8.24) \end{aligned}$$

etc. In consequence, it suffices to show that

$$\sup_{\theta_0 \in \Theta_1} \sup_{z \in R^{L_\theta}} \left| P_{\theta_0}(N^{1/2}(\widehat{\theta}_{N,k} - \theta_0) \leq z) - P_{\theta_0}(N^{1/2}(\widehat{\theta}_N - \theta_0) \leq z) \right| = o(N^{-a}),$$
$$\sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(T_{N,k}(\theta_{0,r}) \leq z) - P_{\theta_0}(T_N(\theta_{0,r}) \leq z)| = o(N^{-a}), \text{ and}$$
$$\sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(\mathcal{W}_{N,k}(\beta_0) \leq z) - P_{\theta_0}(\mathcal{W}_N(\beta_0) \leq z)| = o(N^{-a}). \quad (8.25)$$

We apply Lemma 6 three times with $\omega_N = N^{1/2}\mu_{N,k}$ and with $(A_N(\theta_0), \xi_N(\theta_0))$ equal to $(N^{1/2}(\widehat{\theta}_N - \theta_0),\ N^{1/2}(\widehat{\theta}_{N,k} - \widehat{\theta}_N))$, $(T_N(\theta_{0,r}),\ T_{N,k}(\theta_{0,r}) - T_N(\theta_{0,r}))$, and $(H_N(\widehat{\theta}_N),\ H_N(\widehat{\theta}_{N,k}) - H_N(\widehat{\theta}_N))$. In the third application, we consider the convex sets $B_z = \{x \in R^{L_\beta} : x'x \leq z\}$ and use the fact that $\mathcal{W}_{N,k} = H_N(\widehat{\theta}_{N,k})' H_N(\widehat{\theta}_{N,k})$. By the assumption that $\mu_{N,k} = o(N^{-(a+1/2)})$, we have $\omega_N = o(N^{-a})$, as required by Lemma 6. The condition of Lemma 6 on $\xi_N(\theta_0)$ holds by (8.9). As required by Lemma 6, the random vector $T_N(\theta_{0,r})$ has an Edgeworth expansion with remainder $o(N^{-a})$ by Lemma 9(a). The same is true for $\Sigma^{-1/2}N^{-1/2}(\widehat{\theta}_N - \theta_0)$ and $H_N(\widehat{\theta}_N)$ by an argument analogous to that used to prove Lemma 9(a). □

8.2.3 Proof of Theorem 3

The proof of Theorem 3 is the same as that of Theorem 1 except that the results of Theorem 2(b) allow one to replace $T^*_N(\widehat{\theta}_{N,r})$, $z^*_{T,\alpha}$, and $z^*_{|T|,\alpha}$ by $T^*_{N,k}(\widehat{\theta}_{N,r})$, $z^*_{T,k,\alpha}$, and $z^*_{|T|,k,\alpha}$ throughout. In particular, the results of Theorem 2(b) allow one to replace $T^*_N(\widehat{\theta}_{N,r})$ by $T^*_{N,k}(\widehat{\theta}_{N,r})$ in the first line of (8.1) and the replacements elsewhere all follow. □

8.3 Proofs of Lemmas

8.3.1 Proof of Lemma 1

The NR result of the Lemma holds by definition of $Q^{NR,*}_{N,j-1}$.

To prove the other results of the Lemma, let $Q_{N,j-1}$, $Q^s_{N,j-1}$ for $s = NR, D, LS$, and GN, and $\widehat{\theta}_{N,j}$ for $j = 1, ..., k$ be defined as $Q^*_{N,j-1}$, $Q^{*,s}_{N,j-1}$, and $\theta^*_{N,j}$ are defined, respectively, but with the bootstrap sample $\{\widetilde{W}^*_i : i = 1, ..., N\}$ and estimator θ^*_N replaced by the original sample $\{\widetilde{W}_i : i = 1, ..., N\}$ and estimator $\widehat{\theta}_N$ and with the initial value $\widehat{\theta}_{N,0}$ replaced by the true parameter value θ_0. Then,

$$P^*_{\theta_0}(||Q^*_{N,j-1} - D^*_N(\theta^*_{N,j-1})|| > \psi_N) = P_{\theta_0}(||Q_{N,j-1} - D_N(\widehat{\theta}_{N,j-1})|| > \psi_N). \quad (8.26)$$

Hence, it suffices to show that the following holds for $Q_{N,j-1} = Q^s_{N,j-1}$ for $s = D, LS$, and GN:

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||Q_{N,j-1} - D_N(\widehat{\theta}_{N,j-1})|| > \psi_N) = o(N^{-a}). \quad (8.27)$$

We now establish (8.27) for the default NR matrix. Let $\widehat{\theta}_{N,j}$ denote the NR j-step estimator for $j = 1, ..., k$. Equation (8.27) holds with $Q_{N,j-1} = Q^D_{N,j-1}$ if

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\rho_N(\widehat{\theta}_{N,j}) - \rho_N(\widehat{\theta}_{N,j-1}) > 0) = o(N^{-a}) \quad (8.28)$$

for all $j = 1, ..., k$, because this implies that $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(Q^D_{N,j-1} \neq Q^{NR}_{N,j-1}$ for some $j = 1, ..., k) = o(N^{-a})$ and, by definition, $Q^{NR}_{N,j-1} = D_N(\widehat{\theta}_{N,j-1})$. When $\widehat{\theta}_{N,j} \neq \widehat{\theta}_{N,j-1}$, a Taylor expansion of $\rho_N(\widehat{\theta}_{N,j})$ about $\widehat{\theta}_{N,j-1}$ gives

$$\begin{aligned}
& \rho_N(\widehat{\theta}_{N,j}) - \rho_N(\widehat{\theta}_{N,j-1}) \\
&= \frac{\partial}{\partial \theta'} \rho_N(\widehat{\theta}_{N,j-1}) \zeta_{N,j} \phi_{N,j} + \frac{1}{2} \zeta'_{N,j} \frac{\partial^2}{\partial \theta \partial \theta'} \rho_N(\widehat{\theta}_{N,j-1}) \zeta_{N,j} \phi^2_{N,j} + \Gamma_{N,j} \phi^3_{N,j} \\
&= -\frac{1}{2} \zeta'_{N,j} \frac{\partial^2}{\partial \theta \partial \theta'} \rho_N(\widehat{\theta}_{N,j-1}) \zeta_{N,j} \phi^2_{N,j} + \Gamma_{N,j} \phi^3_{N,j}, \text{ where} \\
\Gamma_{N,j} &= \frac{1}{6} \sum_{u=1}^{L_\theta} \zeta_{N,j,u} \zeta'_{N,j} \frac{\partial^3}{\partial \theta_u \partial \theta \partial \theta'} \rho_N(\theta^+_{N,j-1}) \zeta_{N,j}, \\
\zeta_{N,j} &= (\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1}) / ||\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1}||, \ \phi_{N,j} = ||\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1}||, \quad (8.29)
\end{aligned}$$

$\zeta_{N,j,u}$ denotes the u-th element of $\zeta_{N,j}$, and $\theta^+_{N,j-1}$ lies between $\widehat{\theta}_{N,j}$ and $\widehat{\theta}_{N,j-1}$. The second equality holds by the definition of $\widehat{\theta}_{N,j}$. Using (8.29), the left-hand side of (8.28) is less than or equal to

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0} \left(-\lambda_{\min}(\frac{\partial^2}{\partial \theta \partial \theta'} \rho_N(\widehat{\theta}_{N,j-1}))/2 + \Gamma_{N,j} \phi_{N,j} > 0 \right). \quad (8.30)$$

The expression in (8.30) is $o(N^{-a})$, because

$$\begin{aligned}
& \sup_{\theta_0 \in \Theta_1} P_{\theta_0} \left(\lambda_{\min}(\frac{\partial^2}{\partial \theta \partial \theta'} \rho_N(\widehat{\theta}_{N,j-1})) < \lambda_{\min}(D(\theta_0))/2 \right) = o(N^{-a}), \\
& \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(|\Gamma_{N,j}| > K) = o(N^{-a}) \text{ for some } K < \infty, \text{ and} \\
& \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\phi_{N,j} > \varepsilon) = o(N^{-a}) \text{ for all } \varepsilon > 0, \quad (8.31)
\end{aligned}$$

where the first result holds by the second result of Lemma 4 with $\overline{\theta}_N = \widehat{\theta}_{N,j-1}$ and Assumption 3(c), the second holds by the third result of Lemma 4, and the third holds by two applications of the first result of (8.9) in the proof of part (a) of Theorem 2 for the NR estimator—one with $k = j-1$ and one with $k = j$. This completes the proof.

