HIGHER-ORDER IMPROVEMENTS OF A COMPUTATIONALLY ATTRACTIVE *k*-STEP BOOTSTRAP FOR EXTREMUM ESTIMATORS

BY

DONALD W. K. ANDREWS

COWLES FOUNDATION PAPER NO. 1031

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

2002

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

Econometrica, Vol. 70, No. 1 (January, 2002), 119–162

HIGHER-ORDER IMPROVEMENTS OF A COMPUTATIONALLY ATTRACTIVE k-STEP BOOTSTRAP FOR EXTREMUM ESTIMATORS

By Donald W. K. Andrews[1]

This paper establishes the higher-order equivalence of the k-step bootstrap, introduced recently by Davidson and MacKinnon (1999), and the standard bootstrap. The k-step bootstrap is a very attractive alternative computationally to the standard bootstrap for statistics based on nonlinear extremum estimators, such as generalized method of moment and maximum likelihood estimators. The paper also extends results of Hall and Horowitz (1996) to provide new results regarding the higher-order improvements of the standard bootstrap and the k-step bootstrap for extremum estimators (compared to procedures based on first-order asymptotics).

The results of the paper apply to Newton-Raphson (NR), default NR, line-search NR, and Gauss-Newton k-step bootstrap procedures. The results apply to the nonparametric iid bootstrap and nonoverlapping and overlapping block bootstraps. The results cover symmetric and equal-tailed two-sided t tests and confidence intervals, one-sided t tests and confidence intervals, Wald tests and confidence regions, and J tests of over-identifying restrictions.

Keywords: Block bootstrap, Edgeworth expansion, generalized method of moments estimator, k-step bootstrap, Newton-Raphson method.

1. INTRODUCTION

This paper analyzes the higher-order properties of a computationally attractive k-step bootstrap procedure for extremum estimators, such as generalized method of moments (GMM) and maximum likelihood (ML) estimators. The method was proposed first by Davidson and MacKinnon (1999). It is closely related to the one-step and k-step estimators considered by many authors, including Fisher (1925), LeCam (1956), 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 extremum estimator requires that one solve B nonlinear optimization problems to obtain B bootstrap estimators. These estimators are then used to construct bootstrap confidence intervals (CI's), 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

[1] The author is very grateful to Joel Horowitz for his helpful comments. He also thanks a co-editor, three referees, and Jong Kim, for their comments and suggestions. 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.

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, where N denotes the sample size. 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 the k-step bootstrap yields higher-order improvements over procedures based on first-order asymptotics. The results cover the nonparametric iid bootstrap and overlapping and nonoverlapping block bootstraps for time series.

This paper also provides a number of new results concerning the higher-order properties of the standard (i.e., non-k-step) nonparametric iid and block bootstraps for extremum estimators. These results extend results of Hall and Horowitz (1996) (denoted HH hereinafter).

The paper establishes that the standard and k-step bootstraps reduce the error in test rejection probability and CI coverage probability (relative to standard first-order asymptotics) by a factor of $N^{-\xi}$ for some $\xi > 0$. The error is shown to be of magnitude $o(N^{-(1+\xi)})$ for symmetric t, Wald, and J tests and corresponding CI's and $o(N^{-(1/2+\xi)})$ for equal-tailed and one-sided t tests and corresponding CI's. When the data are dependent and the block bootstrap employed, the value of ξ depends on the block length parameter γ, where the block length ℓ is proportional to N^{γ} for some $0 \leq \gamma < 1/2$. For the block length $\ell \propto N^{1/4}$, which maximizes the upper bound on ξ, ξ is bounded above by 1/4. For equal-tailed and one-sided tests and corresponding CI's, this upper bound on ξ is sharp. The source of this bound is the combination of (i) the large sample bias of the block bootstrap Edgeworth coefficients, viewed as estimators of the corresponding nonbootstrap Edgeworth coefficients, which goes to zero sufficiently fast only if $\xi < \gamma$, and (ii) the variability of the block bootstrap Edgeworth coefficients, which goes to zero sufficiently fast only if $\xi + \gamma < 1/2$. In contrast, the results of HH show that the error in test rejection probability is reduced from $O(N^{-1})$ to $o(N^{-1})$ for symmetric two-sided t tests and J tests.

When the data are iid and the standard or k-step nonparametric iid bootstrap is employed, our results establish a reduction in the error of test rejection probability and CI coverage probability (relative to standard first-order asymptotics) by a factor of N^{-1} for two-sided t tests and symmetric percentile t CI's. The magnitude of these errors is shown to be $O(N^{-2})$, which is sharp. These results use an argument of Hall (1988, Sec. 3). Note the difference between the results for the improvements due to the block bootstrap for time series, viz. $N^{-\xi}$ for all $\xi < 1/4$, and the results for the nonparametric iid bootstrap, viz., N^{-1}. The block bootstrap is much less effective than the nonparametric iid bootstrap.

The results given here allow for a great deal of flexibility in the choice of the block length parameter in the case of time series data. Specifically, if the block length is $\ell \propto N^{\gamma}$, then we just require $0 < \gamma < 1/2$. The results of HH are somewhat restrictive in this dimension and only cover the block length $\ell \propto N^{1/5}$.

The results given here apply to both the overlapping and nonoverlapping block bootstraps. The results of HH apply to the nonoverlapping block bootstrap, whereas much of the literature on the block bootstrap focuses on overlapping blocks; e.g., see Künsch (1989), Lahiri (1992, 1996), and Götze and Künsch (1996). The overlapping block bootstrap is slightly more efficient asymptotically for estimating the distribution function of a t statistic than the nonoverlapping block bootstrap; see Hall, Horowitz, and Jing (1995). The overlapping block bootstrap also is more efficient asymptotically for estimating the variance of an estimator than the nonoverlapping block bootstrap; see Lahiri (1999). Given these superior asymptotic properties of the overlapping block bootstrap, it is desirable to have results for extremum estimators that cover it.

A key assumption made throughout the paper is that the estimator moment conditions are uncorrelated beyond some finite integer $\kappa \geq 0$, which implies that the covariance matrix of the estimator can be estimated using at most κ correlation estimates. This assumption is satisfied with $\kappa = 0$ in many time series models in which the estimator moment conditions form a martingale difference sequence due to optimizing behavior by economic agents, due to inheritance of this property from a regression error term, or due to the martingale difference property of the ML score function. It also holds with $0 < \kappa < \infty$ in many models with rational expectations and/or overlapping forecast errors, such as McCallum (1979), Hansen and Hodrick (1980), Brown and Maital (1981), and Hansen and Singleton (1982). For additional references, see Hansen and Singleton (1996). This assumption is also employed in HH.

Two papers in the literature concerning the overlapping block bootstrap are Götze and Künsch (1996) and Lahiri (1996). They consider statistics that are smooth functions of sample averages and regression parameter t statistics, respectively. They allow the asymptotic variances of the statistics of interest to depend on an infinite number of correlations, which is less restrictive than the assumption employed here. On the other hand, they obtain accuracy of CI coverage probabilities only up to $o(N^{-1/2})$, whereas we obtain accuracy up to $o(N^{-(1+\xi)})$ for symmetric t, Wald, and J tests and corresponding CI's, as outlined above. Two recent papers that consider the block bootstrap for econometric models are Zvingelis (2001) and Inoue and Shintani (2000).

We note that Davidson and MacKinnon (1999) provide an argument for higher-order improvements of the k-step bootstrap, based on Robinson's (1988) stochastic difference results for k-step estimators. However, their argument is heuristic. They "simply assume that rejection probabilities differ at the same order as the order in probability of the difference between the statistics themselves." They do not provide any regularity conditions, but they point to Robinson (1988) for the type of conditions needed. Robinson (1988), however, does not deal with bootstrapping.

In this paper, we make use of a moment inequality of Yokoyama (1980, equation (4.1)) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) rather than the weaker inequality of Lahiri (1992, Lemma 5.1), which is used in HH. We rely heavily on the methods used by HH in our proofs. In turn, the methods used by HH build on those of Bhattacharya and Ghosh (1978), Chandra and Ghosh (1979), Götze and Hipp (1983, 1994), Hall (1985), Carlstein (1986), Bhattacharya (1987), and Lahiri (1992). For part of our proofs, our methods are similar to those of Robinson (1988). The methods of Robinson (1988) are related to those of Pfanzagl (1974) and utilize results from the numerical analysis literature. Our results for the nonparametric iid bootstrap utilize results in Hall (1988).

Andrews (2001b) provides results analogous to those of this paper for the parametric bootstrap. Andrews (2000) provides results on the higher-order equivalence of nonbootstrap k-step estimators and corresponding extremum estimators.

The remainder of the paper is organized as follows: Section 2 gives an outline of the main results of the paper and their proofs. Section 3 defines the extremum estimators. Section 4 defines the overlapping and nonoverlapping block bootstraps. Section 5 defines the k-step block bootstraps. Section 6 states the assumptions. Section 7 establishes the higher-order asymptotic equivalence of the k-step and standard block bootstraps. Section 8 establishes the higher-order improvements of the k-step and standard block bootstraps. An Appendix contains proofs of the results. Andrews (2001a) gives more details than are given in the Appendix for some of the proofs.

2. OUTLINE OF THE RESULTS

In this section, we provide an outline of the methods and results established in detail in the sections below. We start by discussing the usual asymptotic t tests and CI's and the higher-order improvements that can be obtained by bootstrap t tests and CI's. Then, we outline the argument establishing the higher-order improvements of k-step bootstrap t tests and CI's.

The observed sample is $\chi_N = \{X_i : i \leq N\}$. It is a stationary time series or an iid cross-section of observations that is distributed according to a probability distribution P. An extremum estimator $\hat{\theta}_N$ of a parameter $\theta \in \Theta$ is defined to minimize a criterion function $J_N(\theta)$, which depends on the sample χ_N, over Θ. For example, $J_N(\theta)$ could be a GMM or (the negative of an) ML criterion function.

We are interested in either a two-sided test of the null hypothesis $H_0 : \theta_r = \theta_{0,r}$, where θ_r is the rth element of θ, or a CI for θ_r. The t statistic for H_0 is $T_N = N^{1/2}(\hat{\theta}_{N,r} - \theta_{0,r})/(\sigma_N)_{rr}^{1/2}$, where $\hat{\theta}_{N,r}$ denotes the rth element of $\hat{\theta}_N$ and $(\sigma_N)_{rr}$ denotes an estimator of the asymptotic variance of $N^{1/2}(\hat{\theta}_{N,r} - \theta_{0,r})$. The usual t test with asymptotic significance level α rejects H_0 if $|T_N| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of a standard normal distribution. Correspondingly, the usual CI for the true parameter $\theta_{0,r}$ with asymptotic confidence level $100(1-\alpha)\%$ is $CI_N = [\hat{\theta}_{N,r} - z_{\alpha/2}(\sigma_N)_{rr}^{1/2}/N^{1/2}, \hat{\theta}_{N,r} + z_{\alpha/2}(\sigma_N)_{rr}^{1/2}/N^{1/2}]$. The error in the rejection probability and coverage probability of these procedures can be shown to be $O(N^{-1})$. That is,

$$\text{(2.1)} \qquad P(|T_N| > z_{\alpha/2}) = \alpha + O(N^{-1}) \quad \text{under } H_0 \text{ and}$$
$$P(\theta_{0,r} \in CI_N) = 1 - \alpha + O(N^{-1}) \quad \text{when } \theta_{0,r} \text{ is the true value.}$$

Bootstrap t tests and CI's are alternatives to the t tests and CI's described above. They have errors in rejection probability and coverage probability that are smaller than $O(N^{-1})$. Bootstrap procedures are based on an estimator P^* of the probability P that generates the sample χ_N. The estimator P^* depends on the sample χ_N. For example, if χ_N consists of iid random vectors X_i, each with distribution function (df) F, then P^* could be the distribution of iid random vectors each with df $\widehat{F}_N$, where $\widehat{F}_N$ is the empirical df χ_N. This is the standard nonparametric iid bootstrap. It works because $\widehat{F}_N$ is a uniformly consistent estimator of F by the Glivenko-Cantelli Theorem. For dependent data, P^* could be a block bootstrap that resamples blocks of observations in χ_N in order to mimic the time series structure of χ_N. (The block bootstrap is described in detail below.) Alternatively, P^* could be an estimator of P that utilizes an assumed parametric model.

Let $\chi_N^* = \{X_i^* : i \leq N\}$ be a sample of random vectors that are distributed according to P^* conditional on χ_N. Define $J_N^*(\theta)$ as $J_N(\theta)$ is defined, but with χ_N^* in place of χ_N; define θ_N^* to minimize $J_N^*(\theta)$ over Θ; define $(\sigma_N^*)_{rr}^{1/2}$ as $(\sigma_N)_{rr}^{1/2}$ is defined, but with χ_N^* in place of χ_N; and define the bootstrap t statistic $T_N^* = N^{1/2}(\theta_{N,r}^* - \hat{\theta}_{N,r})/(\sigma_N^*)_{rr}^{1/2}$. (Depending on the criterion function $J_N(\theta)$ and the bootstrap distribution P^*, the above definitions of $J_N^*(\theta)$ and T_N^* need to be adjusted somewhat from those just given. For example, this is true for the GMM criterion function and the block bootstrap; see below.) The distribution of T_N^* under P^* mimics that of T_N under P, provided P^* is a reasonable estimator of P. Let $z_{|T|,\alpha}^*$ denote the $1-\alpha$ quantile of $|T_N^*|$.

The bootstrap test of asymptotic significance level α rejects H_0 if $|T_N| > z_{|T|,\alpha}^*$. The symmetric percentile t bootstrap CI for $\theta_{0,r}$ of asymptotic confidence level $100(1-\alpha)\%$ is $CI_N^* = [\hat{\theta}_{N,r} - z_{|T|,\alpha}^*(\sigma_N)_{rr}^{1/2}/N^{1/2}, \hat{\theta}_{N,r} + z_{|T|,\alpha}^*(\sigma_N)_{rr}^{1/2}/N^{1/2}]$. The error in rejection probability and coverage probability of these bootstrap procedures is shown below to be smaller than that for the usual procedures. In particular, we show that

$$\text{(2.2)} \qquad P(|T_N| > z_{|T|,\alpha}^*) = \alpha + O(N^{-(1+\xi)}) \quad \text{under } H_0 \text{ and}$$
$$P(\theta_{0,r} \in CI_N^*) = 1 - \alpha + O(N^{-(1+\xi)}) \quad \text{when } \theta_{0,r} \text{ is the true value,}$$

for all $\xi \leq 1$ when the data are iid and the nonparametric iid bootstrap is used and for all $\xi < 1/4$ when the data are stationary and the block bootstrap is used (given a suitable choice of the block length parameter). (Note that $O(N^{-(1+\xi)})$ for all $\xi < 1/4$ is equivalent to $o(N^{-(1+\xi)})$ for all $\xi < 1/4$, which is the error stated in Theorem 2 below.)