We now establish (8.27) for the line-search NR matrix. Let $\widehat{\theta}_{N,j}$ be the NR j-step estimator:

$$\widehat{\theta}_{N,j} = \widehat{\theta}_{N,j-1} - \varphi_{N,j-1}\pi_{N,j-1}, \text{ where}$$
$$\varphi_{N,j-1} = ||(Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j-1})|| \text{ and } \pi_{N,j-1} = (Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j-1})/\varphi_{N,j-1}. \quad (8.32)$$

Let

$$\widehat{\theta}^{\alpha}_{N,j} = \widehat{\theta}_{N,j-1} - \alpha(Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j-1}) = \widehat{\theta}_{N,j} + (1-\alpha)\varphi_{N,j-1}\pi_{N,j-1}. \quad (8.33)$$

It suffices to show that

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(\inf_{\alpha\in A,\alpha\neq 1}\rho_N(\widehat{\theta}^{\alpha}_{N,j}) - \rho_N(\widehat{\theta}_{N,j}) < 0) = o(N^{-a}) \quad (8.34)$$

for all $j = 1, ..., k$, because this implies that $\sup_{\theta_0\in\Theta_1} P_{\theta_0}(Q^{LS}_{N,j-1} \neq Q^{NR}_{N,j-1}$ for some $j = 1, ..., k) = o(N^{-a})$.

A Taylor expansion of $\rho_N(\widehat{\theta}^{\alpha}_{N,j})$ about $\widehat{\theta}_{N,j}$ gives

$$\begin{aligned}
&\rho_N(\widehat{\theta}^{\alpha}_{N,j}) - \rho_N(\widehat{\theta}_{N,j}) = (1-\alpha)\varphi_{N,j-1}\pi'_{N,j-1}\frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j})\\
&+\frac{1}{2}(1-\alpha)^2\varphi^2_{N,j-1}\pi'_{N,j-1}\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,j})\pi_{N,j-1}\\
&+\frac{1}{6}(1-\alpha)^3\varphi^3_{N,j-1}\sum_{u=1}^{L_\theta}\pi_{N,j-1,u}\pi'_{N,j-1}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}\rho_N(\theta^{+}_{N,j})\pi_{N,j-1}, \quad (8.35)
\end{aligned}$$

where $\theta^{+}_{N,j}$ lies between $\widehat{\theta}^{\alpha}_{N,j}$ and $\widehat{\theta}_{N,j}$ and $\pi_{N,j-1,u}$ denotes the u-th element of $\pi_{N,j-1}$.

Element by element Taylor expansions of $(\partial/\partial\theta)\rho_N(\widehat{\theta}_{N,j})$ about $\widehat{\theta}_{N,j-1}$ give

$$\begin{aligned}
\frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j}) &= \frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_{N,j-1}) + \frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,j-1})(\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1})\\
&\quad+\frac{1}{2}[(\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1})'\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}\rho_N(\theta^{++}_{N,j-1,u})(\widehat{\theta}_{N,j} - \widehat{\theta}_{N,j-1})]_{L_\theta}\\
&= 0 + \frac{1}{2}\varphi^2_{N,j-1}[\pi'_{N,j-1}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}\rho_N(\theta^{++}_{N,j-1,u})\pi_{N,j-1}]_{L_\theta}, \quad (8.36)
\end{aligned}$$

where $\theta^{++}_{N,j-1,u}$ lies between $\widehat{\theta}_{N,j}$ and $\widehat{\theta}_{N,j-1}$, $[A_u]_{L_\theta}$ denotes the L_θ-vector whose u-th element is A_u, and the second equality holds using the definition of $\widehat{\theta}_{N,j}$.

The following properties hold: for all $\varepsilon > 0$,

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(\lambda_{\min}(\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\widehat{\theta}_{N,j-1})) < \lambda_{\min}(D(\theta_0))/2) = o(N^{-a}),$$
$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||\frac{\partial^3}{\partial\theta^3}\rho_N(\theta^{++}_{N,j-1})|| > K) = o(N^{-a}) \text{ for some } K < \infty, \text{ and}$$
$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(\varphi_{N,j} > \varepsilon) = o(N^{-a}) \text{ for some } \varepsilon > 0 \tag{8.37}$$

for $j = 1, ..., k$, where the first result of (8.37) holds by the second result of Lemma 4 with $\overline{\theta}_N = \widehat{\theta}_{N,j-1}$, Assumption 3(c), and the first result of (8.9) of the proof of part (a) of Theorem 2 (which ensures that $\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||\widehat{\theta}_{N,j-1} - \theta_0|| > \varepsilon) = o(N^{-a})$), the second holds by the third result of Lemma 4 with $\overline{\theta}_N = \theta^{++}_{N,j-1}$, and the third holds by (i) the second result of Lemma 4 with $\overline{\theta}_N = \widehat{\theta}_{N,j-1}$ and Assumption 3(c) (which ensure that $(Q^{NR}_{N,j})^{-1}$ is well-behaved) and (ii) the fourth result of Lemma 4 with $\overline{\theta}_N = \widehat{\theta}_{N,j-1}$. The second result of (8.37) also holds with $\theta^{++}_{N,j-1}$ replaced by $\theta^{+}_{N,j-1}$.

Substituting (8.36) into the right-hand side of (8.35), dividing (8.35) by $\varphi^2_{N,j-1}$ (when $\varphi_{N,j-1} > 0$), and applying (8.37) yields the resultant first and third terms on the right-hand side of (8.35) to have norm greater than $\varepsilon > 0$ with probability $o(N^{-a})$ and the second term to be strictly positive with probability $1 - o(N^{-a})$ (uniformly over $\alpha \in A$ with $\alpha \neq 1$), which gives (8.34). This completes the proof.

Lastly, we establish (8.27) for the GN matrix. It suffices to show that

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{W}_i,\widehat{\theta}_{N,j-1}) - \frac{\partial}{\partial\theta'}g(\widetilde{W}_i,\widehat{\theta}_{N,j-1}))|| > N^{-1/2}\ln(N)) = o(N^{-a}). \tag{8.38}$$

By mean value expansions about θ_0 and the triangle inequality,

$$\begin{aligned} &||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{W}_i,\widehat{\theta}_{N,j-1}) - \frac{\partial}{\partial\theta'}g(\widetilde{W}_i,\widehat{\theta}_{N,j-1}))|| \\ \leq\ &||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{W}_i,\theta_0) - \frac{\partial}{\partial\theta'}g(\widetilde{W}_i,\theta_0))|| \\ &+N^{-1}\sum_{i=1}^N \sup_{\theta\in B(\theta_0,\varepsilon),u\leq L_\theta} ||\frac{\partial}{\partial\theta_u}\Delta(\widetilde{W}_i,\theta) - \frac{\partial^2}{\partial\theta_u\partial\theta'}g(\widetilde{W}_i,\theta)|| \cdot ||\widehat{\theta}_{N,j-1} - \theta_0||. \end{aligned} \tag{8.39}$$

In addition, $||\widehat{\theta}_{N,j-1} - \theta_0|| \leq ||\widehat{\theta}_{N,j-1} - \widehat{\theta}_N|| + ||\widehat{\theta}_N - \theta_0||$. Hence, it suffices to show that

$$\text{(i)}\ \sup_{\theta_0\in\Theta_1} P_{\theta_0}\left(||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{W}_i,\theta_0) - \frac{\partial}{\partial\theta'}g(\widetilde{W}_i,\theta_0))|| > N^{-1/2}\ln(N)\right) = o(N^{-a}),$$

(ii) $\sup_{\theta_0 \in \Theta_1} P_{\theta_0} \left(N^{-1} \sum_{i=1}^{N} \sup_{\theta \in B(\theta_0, \varepsilon), u \leq L_\theta} || \frac{\partial}{\partial \theta_u} \Delta(\widetilde{W}_i, \theta) - \frac{\partial^2}{\partial \theta_u \partial \theta'} g(\widetilde{W}_i, \theta) || > K \right)$
$= o(N^{-a}),$

(iii) $\sup_{\theta_0 \in \Theta_1} P_{\theta_0} \left(||\widehat{\theta}_{N,j-1} - \widehat{\theta}_N|| > N^{-1/2} \ln(N) \right) = o(N^{-a}),$ and

(iv) $\sup_{\theta_0 \in \Theta_1} P_{\theta_0} \left(||\widehat{\theta}_N - \theta_0|| > N^{-1/2} \ln(N) \right) = o(N^{-a})$ (8.40)

for all $j = 1, ..., k$ and some $K < \infty$. Condition (i) holds by Lemma 3(c), (ii) holds by Lemma 3(b) with $p = \min\{q_1, q_2\}$, (iv) holds by Lemma 5, (iii) holds for $j = 1$ because $\widehat{\theta}_{N,0} = \widehat{\theta}_N$, and (iii) holds for $j = 2, ..., k$ by recursively applying the first result of (8.9) in the proof of part (a) of Theorem 2 with $k = j - 1$, which holds without assuming Assumption 5 by the present proof that the result of Assumption 5 holds for $Q_{N,i}$ for $i \leq j - 1$ under the assumptions. □

8.3.2 Proof of Lemma 2

We have

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(|\lambda_N(\widehat{\theta}_N)| > N^{-a}\varepsilon) \\
\leq &\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(|\lambda_N(\widehat{\theta}_N)| > N^{-a}\varepsilon, \widehat{\theta}_N \in B(\theta_0, \delta/2)) + \sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\widehat{\theta}_N \notin B(\theta_0, \delta/2)) \\
\leq &\sup_{\theta_0 \in \Theta_0} P_{\theta_0}(\sup_{\theta \in \Theta_1} |\lambda_N(\theta)| > N^{-a}\varepsilon) + o(N^{-a}) \\
= &\, 1(o(N^{-a}) > N^{-a}\varepsilon) + o(N^{-a}) \\
= &\, o(N^{-a}), \qquad (8.41)
\end{aligned}$$

where the second inequality uses the fact that when $\widehat{\theta}_N \in B(\theta_0, \delta/2)$ and $\theta_0 \in \Theta_0$ one has $\widehat{\theta}_N \in \Theta_1$. □

8.3.3 Proof of Lemma 3

A strong mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) gives $\sup_{\theta_0 \in \Theta_1} E_{\theta_0} || \sum_{i=1}^N m(\widetilde{W}_i, \theta_0)||^p < CN^{p/2}$ provided $p \geq 2$. Application of Markov's inequality and the Yokoyama–Doukhan inequality yields the left-hand side in part (a) of the Lemma to be less than or equal to

$$\varepsilon^{-p} N^{-p} \sup_{\theta_0 \in \Theta_1} E_{\theta_0} || \sum_{i=1}^{N} m(\widetilde{W}_i, \theta_0)||^p \leq \varepsilon^{-p} C N^{-p/2} = o(N^{-a}). \qquad (8.42)$$

Part (b) follows from part (a) applied to $m(\widetilde{W}_i, \theta_0) - E_{\theta_0} m(\widetilde{W}_i, \theta_0)$ and the triangle inequality.

To establish part (c), we use the Edgeworth expansion given in Theorem 2.3 of Lahiri (1993) (also see Corollary 2.9 of Götze and Hipp (1983)) with their $s = 2a + 2$.

Conditions 1 and 3–6 of Lahiri (1993) hold uniformly over $\theta_0 \in \Theta_1$ by Assumption 4. Their condition 2 holds uniformly over $\theta_0 \in \Theta_1$ by Assumption 3(b). Because the result of the Lemma can be proved element by element, we consider an arbitrary element $f_v(\cdot,\theta_0)$ of $f(\cdot,\theta_0)$. Let $\Phi(\cdot)$ denote the standard normal distribution function. By the Edgeworth expansion, for each $\theta_0 \in \Theta_1$ there are homogeneous polynomials $\pi_i(\delta,\theta_0)$ in $\delta = \partial/\partial z$ for $i = 1, ..., 2a$ such that

$$\begin{aligned}
&\sup_{z\in R} |P_{\theta_0}(N^{-1/2}\sum_{i=1}^{N}(f_v(\widetilde{W}_i,\theta_0) - E_{\theta_0} f_v(\widetilde{W}_i,\theta_0)) \le z) \\
&\qquad\qquad -(1+\sum_{i=1}^{2a} N^{-i/2}\pi_i(\delta,\theta_0))\Phi(z)| \\
&= o(N^{-a}). \qquad (8.43)
\end{aligned}$$

The error $o(N^{-a})$ holds uniformly over $\theta_0 \in \Theta_1$ because Assumptions 3(b) and 4 hold uniformly over $\theta_0 \in \Theta_1$. Equation (8.43) implies that for any constant z_N

$$\begin{aligned}
&P_{\theta_0}(|N^{-1/2}\sum_{i=1}^{N}(f_v(\widetilde{W}_i,\theta_0) - E_{\theta_0} f_v(\widetilde{W}_i,\theta_0))| > z_N) \\
&= 1 - (1+\sum_{i=1}^{2a} N^{-i/2}\pi_i(\delta,\theta_0))(\Phi(z_N) - \Phi(-z_N)) + o(N^{-a}) \\
&= 2\Phi(-z_N) - (\sum_{i=1}^{2a} N^{-i/2}\pi_i(\delta,\theta_0))(\Phi(z_N) - \Phi(-z_N)) + o(N^{-a}), \qquad (8.44)
\end{aligned}$$

where the error holds uniformly over $\theta_0 \in \Theta_1$. Let $z_N = \varepsilon \ln(N)$. Using $\Phi(-z) \le C\exp(-z^2/2)$ for $z > 1$, we have

$$\Phi(-z_N) \le C\exp(-\varepsilon^2 \ln^2(N)/2) \le C\exp(-(a+1)\ln(N)) = CN^{-(a+1)} = o(N^{-a}), \qquad (8.45)$$

where the second inequality holds for any given $a \ge 0$ and $\varepsilon > 0$ for N sufficiently large. The expression $\pi_i(\delta,\theta_0)\Phi(z_N)$ is a finite sum of terms of the form $b(\theta_0)z_N^j\phi(z_N)$ for some integer j and some function $b(\theta_0)$ that satisfies $\sup_{\theta_0\in\Theta_1}|b(\theta_0)| < \infty$ (which holds by the uniform moment bound over $\theta_0 \in \Theta_1$ given in Assumption 3(b)), where $\phi(\cdot)$ denotes the standard normal density. By an analogous calculation to that in (8.45), $z_N^j\phi(z_N) = \varepsilon^j \ln^j(N)(2\pi)^{-1/2}\exp(-\varepsilon^2\ln^2(N)/2) = o(N^{-a})$. This completes the proof. □

8.3.4 Proof of Lemma 4

The first result of the Lemma follows from

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||V_N(\overline{\theta}_N) - V_N(\theta_0)|| > \varepsilon) = o(N^{-a}), \qquad (8.46)$$

$$\sup_{\theta_0\in\Theta_1} P_{\theta_0}(||V_N(\theta_0) - E_{\theta_0}V_N(\theta_0)|| > \varepsilon) = o(N^{-a}), \text{ and} \qquad (8.47)$$

$$\sup_{\theta_0\in\Theta_1} |E_{\theta_0}V_N(\theta_0) - V(\theta_0)| = o(1). \qquad (8.48)$$

To establish (8.46), we take mean value expansions about θ_0, apply Lemma 3(b) with $m(\widetilde{W}_i, \theta_0) = \sup_{\theta \in \Theta_2} ||g(\widetilde{W}_i, \theta)|| \cdot ||(\partial/\partial\theta')g(\widetilde{W}_i, \theta)||$ and $p = q_1$, where the sup is over $\theta \in \Theta_2$ because $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\overline{\theta}_N \notin \Theta_2) = o(N^{-a})$, and use the assumption on $\overline{\theta}_N$. To establish (8.47), we use Lemma 3(a) with $m(\widetilde{W}_i, \theta_0) = g(\widetilde{W}_i, \theta_0)g(\widetilde{W}_i, \theta_0)'$ $-E_{\theta_0} g(\widetilde{W}_i, \theta_0)g(\widetilde{W}_i, \theta_0)'$ and $p = q_1$. Equation (8.48) holds by Assumption 3(c).