The proof of (2.2) uses Edgeworth expansions to show that the df's of T_N^* and T_N are sufficiently close that

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R}|P^*(T_N^* \le z) - P(T_N \le z)| > N^{-a}\right) = 0 \tag{2.3}$$

for some $a \ge 1+\xi$, where $2a$ is an integer. In the independent case, $a = 2$; in the dependent case, $a = 3/2$. Hence, the random df $P^*(T_N^* \le z)$, which depends on χ_N, differs from the nonrandom df $P(T_N \le z)$ by a small amount, N^{-a}, except on a set with small probability N^{-a}.

Typically, the analytic calculation of $z^*_{|T|,\alpha}$ is intractable, but the simulation of samples χ_N^* with distribution P^* is easy and fast. In consequence, the bootstrap is carried out by (i) simulating a large number, B, of independent bootstrap samples $\chi_N^*(b) = \{X_i^*(b) : i \le N\}$ for $b = 1, \dots, B$, each with distribution P^*; (ii) computing the B bootstrap criterion functions $J_N^*(\theta, b)$, estimators $\theta_N^*(b)$, and t statistics $T_N^*(b)$ for the bootstrap samples $\chi_N^*(b)$ for $b = 1, \dots, B$; and (iii) approximating the population $1-\alpha$ quantile $z^*_{|T|,\alpha}$ of $|T_N^*|$ by the sample $1-\alpha$ quantile $z^*_{|T|,\alpha}(B)$ of $\{|T_N^*(b)| : b = 1, \dots, B\}$. As $B \to \infty$, $z^*_{|T|,\alpha}(B)$ converges in probability to $z^*_{|T|,\alpha}$, because a sample quantile of iid random variables converges in probability to the corresponding population quantile. Andrews and Buchinsky (2000) provide a three-step method for determining a value of B so that $z^*_{|T|,\alpha}(B)$ is close to $z^*_{|T|,\alpha}$ with high probability. Often, B needs to be in the range of 750–1000.

A computational problem with the bootstrap procedure described above is that one needs to compute the minima of B nonlinear functions, where B is fairly large. In particular, one needs to minimize $J_N^*(\theta, b)$ over $\theta \in \Theta$ for $b = 1, \dots, B$. This can be a very time consuming task unless the minimization problem is easy. To circumvent this problem, one can use the fact that θ_N^* is close to $\hat\theta_N$ with high probability (because $N^{1/2}(\theta_N^* - \hat\theta_N) = O_p(1)$) and $\hat\theta_N$ is known. One can start at $\hat\theta_N$ and take a small number k of steps, such as NR steps, towards θ_N^*. Denote the resulting k-step bootstrap estimator by $\theta^*_{N,k}$. Then, one can use $\theta^*_{N,k}$ in place of θ_N^* in forming the bootstrap t statistic.

By definition,

$$\begin{aligned} &\theta^*_{N,0} = \hat\theta_N \quad \text{and} \\ &\theta^*_{N,j} = \theta^*_{N,j-1} - (Q^*_{N,j-1})^{-1}\frac{\partial}{\partial\theta}J_N^*(\theta^*_{N,j-1}) \quad \text{for} \quad j = 1, \dots, k, \end{aligned} \tag{2.4}$$

where $Q^*_{N,j-1}$ is a matrix that depends on $\theta^*_{N,j-1}$. For NR steps, $Q^*_{N,j-1} = (\partial^2/\partial\theta\,\partial\theta')J_N^*(\theta^*_{N,j-1})$. In this case, the definition of $\theta^*_{N,k}$ is motivated by the approximation of $(\partial/\partial\theta)J_N^*(\theta)$ at the $k-1$ step by the affine function

$$A^*_{N,k-1}(\theta) = (\partial/\partial\theta)J_N^*(\theta^*_{N,k-1}) + (\partial^2/\partial\theta\,\partial\theta')J_N^*(\theta^*_{N,k-1})(\theta - \theta^*_{N,k-1}). \tag{2.5}$$

The value of θ that solves the approximate first-order conditions $A^*_{N,k-1}(\theta) = 0$ is easily seen to be $\theta^*_{N,k}$.

Let $T^*_{N,k}$ denote the bootstrap t statistic defined with the k-step estimator $\theta^*_{N,k}$ in place of θ^*_N. Let $z^*_{|T|,k,\alpha}$ denote the $1-\alpha$ quantile of $|T^*_{N,k}|$. The k-step bootstrap t test and CI are the same as the bootstrap procedures defined above, but with $z^*_{|T|,k,\alpha}$ in place of $z^*_{|T|,\alpha}$. The computational advantage of the k-step procedures is that the approximation of $z^*_{|T|,k,\alpha}$ by the $1-\alpha$ sample quantile of $\{|T^*_{N,k}(b)| : b = 1, \dots, B\}$ only requires calculation of the k-step estimators $\{\theta^*_{N,k}(b) : b = 1, \dots, B\}$, which have closed form solutions, rather than the calculation of the extremum estimators $\{\theta^*_N(b) : b = 1, \dots, B\}$, which require B nonlinear minimizations.

If $\theta^*_{N,k}$ is close enough to θ^*_N, one hopes that the higher-order improvements given in (2.2) still hold with $z^*_{|T|,k,\alpha}$ in place of $z^*_{|T|,\alpha}$. In this paper, we show that this is true. We now outline the proof for the the NR choice of $Q^*_{N,j-1}$. In this case, we show that it suffices to have the number of steps $k \geq 3$.

The higher-order improvements of the k-step bootstrap procedures can be shown by establishing that for a as in (2.3) and for $k \geq 3$,

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R} |P^*(T^*_{N,k} \leq z) - P^*(T^*_N \leq z)| > N^{-a} \right) = 0, \tag{2.6}$$

because this implies that (2.3) holds with $T^*_{N,k}$ in place of T^*_N. First, we show that the df of T^*_N possesses a well-behaved Edgeworth expansion with remainder of order $o(N^{-a})$ conditional on χ_N except possibly on a set with P-probability $o(N^{-a})$. In consequence, a small change in z yields a small change in $P^*(T^*_N \leq z)$. This is used to show that (2.6) holds provided $T^*_{N,k}$ and T^*_N are close in the sense that

$$\lim_{N\to\infty} N^a P(P^*(|T^*_{N,k} - T^*_N| > \vartheta_N) > N^{-a}) = 0 \tag{2.7}$$

for some constants $\vartheta_N = o(N^{-a})$. Equation (2.7) is established by showing that

$$\lim_{N\to\infty} N^a P(P^*(N^{1/2}\|\theta^*_{N,k} - \theta^*_N\| > \vartheta_N) > N^{-a}) = 0 \tag{2.8}$$

and then extending this result to obtain (2.7). Hence, we focus on the proof of (2.8).

We show that (2.8) holds for the NR choice of $Q^*_{N,j-1}$ with $\vartheta_N = N^{1/2}N^{-2^k c}$, where c is a constant in $(0, 1/2)$. This corresponds to quadratic convergence of $\theta^*_{N,k}$ to θ^*_N as the number of steps k increases, which is very fast. For this definition of ϑ_N, $\vartheta_N = o(N^{-a})$ provided $2^k c \geq a + 1/2$. For $a = 2$ or $a = 3/2$ and c close to $1/2$, the latter holds provided $k \geq 3$, as stated above. For a different choice of $Q^*_{N,j-1}$, such as the GN choice, the convergence of $\theta^*_{N,k}$ to θ^*_N may be slower and k may need to be larger to obtain the desired higher-order improvements.

The first step in showing (2.8) is to show that θ^*_N is close to $\hat{\theta}_N$ with high probability. Using Taylor expansions and good moment inequalities for sums of strong mixing rv's, we show that for all $c < 1/2$ and all $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|\theta^*_N - \hat{\theta}_N\| > N^{-c}\varepsilon\big) > N^{-a}\big) = 0. \tag{2.9}$$

This result relies on sufficient smoothness of $J_N(\theta)$ and sufficient moment conditions on terms that arise in the Taylor expansion of $J_N(\theta)$ about θ_0.

Next, we show that the difference between $\theta^*_{N,k}$ and θ^*_N depends on the difference between $(\partial/\partial\theta)J^*_N(\theta)$ and the affine approximation $A^*_{N,k-1}(\theta)$ at $\theta=\theta^*_N$ (defined in (2.5)) and that the latter is a quadratic function of the difference between $\theta^*_{N,k-1}$ and θ^*_N. Our proof parallels the standard proof in the numerical analysis literature of the quadratic convergence of the NR algorithm; e.g., see Dennis and Schnabel (1983, Sec. 5.2). For notational simplicity, let $\nabla^2 J^*_{N,k-1}$ denote $(\partial^2/\partial\theta\,\partial\theta')J^*_N(\theta^*_{N,k-1})$. By the definition of $\theta^*_{N,k}$,

$$\begin{aligned}
\text{(2.10)}\qquad \theta^*_{N,k}-\theta^*_N &= \theta^*_{N,k-1}-(\nabla^2 J^*_{N,k-1})^{-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,k-1})-\theta^*_N\\
&=(\nabla^2 J^*_{N,k-1})^{-1}\bigg(\frac{\partial}{\partial\theta}J^*_N(\theta^*_N)-\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,k-1})\\
&\qquad -\nabla^2 J^*_{N,k-1}(\theta^*_N-\theta^*_{N,k-1})\bigg)\\
&=(\nabla^2 J^*_{N,k-1})^{-1}\bigg(\frac{\partial}{\partial\theta}J^*_N(\theta^*_N)-A^*_{N,k-1}(\theta^*_N)\bigg),
\end{aligned}$$

where the second equality holds because $(\partial/\partial\theta)J^*_N(\theta^*_N)=0$ with P^*-probability $1-o(N^{-a})$ on a set with P-probability $1-o(N^{-a})$ by the first-order conditions for θ^*_N and (2.9). A Taylor expansion of $(\partial/\partial\theta)J^*_N(\theta^*_N)$ about $\theta^*_{N,k-1}$ gives

$$\begin{aligned}
\text{(2.11)}\qquad &\frac{\partial}{\partial\theta}J^*_N(\theta^*_N)-A^*_{N,k-1}(\theta^*_N)\\
&\quad=\bigg[(\theta^*_N-\theta^*_{N,k-1})'\frac{\partial^3}{\partial\theta_u\,\partial\theta\,\partial\theta'}J^*_N(\theta^+_{N,k-1,u})(\theta^*_N-\theta^*_{N,k-1})/2\bigg]_{\text{vec}},
\end{aligned}$$

where $[\gamma_u]_{\text{vec}}$ denotes a vector whose uth element is γ_u and $\theta^+_{N,k-1,u}$ lies between θ^*_N and $\theta^*_{N,k-1}$.

Combining (2.10) and (2.11) gives

$$\begin{aligned}
\text{(2.12)}\qquad &\|\theta^*_{N,k}-\theta^*_N\|\le \zeta^*_N\|\theta^*_{N,k-1}-\theta^*_N\|^2,\quad\text{where}\\
&\zeta^*_N=\max_{j=1,\dots,k}\|(\nabla^2 J^*_{N,j-1})^{-1})\|\cdot\sum_{u=1}^{L_\theta}\bigg\|\frac{\partial^3}{\partial\theta_u\,\partial\theta\,\partial\theta'}J^*_N(\theta^+_{N,j-1,u})/2\bigg\|.
\end{aligned}$$

We show that there exists a constant $K<\infty$ such that

$$\text{(2.13)}\qquad \lim_{N\to\infty}N^aP\big(P^*(\zeta^*_N>K)>N^{-a}\big)=0.$$

Repeated substitution into the right-hand side of (2.12) gives

$$\text{(2.14)}\qquad \|\theta^*_{N,k}-\theta^*_N\|\le(\zeta^*_N)^\phi\|\theta^*_{N,0}-\theta^*_N\|^{2^k}=(\zeta^*_N)^\phi(N^c\|\hat\theta_N-\theta^*_N\|)^{2^k}N^{-2^kc},$$

where $\phi=\sum_{j=1}^k 2^{j-1}$ and the equality uses $\theta^*_{N,0}=\hat\theta_N$.

Combining (2.9), (2.13), and (2.14) gives the result that, for $\varepsilon > 0$ sufficiently small,

$$\begin{aligned}
(2.15)\qquad & \lim_{N\to\infty} N^a P\big(P^*\big(\|\theta^*_{N,k} - \theta^*_N\| > N^{-2^k c}\big) > N^{-a}\big) \\
& \le \lim_{N\to\infty} N^a P\big(P^*\big((\zeta^*_N)^\phi (N^c \|\hat{\theta}_N - \theta^*_N\|)^{2^k c} > 1\big) > N^{-a}\big) \\
& \le \lim_{N\to\infty} N^a P\big(P^*\big(K^\phi \varepsilon^{2^k c} > 1\big) > N^{-a}\big) \\
& = 0.
\end{aligned}$$

This establishes (2.8) with $\vartheta_N = N^{1/2} N^{-2^k c}$ and implies that NR k-step bootstrap t tests and CI's have the same errors as given in (2.2) for the standard bootstrap procedures, provided $k \geq 3$. The proof of analogous results for GN k-step bootstrap procedures is similar, though more complicated, and requires k to be larger.