The remaining results of the Lemma hold by mean value expansions about θ_0, multiple applications of Lemma 3(b) with $m(\widetilde{W}_i, \theta_0) = (\partial^j/\partial\theta^j)g(\widetilde{W}_i, \theta_0)$ for $j = 0, ..., 3$, multiple applications of Lemma 3(a) with $m(\widetilde{W}_i, \theta_0) = (\partial^j/\partial\theta^j)\ g(\widetilde{W}_i, \theta_0)$ $-E_{\theta_0}(\partial^j/\partial\theta^j)\ g(\widetilde{W}_i, \theta_0)$ for $j = 0, 1$ and $p = q_1$, the assumption on $\overline{\theta}_N$, and Assumption 3(c). □

8.3.5 Proof of Lemma 5

First, we show that for all $\varepsilon > 0$,

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\sup_{\theta \in \Theta} |N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i, \theta) - E_{\theta_0}\rho(\widetilde{W}_i, \theta)| > \varepsilon) = o(N^{-a}). \tag{8.49}$$

By Assumption 2(a), Θ is compact. Hence, for any $\eta > 0$, there exist points $\{\theta_j \in \Theta : 2 \leq j \leq J\}$ such that $\cup_{j=2}^{J} B(\theta_j, \eta)$ contains Θ (where $B(\theta_j, \varepsilon)$ denotes the open ball centered at θ_j with radius ε). The left-hand side of (8.49) is less than or equal to

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\max_{2 \leq j \leq J} \sup_{\theta \in B(\theta_j, \eta)} \left(|N^{-1} \sum_{i=1}^{N} [\rho(\widetilde{W}_i, \theta) - E_{\theta_0}\rho(\widetilde{W}_i, \theta)\right.\right.\\
&\quad \left.\left. -(\rho(\widetilde{W}_i, \theta_j) - E_{\theta_0}\rho(\widetilde{W}_i, \theta_j))]| + |N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i, \theta_j) - E_{\theta_0}\rho(\widetilde{W}_i, \theta_j)|\right) > \varepsilon\right)\\
&\leq \sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\max_{2 \leq j \leq J} \sup_{\theta \in B(\theta_j, \eta)} N^{-1} \sum_{i=1}^{N} \left(\sup_{\overline{\theta} \in \Theta} ||(\partial/\partial\theta)\rho(\widetilde{W}_i, \overline{\theta})||\right.\right.\\
&\quad \left.\left. +E_{\theta_0} \sup_{\overline{\theta} \in \Theta} ||(\partial/\partial\theta)\rho(\widetilde{W}_i, \overline{\theta})||\right) ||\theta - \theta_j|| > \varepsilon/2\right)\\
&\quad + \sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\max_{2 \leq j \leq J} |N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i, \theta_j) - E_{\theta_0}\rho(\widetilde{W}_i, \theta_j)| > \varepsilon/2\right)\\
&\leq \sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(N^{-1} \sum_{i=1}^{N} \left(\sup_{\overline{\theta} \in \Theta} ||(\partial/\partial\theta)\rho(\widetilde{W}_i, \overline{\theta})|| + E_{\theta_0} \sup_{\overline{\theta} \in \Theta} ||(\partial/\partial\theta)\rho(\widetilde{W}_i, \overline{\theta})||\right) \eta > \varepsilon/2\right)\\
&\quad + \sum_{j=2}^{J} \sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(|N^{-1} \sum_{i=1}^{N} \rho(\widetilde{W}_i, \theta_j) - E_{\theta_0}\rho(\widetilde{W}_i, \theta_j)| > \varepsilon/2\right)\\
&= o(N^{-a}),
\end{aligned} \tag{8.50}$$

where the first inequality uses mean value expansions and the equality holds using Assumption 2(e) by Lemma 3(b) with $p = q_0$ by taking η sufficiently small and by Lemma 3(a) with $p = q_0$.

Next, we prove that $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\widehat{\theta}_N - \theta_0|| > \varepsilon) = o(N^{-a})$. By Assumption 2(d), given $\varepsilon > 0$, there exists a $\delta > 0$ such that $||\theta - \theta_0|| > \varepsilon$ implies that $\rho(\theta, \theta_0) - \rho(\theta_0, \theta_0) \geq \delta > 0$. Thus,

$$\begin{aligned}
\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\widehat{\theta}_N - \theta_0|| > \varepsilon) &\leq \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\rho(\widehat{\theta}_N, \theta_0) - \rho_N(\widehat{\theta}_N) + \rho_N(\widehat{\theta}_N) - \rho(\theta_0, \theta_0) > \delta) \\
&\leq \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(\rho(\widehat{\theta}_N, \theta_0) - \rho_N(\widehat{\theta}_N) + \rho_N(\theta_0) - \rho(\theta_0, \theta_0) > \delta) \\
&\leq \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(2 \sup_{\theta \in \Theta} |\rho_N(\theta) - E_{\theta_0} \rho_N(\theta)| \\
&\qquad + 2 \sup_{\theta \in \Theta} |E_{\theta_0} \rho_N(\theta) - \rho(\theta, \theta_0)| > \delta) \\
&= o(N^{-a}) \qquad (8.51)
\end{aligned}$$

using (8.49) and Assumption 2(c).

The result of (8.51) and the assumption that all $\theta_0 \in \Theta_1$ are in the interior of Θ imply that $\inf_{\theta_0 \in \Theta_1} P_{\theta_0}(\widehat{\theta}_N$ is in the interior of $\Theta) = 1 - o(N^{-a})$ and $\inf_{\theta_0 \in \Theta_1} P_{\theta_0}((\partial/\partial\theta)\rho_N(\widehat{\theta}_N) = 0) = 1 - o(N^{-a})$. Hence, element by element mean value expansions of $(\partial/\partial\theta)\rho_N(\widehat{\theta}_N)$ about θ_0 and rearrangement give

$$\sup_{\theta_0 \in \Theta_1} P_{\theta_0}\left(\widehat{\theta}_N - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\theta_N^+)\right)^{-1} \frac{\partial}{\partial\theta}\rho_N(\theta_0)\right) = 1 - o(N^{-a}), \qquad (8.52)$$

where θ_N^+ lies between $\widehat{\theta}_N$ and θ_0 and may differ across rows. In consequence, the result of the Lemma follows from the second result of Lemma 4 with $\overline{\theta}_N = \theta_N^+$ and $\sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||N^{-1/2}\sum_{i=1}^N g(\widetilde{W}_i, \theta_0)|| > \ln(N)\varepsilon) = o(N^{-a})$, which holds by Lemma 3(c) with $m(\widetilde{W}_i, \theta_0) = g(\widetilde{W}_i, \theta_0)$ using the assumption that $q_1 \geq 2a + 3$. □

8.3.6 Proof of Lemma 6

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

$$\begin{aligned}
&\sup_{\theta_0 \in \Theta_1, B \in \mathcal{B}_{L_A}} (P_{\theta_0}(A_N(\theta_0) + \xi_N(\theta_0) \in B) - P_{\theta_0}(A_N(\theta_0) \in B)) \\
= &\sup_{\theta_0 \in \Theta_1, B \in \mathcal{B}_{L_A}} (P_{\theta_0}(A_N(\theta_0) + \xi_N(\theta_0) \in B, ||\xi_N(\theta_0)|| \leq \omega_N) - P_{\theta_0}(A_N(\theta_0) \in B) \\
&\quad + P_{\theta_0}(A_N(\theta_0) + \xi_N(\theta_0) \in B, ||\xi_N(\theta_0)|| > \omega_N)) \\
\leq &\sup_{\theta_0 \in \Theta_1, B \in \mathcal{B}_{L_A}} \left(P_{\theta_0}(A_N(\theta_0) \in B_{\omega_N}^+) - P_{\theta_0}(A_N(\theta_0) \in B)\right) \\
&\quad + \sup_{\theta_0 \in \Theta_1} P_{\theta_0}(||\xi_N(\theta_0)|| > \omega_N). \qquad (8.53)
\end{aligned}$$

The second term on the right-hand side is $o(N^{-a})$ by assumption. Under the assumption that $A_N(\theta_0)$ has an Edgeworth expansion with remainder $o(N^{-a})$ uniformly over

$\theta_0 \in \Theta_1$, the first term on the right-hand side of (8.53) is less than or equal to

$$\sup_{\theta_0\in\Theta_1, B\in\mathcal{B}_{L_A}} \left(\int_{B^+_{\omega_N}} (1+\sum_{i=1}^{[2a]} N^{-i/2}\pi_{N,i}(z,\theta_0))\phi_{\Omega_N(\theta_0)}(z)dz \right. $$
$$\left. - \int_B (1+\sum_{i=1}^{[2a]} N^{-i/2}\pi_{N,i}(z,\theta_0))\phi_{\Omega_N(\theta_0)}(z)dz \right) + o(N^{-a}). \quad (8.54)$$

The expression in (8.54) is $O(\omega_N) = o(N^{-a})$ because $\phi_{\Omega_N}(z)$ and its derivatives of all orders are bounded over $z \in R^{L_A}$ given the assumptions on $\Omega_N(\theta_0)$ and the polynomials $\{\pi_{N,i}(z,\theta_0) : i = 1,...,2a\}$ have coefficients that are $O(1)$ uniformly over $\theta_0 \in \Theta_1$. Hence, the left-hand side of (8.53) is less than or equal to $o(N^{-a})$.