3. EXTREMUM ESTIMATORS AND TESTS

In this section and those that follow, we provide a rigorous and detailed treatment of the results that are outlined in the previous section as well as additional results. As much as possible, we use the same notation as HH. We consider extremum estimators that are either GMM estimators or estimators that minimize a sample average. We call the latter "minimum ρ estimators," because the sample average is taken to be $N^{-1}\sum_{i=1}^{N} \rho(X_i, \theta)$, where $X_i \in R^{L_x}$ is a random vector, $\theta \in \Theta \subset R^{L_\theta}$ is an unknown parameter, and $\rho(\cdot,\cdot)$ is a known real function. Maximum likelihood (ML), least squares (LS), and regression M estimators are examples of minimum ρ estimators. GMM estimators are based on the moment conditions $Eg(X_i, \theta_0) = 0$, where $g(\cdot,\cdot)$ is a known L_g-valued function, X_i is as above, $\theta_0 \in \Theta \subset R^{L_\theta}$ is the true unknown parameter, and $L_g \geq L_\theta$.

Minimum ρ estimators can be written as GMM estimators with $g(X_i, \theta) = (\partial/\partial\theta)\rho(X_i, \theta)$. It is useful to consider minimum ρ estimators separately, however, because the identification condition for consistency of a minimum ρ estimator requires that there be a unique minimum of $E\rho(X_i, \theta)$ over $\theta \in \Theta$, whereas the identification condition for consistency of the GMM estimator based on the first-order conditions of the minimum ρ estimator requires that there be a unique solution to the equations $E(\partial/\partial\theta)\rho(X_i, \theta) = 0$ over $\theta \in \Theta$. The latter may have multiple solutions even though the former has a unique minimum.

The observations are $\{X_i : i = 1, \ldots, n\}$. They are assumed to be from a (strictly) stationary ergodic sequence of random vectors. We assume that the true moment conditions (for a GMM or minimum ρ estimator) are uncorrelated beyond lags of length κ for some $0 \leq \kappa < \infty$. That is, $Eg(X_i, \theta_0)g(X_{i+j}, \theta_0)' = 0$ for all $j > \kappa$. In consequence, the covariance matrix estimator and the asymptotically optimal weight matrix for the GMM estimator only depend on terms of the form $g(X_i, \theta)g(X_{i+j}, \theta)'$ for $0 \leq j \leq \kappa$. This means that the covariance matrix estimator and the weight matrix can be written as sample averages, which allows us to use the Edgeworth expansion results of Götze and Hipp (1983, 1994) for sample averages of stationary dependent random vectors, as in HH. To this end, we let

$$\widetilde{X}_i = \left(X_i', X_{i+1}', \ldots, X_{i+\kappa}'\right)' \quad \text{for} \quad i = 1, \ldots, n-\kappa. \tag{3.1}$$

All of the statistics considered below can be closely approximated by sample averages of functions of the random vectors $\widetilde{X}_i$ in the sample χ_N:

$$\chi_N = \{\widetilde{X}_i : i = 1, \ldots, N\}, \tag{3.2}$$

where $N = [(n-\kappa)/\ell]\ell$ for block bootstraps with block length ℓ, $[\cdot]$ denotes the integer part of $\cdot$, and $\kappa = 0$, $\ell = 1$, and $N = n$ for the nonparametric iid bootstrap. Thus, as in HH and Götze and Künsch (1996), some observations $\widetilde{X}_i$ are dropped if $(n-\kappa)/\ell$ is not an integer to ensure that the sample size N is an integer multiple of the block length ℓ.[2]

We consider two forms of GMM estimator. The first is a one-step GMM estimator that utilizes an $L_g \times L_g$ nonrandom positive-definite symmetric weight matrix Ω. In practice, Ω is often taken to be the identity matrix I_{L_g}. The second is a two-step GMM estimator that utilizes an asymptotically optimal weight matrix. It relies on a one-step GMM estimator to define its weight matrix.

The one-step GMM estimator, $\hat{\theta}_N$, solves

$$\min_{\theta \in \Theta} J_N(\theta) = \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta)\right)' \Omega \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta)\right). \tag{3.3}$$

The two-step GMM estimator which, for economy of notation, we also denote by $\hat{\theta}_N$, solves

$$\min_{\theta \in \Theta} J_N(\theta, \tilde{\theta}_N) = \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta)\right)' \Omega_N(\tilde{\theta}_N) \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta)\right), \quad \text{where} \tag{3.4}$$

$$\Omega_N(\theta) = \overline{W}_N^{-1}(\theta),$$

$$\overline{W}_N(\theta) = N^{-1} \sum_{i=1}^{N} \left(g(X_i, \theta) g(X_i, \theta)' + \sum_{j=1}^{\kappa} H(X_i, X_{i+j}, \theta)\right),$$

$$H(X_i, X_{i+j}, \theta) = g(X_i, \theta) g(X_{i+j}, \theta)' + g(X_{i+j}, \theta) g(X_i, \theta)',$$

and $\tilde{\theta}_N$ solves (3.3).

The minimum ρ estimator, which we also denote by $\hat{\theta}_N$, solves

$$\min_{\theta \in \Theta} N^{-1} \sum_{i=1}^{N} \rho(X_i, \theta). \tag{3.5}$$

[2] For convenience, we state that limits are as $N \to \infty$ below, although, strictly speaking, they are limits as $n \to \infty$.

For this estimator, we let $g(X_i, \theta)$ denote $(\partial/\partial\theta)\rho(X_i, \theta)$. Except for consistency properties, the minimum ρ estimator can be analyzed simultaneously with the GMM estimators. The reason is that with probability that goes to one (at an appropriate rate) the solution $\hat{\theta}_N$ to the minimization problem (3.5) is an interior solution and, hence, is also a solution to the problem of minimizing a quadratic form in the first-order conditions from this problem with weight matrix given by the identity matrix, which is just the one-step GMM criterion function.

The asymptotic covariance matrix, σ, of the extremum estimator $\hat{\theta}_N$ is

$$\sigma = \begin{cases} (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.3)}, \\ (D'\Omega_0 D)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.4)}, \\ D^{-1}\Omega_0^{-1}D^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.5)}, \text{ where} \end{cases} \tag{3.6}$$

$$\Omega_0 = (E\overline{W}_N(\theta_0))^{-1} \quad \text{and} \quad D = E\frac{\partial}{\partial\theta'}g(X_i, \theta_0).$$

A consistent estimator of σ is

$$\sigma_N = \begin{cases} (D_N'\Omega D_N)^{-1}D_N'\Omega\Omega_N^{-1}(\hat{\theta}_N) & \\ \qquad \times \Omega D_N(D_N'\Omega D_N)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.3)}, \\ (D_N'\Omega_N(\hat{\theta}_N)D_N)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.4)}, \\ D_N^{-1}\Omega_N^{-1}(\hat{\theta}_N)D_N^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.5)}, \text{ where} \end{cases} \tag{3.7}$$

$$D_N = N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}g(X_i, \hat{\theta}_N).$$

Let θ_r, $\theta_{0,r}$, and $\hat{\theta}_{N,r}$ denote the rth elements of θ, θ_0, and $\hat{\theta}_N$ 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

$$T_N = N^{1/2}(\hat{\theta}_{N,r} - \theta_{0,r})/(\sigma_N)_{rr}^{1/2}. \tag{3.8}$$

Let $\eta(\theta)$ be an R^{L_η}-valued function (for some integer $L_\eta \geq 1$) that is continuously differentiable at θ_0. The Wald statistic for testing $H_0 : \eta(\theta_0) = 0$ versus $H_1 : \eta(\theta_0) \neq 0$ is

$$\mathscr{W}_N = N\eta(\hat{\theta}_N)'\left(\frac{\partial}{\partial\theta'}\eta(\hat{\theta}_N)\sigma_N\left(\frac{\partial}{\partial\theta'}\eta(\hat{\theta}_N)\right)'\right)^{-1}\eta(\hat{\theta}_N). \tag{3.9}$$

The J statistic for testing overidentifying restrictions is

$$J_N = K_N(\hat{\theta}_N)'K_N(\hat{\theta}_N), \quad \text{where} \tag{3.10}$$

$$K_N(\theta) = \Omega_N^{1/2}(\theta)N^{-1/2}\sum_{i=1}^{N}g(X_i, \theta)$$

and $\hat{\theta}_N$ is the two-step GMM estimator. Under H_0, T_N has an asymptotic $N(0, 1)$ distribution. If $L_g > L_\theta$ and the overidentifying restrictions hold, then J_N has an asymptotic chi-squared distribution with $L_g - L_\theta$ degrees of freedom. (This is not true if the one-step GMM estimator is used to define the J statistic.)

4. NONPARAMETRIC BLOCK BOOTSTRAP

We consider both the overlapping and the nonoverlapping block bootstraps. The former is often called the Künsch (1989) blocking scheme, and the latter, the Carlstein (1986) scheme, although Hall (1985) considers both of these schemes in a related context.

The observations to be bootstrapped are $\{\widetilde{X}_i : 1 \leq i \leq N\}$. Let ℓ denote the length of the blocks. We assume that $\ell \propto N^\gamma$ for some $0 \leq \gamma < 1$. For dependent data, one takes $\gamma > 0$. (Note that this is necessary even if the data are m-dependent, because the independence between the bootstrap blocks requires the number of blocks to increase more slowly than N to properly capture the m-dependence.) For the nonoverlapping block bootstrap, the first block is $\widetilde{X}_1, \ldots, \widetilde{X}_\ell$, the second block is $\widetilde{X}_{\ell+1}, \ldots, \widetilde{X}_{2\ell}$, etc. There are b different blocks, where $b\ell = N$. For the overlapping block bootstrap, the first block is $\widetilde{X}_1, \ldots, \widetilde{X}_\ell$, the second is $\widetilde{X}_2, \ldots, \widetilde{X}_{\ell+1}$, etc. There are $N-\ell+1$ different blocks.

For iid data, $\kappa = 0$, $\widetilde{X}_i = X_i$, $N = n$, the block length ℓ equals one, and $\gamma = 0$. In this case, the bootstrap is referred to as the *nonparametric iid* bootstrap.

The bootstrap is implemented by sampling b blocks randomly with replacement from either the b nonoverlapping or the $N-\ell+1$ overlapping blocks. Let $\widetilde{X}_1^*, \ldots, \widetilde{X}_N^*$ denote the bootstrap sample obtained from this sampling scheme. Note that $\widetilde{X}_1^*, \ldots, \widetilde{X}_N^*$ is comprised of b randomly selected blocks, each of length ℓ, whether the overlapping or the nonoverlapping block bootstrap is used. The difference between the two blocking schemes is in the different collection of blocks from which blocks are randomly selected.

The bootstrap one-step GMM estimator, θ_N^*, solves

$$\text{(4.1)} \qquad \min_{\theta \in \Theta} J_N^*(\theta) = \left(N^{-1}\sum_{i=1}^N g^*(X_i^*, \theta)\right)' \Omega \left(N^{-1}\sum_{i=1}^N g^*(X_i^*, \theta)\right), \quad \text{where}$$
$$g^*(X_i^*, \theta) = g(X_i^*, \theta) - E^* g(X_i^*, \hat{\theta}_N),$$

X_i^* denotes the first element of $\widetilde{X}_i^*$, and E^* denotes expectation with respect to the distribution of the bootstrap sample conditional on the original sample. For the nonoverlapping and overlapping block bootstraps, respectively, we have

$$\text{(4.2)} \qquad N^{-1}\sum_{i=1}^N E^* g(X_i^*, \theta) = N^{-1}\sum_{i=1}^N g(X_i, \theta) \quad \text{and}$$
$$N^{-1}\sum_{i=1}^N E^* g(X_i^*, \theta) = (N-\ell+1)^{-1}\sum_{i=1}^N w(i, \ell, N) g(X_i, \theta), \quad \text{where}$$
$$w(i, \ell, N) = \begin{cases} i/\ell & \text{if } i \in [1, \ell-1], \\ 1 & \text{if } i \in [\ell, N-\ell+1], \\ (N-i+1)/\ell & \text{if } i \in [N-\ell+2, N]. \end{cases}$$

The bootstrap sample moments $N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta)$ in (4.1) are recentered (by subtracting off $E^*g(X_i^*,\hat\theta_N)$) to ensure that their expectation $E^*N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta)$ equals zero when $\theta=\hat\theta_N$, which mimics the population moments $Eg(X_i,\theta)$, which equal zero when $\theta=\theta_0$. Note that recentering also appears in Shorack (1982), who considers bootstrapping robust regression estimators, as well as in HH.

The bootstrap two-step estimator, also denoted by θ_N^*, solves

$$(4.3)\qquad \min_{\theta\in\Theta} J_N^*(\theta,\tilde\theta_N^*)=\left(N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta)\right)'\Omega_N^*(\tilde\theta_N^*)\left(N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta)\right),$$

where

$$\Omega_N^*(\theta)=\overline{W}_N^*(\theta)^{-1},$$
$$\overline{W}_N^*(\theta)=N^{-1}\sum_{i=1}^N\left(g^*(X_i^*,\theta)g^*(X_i^*,\theta)'+\sum_{j=1}^{\kappa}H^*(X_i^*,X_{i+j}^*,\theta)\right),$$
$$H^*(X_i^*,X_{i+j}^*,\theta)=g^*(X_i^*,\theta)g^*(X_{i+j}^*,\theta)'+g^*(X_{i+j}^*,\theta)g^*(X_i^*,\theta)',$$

$\tilde\theta_N^*$ denotes the one-step bootstrap estimator that solves (4.1), and, with some abuse of notation, X_{i+j}^* denotes the $(j+1)$st element of $\widetilde{X}_i^*$.[3]

The bootstrap minimum ρ estimator, also denoted by θ_N^*, solves

$$(4.4)\qquad \min_{\theta\in\Theta} N^{-1}\sum_{i=1}^N(\rho(X_i^*,\theta)-E^*g(X_i^*,\hat\theta_N)'\theta),$$

where $g(\cdot,\theta)=(\partial/\partial\theta)\rho(\cdot,\theta)$. For the nonoverlapping block bootstrap, the term $E^*g(X_i^*,\hat\theta_N)'\theta$ is zero, because $E^*g(X_i^*,\hat\theta_N)=N^{-1}\sum_{i=1}^N g(X_i,\hat\theta_N)=0$, where the second equality holds by the first-order conditions for $\hat\theta_N$ using the fact that the dimensions of $g(\cdot,\cdot)$ and θ are equal. For the overlapping block bootstrap, however, $E^*g(X_i^*,\hat\theta_N)\neq N^{-1}\sum_{i=1}^N g(X_i,\hat\theta_N)=0$ and the extra term in (4.4) is nonzero. In this case, the term $E^*g(X_i^*,\hat\theta_N)'\theta$ properly recenters the minimum ρ bootstrap criterion function. It yields bootstrap population first-order conditions that equal zero at $\hat\theta_N$, as desired. That is, $E^*(\partial/\partial\theta)(N^{-1}\sum_{i=1}^N(\rho(X_i^*,\theta)-E^*g(X_i^*,\hat\theta_N)'\theta))=E^*N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta)=0$ when $\theta=\hat\theta_N$. With this recentering, the first-order conditions for θ_N^* are $N^{-1}\sum_{i=1}^N g^*(X_i^*,\theta_N^*)=0$, rather than $N^{-1}\sum_{i=1}^N g(X_i^*,\theta_N^*)=0$, which means that θ_N^* minimizes the one-step GMM bootstrap criterion function $J_N^*(\theta)$ with $g(\cdot,\theta)=(\partial/\partial\theta)\rho(\cdot,\theta)$ and arbitrary positive definite weight matrix Ω.

[3] This abuse of notation can be avoided by writing X_i^* as $X_{i,0}^*$ and X_{i+j}^* as $X_{i,j}^*$, where $\widetilde{X}_i^*=(X_{i,0}^*,\ldots,X_{i,\kappa}^*)'$ wherever they appear. To maintain consistency of notation with HH, we do not do so.

The bootstrap covariance matrix estimator is

$$\sigma_N^* = \sigma_N^*(\theta_N^*), \quad \text{where} \tag{4.5}$$

$$\sigma_N^*(\theta) = \begin{cases} (D_N^*(\theta)'\Omega D_N^*(\theta))^{-1} & \\ \quad \times D_N^*(\theta)\Omega\Omega_N^*(\theta)^{-1}\Omega D_N^*(\theta) & \\ \quad \times (D_N^*(\theta)'\Omega D_N^*(\theta))^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.3)}, \\ (D_N^*(\theta)'\Omega_N^*(\theta) D_N^*(\theta))^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.4)}, \\ D_N^*(\theta)^{-1}\Omega_N^*(\theta)^{-1}D_N^*(\theta)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.5), and} \end{cases}$$

$$D_N^*(\theta) = N^{-1}\sum_{i=1}^N \frac{\partial}{\partial\theta'} g(X_i^*, \theta).$$

The bootstrap t, Wald, and J statistics are defined using correction factors, $\tau_{N,r}$, Ξ_N, and V_N, respectively, to correct for the fact that the independence between the bootstrap blocks does not properly mimic the dependence in the original sample. See HH for further discussion. These correction factors are not used in the case where the observations are iid. The bootstrap t, Wald, and J statistics are

$$\begin{aligned} & T_N^* = \tau_{N,r} N^{1/2}((\theta_N^*)_r - \hat{\theta}_{N,r})/\sigma_N^*(\theta_N^*)_{rr}^{1/2}, \\ & \mathcal{W}_N^* = H_N^*(\theta_N^*)' H_N^*(\theta_N^*), \quad \text{and} \\ & J_N^* = K_N^*(\theta_N^*)' K_N^*(\theta_N^*), \quad \text{where} \\ & H_N^*(\theta) = \Xi_N\left(\left(\frac{\partial}{\partial\theta'}\eta(\theta)\right)\sigma_N^*(\theta)\left(\frac{\partial}{\partial\theta'}\eta(\theta)\right)'\right)^{-1/2} N^{1/2}(\eta(\theta) - \eta(\hat{\theta}_N)), \\ & K_N^*(\theta) = (V_N^+)^{1/2}\Omega_N^*(\theta)^{1/2} N^{-1/2}\sum_{i=1}^N g^*(X_i^*, \theta), \end{aligned} \tag{4.6}$$

$(\theta_N^*)_r$ denotes the rth element of θ_N^*,[4] $\sigma_N^*(\theta_N^*)_{rr}$ denotes the (r, r)th element of $\sigma_N^*(\theta_N^*)$, and V_N^+ denotes the Moore-Penrose inverse of V_N. The correction factor $\tau_{N,r}$ is defined as follows:

$$\tau_{N,r} = (\sigma_N)_{rr}^{1/2}/(\tilde{\sigma}_N)_{rr}^{1/2}, \quad \text{where} \tag{4.7}$$

$$\tilde{\sigma}_N = \begin{cases} (D_N'\Omega D_N)^{-1}D_N'\Omega\widetilde{W}_N & \\ \quad \times \Omega D_N(D_N'\Omega D_N)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.3)}, \\ (D_N'\Omega_N(\hat{\theta}_N)D_N)^{-1}D_N'\Omega_N(\hat{\theta}_N) & \\ \quad \times \widetilde{W}_N\Omega_N(\hat{\theta}_N)D_N(D_N'\Omega_N(\hat{\theta}_N)D_N)^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.4)}, \\ D_N^{-1}\widetilde{W}_N D_N{}^{-1} & \text{if } \hat{\theta}_N \text{ solves (3.5), and} \end{cases}$$

[4] The rth 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 5.

$$\widetilde{W}_N = E^* N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} g^*(X_i^*, \hat{\theta}_N) g^*(X_j^*, \hat{\theta}_N)'$$

$$= \begin{cases} N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} g^*(X_{i\ell+j}, \hat{\theta}_N) g^*(X_{i\ell+m}, \hat{\theta}_N)' \\ \qquad\qquad \text{for non-overlapping blocks,} \\ bN^{-1}(N-\ell+1)^{-1} \sum_{i=0}^{N-\ell} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} g^*(X_{i+j}, \hat{\theta}_N) g^*(X_{i+m}, \hat{\theta}_N)' \\ \qquad\qquad \text{for overlapping blocks.} \end{cases}$$

The correction factor Ξ_N is defined to be

$$\Xi_N = \left(\frac{\partial}{\partial \theta'} \eta(\hat{\theta}_N) \tilde{\sigma}_N \left(\frac{\partial}{\partial \theta'} \eta(\hat{\theta}_N) \right)' \right)^{-1/2} \left(\frac{\partial}{\partial \theta'} \eta(\hat{\theta}_N) \sigma_N \left(\frac{\partial}{\partial \theta'} \eta(\hat{\theta}_N) \right)' \right)^{1/2}. \tag{4.8}$$

The correction factor V_N is defined to be

$$V_N = M_N \overline{W}_N^{-1/2}(\hat{\theta}_N) \widetilde{W}_N \overline{W}_N^{-1/2}(\hat{\theta}_N) M_N, \quad \text{where} \tag{4.9}$$
$$M_N = I_{L_g} - \overline{W}_N^{-1/2}(\hat{\theta}_N) D_N (D_N' \overline{W}_N^{-1}(\hat{\theta}_N) D_N)^{-1} D_N' \overline{W}_N^{-1/2}(\hat{\theta}_N).$$

Let $z^*_{|T|,\alpha}$, $z^*_{T,\alpha}$, $z^*_{\mathscr{W},\alpha}$, and $z^*_{J,\alpha}$ denote the $1-\alpha$ quantiles of $|T_N^*|$, T_N^*, $\mathscr{W}_N^*$, and J_N^* respectively. To be precise, since the distributions of $|T_N^*|$ etc. are discrete, we define $z^*_{|T|,\alpha}$ to be a value that minimizes $|P^*(|T_N^*| \le z) - (1-\alpha)|$ over $z \in R$. The precise definitions of $z^*_{T,\alpha}$, $z^*_{\mathscr{W},\alpha}$, and $z^*_{J,\alpha}$ are analogous.

Each of the following tests is of asymptotic significance level α. The symmetric two-sided bootstrap t test of $H_0 : \theta_r = \theta_{0,r}$ versus $H_1 : \theta_r \neq \theta_{0,r}$ rejects H_0 if $|T_N| > z^*_{|T|,\alpha}$. The equal-tailed two-sided bootstrap t test for the same hypotheses rejects H_0 if $T_N < z^*_{T,1-\alpha/2}$ or $T_N > z^*_{T,\alpha/2}$. The one-sided bootstrap t test of $H_0 : \theta_r \le \theta_{0,r}$ versus $H_1 : \theta_r > \theta_{0,r}$ rejects H_0 if $T_N > z^*_{T,\alpha}$. The bootstrap Wald test of $H_0 : \eta(\theta_0) = 0$ versus $H_1 : \eta(\theta_0) \neq 0$ rejects the null hypothesis if $\mathscr{W}_N > z^*_{\mathscr{W},\alpha}$. The bootstrap J test of overidentifying restrictions rejects the null if $J_N > z^*_{J,\alpha}$.

Each of the following CI's is of asymptotic confidence level $100(1-\alpha)\%$. The symmetric two-sided bootstrap CI for $\theta_{0,r}$ is $\left[\hat{\theta}_{N,r} - z^*_{|T|,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2}, \hat{\theta}_{N,r} + z^*_{|T|,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2}\right]$. The equal-tailed two-sided bootstrap CI for $\theta_{0,r}$ is $\left[\hat{\theta}_{N,r} - z^*_{T,\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2}, \ \hat{\theta}_{N,r} + z^*_{T,1-\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2}\right]$. The upper one-sided bootstrap CI for $\theta_{0,r}$ is $\left[\hat{\theta}_{N,r} - z^*_{T,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2}, \infty\right)$. The Wald-based bootstrap confidence region for $\eta(\theta_0)$ is $\{\eta \in R^{L_\eta} : N(\eta(\hat{\theta}_N) - \eta)'((\partial\eta(\hat{\theta}_N)/\partial\theta')\sigma_N(\partial\eta(\hat{\theta}_N)/\partial\theta')')^{-1}(\eta(\hat{\theta}_N) - \eta) \le z^*_{\mathscr{W},\alpha}\}$.

5. k-STEP BOOTSTRAP

Here, we define the k-step bootstrap estimators and corresponding t, Wald, and J statistics. The k-step bootstrap estimator is denoted $\theta^*_{N,k}$. For the one-step GMM estimator for which Ω is fixed, we define recursively

$$\theta^*_{N,j} = \theta^*_{N,j-1} - (Q^*_{N,j-1})^{-1} \frac{\partial}{\partial \theta} J_N^*(\theta^*_{N,j-1}) \quad \text{for} \quad 1 \le j \le k \tag{5.1}$$

and $\theta^*_{N,0} = \hat{\theta}_N$, where $\hat{\theta}_N$ denotes the one-step GMM estimator. For two-step GMM and minimum ρ estimators, $\theta^*_{N,k}$ is defined in the same way with $(\partial/\partial\theta)J^*_N(\theta^*_{N,j-1})$ replaced by $(\partial/\partial\theta)J^*_N(\theta^*_{N,j-1}, \tilde{\theta}^*_{N,k_1})$ and $N^{-1}\sum_{i=1}^N g^*(X^*_i, \theta^*_{N,j-1})$, respectively, and with $\hat{\theta}_N$ denoting the two-step GMM estimator and the minimum ρ estimator, respectively, where the derivative is taken with respect to the first argument of $J^*_N(\cdot,\cdot)$ and $\tilde{\theta}^*_{N,k_1}$ denotes the k_1-step bootstrap one-step GMM estimator defined in (5.1). We assume that $k_1 \geq k$.

The $L_\theta \times L_\theta$ random matrix $Q^*_{N,j-1}$ depends on $\theta^*_{N,j-1}$. It determines whether the k-step estimator is a NR, default NR, line-search NR, or GN k-step 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 extremum bootstrap estimator θ^*_N to a specified accuracy than does the GN k-step estimator. The NR choice of $Q^*_{N,j-1}$ is

$$(5.2) \qquad Q^{*,NR}_{N,j-1} = \begin{cases} \dfrac{\partial^2}{\partial\theta\partial\theta'}J^*_N(\theta^*_{N,j-1}) & \text{for the one-step GMM estimator,} \\[2ex] \dfrac{\partial^2}{\partial\theta\partial\theta'}J^*_N(\theta^*_{N,j-1}, \tilde{\theta}^*_{N,k_1}) & \text{for the two-step GMM estimator,} \\[2ex] D^*_N(\theta^*_{N,j-1}) & \text{for the minimum } \rho \text{ estimator, where} \end{cases}$$

$$D^*_N(\theta^*_{N,j-1}) = N^{-1}\sum_{i=1}^N \frac{\partial}{\partial\theta'} g(X^*_i, \theta^*_{N,j-1}).$$

Note that the expression for $\theta^*_{N,k}$ for a minimum ρ estimator with NR matrix $Q^{*,NR}_{N,j-1}$ is just the bootstrap version of the usual one-step scoring estimator starting from $\theta^*_{N,k-1}$ in the case of the ML estimator with score function $g(x,\theta)$ $(= (\partial/\partial\theta)\rho(x,\theta))$.

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 (5.1) for which $J^*_N(\theta^*_{N,j}) \leq J^*_N(\theta^*_{N,j-1})$ for the one-step GMM estimator, but equals some other matrix otherwise. In practice, one wants this other matrix to be such that $J^*_N(\theta^*_{N,j}) < J^*_N(\theta^*_{N,j-1})$ (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.) For the two-step GMM and minimum ρ estimators, $J^*_N(\cdot)$ above is replaced by $J^*_N(\cdot, \tilde{\theta}^*_{N,k_1})$ and $\rho^*_N(\cdot)$ respectively.

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 $\theta^*_{N,j}$ via (5.1) 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 $(1/\alpha)Q^{*,NR}_{N,j-1}$ that minimizes $J^*_N(\theta^*_{N,j})$ over all $\alpha \in A$ for the one-step GMM estimator. (If the minimizing of value of α is not unique, one takes the largest minimizing value of α in A.) For the two-step GMM and the minimum ρ estimators, one replaces $J_N^*(\theta_{N,j}^*)$ by $J_N^*(\theta_{N,j}^*, \tilde{\theta}_{N,k_1}^*)$ and $\rho_N^*(\theta_{N,j}^*)$ respectively.