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

$$P_{\theta_0}(A_N(\theta_0)+\xi_N(\theta_0) \in B) \geq P_{\theta_0}(A_N(\theta_0) \in B^-_{\omega_N}, ||\xi_N(\theta_0)|| \leq \omega_N). \quad (8.55)$$

Using this, an analogous argument to that of (8.53) and (8.54) shows that

$$\sup_{\theta_0\in\Theta_1, B\in\mathcal{B}_{L_A}} (P_{\theta_0}(A_N(\theta_0) \in B) - P_{\theta_0}(A_N(\theta_0)+\xi_N(\theta_0) \in B)) \leq o(N^{-a}), \quad (8.56)$$

which completes the proof. □

8.3.7 Proof of Lemma 7

Suppose $\Delta_N(\theta_0) = N^{1/2}(\widehat{\theta}_N - \theta_0)$. By Lemma 5 and Assumption 2(a), we have $\inf_{\theta_0\in\Theta_1} P_{\theta_0}(\widehat{\theta}_N \text{ is in the interior of } \Theta) = 1-o(N^{-a})$ and $\inf_{\theta_0\in\Theta_1} P_{\theta_0}((\partial/\partial\theta)\rho_N(\widehat{\theta}_N) = 0) = 1-o(N^{-a})$. Element by element Taylor expansions of $(\partial/\partial\theta)\rho_N(\widehat{\theta}_N)$ about θ_0 of order $d-1$ give

$$0 = \frac{\partial}{\partial\theta}\rho_N(\widehat{\theta}_N) = \frac{\partial}{\partial\theta}\rho_N(\theta_0) + \sum_{j=1}^{d-1}\frac{1}{j!}D^j\frac{\partial}{\partial\theta}\rho_N(\theta_0)(\widehat{\theta}_N-\theta_0,...,\widehat{\theta}_N-\theta_0) + \zeta_N(\theta_0),$$

where

$$\zeta_N(\theta_0) = \frac{1}{j!}(D^{d-1}\frac{\partial}{\partial\theta}\rho_N(\theta^+_N) - D^{d-1}\frac{\partial}{\partial\theta}\rho_N(\theta_0))(\widehat{\theta}_N-\theta_0,...,\widehat{\theta}_N-\theta_0), \quad (8.57)$$

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 $(\widehat{\theta}_N-\theta_0,...,\widehat{\theta}_N-\theta_0)$. Let $R_N(\theta_0)$ denote the column vector whose elements are the unique components of $(\partial/\partial\theta)\rho_N(\theta_0)$, $D^1(\partial/\partial\theta)\rho_N(\theta_0)$, ..., $D^{d-1}(\partial/\partial\theta)\rho_N(\theta_0)$. Each element of $R_N(\theta_0)$ is an element of $S_N(\theta_0)$. Let $e_N(\theta_0) = (\zeta_N(\theta_0)', 0, ..., 0)'$ be conformable to $R_N(\theta_0)$. The first equation in (8.57) can be written as $\nu(R_N(\theta_0)+e_N(\theta_0), \widehat{\theta}_N-\theta_0) = 0$, where $\nu(\cdot,\cdot)$ is an infinitely differentiable function, $\nu(E_{\theta_0}R_N(\theta_0), 0) = 0$ for all $N \geq 1$, and

$(\partial/\partial x)\nu(E_{\theta_0}R_N(\theta_0),x)|_{x=0} = N^{-1}\sum_{i=1}^N E_{\theta_0}g(\widetilde{W}_i,\theta_0)g(\widetilde{W}_i,\theta_0)'$ is positive definite for N large by Assumption 3(c). Hence, the implicit function theorem can be applied to $\nu(\cdot,\cdot)$ at the point $(E_{\theta_0}R_N(\theta_0),0)$ to obtain

$$\inf_{\theta_0\in\Theta_1} P_{\theta_0}(\widehat{\theta}_N-\theta_0=\Lambda(R_N(\theta_0)+e_N(\theta_0)))=1-o(N^{-a}), \tag{8.58}$$

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

We apply Lemma 6 with $A_N(\theta_0)=N^{1/2}\Lambda(R_N(\theta_0))$ and $\xi_N(\theta_0)=N^{1/2}(\Lambda(R_N(\theta_0)$ $+e_N(\theta_0))-\Lambda(R_N(\theta_0)))$ to obtain

$$\sup_{\theta_0\in\Theta_1,B\in\mathcal{B}_{L_\theta}}|P_{\theta_0}(N^{1/2}\Lambda(R_N(\theta_0)+e_N(\theta_0))\in B)-P_{\theta_0}(N^{1/2}\Lambda(R_N(\theta_0))\in B)|=o(N^{-a}). \tag{8.59}$$

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) $||e_N(\theta_0)|| = ||\zeta_N(\theta_0)||$, (iii) $\zeta_N(\theta_0)$ satisfies $\inf_{\theta_0\in\Theta_1}P_{\theta_0}(||\zeta_N(\theta_0)||\leq C||\widehat{\theta}_N-\theta_0||^d)=1-o(N^{-a})$, (iv) ω_N, which is defined to equal $N^{1/2-d/2}\ln^d(N)$, is $o(N^{-a})$ because $d\geq 2a+2$ by Assumption 3(a), (v) $\sup_{\theta_0\in\Theta_1}P_{\theta_0}(N^{1/2}||e_N(\theta_0)||>\omega_N)\leq\sup_{\theta_0\in\Theta_1}P_{\theta_0}(CN^{1/2}||\widehat{\theta}_N-\theta_0||^d>\omega_N)$ $+o(N^{-a})=o(N^{-a})$ by Lemma 5, (vi) $\Lambda(R_N(\theta_0))$ can be written as $G(S_N(\theta_0))$, where $G(\cdot)$ is infinitely differentiable and $G(E_{\theta_0}S_N(\theta_0))=0$ for all N large and (vii) $A_N(\theta_0)=N^{1/2}\Lambda(R_N(\theta_0))=N^{1/2}G(S_N(\theta_0))$ has an Edgeworth expansion (with remainder $o(N^{-a})$ uniformly over $\theta_0\in\Theta_1$) by the proof of Lemma 9 below.

Equations (8.58) and (8.59) and $\Lambda(R_N(\theta_0))=G(S_N(\theta_0))$ yield the result of the Lemma.

Each of the remaining forms of $\Delta_N(\theta_0)$ (viz., $T_N(\theta_{0,r})$ and $H_N(\widehat{\theta}_N,\beta_0)$) 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 $d-1$ to obtain

$$\Delta_N(\theta_0)=N^{1/2}(\Lambda^{**}(S_N(\theta_0),\widehat{\theta}_N-\theta_0)+\zeta_N^{**}(\theta_0)), \tag{8.60}$$

where Λ^{**} is an infinitely differentiable function that does not depend on θ_0, $\Lambda^{**}(E_{\theta_0}S_N(\theta_0),0)=0$ for N large, $\zeta_N^{**}(\theta_0)$ is the remainder term in the Taylor expansion, and $||\zeta_N^{**}(\theta_0)||=O(||\widehat{\theta}_N-\theta_0||^d)$. Combining (8.58) with (8.60) gives $\Delta_N(\theta_0)=N^{1/2}(\Lambda^{**}(S_N(\theta_0),\ \Lambda(R_N(\theta_0)+e_N(\theta_0)))+\zeta_N^{**}(\theta_0))$. We apply Lemma 6 again, using the result above for $||\zeta_N^{**}(\theta_0)||$, to obtain an analogue of (8.59) with $A_N(\theta_0)=N^{1/2}\Lambda^{**}(S_N(\theta_0),\Lambda(R_N(\theta_0)))$. We can write $G(S_N(\theta_0))=\Lambda^{**}(S_N(\theta_0),$ $\Lambda(R_N(\theta_0)))$, where $G(\cdot)$ is infinitely differentiable and $G(E_{\theta_0}S_N(\theta_0))=\Lambda^{**}(E_{\theta_0}S_N(\theta_0),$ $\Lambda(E_{\theta_0}R_N(\theta_0)))=\Lambda^{**}(E_{\theta_0}S_N(\theta_0),0)=0$ for all N large. Combining this, the analogue of (8.59), and (8.60) gives the result of the Lemma for $\Delta_N(\theta_0)$ equal to $T_N(\theta_{0,r})$ and $H_N(\widehat{\theta}_N,\beta_0)$. $\square$

8.3.8 Proof of Lemma 8

We show below that for all $\theta_0 \in \Theta_1$ and all $\theta \in \Theta_2$ such that $||\theta - \theta_0|| < \delta$ (where δ is as in the definition of Θ_1),

$$|N^{\alpha(m)} E_\theta \prod_{\mu=1}^{m} \Psi_{N,j_\mu} - N^{\alpha(m)} E_{\theta_0} \prod_{\mu=1}^{m} \Psi_{N,j_\mu}| \leq B_N ||\theta - \theta_0||, \tag{8.61}$$

where $\limsup_{N\to\infty} B_N < \infty$. Let $\eta > 0$ satisfy $\eta < \varepsilon/(L_\nu^{1/2} \limsup_{N\to\infty} B_N)$, where L_ν denotes the dimension of $\nu_{N,a}(\theta_0)$. Then,

$$\begin{aligned}
&\sup_{\theta_0\in\Theta_1} P_{\theta_0}(N^{1/2}||\nu_{N,a}(\widehat{\theta}_N) - \nu_{N,a}(\theta_0)|| > \ln(N)\varepsilon) \\
&\leq \sup_{\theta_0\in\Theta_1} P_{\theta_0}(N^{1/2}||\nu_{N,a}(\widehat{\theta}_N) - \nu_{N,a}(\theta_0)|| > \ln(N)\varepsilon,\ N^{1/2}||\widehat{\theta}_N - \theta_0|| \leq \ln(N)\eta) \\
&\quad + \sup_{\theta_0\in\Theta_1} P_{\theta_0}(N^{1/2}||\widehat{\theta}_N - \theta_0|| > \ln(N)\eta) \\
&\leq \sup_{\theta_0\in\Theta_1} P_{\theta_0}(L_\nu^{1/2} B_N N^{1/2}||\widehat{\theta}_N - \theta_0|| > \ln(N)\varepsilon,\ N^{1/2}||\widehat{\theta}_N - \theta_0|| \leq \ln(N)\eta) + o(N^{-a}) \\
&= o(N^{-a}),
\end{aligned} \tag{8.62}$$

where the second inequality uses (8.61) Lemma 5.