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

$$
(5.3) \qquad Q_{N,j-1}^{*,GN} = \begin{cases} 2D_{N,j-1}^{*\prime}\Omega D_{N,j-1}^* & \text{for the one-step GMM estimator,} \\ 2D_{N,j-1}^{*\prime}\Omega_N^*(\tilde{\theta}_{N,k_1}^*)D_{N,j-1}^* & \text{for the two-step GMM estimator,} \\ D_{N,j-1}^* & \text{for the minimum } \rho \text{ estimator,} \end{cases}
$$

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

$$
(5.4) \qquad D_{N,j-1}^* = N^{-1}\sum_{i=1}^N \Delta(\widetilde{X}_i^*, \theta_{N,j-1}^*) \in R^{L_g \times L_\theta} \quad \text{and}
$$

$$
E\Delta(\widetilde{X}_i, \theta_0) = E\frac{\partial}{\partial\theta'}g(X_i, \theta_0).
$$

The latter condition is responsible for $D_{N,j-1}^*$ being a close approximation to $D_N^*(\theta_{N,j-1}^*)$, which appears in $Q_{N,j-1}^{*,NR}$. Note that, for the one-step and two-step GMM estimators, $Q_{N,j-1}^{*,NR}$ is the sum of two terms, one of which contains $N^{-1}\sum_{i=1}^N(\partial^2/\partial\theta\,\partial\theta')\ g^*(X_i^*, \theta_{N,j-1}^*)$. The latter term is omitted in $Q_{N,j-1}^{*,GN}$. It is close to zero, because it is multiplied by the factor $N^{-1}\sum_{i=1}^N g^*(X_i^*, \theta_{N,j-1}^*)$, which is close to zero.

For an example of a GN matrix for one-step or two-step GMM estimators, consider a nonlinear instrumental variables (IV) estimator for which

$$
(5.5) \qquad g(X_i, \theta) = U(X_i, \theta)L(Z_i, \theta) \quad \text{and} \quad E(U(X_i, \theta_0)|Z_i) = 0 \quad \text{a.s.,}
$$

where $U(X_i, \theta) \in R$ is a residual, $L(Z_i, \theta) \in R^{L_g}$ is a function of some IV's Z_i, and Z_i is a subvector of X_i. In this case,

$$
(5.6) \qquad \frac{\partial}{\partial\theta'}g(X_i, \theta) = L(Z_i, \theta)\frac{\partial}{\partial\theta'}U(X_i, \theta) + U(X_i, \theta)\frac{\partial}{\partial\theta'}L(Z_i, \theta).
$$

The GN choice of $Q_{N,j-1}^*$ omits the second summand of the bootstrap version of $(\partial/\partial\theta')g(X_i, \theta)$ in $D_{N,j-1}^*$ because $EU(X_i, \theta_0)(\partial/\partial\theta')L(Z_i, \theta_0) = 0$. That is, $Q_{N,j-1}^{*,GN}$ is as in (5.3) and (5.4) with

$$
(5.7) \qquad \Delta(\widetilde{X}_i^*, \theta) = L(Z_i^*, \theta)\frac{\partial}{\partial\theta'}U(X_i^*, \theta).
$$

An example of a GN matrix $Q_{N,j-1}^{*,GN}$ for a minimum ρ estimator is the sample outer-product estimator of the bootstrap information matrix in a ML scenario. Suppose that $\rho_N(\theta)$ is a normalized negative log likelihood function and

$g(X_i, \theta) = (\partial/\partial\theta)\rho(X_i, \theta)$ is the negative score (or conditional score) function for the X_ith observation. By the information matrix equality,

$$(5.8) \qquad E\frac{\partial}{\partial\theta'}g(X_i, \theta_0) = Eg(X_i, \theta_0)g(X_i, \theta_0)'$$

when the model is correctly specified. In this case, the NR matrix $Q^{*,NR}_{N,j-1}$ is the bootstrap version of the sample analogue of the expectation on the left-hand side of (5.8): $Q^{*,NR}_{N,j-1} = N^{-1}\sum_{i=1}^{N}(\partial/\partial\theta')g(X_i^*, \theta^*_{N,j-1})$. The GN matrix $Q^{*,GN}_{N,j-1}$ is the bootstrap version of the sample analogue of the expectation on the right-hand side of (5.8). Thus, $Q^{*,GN}_{N,j-1}$ is as in (5.3) (for minimum ρ estimators) and (5.4) with

$$(5.9) \qquad \Delta(\widetilde{X}_i^*, \theta) = g(X_i^*, \theta)g(X_i^*, \theta)'.$$

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

A second example of a GN matrix for a minimum ρ estimator arises with the least squares (LS) estimator of a nonlinear regression model. The GN matrix $Q^{*,GN}_{N,j-1}$ omits a term (whose expectation at $\theta = \theta_0$ is zero) that involves the second partial derivatives of the nonlinear regression function. For brevity, we do not give details.

We define the k-step bootstrap t statistic, $T^*_{N,k}$, Wald statistic, $\mathscr{W}^*_{N,k}$, and J statistic, $J^*_{N,k}$, as in (4.6) but with $(\theta^*_N)_r$ and θ^*_N replaced by $\theta^*_{N,k,r}$ and $\theta^*_{N,k}$, respectively, where $\theta^*_{N,k,r}$ denotes the rth element of $\theta^*_{N,k}$. Let $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, $z^*_{\mathscr{W},k,\alpha}$, and $z^*_{J,k,\alpha}$ denote the $1-\alpha$ quantiles of $|T^*_{N,k}|$, $T^*_{N,k}$, $\mathscr{W}^*_{N,k}$, and $J^*_{N,k}$ respectively.

The k-step bootstrap t, Wald, and J tests and corresponding CI's and regions are defined in the same way as the standard bootstrap tests and confidence intervals (given in the last two paragraphs of Section 4), but with $z^*_{|T|,\alpha}$, $z^*_{T,\alpha}$, $z^*_{\mathscr{W},\alpha}$, and $z^*_{J,\alpha}$ replaced by $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, $z^*_{\mathscr{W},k,\alpha}$, and $z^*_{J,k,\alpha}$ respectively (the precise definitions of which are analogous to that of $z^*_{|T|,\alpha}$ given above).

6. ASSUMPTIONS

We now introduce the assumptions. They are similar to those of HH.

Let $f(\widetilde{X}_i, \theta)$ denote the vector containing the unique components of $g(X_i, \theta)$ and $g(X_i, \theta)g(X_{i+j}, \theta)'$ for $j = 0, \ldots, \kappa$, and their derivatives through order $d_1 \geq 3$ with respect to θ. Let $(\partial^j/\partial\theta^j)g(X_i, \theta)$ and $(\partial^j/\partial\theta^j)f(\widetilde{X}_i, \theta)$ denote the vectors of partial derivatives with respect to θ of order j of $g(X_i, \theta)$ and $f(\widetilde{X}_i, \theta)$, respectively.

The following assumptions apply to the one-step GMM, two-step GMM, or minimum ρ estimator.

ASSUMPTION 1: *There is a sequence of iid vectors* $\{\varepsilon_i : i = -\infty, \ldots, \infty\}$ *of dimension* $L_\varepsilon \geq L_x$ *and an* $L_x \times 1$ *function* h *such that* $X_i = h(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$. *There are constants* $K < \infty$ *and* $\xi > 0$ *such that for all* $m \geq 1$

$$E\|h(\varepsilon_i, \varepsilon_{i-1}, \ldots) - h(\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m}, 0, 0, \ldots)\| \leq K\exp(-\xi m).$$

ASSUMPTION 2: (a) *Θ is compact and θ_0 is an interior point of Θ.* (b) *Either* (i) *θ_N minimizes $J_N(\theta)$ or $J_N(\theta, \tilde{\theta}_N)$ over $\theta \in \Theta$; θ_0 is the unique solution in Θ to $Eg(X_1, \theta) = 0$; for some function $C_g(x)$, $\|g(x, \theta_1) - g(x, \theta_2)\| < C_g(x)\|\theta_1 - \theta_2\|$ for all x in the support of X_1 and all $\theta_1, \theta_2 \in \Theta$; and $EC_g^{q_1}(X_1) < \infty$ and $E\|g(X_1, \theta)\|^{q_1} < \infty$ for all $\theta \in \Theta$ for all $0 < q_1 < \infty$, or* (ii) *$\hat{\theta}_N$ minimizes $N^{-1}\sum_{i=1}^N \rho(X_i, \theta)$ over $\theta \in \Theta$ for some function $\rho(x, \theta)$ such that $(\partial/\partial\theta)\rho(x, \theta) = g(x, \theta)$ for all x in the support of X_1; θ_0 is the unique minimum of $E\rho(X_1, \theta)$ over $\theta \in \Theta$; and $E\sup_{\theta\in\Theta}\|g(X_1, \theta)\|^{q_1} < \infty$ and $E|\rho(X_1, \theta)|^{q_1} < \infty$ for all $\theta \in \Theta$ for all $0 < q_1 < \infty$.*

ASSUMPTION 3: (a) *$Eg(X_1, \theta_0)g(X_{1+j}, \theta_0)' = 0$ for all $j > \kappa$ for some $0 \le \kappa < \infty$.* (b) *Ω and Ω_0 are positive definite and D is full rank L_θ.* (c) *$g(x, \theta)$ is $d = d_1 + d_2$ times differentiable with respect to θ on N_0, some neighborhood of θ_0, for all x in the support of X_1, where $d_1 \ge 3$ and $d_2 \ge 0$.* (d) *There is a function $C_{\partial f}(\widetilde{X}_1)$ such that $\|(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta) - (\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)\| \le C_{\partial f}(\widetilde{X}_1)\|\theta - \theta_0\|$ for all $\theta \in N_0$ for all $j = 0, \ldots, d_2$.* (e) *$EC_{\partial f}^{q_2}(\widetilde{X}_1) < \infty$ and $E\|(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)\|^{q_2} \le C_f < \infty$ for all $j = 0, \ldots, d_2$ for some constant C_f (that may depend on q_2) and all $0 < q_2 < \infty$.* (f) *$f(\widetilde{X}_1, \theta_0)$ is once differentiable with respect to $\widetilde{X}_1$ with uniformly continuous first derivative.* (g) *If the Wald statistic is considered, the R^{L_η}-valued function $\eta(\cdot)$ is d_1 times continuously differentiable at θ_0 and $(\partial/\partial\theta')\eta(\theta_0)$ is full rank $L_\eta \le L_\theta$.*

ASSUMPTION 4: *There exist constants $K_1 < \infty$ and $\delta > 0$ such that for arbitrarily large $\zeta > 1$ and all integers $m \in (\delta^{-1}, N)$ and $t \in R^{\dim(f)}$ with $\delta < \|t\| < N^\zeta$,*

$$E\left|E\left(\exp\left(it'\sum_{s=1}^{2m+1} f(\widetilde{X}_s, \theta_0)\right)\middle|\{\varepsilon_j : |j - m| > K_1\}\right)\right| \le \exp(-\delta),$$

where $i = \sqrt{-1}$ here.

ASSUMPTION 5: *The function $\Delta(\cdot, \cdot)$ in* (5.4) *satisfies*

$$\begin{aligned}
&E\Delta(\widetilde{X}_i, \theta_0) = E(\partial/\partial\theta')g(X_i, \theta_0),\\
&E\|\Delta(\widetilde{X}_i, \theta_0)\|^{q_3} < \infty \quad \text{for all} \quad 0 < q_3 < \infty, \quad \text{and}\\
&E\sup_{\theta\in N_0}\|(\partial/\partial\theta_u)\Delta(\widetilde{X}_i, \theta)\|^{q_4} < \infty\\
&\qquad \text{for all } u = 1, \ldots, L_\theta \text{ and all } 0 < q_4 < \infty.
\end{aligned}$$

Assumption 1 is the same as Assumption 1 of HH.[5] The lower bounds on d_1 and d_2 in Assumption 3 are minimal bounds. The results stated below specify more stringent lower bounds that vary depending upon the result. Assumption 4 is the same as condition (4) of Götze and Hipp (1994). It reduces to the standard Cramér condition if $\{X_i : i \ge 1\}$ are iid. The moment conditions in Assumptions 2,

[5] Assumption 1 of HH is missing the expectation operator E in its statement.

3, and 5 are stronger than necessary, but lead to relatively simple results. See Andrews (2001a) for similar results to those given below, but established under weaker moment conditions.

7. HIGHER-ORDER ASYMPTOTIC EQUIVALENCE OF THE k-STEP AND STANDARD BOOTSTRAPS

The higher-order asymptotic equivalence of the k-step and standard bootstraps is established in the following theorem. This theorem holds when $\theta^*_{N,k}$ is the one-step GMM, two-step GMM, or minimum ρ k-step bootstrap estimator. The bootstrap employed may use nonoverlapping blocks or overlapping blocks. The standard nonparametric iid bootstrap is a special case of each of these.

In the following theorem, the constant $a \geq 0$ indexes the order of magnitude of the probabilities that the k-step and standard bootstrap statistics are not close. These probabilities are $o(N^{-a})$. The larger is a, the stronger are the results. On the other hand, the larger is a, the stronger are the requisite smoothness assumptions.

THEOREM 1: *Suppose Assumptions* 1–3 *hold with* $d_1 \geq 2a+2$ *and* $d_2 \geq 0$ *for some* $a > 0$; $\ell \propto N^{\gamma}$ *for some* $0 \leq \gamma < 1$; *and k is a positive integer.*

(a) *When* $Q^*_{N,j-1}$ *is the NR, default NR, or line search NR matrix, we have, for all* $0 < c < 1/2$,

$$P^*(\|\theta^*_{N,k} - \theta^*_N\| > N^{-2^k c}) < N^{-a},$$
$$P^*(|T^*_{N,k} - T^*_N| > N^{-2^k c} N^{1/2}) < N^{-a},$$
$$P^*(|\mathcal{W}^*_{N,k} - \mathcal{W}^*_N| > N^{-2^k c} N^{1/2}) < N^{-a}, \quad \textit{and}$$
$$P^*(|J^*_{N,k} - J^*_N| > N^{-2^k c} N^{1/2}) < N^{-a},$$

except if $\{\chi_N : N \geq 1\}$ *are in a sequence of sets with probability* $o(N^{-a})$. *When* $Q^*_{N,j-1}$ *is the GN matrix and Assumption* 5 *holds, each of the results above holds with* $N^{-2^k c}$ *replaced by* $N^{-(k+1)c}$.

(b) *Suppose the following additional conditions hold: k satisfies* $2^k \geq 2a+2$ *when* $Q^*_{N,j-1}$ *is the NR, default NR, or line search NR matrix*; $k \geq 2a+1$ *and Assumption* 5 *holds when* $Q^*_{N,j-1}$ *is the GN matrix*; $d_2 > 2a+1$; $0 < \gamma < 1/2$ (*where* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent*); $2a$ *is an integer*; *and Assumption* 4 *holds. Then,*

$$\sup_{z \in R^{L_\theta}} \left| P^*(N^{1/2}(\theta^*_{N,k} - \hat{\theta}_N) \leq z) - P^*(N^{1/2}(\theta^*_N - \hat{\theta}_N) \leq z) \right| < N^{-a},$$
$$\sup_{z \in R} \left| P^*(T^*_{N,k} \leq z) - P^*(T^*_N \leq z) \right| < N^{-a} \quad \textit{under } H_0,$$
$$\sup_{z \in R} \left| P^*(\mathcal{W}^*_{N,k} \leq z) - P^*(\mathcal{W}^*_N \leq z) \right| < N^{-a} \quad \textit{under } H_0,$$
$$\sup_{z \in R} \left| P^*(J^*_{N,k} \leq z) - P^*(J^*_N \leq z) \right| < N^{-a} \quad \textit{under } H_0,$$

except if $\{\chi_N : N \geq 1\}$ *are in a sequence of sets with probability* $o(N^{-a})$.

COMMENTS: 1. Another way to express the first result of part (a) of the theorem (and analogously the other results of part (a) and part (b)) is

$$\lim_{N\to\infty} N^a P(P^*(\|\theta^*_{N,k}-\theta^*_N\| > N^{-2^k c}) > N^{-a}) = 0. \tag{7.1}$$

2. In the usual terminology, part (a) gives the *stochastic differences* between the bootstrap statistics $(\theta^*_{N,k}, T^*_{N,k}, \mathcal{W}^*_{N,k}, J^*_{N,k})$ and $(\theta^*_N, T^*_N, \mathcal{W}^*_N, J^*_N)$, whereas part (b) gives *distributional approximations* of $(\theta^*_{N,k}, T^*_{N,k}, \mathcal{W}^*_{N,k}, J^*_{N,k})$ by $(\theta^*_N, T^*_N, \mathcal{W}^*_N, J^*_N)$ up to order $o(N^{-a})$ respectively. Results of the latter sort are used in higher-order efficiency comparisons of estimators and tests. For example, see Pfanzagl (1974) and Rothenberg (1984).

3. It may seem odd that the results of part (a) of the Theorem hold without requiring the block length parameter γ to be positive when the data are dependent. The explanation is simple. If the data are dependent and $\gamma = 0$, then the bootstrap estimator θ^*_N does not have a distribution that properly mimics the distribution of $\hat{\theta}_N$, but the distributions of θ^*_N and $\theta^*_{N,k}$ are still close.

8. HIGHER-ORDER IMPROVEMENTS OF THE k-STEP AND STANDARD BLOCK BOOTSTRAPS

In this section, we show that the k-step and standard block bootstrap procedures lead to higher-order improvements in test rejection probabilities and confidence interval coverage probabilities when compared to procedures based on standard first-order asymptotics.

The following Theorem shows that the k-step and standard symmetric two-sided block bootstrap t, Wald, and J tests have rejection probabilities that are correct up to $o(N^{-(1+\xi)})$ for all $\xi < 1/4$ when the block length ℓ is chosen proportional to $N^{1/4}$, which maximizes the upper bound on ξ. It shows that the k-step and standard block bootstrap equal-tailed two-sided t and one-sided t tests have rejection probabilities that are correct up to $o(N^{-(1/2+\xi)})$ for all $\xi < 1/4$ when ℓ is chosen proportional to $N^{1/4}$, which again maximizes the upper bound on ξ. For iid data, the Theorem shows that the k-step and standard nonparametric iid bootstrap symmetric two-sided t tests (which have block length $\ell = 1$) have rejection probabilities that are correct up to $O(N^{-2})$. The latter result is sharp; see Hall (1988, Sec. 3). The coverage probabilities of the corresponding CI's are correct to the same orders.

The following results hold for statistics based on one-step GMM, two-step GMM, and minimum ρ estimators.

THEOREM 2: (a) *Suppose Assumptions* 1–4 *hold with* $d_1 \geq 5$ *and* $d_2 \geq 4$; *Assumption* 5 *holds when the GN matrix is employed*; k *is an integer that satisfies* $k \geq 3$ *when the NR, default NR, or line search NR matrix is employed and* $k \geq 4$ *when the GN matrix is employed*; $0 \leq \xi < 1/2 - \gamma$; *and either* (i) $\xi < \gamma$ *and*

$0 < \gamma < 1/2$ *or* (ii) $\{X_i : i \geq 1\}$ *are independent*. *Then*, *under* $H_0 : \theta_r = \theta_{0,r}$,

$$P(|T_N| > z^*_{|T|,k,\alpha}) = \alpha + o(N^{-(1+\xi)}) \quad \textit{and}$$
$$P(|T_N| > z^*_{|T|,\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

Under $H_0 : \eta(\theta_0) = 0$,

$$P(\mathscr{W}_N > z^*_{\mathscr{W},k,\alpha}) = \alpha + o(N^{-(1+\xi)}) \quad \textit{and}$$
$$P(\mathscr{W}_N > z^*_{\mathscr{W},\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

In addition, *if* $L_g > L_\theta$, *then*

$$P(J_N > z^*_{J,k,\alpha}) = \alpha + o(N^{-(1+\xi)}) \quad \textit{and} \quad P(J_N > z^*_{J,\alpha}) = \alpha + o(N^{-(1+\xi)}).$$

(b) *Suppose Assumptions* 1–4 *hold with* $d_1 > 4$ *and* $d_2 \geq 3$; *Assumption* 5 *holds when the GN matrix is employed*; k *is an integer that satisfies* $k \geq 2$ *when the NR*, *default NR*, *or line search NR matrix is employed and* $k \geq 3$ *when the GN matrix is employed*; $0 \leq \xi < 1/2 - \gamma$; *and either* (i) $\xi < \gamma$ *and* $0 < \gamma < 1/2$ *or* (ii) $\{X_i : i \geq 1\}$ *are independent*. *Then*, *under* $H_0 : \theta_r = \theta_{0,r}$,

$$P(T_N < z^*_{T,k,\alpha/2} \textit{ or } T_N > z^*_{T,k,1-\alpha/2}) = \alpha + o(N^{-(1/2+\xi)}),$$
$$P(T_N < z^*_{T,\alpha/2} \textit{ or } T_N > z^*_{T,1-\alpha/2}) = \alpha + o(N^{-(1/2+\xi)}),$$
$$P(T_N > z^*_{T,k,\alpha}) = \alpha + o(N^{-(1/2+\xi)}), \quad \textit{and}$$
$$P(T_N > z^*_{T,\alpha}) = \alpha + o(N^{-(1/2+\xi)}).$$

(c) *Suppose* $\{X_i : i \geq 1\}$ *are iid*; *Assumptions* 1–4 *hold with* $d_1 \geq 6$ *and* $d_2 \geq 5$; *Assumption* 5 *holds when the GN matrix is employed*; k *is an integer that satisfies* $k \geq 3$ *when the NR*, *default NR*, *or line search NR matrix is employed*, *and* $k \geq 5$ *when the GN matrix is employed*; *and* $\ell = 1$; *and* $\gamma = 0$. *Then*, *under* $H_0 : \theta_r = \theta_{0,r}$,

$$P(|T_N| > z^*_{|T|,k,\alpha}) = \alpha + O(N^{-2}) \quad \textit{and} \quad P(|T_N| > z^*_{|T|,\alpha}) = \alpha + O(N^{-2}).$$

COMMENTS: 1. The errors in part (a), (b), and (c) of the Theorem when the critical values are based on standard first-order asymptotics (using the normal distribution or the chi-square distribution) are $O(N^{-1})$, $O(N^{-1/2})$, and $O(N^{-1})$ respectively. Thus, parts (a) and (b) of the Theorem show that the bootstrap critical values reduce the error in test rejection probabilities (and in CI coverage probability) relative to first-order asymptotics by a factor of at least $N^{-\xi}$. When the data are dependent, the choice of $\gamma = 1/4$ maximizes ξ subject to the requirements of the Theorem that $\xi < \gamma$ and $\xi + \gamma < 1/2$. For this choice of γ, the results of parts (a) and (b) hold for all $\xi < 1/4$. When the data are independent, one takes $\gamma = 0$ and the results of parts (a) and (b) hold for all $\xi < 1/2$.

When the data are independent, the results of part (c) of the Theorem show that the bootstrap critical values reduce the error in test rejection probability for symmetric two-sided t tests relative to first-order asymptotics by a factor of N^{-1}, as in Hall (1988).

In contrast to Theorem 2, the results of HH show that the use of the standard bootstrap in place of first-order asymptotics reduces the error in test rejection probability from $O(N^{-1})$ to $o(N^{-1})$ for the tests of part (a) and from $O(N^{-1/2})$ to $o(N^{-1/2})$ for the tests of part (b).

2. The reason that symmetric two-sided t tests, Wald tests, and J tests are correct to a higher order than equal-tailed two-sided t tests and one-sided t tests is that the $O(N^{-1/2})$ terms of the Edgeworth expansions of $|T_N|$, $\mathcal{W}_N$, and J_N are zero by a symmetry property. See Hall (1992), HH, or the proof of Theorem 2 for details.

3. The proof of Theorem 2(c) relies on the argument given in Hall (1988). Analogous results to those of Theorem 2(c) are likely to hold for Wald and J statistics in the iid case. See Hall (1992, Sec. 4.2, pp. 165–166) for related results.

4. The possibility of improving the result of Theorem 2(a) for $|T_N|$ when the data are dependent via the symmetry argument of Hall (1988), as is done in Theorem 2(c) for iid data, is unclear. At best, this would lead to an error of $O(N^{-3/2})$, because the bootstrap moments and population moments differ by at least $N^{-1/4}$ in the dependent case, rather than $N^{-1/2}$ in the independent case; see Lemma 14 in the Appendix. But, any improvement is difficult to establish for the following reason.

Hall's (1988) argument for the $O(N^{-2})$ error in the iid case relies on determining Edgeworth expansions of $T_N \pm \Delta$, where Δ denotes the difference between the exact critical value and the bootstrap critical value. This is done by establishing Cornish-Fisher expansions for these two critical values, approximating the difference of these two expansions by a linear combination of sample averages of the data, and utilizing the smooth function of sample averages approach to Edgeworth expansions to get the Edgeworth expansions for $T_N \pm \Delta$; see Hall (1988, Sec. 3; 1992, Sec. 5.3).

This method relies on the fact that the coefficients of the Cornish-Fisher expansion of the bootstrap critical value depend on bootstrap moments that are sample averages. With dependent observations, the coefficients of the Cornish-Fisher expansion of the bootstrap critical value depend on bootstrap moments, but the bootstrap moments are not sample averages. The bootstrap moments depend on terms of the form $\sum_{j=1}^{\mathcal{N}_\ell}(\sum_{i\in b_j} h(\widetilde{X}_i))^s$, where s is an integer that depends on the order of the moment, $h(\cdot)$ is some function that depends on the criterion function or its derivatives, b_j is a set of indices for the jth block, and $\mathcal{N}_\ell$ is the number of blocks. The number of indices in b_j increases with N. There are no Edgeworth expansion results in the literature that cover terms of the above type when the random variables $\{\widetilde{X}_i : i \geq 1\}$ are dependent. Thus, without new Edgeworth expansions for statistics involving terms of the above type, one cannot prove results utilizing Hall's (1988) symmetry argument.

Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, 30 Hillhouse Ave., New Haven, CT 06520-8281, U.S.A.

Manuscript received August, 1999; final revision received January, 2001.

APPENDIX A: PROOFS

In the first subsection of this Appendix, we state Lemmas 1–17 that are used in the proofs of Theorems 1 and 2. A number of these Lemmas are similar to Lemmas in HH, but in the Lemmas given here the rates of convergence to the limits are much faster. In the second subsection, we prove Theorems 1 and 2. In the third subsection, we prove Lemmas 1–17.

We use the following notation. Let $\mathcal{N}_\ell$ denote the number of different blocks of length ℓ. For nonoverlapping blocks, $\mathcal{N}_\ell = b$. For overlapping blocks, $\mathcal{N}_\ell = N-\ell+1$. Let $\{b_j : j = 1, \ldots, \mathcal{N}_\ell\}$ denote the $\mathcal{N}_\ell$ sets of indices of the observations in each of the $\mathcal{N}_\ell$ blocks. For nonoverlapping blocks, $b_1 = \{1, \ldots, \ell\}$, $b_2 = \{\ell+1, \ldots, 2\ell\}$, etc. For overlapping blocks, $b_1 = \{1, \ldots, \ell\}$, $b_2 = \{2, \ldots, \ell+1\}$, etc. For either the overlapping or the nonoverlapping block bootstrap, let $\{b_j^* : j = 1, \ldots, b\}$ denote the b iid bootstrap blocks used to construct the bootstrap sample $\widetilde{X}_1^*, \ldots, \widetilde{X}_N^*$. By definition of the block bootstrap, $\{b_j^* : j = 1, \ldots, b\}$ are iid and each b_j^* has a discrete distribution with probability $1/\mathcal{N}_\ell$ of equaling each element in $\{b_j : j = 1, \ldots, \mathcal{N}_\ell\}$.

Throughout the Appendix, C denotes a generic constant that may change from one equality or inequality to another.