Under the assumptions, (8.61) holds provided: for all $\theta_0 \in \Theta_1$ and all $\theta \in \Theta_2$ such that $||\theta - \theta_0|| < \delta$,

$$|E_\theta \prod_{\mu=1}^{m} f_{j_\mu}(\widetilde{W}_i, \theta) - E_{\theta_0} \prod_{\mu=1}^{m} f_{j_\mu}(\widetilde{W}_i, \theta_0)| \leq B_{1,N} ||\theta - \theta_0||, \tag{8.63}$$

for all $m \leq 2a + 2$, all $i \geq 1$, and all $j_\mu \leq L_f$, where $f_{j_\mu}(\widetilde{W}_i, \theta)$ denotes the j_μ–th element of $f(\widetilde{W}_i, \theta)$ and $\limsup_{N\to\infty} B_{1,N} < \infty$. The triangle inequality, a mean-value expansion, and some calculations show that (8.63) holds if

$$\sup_{\theta_0\in\Theta_1, i\geq 1} E_{\theta_0} ||C_f^j(\widetilde{W}_i) f_{j_\mu}^{2a+3-j}(\widetilde{W}_i, \theta_0)|| < \infty \text{ for all } j = 0, ..., 2a + 2 \tag{8.64}$$

and for all elements j_μ of $f(\widetilde{W}_i, \theta_0)$. This holds if $q_1 \geq 2a + 3$, as is assumed. $\square$

8.3.9 Proof of Lemma 9

We establish the first result of part (a) first. By Lemma 7, it suffices to show that the random variable $N^{1/2} G(S_N(\theta_0))$ of Lemma 7(a) possesses an Edgeworth expansion with remainder $o(N^{-a})$ uniformly over $\theta_0 \in \Theta_1$. We obtain an Edgeworth expansion for $N^{1/2}(S_N(\theta_0) - E_{\theta_0} S_N(\theta_0))$ for each $\theta_0 \in \Theta_1$ via Theorem 2.1 of Lahiri (1993) (also see Corollary 2.9 of Götze and Hipp (1983)), as in the proof of Lemma 3(c). The remainder is uniform in $\theta_0 \in \Theta_1$ because the conditions in Assumptions 3(b), 3(c), and 4 hold uniformly over $\theta_0 \in \Theta_1$. Edgeworth expansions for

$N^{1/2}G(S_N(\theta_0))$ are now obtained from those of $N^{1/2}(S_N(\theta_0) - E_{\theta_0}S_N(\theta_0))$ by the argument in Bhattacharya (1985, Pf. of Thm. 1) or Bhattacharya and Ghosh (1978, Pf. of Thm. 2) using the smoothness of $G(\cdot)$, $G(E_{\theta_0}S_N(\theta_0)) = 0$ for all $N \geq 1$ and all $\theta_0 \in \Theta_1$, and Assumption 3(c).

To establish the second result of part (a), we consider the convex sets $B_z = \{x \in R^{L_\beta} : x'x \leq z\}$ for $z \in R$. By Lemma 7(a) with $\Delta_N(\theta_0) = H_N(\widehat{\theta}_N, \beta_0)$, we have

$$\begin{aligned} o(N^{-a}) &= \sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(H_N(\widehat{\theta}_N, \beta_0) \in B_z) - P_{\theta_0}(N^{1/2}G(S_N(\theta_0)) \in B_z)| \\ &= \sup_{\theta_0 \in \Theta_1} \sup_{z \in R} |P_{\theta_0}(\mathcal{W}_N(\beta_0) \leq z) - P_{\theta_0}(NG(S_N(\theta_0))'G(S_N(\theta_0)) \leq z)|. \end{aligned} \tag{8.65}$$

Hence, it suffices to show that the second result of part (a) holds with $\mathcal{W}_N(\beta_0)$ replaced by $NG(S_N(\theta_0))'G(S_N(\theta_0))$. By the same argument as in the previous paragraph, $N^{1/2}G(S_N(\theta_0))$ has a multivariate Edgeworth expansion with remainder $o(N^{-a})$ uniform in $\theta_0 \in \Theta_1$, when $N^{1/2}G(S_N(\theta_0))$ corresponds to $H_N(\widehat{\theta}_N, \beta_0)$. This Edgeworth expansion, coupled with Theorem 1 and Remark 2.2 of Chandra and Ghosh (1979), yields an Edgeworth expansion for $NG(S_N(\theta_0))'G(S_N(\theta_0))$ equal to that given for $\mathcal{W}_N(\beta_0)$ in the Lemma.

The first result of part (b) follows from Lemma 2 with

$$\begin{aligned} \lambda_N(\theta_0) &= \sup_{z \in R} |P^*_{\theta_0}(T^*_N(\theta_{0,r}) \leq z) - [1 + \sum_{i=1}^{2a} N^{-i/2}\pi_{Ti}(\delta, \nu_{N,a}(\theta_0))]\Phi(z)| \\ &= \sup_{z \in R} |P_{\theta_0}(T_N(\theta_{0,r}) \leq z) - [1 + \sum_{i=1}^{2a} N^{-i/2}\pi_{Ti}(\delta, \nu_{N,a}(\theta_0))]\Phi(z)|. \end{aligned} \tag{8.66}$$

The first condition of Lemma 2 holds by Lemma 5 and the second condition of Lemma 2 holds by part (a) of the present Lemma. The proof of the second result of part (b) is analogous. □

Footnotes

[1] The author thanks Carol Copeland for proofreading the manuscript. The author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277.

[2] This specification of the log likelihood does not utilize the first κ observations except as conditioning variables.

[3] The r-th element of θ_N^* is denoted $(\theta_N^*)_r$, rather than $\theta_{N,r}^*$, to distinguish it from the k-step bootstrap estimator, $\theta_{N,k}^*$ defined in Section 6.

[4] The latter results only require strong mixing coefficients that decline polynomially fast. In this case, it is useful to weaken the conditions on the mixing numbers in Assumption 1(b) to $\sum_{m=1}^{\infty}(m+1)^{\lambda/2-1}\alpha^{\delta/(\lambda+\delta)}(m) < \infty$ for some $\lambda > \max\{2a, 2\}$ and some $\delta > 0$, where $\alpha(m) = \sup_{\theta_0 \in \Theta_1} \alpha(m, \theta_0)$. This weakening is possible because one can establish the results of Lemma 3(a) and (b) in the Appendix using the given condition and results of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30).

[5] Stationarity of an AR(2) process with AR parameters (ρ_1, ρ_2) requires that (i) $-1 < \rho_2 < 1$, (ii) $\rho_1 + \rho_2 < 1$, and (iii) $\rho_2 - \rho_1 < 1$. To ensure that the parametric bootstrap distribution of the AR(2) process is stationary, we adjust the LS estimators $(\widehat{\rho}_1, \widehat{\rho}_2)$ (only when generating bootstrap samples and not in the expressions for the CIs given in (3.3) and (3.4)) so that they necessarily satisfy the stationarity conditions. In particular, the parametric bootstrap distribution is based on the estimators $(\widetilde{\rho}_1, \widetilde{\rho}_2)$, where $\widetilde{\rho}_2 = sgn(\widehat{\rho}_2)\min\{|\widehat{\rho}_2|, .98\}$ and $\widetilde{\rho}_1 = 1(\widehat{\rho}_1 \geq 0)\min\{\widehat{\rho}_1, .98 - \widetilde{\rho}_2\} + 1(\widehat{\rho}_1 < 0)\min\{\widehat{\rho}_1, \widetilde{\rho}_2 - .98\}$. These alterations have no effect on the asymptotic properties of the bootstrap CIs (for the true parameter values that we consider) because $\widetilde{\rho}_1 = \widehat{\rho}_1$ and $\widetilde{\rho}_2 = \widehat{\rho}_2$ with probability that goes to one at a sufficiently fast rate as $N \to \infty$. In fact, these adjustments very rarely come into play in the simulations and, hence, have no noticeable impact on the results.