9.1. *Lemmas*

LEMMA 1: *Suppose Assumption* 1 *holds.*

(a) *Let $h(\cdot)$ be a matrix-valued function that satisfies $Eh(\widetilde{X}_i) = 0$ and $E\|h(\widetilde{X}_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a/(1-2c)$ for some $c \in [0, 1/2)$ and $a \geq 0$. Then, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i)\right\| > N^{-c}\varepsilon\right) = 0.$$

(b) *Let $h(\cdot)$ be a matrix-valued function that satisfies $E\|h(\widetilde{X}_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a$ for some $a \geq 0$. Then, there exists a constant $K < \infty$ such that*

$$\lim_{N\to\infty} N^a P\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i)\right\| > K\right) = 0.$$

LEMMA 2: *Suppose Assumptions* 1–3 *hold. Define $G_1(X_i, \theta) = g(X_i, \theta) - Eg(X_1, \theta)$ and $G_2(X_i, \theta) = \rho(X_i, \theta) - E\rho(X_1, \theta)$. Then, for all $a \geq 0$ and all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P\left(\sup_{\theta\in\Theta}\left\|N^{-1}\sum_{i=1}^N G_j(X_i, \theta)\right\| > \varepsilon\right) = 0 \quad \textit{for} \quad j = 1, 2.$$

LEMMA 3: *Let $\hat{\theta}_N$ denote the one-step GMM or minimum ρ estimator. Suppose Assumptions* 1–3 *hold. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

$$\lim_{N\to\infty} N^a P(\|\hat{\theta}_N - \theta_0\| > N^{-c}) = 0.$$

LEMMA 4: *Let $\tilde{\theta}_N$ denote the one-step GMM estimator based on the weight matrix Ω. Let $\hat{\theta}_N$ denote the two-step GMM estimator based on the weight matrix $\Omega_N(\tilde{\theta}_N)$. Suppose Assumptions* 1–3 *hold. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

$$\lim_{N\to\infty} N^a P(\|\hat{\theta}_N - \theta_0\| > N^{-c}) = 0.$$

LEMMA 5: (a) *Let $\{A_N : N \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with either* (i) *uniformly bounded densities over $N > 1$ or* (ii) *an Edgeworth expansion with coefficients of order $O(1)$ and remainder of order $o(N^{-a})$ for some $a \geq 0$ (i.e., for some polynomials $\pi_i(\delta)$ in $\delta = \partial/\partial z$ whose coefficients are $O(1)$ for $i = 1, \ldots, 2a$, $\lim_{N\to\infty} N^a \sup_{z\in R^{L_A}} |P(A_N \leq z) - [1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\partial/\partial z)]\Phi_{\Sigma_N}(z)| = 0$,* *where $\Phi_{\Sigma_N}(z)$ is the distribution function of a $N(0, \Sigma_N)$ random variable and the eigenvalues of Σ_N are bounded away from 0 and ∞ for $N \geq 1$). Let $\{\xi_N : N \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with $P(\|\xi_N\| > \vartheta_N) = o(N^{-a})$ for some constants $\vartheta_N = o(N^{-a})$ and some $a \geq 0$. Then,*

$$\lim_{N\to\infty} \sup_{z \in R^{L_A}} N^a |P(A_N + \xi_N \leq z) - P(A_N \leq z)| = 0.$$

(b) *Let $\{A_N^* : N \geq 1\}$ be a sequence of $L_A \times 1$ bootstrap random vectors that possesses an Edgeworth expansion with coefficients of order $O(1)$ and remainder of order $o(N^{-a})$ that holds except if $\{\chi_N : N \geq 1\}$ are in a sequence of sets with probability $o(N^{-a})$ for some $a \geq 0$. (That is, for all $\varepsilon > 0$, $\lim_{N\to\infty} N^a P(N^a \sup_{z\in R^{L_A}} |P^*(A_N^* \leq z) - [1 + \sum_{i=1}^{[2a]} N^{-i/2} \pi_i^*(\partial/\partial z)]\Phi_{\Sigma_N^*}(z)| > \varepsilon) = 0$, where $\pi_i^*(\delta)$ are polynomials in $\delta = \partial/\partial z$ whose coefficients, $C_{N,\ell}^*$, satisfy: for all $\rho > 0$, there exists $K_\rho < \infty$ such that $\lim_{N\to\infty} N^a P(P^*(|C_{N,\ell}^*| > K_\rho) > \rho) = 0$ for all ℓ and all $i = 1, \ldots, 2a$, $\Phi_{\Sigma_N^*}(z)$ is the distribution function of a $N(0, \Sigma_N^*)$ random variable conditional on Σ_N^* and Σ_N^* is a possibly random matrix whose eigenvalues are bounded away from 0 and ∞ with probability $1 - o(N^{-a})$ as $N \to \infty$.) Let $\{\xi_N^* : N \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with $\lim_{N\to\infty} N^a P(P^*(\|\xi_N^*\| > \vartheta_N) > N^{-a}) = 0$ for some constants $\vartheta_N = o(N^{-a})$. Then,*

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R^{L_A}} |P^*(A_N^* + \xi_N^* \leq z) - P^*(A_N^* \leq z)| > N^{-a}\right) = 0.$$

LEMMA 6: *Suppose Assumption 1 holds. Let $h(\cdot)$ be a matrix-valued function that satisfies $Eh(\widetilde{X}_i) = 0$ and $E\|h(\widetilde{X}_i)\|^p < \infty$ for $p \geq 2$ and $p > 4a/(1-2c)$ for some $c \in [0, 1/2)$ and $a \geq 0$. Assume $\ell \propto N^\gamma$ for some $0 \leq \gamma \leq 1$. Then,*

(a) $$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*) - E^* h(\widetilde{X}_i^*)\right\| > N^{-c}\right) > N^{-a}\right) = 0,$$

(b) $$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N E^* h(\widetilde{X}_i^*)\right\| > N^{-c}\right) > N^{-a}\right) = 0,$$

(c) $$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*)\right\| > N^{-c}\right) > N^{-a}\right) = 0, \quad \text{and}$$

(d) $$\text{for some } K < \infty, \quad \lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N E^* h(\widetilde{X}_i^*)\right\| > K\right) > N^{-a}\right) = 0 \quad \text{and}$$

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*)\right\| > K\right) > N^{-a}\right) = 0,$$

even if $Eh(X_i) \neq 0$ and p only satisfies $p \geq 2$ and $p > 4a$ in part (d).

LEMMA 7: *Suppose Assumptions 1–3 hold and $\ell \propto N^\gamma$ for some $0 \leq \gamma \leq 1$. Define $G_1^*(X_i^*, \theta) = g(X_i^*, \theta) - E^* g(X_i^*, \theta)$ and $G_2^*(X_i^*, \theta) = \rho(X_i^*, \theta) - E^* \rho(X_i^*, \theta)$. Then, for all $a \geq 0$ and all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P\left(P^*\left(\sup_{\theta\in\Theta}\left\|N^{-1}\sum_{i=1}^N G_j^*(X_i^*, \theta)\right\| > \varepsilon\right) > N^{-a}\right) = 0 \quad \text{for} \quad j = 1, 2.$$

LEMMA 8: *Suppose Assumptions 1–3 hold and $\ell \propto N^\gamma$ for some $0 \leq \gamma \leq 1$. Let $\hat{\theta}_N^*$ denote any bootstrap estimator that satisfies: For all $\varepsilon > 0$, $\lim_{N\to\infty} N^a P\big(P^*\big(\|\hat{\theta}_N^* - \theta_0\| > \varepsilon\big) > N^{-a}\big) = 0$ for some $a \geq 0$. Then, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|\Omega_N^*(\hat{\theta}_N^*) - \Omega_0\| > \varepsilon\big) > N^{-a}\big) = 0.$$

LEMMA 9: *Let $\hat{\theta}_N$ denote the one-step GMM or minimum ρ estimator. Let θ_N^* denote the corresponding bootstrap estimator. Suppose Assumptions* 1–3 *hold and $\ell \propto N^{\gamma}$ for some $0 \leq \gamma \leq 1$. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|\theta_N^* - \hat{\theta}_N\| > N^{-c}\big) > N^{-a}\big) = 0.$$

LEMMA 10: *Let $\tilde{\theta}_N$ and $\tilde{\theta}_N^*$ denote the one-step GMM estimator and its bootstrap analogue based on the weight matrix Ω. Let $\hat{\theta}_N$ and θ_N^* denote the two-step GMM estimator and its bootstrap analogue based on the weight matrices $\Omega_N(\tilde{\theta}_N)$ and $\Omega_N^*(\tilde{\theta}_N^*)$ respectively. Suppose Assumptions* 1–3 *hold and $\ell \propto N^{\gamma}$ for some $0 \leq \gamma \leq 1$. Then, for all $c \in [0, 1/2)$ and all $a \geq 0$,*

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|\theta_N^* - \hat{\theta}_N\| > N^{-c}\big) > N^{-a}\big) = 0.$$

LEMMA 11: *Suppose Assumptions* 1–3 *hold and $\ell \propto N^{\gamma}$ for some $0 \leq \gamma \leq 1$. Let $\tilde{\theta}_N^*$ denote the bootstrap one-step GMM estimator based on the weight matrix Ω. Let $\hat{\theta}_N^*$ denote any bootstrap estimator that satisfies*: *For all $\varepsilon > 0$, $\lim_{N\to\infty} N^a P\big(P^*\big(\|\hat{\theta}_N^* - \theta_0\| > \varepsilon\big) > N^{-a}\big) = 0$. Then, for all $a \geq 0$ and all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|D_N^*(\hat{\theta}_N^*) - D\| > \varepsilon\big) > N^{-a}\big) = 0,$$

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|\frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\hat{\theta}_N^*, \tilde{\theta}_N^*) - 2D'\Omega_0 D\right\| > \varepsilon\right) > N^{-a}\right) = 0,$$

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|\frac{\partial^3}{\partial\theta^3} J_N^*(\hat{\theta}_N^*, \tilde{\theta}_N^*)\right\| > K\right) > N^{-a}\right) = 0 \quad \textit{for some} \quad K < \infty,$$

and analogous results hold for $(\partial^2/\partial\theta\,\partial\theta')J_N^*(\hat{\theta}_N^*) - 2D'\Omega D$ and $(\partial^3/\partial\theta^3)J_N^*(\hat{\theta}_N^*)$.

For any function $m(\widetilde{X}_i, \theta)$, let $m_N^*(\theta) = N^{-1}\sum_{i=1}^N (m(\widetilde{X}_i^*, \theta) - E^* m(\widetilde{X}_i^*, \theta))$.

LEMMA 12: *Suppose Assumption* 1 *holds, $m(\widetilde{X}_i, \theta)$ is differentiable with respect to θ, and $E\sup_{\theta\in N_0} \|(\partial/\partial\theta)m(\widetilde{X}_1, \theta)\|^p < \infty$ for some $p > 4a$ and $p \geq 2$ for some $a \geq 0$. Suppose $\lim_{N\to\infty} N^a P(P^*(\|\theta_N^* - \theta_0\| > \varepsilon) > N^{-a}) = 0$ for all $\varepsilon > 0$ and $\lim_{N\to\infty} N^a P(P^*(\|\theta_{N,k}^* - \theta_N^*\| > \vartheta_N) > N^{-a}) = 0$ for some sequence of constants $\{\vartheta_N : N \geq 1\}$ for which $\vartheta_N \to 0$. Then,*

$$\lim_{N\to\infty} N^a P(P^*(\|m_N^*(\theta_{N,k}^*) - m_N^*(\theta_N^*)\| > \vartheta_N) > N^{-a}) = 0.$$

We now introduce some additional notation. Let $f^*(\widetilde{X}_i^*, \theta)$ denote the vector containing the unique components of $g^*(X_i^*, \theta)$ and $g^*(X_i^*, \theta)g^*(X_{i+j}^*, \theta)'$ for all $j = 0, \ldots, \kappa$ and their derivatives with respect to θ through order d_1. Let $S_N = N^{-1}\sum_{i=1}^N f(\widetilde{X}_i, \theta_0)$, $S = ES_N$, $S_N^* = N^{-1}\sum_{i=1}^N f^*(\widetilde{X}_i^*, \hat{\theta}_N)$, and $S^* = E^* S_N^*$. Let $\widetilde{T}_N$ and $\widetilde{K}_N(\theta)$ denote T_N^* and $K_N^*(\theta)$, respectively, without the correction factors $\tau_{N,r}$ and $(V_N^+)^{1/2}$, i.e., $\widetilde{T}_N = N^{1/2}((\theta_N^*)_r - \hat{\theta}_{N,r})/\sigma_N^*(\theta_N^*)_{rr}^{1/2}$ and $\widetilde{K}_N(\theta) = \Omega_N^*(\theta)^{1/2} N^{-1/2}\sum_{i=1}^N g^*(X_i^*, \theta)$. Let $H_N(\theta) = ((\partial/\partial\theta')\eta(\theta)\sigma_N((\partial/\partial\theta')\eta(\theta))')^{-1/2} N^{1/2}\eta(\theta)$ and $\widetilde{H}_N(\theta) = ((\partial/\partial\theta')\eta(\theta)\sigma_N^*((\partial/\partial\theta')\eta(\theta))')^{-1/2} N^{1/2}(\eta(\theta) - \eta(\hat{\theta}_N))$.

LEMMA 13: *Let Δ_N and Δ_N^* denote $N^{1/2}(\hat{\theta}_N - \theta_0)$ and $N^{1/2}(\theta_N^* - \hat{\theta}_N)$, or T_N and $\widetilde{T}_N$, or $H_N(\hat{\theta}_N)$ and $\widetilde{H}_N(\theta_N^*)$, or $K_N(\hat{\theta}_N)$ and $\widetilde{K}_N(\theta_N^*)$* (*where the statistics may be defined using one-step GMM, two-step GMM, or minimum ρ estimators in each case except the last, in which case $\hat{\theta}_N$ and θ_N^* are two-step GMM estimators*). *For each definition of Δ_N and Δ_N^*, there is an infinitely differentiable function $G(\cdot)$ with $G(S) = 0$ and $G(S^*) = 0$ such that the following results hold*.

(a) *Suppose Assumptions* 1–4 *hold with $d_1 \geq 2a + 2$, where $2a$ is some nonnegative integer. Then*,

$$\lim_{N\to\infty} \sup_z N^a |P(\Delta_N \leq z) - P(N^{1/2}G(S_N) \leq z)| = 0.$$

(b) *Suppose Assumptions* 1–4 *hold with* $d_1 \geq 2a+2$, *where* $2a$ *is some nonnegative integer, and* $0<\gamma<1$ (*where* $\gamma=0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent*). *Then*,

$$\lim_{N\to\infty} N^a P\left(\sup_z |P^*(\Delta_N^* \leq z) - P^*(N^{1/2}G(S_N^*) \leq z)| > N^{-a}\right) = 0.$$

We now define the components of the Edgeworth expansions of the test statistics T_N, $\mathscr{W}_N$, and J_N, as well as their bootstrap analogues T_N^*, $\mathscr{W}_N^*$, and J_N^*. Let $\Psi_N = N^{1/2}(S_N - S)$ and $\Psi_N^* = N^{1/2}(S_N^* - S^*)$. Let $\Psi_{N,j}$ and $\Psi_{N,j}^*$ denote the jth elements of Ψ_N and Ψ_N^* respectively. Let $\nu_{N,a}$ and $\tilde{\nu}_{N,a}$ denote vectors of moments of the form $N^{\alpha(m)} E\prod_{\mu=1}^m \Psi_{N,j_\mu}$ and $N^{\alpha(m)} E^*\prod_{\mu=1}^m \Psi_{N,j_\mu}^*$, respectively, where $2 \leq m \leq 2a+2$, $\alpha(m)=0$ if m is even, and $\alpha(m)=1/2$ if m is odd. Let $\nu_a = \lim_{N\to\infty} \nu_{N,a}$. (The limit exists under Assumption 1.)