References

Andrews, D. W. K. (2001): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap," *Econometrica,* 69, forthcoming. Available at www.cowles.econ.yale.edu.

Andrews, D. W. K. and H.-Y. Chen (1994): "Approximately Median-unbiased Estimation of Autoregressive Models," *Journal of Business and Economic Statistics,* 12, 187–204.

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

———— (1987): "Some Aspects of Edgeworth Expansions in Statistics and Probability," in *New Perspectives in Theoretical and Applied Statistics*, ed. by M. L. Puri, J. P. Vilaploma, and W. Wertz. New York: Wiley, 157–170.

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

Bose, A. (1988): "Edgeworth Correction by Bootstrap in Autoregression," *Annals of Statistics,* 16, 1709–1722.

Bühlmann, P. (1997): "Sieve Bootstrap for Time Series," *Bernoulli,* 3, 123–148.

————(1998): "Sieve Bootstrap for Smoothing in Nonstationary Time Series," *Annals of Statistics,* 26, 48–83.

Carlstein, E. (1986): "The Use of Subseries Methods for Estimating the Variance of a General Statistic from a Stationary Time Series," *Annals of Statistics,* 14, 1171–1179.

Chandra, T. K. and J. K. Ghosh (1979): "Valid Asymptotic Expansions for the Likelihood Ratio Statistic and Other Perturbed Chi-square Variables," *Sankhya*, 41, Series A, 22–47.

Chang, Y. and J. Y. Park (1999): "A Sieve Bootstrap for the Test of a Unit Root," unpublished manuscript, Department of Economics, Rice University.

Choi, E. and P. Hall (2000): "Bootstrap Confidence Regions Computed from Autoregressions of Arbitrary Order," *Journal of the Royal Statistical Society,* Series B, 62, 461–477.

Datta, S. and W. P. McCormick (1995): "Some Continuous Edgeworth Expansions for Markov Chains with Applications to Bootstrap," *Journal of Multivariate Analysis,* 52, 83–106.

Davidson, R. and J. G. MacKinnon (1999a): "Bootstrap Testing in Nonlinear Models," *International Economic Review,* 40, 487–508.

————(1999b): "The Size Distortion of Bootstrap Tests," *Econometric Theory,* 15, 361–376.

Doukhan, P. (1995): *Mixing: Properties and Examples.* New York: Springer-Verlag.

Fisher, R. A. (1925): "Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society,* 22, 700–725.

Gonçalves, S. and H. White (2000): "Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models," unpublished manuscript, Department of Economics, University of California, San Diego.

Götze, F. and C. Hipp (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 64, 211–239.

————(1994): "Asymptotic Distribution of Statistics in Time Series," *Annals of Statistics*, 22, 2062–2088.

Götze, F. and H. R. Künsch (1996): "Second-order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics,* 24, 1914-1933.

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–Verlag.

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

Horowitz, J. L. (2001): "Bootstrap Methods for Markov Processes," unpublished manuscript, Department of Economics, Northwestern University.

Inoue, A. and L. Kilian (1999): "Bootstrapping Autoregressive Processes with Possible Unit Roots," unpublished manuscript, Department of Economics, North Carolina State University.

Inoue, A. and M. Shintani (2001): "Bootstrapping GMM Estimators for Time Series," unpublished manuscript, Department of Agricultural Economics and Resource Economics, North Carolina State University.

Janssen, P., J. Jureckova, and N. Veraverbeke (1985): "Rate of Convergence of One- and Two-step M-estimators with Applications to Maximum Likelihood and Pitman Estimators," *Annals of Statistics,* 13, 1222-1229.

Jensen, J. L. (1989): "Asymptotic Refinements for Strongly Mixing Harris Recurrent Markov Chains," *Scandinavian Journal of Statistics,* 16, 47–63.

Künsch, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.

Lahiri, S. N. (1992): "Edgeworth Correction by 'Moving Block' Bootstrap for Stationary and Nonstationary Data," *Exploring the Limits of the Bootstrap*, ed. by R. Lepage and L. Billard. New York: Wiley, 182–214.

————(1993): "Refinements in Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Annals of Probability,* 21, 791–799.

————(1996): "On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-estimators in Multiple Linear Regression Models," *Journal of Multivariate Analysis,* 56, 42-59.

————(1999): "Theoretical Comparisons of Block Bootstrap Methods," *Annals of Statistics,* 27, 386–404.

LeCam, L. (1956): "On the Asymptotic Theory of Estimation and Testing Hypotheses," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability,* 1, 129-156.

Malinovskii, V. K. (1987): "Limit Theorems for Harris Markov Chains, I" *Theory of Probability and Its Applications,* 31, 269–285.

Paparoditis, E. (1996): "Bootstrapping Autoregressive and Moving Average Parameter Estimates of Infinite Order Vector Autoregressive Processes," *Journal of Multivariate Analysis,* 57, 277-296.

Park, J. Y. (1999): "An Invariance Principle for Sieve Bootstrap in Time Series," unpublished manuscript, School of Economics, Seoul National University, Seoul, Korea.

Pfanzagl, J. (1974): "Asymptotic Optimum Estimation and Test Procedures," in *Proceedings of the Prague Symposium on Asymptotic Statistics 3–6 September 1973*, Vol. I, ed. by J. Hájek. Charles University: Prague.

Rajarshi, M. B. (1995): "Bootstrap in Markov-sequences Based on Estimates of Transition Density," *Annals of the Institute of Statistical Mathematics,* 42, 253–268.

Robinson, P. M. (1988): "The Stochastic Difference Between Econometric Statistics," *Econometrica*, 56, 531–548.

Rothenberg, T. J. and C. T. Leenders (1964): "Efficient Estimation of Simultaneous Equation Systems," *Econometrica,* 32, 57–76.

Yokoyama, R. (1980): "Moment Bounds for Stationary Mixing Sequences," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 52, 45–57.

Zvingelis, J. J. (2000): "On Bootstrap Coverage Probability with Dependent Data," unpublished manuscript, Department of Economics, University of Iowa.

TABLE I

Coverage Probabilities *etc.* of Nominal 95% Confidence Intervals for the Cumulative Impulse Response, $1/(1-\alpha)$, for AR(2) Processes, N(0, 1) Errors, and $N = 100$

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probability	Probability CI Misses Left	Probability CI Misses Right	Average Length of CI
(1.4, -.5)	.9	Delta	.802	.198	.000	11.4
		Sym Boot	.909	.091	.000	24.5
		ET Boot	.886	.058	.056	18.1
(.9, 0)	.9	Delta	.714	.286	.000	15.0
		Sym Boot	.876	.124	.000	50.3
		ET Boot	.847	.087	.067	34.8
(0, .9)	.9	Delta	.591	.409	.000	218
		Sym Boot	.822	.178	.000	4018
		ET Boot	.794	.131	.074	2599
(1.0, -.5)	.5	Delta	.920	.080	.001	1.11
		Sym Boot	.945	.055	.000	1.35
		ET Boot	.930	.033	.038	1.21
(.5, 0)	.5	Delta	.880	.121	.000	1.52
		Sym Boot	.929	.071	.000	2.19
		ET Boot	.915	.041	.045	1.81
(0, .5)	.5	Delta	.855	.145	.000	1.82
		Sym Boot	.921	.079	.000	2.99
		ET Boot	.905	.048	.046	2.36
(0, -.5)	-.5	Delta	.941	.053	.007	.215
		Sym Boot	.947	.050	.003	.223
		ET Boot	.937	.033	.030	.220
(-.5, 0)	-.5	Delta	.931	.067	.002	.301
		Sym Boot	.944	.057	.000	.336
		ET Boot	.933	.035	.033	.316
(-1.0, -.5)	-1.5	Delta	.947	.042	.011	.101
		Sym Boot	.949	.043	.008	.101
		ET Boot	.938	.031	.031	.101

TABLE II
Coverage Probabilities of Nominal 95% Confidence Intervals for α, ρ_1, and ρ_2 for AR(2) Processes, N(0, 1) Errors, and $N = 100$

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probability of Confidence Interval for α	ρ_1	ρ_2
(1.4, -.5)	.9	Delta	.926	.933	.945
		Sym Boot	.943	.947	.946
		ET Boot	.930	.946	.939
(.9, 0)	.9	Delta	.907	.930	.939
		Sym Boot	.934	.946	.944
		ET Boot	.920	.947	.936
(0, .9)	.9	Delta	.880	.908	.853
		Sym Boot	.918	.933	.916
		ET Boot	.907	.932	.912
(1.0, -.5)	.5	Delta	.943	.939	.950
		Sym Boot	.951	.946	.952
		ET Boot	.943	.946	.945
(.5, 0)	.5	Delta	.937	.937	.943
		Sym Boot	.948	.946	.948
		ET Boot	.938	.945	.943
(0, .5)	.5	Delta	.933	.934	.927
		Sym Boot	.944	.945	.944
		ET Boot	.934	.942	.937
(0, -.5)	-.5	Delta	.945	.942	.949
		Sym Boot	.948	.946	.951
		ET Boot	.945	.944	.945
(-.5, 0)	-.5	Delta	.942	.942	.944
		Sym Boot	.947	.947	.948
		ET Boot	.942	.944	.943
(-1.0, -.5)	-1.5	Delta	.945	.942	.949
		Sym Boot	.948	.947	.951
		ET Boot	.943	.944	.945

TABLE III
Coverage Probabilities of Nominal 95% Confidence Intervals for $1/(1-\alpha)$, α, ρ_1, and ρ_2 for AR(2) Processes, t-5 Errors, and $N = 100$

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probabilities of Confidence Intervals for $1/(1-\alpha)$	α	ρ_1	ρ_2	Avg Length of CI for $1/(1-\alpha)$
(1.4, -.5)	.9	Delta	.805	.917	.941	.950	11.3
		Sym Boot	.910	.943	.951	.953	24.1
		ET Boot	.890	.932	.951	.945	17.8
(.9, 0)	.9	Delta	.713	.908	.934	.947	14.9
		Sym Boot	.874	.931	.950	.952	50.4
		ET Boot	.848	.920	.952	.946	34.9
(0, .9)	.9	Delta	.592	.879	.908	.850	328
		Sym Boot	.824	.916	.932	.913	6120
		ET Boot	.794	.906	.934	.917	3916
(1.0, -.5)	.5	Delta	.914	.943	.943	.953	1.10
		Sym Boot	.938	.950	.950	.954	1.34
		ET Boot	.934	.948	.948	.949	1.20
(.5, 0)	.5	Delta	.883	.939	.942	.948	1.51
		Sym Boot	.929	.947	.949	.954	2.17
		ET Boot	.920	.940	.948	.946	1.80
(0, .5)	.5	Delta	.854	.932	.941	.934	1.81
		Sym Boot	.922	.945	.949	.948	2.96
		ET Boot	.906	.936	.946	.943	2.34
(0, -.5)	-.5	Delta	.941	.948	.947	.954	.215
		Sym Boot	.946	.952	.951	.955	.223
		ET Boot	.940	.947	.950	.949	.219
(-.5, 0)	-.5	Delta	.932	.945	.944	.944	.300
		Sym Boot	.944	.949	.950	.950	.334
		ET Boot	.935	.946	.947	.947	.315
(-1.0, -.5)	-1.5	Delta	.949	.948	.948	.950	.100
		Sym Boot	.951	.951	.951	.952	.101
		ET Boot	.941	.946	.948	.946	.101

TABLE IV

Coverage Probabilities of Nominal 95% Confidence Intervals for $1/(1-\alpha)$, α, ρ_1, and ρ_2 for AR(2) Processes, χ^2-1 Errors, and $N = 100$

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probabilities of Confidence Intervals for $1/(1-\alpha)$	α	ρ_1	ρ_2	Avg Length of CI for $1/(1-\alpha)$
(1.4, -.5)	.9	Delta	.814	.939	.952	.960	11.5
		Sym Boot	.925	.954	.962	.962	24.8
		ET Boot	.900	.941	.960	.957	18.3
		Sym RB Boot	.912	.945	.953	.955	23.1
		ET RB Boot	.889	.928	.949	.942	17.0
(.9, 0)	.9	Delta	.714	.918	.950	.949	53.8
		Sym Boot	.887	.944	.963	.955	522
		ET Boot	.862	.929	.958	.953	342
		Sym RB Boot	.870	.930	.954	.950	539
		ET RB Boot	.844	.916	.954	.948	344
(0, .9)	.9	Delta	.587	.884	.915	.860	205
		Sym Boot	.826	.923	.938	.924	3760
		ET Boot	.806	.917	.937	.923	2422
		Sym RB Boot	.812	.906	.929	.908	3619
		ET RB Boot	.788	.900	.928	.908	2340
(1.0, -.5)	.5	Delta	.933	.954	.956	.958	1.10
		Sym Boot	.954	.960	.963	.960	1.34
		ET Boot	.942	.954	.961	.955	1.20
		Sym RB Boot	.943	.951	.954	.956	1.23
		ET RB Boot	.934	.947	.951	.947	1.12
(.5, 0)	.5	Delta	.900	.953	.954	.951	1.51
		Sym Boot	.952	.961	.961	.957	2.18
		ET Boot	.936	.953	.957	.947	1.81
		Sym RB Boot	.932	.949	.952	.952	1.96
		ET RB Boot	.922	.940	.951	.944	1.65

TABLE IV (cont.)

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probabilities of Confidence Intervals for $1/(1-\alpha)$	α	ρ_1	ρ_2	Avg Length of CI for $1/(1-\alpha)$
(0, .5)	.5	Delta	.866	.951	.947	.949	1.81
		Sym Boot	.941	.961	.957	.964	2.98
		ET Boot	.925	.948	.951	.950	2.35
		Sym RB Boot	.920	.951	.949	.953	2.69
		ET RB Boot	.910	.936	.947	.941	2.14
(0, -.5)	-.5	Delta	.952	.956	.953	.952	.215
		Sym Boot	.959	.958	.957	.953	.224
		ET Boot	.947	.954	.956	.946	.220
		Sym RB Boot	.952	.953	.952	.949	.214
		ET RB Boot	.945	.952	.951	.948	.211
(-.5, 0)	-.5	Delta	.949	.955	.956	.957	.300
		Sym Boot	.961	.960	.960	.962	.334
		ET Boot	.946	.954	.957	.954	.315
		Sym RB Boot	.952	.950	.954	.953	.305
		ET RB Boot	.938	.946	.951	.948	.294
(-1.0, -.5)	-1.5	Delta	.955	.954	.952	.957	.100
		Sym Boot	.959	.956	.955	.959	.101
		ET Boot	.946	.951	.954	.954	.101
		Sym RB Boot	.951	.951	.951	.953	.098
		ET RB Boot	.944	.949	.953	.948	.097

TABLE V
Coverage Probabilities for Nominal 95% Confidence Intervals for $1/(1-\alpha)$, α, ρ_1, and ρ_2 for AR(2) Processes with $N = 50$

(ρ_1, ρ_2)	α	Type of Confidence Interval	Coverage Probabilities of Confidence Intervals for $1/(1-\alpha)$	α	ρ_1	ρ_2	Avg Length of CI for $1/(1-\alpha)$
(1.4 , -.5)	.9	Delta	.702	.902	.919	.941	17.7
		Sym Boot	.874	.930	.947	.947	93.1
		ET Boot	.839	.917	.946	.933	63.3
(.9, 0)	.9	Delta	.576	.870	.913	.935	84.0
		Sym Boot	.811	.912	.943	.947	1647
		ET Boot	.783	.909	.943	.930	1089
(0, .9)	.9	Delta	.429	.827	.874	.790	13,355
		Sym Boot	.726	.889	.914	.890	495,905
		ET Boot	.714	.897	.918	.902	322,854
(1.0, -.5)	.5	Delta	.883	.937	.927	.941	1.55
		Sym Boot	.928	.948	.945	.946	2.28
		ET Boot	.908	.934	.944	.932	1.88
(.5, 0)	.5	Delta	.830	.930	.928	.937	2.08
		Sym Boot	.913	.947	.944	.948	3.87
		ET Boot	.892	.931	.945	.935	2.94
(0, .5)	.5	Delta	.787	.919	.924	.912	2.44
		Sym Boot	.898	.940	.944	.935	5.50
		ET Boot	.871	.924	.940	.926	4.01
(0, -.5)	-.5	Delta	.932	.942	.936	.945	.307
		Sym Boot	.938	.948	.946	.948	.340
		ET Boot	.925	.940	.944	.937	.322
(-.5, 0)	-.5	Delta	.907	.936	.935	.936	.422
		Sym Boot	.929	.945	.946	.946	.533
		ET Boot	.917	.938	.941	.936	.468
(-1.0, -.5)	-1.5	Delta	.943	.939	.942	.942	.144
		Sym Boot	.946	.944	.949	.946	.148
		ET Boot	.924	.936	.946	.934	.146