Let $\pi_i(\delta, \nu_a)$ be a polynomial in $\delta = \partial/\partial z$ whose coefficients are polynomials in the elements of ν_a and for which $\pi_i(\delta, \nu_a)\Phi(z)$ is an even function of z when i is odd and is an odd function of z when i is even for $i = 1, \ldots, 2a$, where $2a$ is an integer. The Edgeworth expansion of T_N depends on $\pi_i(\delta, \nu_a)$. In contrast, the Edgeworth expansions of $\mathscr{W}_N$ and J_N depend on $\pi_{\mathscr{W}i}(y, \nu_a)$ and $\pi_{Ji}(y, \nu_a)$, where $\pi_{\mathscr{W}i}(y, \nu_a)$ and $\pi_{Ji}(y, \nu_a)$ denote polynomial functions of y whose coefficients are polynomials in the elements of ν_a for $i = 1, \ldots, [a]$.

The Edgeworth expansion of T_N^* depends on $\pi_i(\delta, \nu_{T,N,a}^*)$, where $\nu_{T,N,a}^*$ is a vector whose jth element is of the form $[\nu_{T,N,a}^*]_j = \tau_{N,r}^{\beta_j}[\tilde{\nu}_{N,a}]_j$ and β_j is some positive integer that depends on the bootstrap moment $[\tilde{\nu}_{N,a}]_j$ being considered. The Edgeworth expansions of $\mathscr{W}_N^*$ and J_N^* depend on $\pi_{\mathscr{W}i}(y, \nu_{\mathscr{W},N,a}^*)$ and $\pi_{Ji}(y, \nu_{J,N,a}^*)$, respectively, where $\nu_{\mathscr{W},N,a}^* = \lambda_{\mathscr{W}}(\Xi_N, \tilde{\nu}_{N,a})$, $\lambda_{\mathscr{W}}(\cdot,\cdot)$ is a function that is continuously differentiable at (I_{L_η}, ν_a), $\lambda_{\mathscr{W}}(I_{L_\eta}, \nu_a) = \nu_a$, $\nu_{J,N,a}^* = \lambda_J((V_N^+)^{1/2}, \tilde{\nu}_{N,a})$, $\lambda_J(\cdot,\cdot)$ is a function that is continuously differentiable at (M_0, ν_a), and $\lambda_J(M_0, \nu_a) = \nu_a$. Here, M_0 is the projection matrix $I_{L_g} - \Omega_0^{1/2}D(D'\Omega_0 D)^{-1}D'\Omega_0^{1/2}$, which is the probability limit of the correction factor $(V_N^+)^{1/2}$. The functions $\lambda_{\mathscr{W}}(\cdot,\cdot)$ and $\lambda_J(\cdot,\cdot)$ are determined by the effect of the correction factors Ξ_N and $(V_N^+)^{1/2}$ on the Edgeworth expansions of the bootstrap Wald and J statistics respectively.

Let χ_λ^2 denote a chi-square random variable with λ degrees of freedom.

The following Lemma shows that the bootstrap moments $\tilde{\nu}_{N,a}$ are close to the population moments ν_a in large samples.

LEMMA 14: *Suppose Assumptions* 1 *and* 3 *hold with* $d_2 \geq 2a+1$ *for some* $a \geq 0$, $0 \leq \xi < 1/2-\gamma$, *and either* (i) $\xi < \gamma$ *and* $0<\gamma<1/2$ *or* (ii) $\{X_i : i \geq 1\}$ *are independent*. *Then*,

$$\lim_{N\to\infty} N^a P(\|\tilde{\nu}_{N,a} - \nu_a\| > N^{-\xi}) = 0.$$

The next Lemma shows that the main components of the correction factors $\tau_{N,r}$, Ξ_N, and $(V_N^+)^{1/2}$ are well behaved.

LEMMA 15: *Suppose Assumptions* 1–3 *hold with* $d = d_1 + d_2 \geq 2a+2$ *for some* $a \geq 0$, $0 \leq \xi < 1/2 - \gamma/2$, *and* $0 < \gamma < 1/2$ (*where* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent*). *Suppose either* $\xi < \gamma$ *or* $\sum_{j=1}^{\kappa} j(Eg_1 g_{1+j}' + Eg_{1+j}g_1') = 0$, *where* $g_i = g(X_i, \theta_0)$.

(a) *Then*,

$$\lim_{N\to\infty} N^a P(\|\widetilde{W}_N - \overline{W}_N(\hat{\theta}_N)\| > N^{-\xi}) = 0.$$

(b) *If, in addition*, $d_2 \geq 2a+1$ *and* $0 \leq \xi < 1/2-\gamma$, *then*

$$\lim_{N\to\infty} N^a P(\|\nu_{s,N,a}^* - \nu_a\| > N^{-\xi}) = 0 \quad \textit{for} \quad s = T, \mathscr{W}, J.$$

LEMMA 16: (a) *Suppose Assumptions* 1–4 *hold with* $d_1 \geq 2a+2$, *where* $2a$ *is some nonnegative integer. Then*,

$$\lim_{N\to\infty} N^a \sup_{z\in R}\left| P(T_N \leq z) - \left[1 + \sum_{i=1}^{2a} N^{-i/2}\pi_i(\delta, \nu_a)\right]\Phi(z)\right| = 0,$$

$$\lim_{N\to\infty} N^a \sup_{z\in R}\left| P(\mathscr{W}_N \leq z) - \int_{-\infty}^{z} d\left[1 + \sum_{i=1}^{[a]} N^{-i}\pi_{\mathscr{W}i}(y, \nu_a)\right] P(\chi_{L_\eta}^2 \leq y)\right| = 0,$$

and

$$\lim_{N\to\infty} N^a \sup_{z\in R}\left|P(J_N \le z) - \int_{-\infty}^{z} d\left[1+\sum_{i=1}^{[a]} N^{-i}\pi_{Ji}(y,\nu_a)\right] P(\chi^2_{L_g-L_\theta} \le y)\right| = 0.$$

(b) *Suppose Assumptions* 1–4 *hold with* $d_1 > 2a+2$ *and* $d_2 \ge 2a+1$, *where* $2a$ *is some nonnegative integer, and* $0<\gamma<1/2$ (*where* $\gamma = 0$ *is permitted if* $\{X_i : i\ge 1\}$ *are independent*). *Then*,

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R}\left|P^*(T_N^* \le z) - \left[1+\sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\delta,\nu^*_{TNa})\right]\Phi(z)\right| > N^{-a}\right) = 0,$$

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R}\left|P^*(\mathscr{W}_N^* \le z) - \int_{-\infty}^{z} d\left[1+\sum_{i=1}^{[a]} N^{-i}\pi_{\mathscr{W}i}(y,\nu^*_{\mathscr{W}Na})\right] P(\chi^2_{L_\eta} \le y)\right| > N^{-a}\right) = 0,$$

and

$$\lim_{N\to\infty} N^a P\left(\sup_{z\in R}\left|P^*(J_N^* \le z) - \int_{-\infty}^{z} d\left[1+\sum_{i=1}^{[a]} N^{-i}\pi_{Ji}(y,\nu^*_{JNa})\right] P(\chi^2_{L_g-L_\theta} \le y)\right| > N^{-a}\right) = 0.$$

(c) *Under the assumptions given, the results of part* (b) *also hold with* T_N^*, $\mathscr{W}_N^*$, *and* J_N^* *replaced by* $T^*_{N,k}$, $\mathscr{W}^*_{N,k}$, *and* $J^*_{N,k}$, *provided* k *satisfies* $2^k \ge 2a+2$ *when* $Q^*_{N,j-1}$ *is the NR, default NR, or line search NR matrix and* $k \ge 2a+1$ *when* $Q^*_{N,j-1}$ *is the GN matrix.*

LEMMA 17: (a) *Suppose Assumptions* 1–3 *hold and* $Q^*_{N,j-1}$ *is the NR, default NR, or line search NR matrix. Then, for one-step GMM estimators, for all* $a \ge 0$,

$$\lim_{N\to\infty} N^a P\left(P^*\left(Q^*_{N,j-1} \ne \frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,j-1})\right) > N^{-a}\right) = 0 \quad \textit{for} \quad j = 1,\dots,k.$$

For two-step GMM and minimum ρ *estimators, analogous results hold with* $(\partial^2/\partial\theta\,\partial\theta')J_N^*(\theta^*_{N,j-1})$ *replaced by* $(\partial^2/\partial\theta\,\partial\theta')J_N^*(\theta^*_{N,j-1},\tilde\theta^*_{N,k_1})$ *and* $D_N^*(\theta^*_{N,j-1})$ *respectively.*

(b) *Suppose Assumptions* 1–3 *and* 5 *hold and* $Q^*_{N,j-1}$ *is the GN matrix. Then, for one-step GMM estimators, for all* $c\in[0,1/2)$ *and* $a\ge 0$,

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|Q^*_{N,j-1} - \frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,j-1})\right\| > N^{-c}\right) > N^{-a}\right) = 0 \quad \textit{for} \quad j = 1,\dots,k.$$

For two-step GMM estimators and minimum ρ *estimators, analogous results hold with*

$$\frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,j-1})$$

replaced by $(\partial^2/\partial\theta\,\partial\theta')J_N^*(\theta^*_{N,j-1},\tilde\theta^*_{N,k_1})$ *and* $D_N^*(\theta^*_{N,j-1})$ *respectively.*

9.2. *Proofs of Theorems*

9.2.1. *Proof of Theorem 1*

We establish the first result of part (a) first. We treat the different choices of $Q^*_{N,j-1}$ simultaneously. To start, suppose θ_N^* is the one-step GMM estimator. A Taylor expansion about $\theta^*_{N,k-1}$ gives

$$\begin{aligned}
(9.1)\qquad 0 &= \frac{\partial}{\partial\theta} J_N^*(\theta_N^*)\\
&= \frac{\partial}{\partial\theta} J_N^*(\theta^*_{N,k-1}) + \frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,k-1})(\theta_N^* - \theta^*_{N,k-1}) + R^*_{N,k}\\
&= \frac{\partial}{\partial\theta} J_N^*(\theta^*_{N,k-1}) + Q^*_{N,k-1}(\theta^*_{N,k} - \theta^*_{N,k-1}) + Q^*_{N,k-1}(\theta^*_N - \theta^*_{N,k})\\
&\quad + \left(\frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,k-1}) - Q^*_{N,k-1}\right)(\theta_N^* - \theta^*_{N,k-1}) + R^*_{N,k}\\
&= Q^*_{N,k-1}(\theta_N^* - \theta^*_{N,k}) + \left(\frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta^*_{N,k-1}) - Q^*_{N,k-1}\right)(\theta_N^* - \theta^*_{N,k-1}) + R^*_{N,k},
\end{aligned}$$

where
$$R^*_{N,k} = \left[(\theta^*_N - \theta^*_{N,k-1})' \frac{\partial^3}{\partial\theta_r\,\partial\theta\,\partial\theta'} J^*_N(\theta^+_{N,k-1,r})(\theta^*_N - \theta^*_{N,k-1})/2\right]_{L_\theta},$$

$[\xi_r]_{L_\theta}$ denotes an L_θ vector whose rth element is ξ_r, $\theta^+_{N,k-1,r}$ lies between θ^*_N and $\theta^*_{N,k-1}$, the first equality holds with P^*-probability $1-o(N^{-a})$ on a set with P-probability $1-o(N^{-a})$ by Lemmas 4 and 9, and the fourth equality holds because $(\partial/\partial\theta)J^*_N(\theta^*_{N,k-1}) + Q^*_{N,k-1}(\theta^*_{N,k} - \theta^*_{N,k-1}) = 0$ by the definition of $\theta^*_{N,k}$.

Define ω^*_N to be the subset of the bootstrap probability space on which $(\partial^2/\partial\theta\,\partial\theta')J^*_N(\theta^*_{N,j-1}) = Q^*_{N,j-1}$ for all $j = 1,\ldots,k$ when $Q^*_{N,j-1}$ is the NR, default NR, or line search NR matrix and define ω^*_N to be the entire bootstrap probability space when $Q^*_{N,j-1}$ is the GN matrix. By Lemma 17(a), $\lim_{N\to\infty} N^a P(P^*(\omega^*_N) < 1 - N^{-a}) = 0$. Let $1_{GN} = 0$ and $\mu_N = N^{-2^k c}$ when $Q^*_{N,j-1}$ is the NR, default NR, or line search NR matrix and $1_{GN} = 1$ and $\mu_N = N^{-(k+1)c}$ when $Q^*_{N,j-1}$ is the GN matrix.

The first inequality below follows from rearranging (9.1). The second inequality holds on the set ω^*_N. We have

$$\begin{aligned}
(9.2)\qquad \|\theta^*_{N,k} - \theta^*_N\| &\le \|(Q^*_{N,k-1})^{-1} R^*_{N,k}\| \\
&\quad + \left\|(Q^*_{N,k-1})^{-1}\left(\frac{\partial^2}{\partial\theta\,\partial\theta'} J^*_N(\theta^*_{N,k-1}) - Q^*_{N,k-1}\right)(\theta^*_{N,k-1} - \theta^*_N)\right\| \\
&\le \zeta^*_N(\|\theta^*_{N,k-1} - \theta^*_N\|^2 + 1_{GN} N^{-c}\|\theta^*_{N,k-1} - \theta^*_N\|), \quad \text{where}
\end{aligned}$$

$$\begin{aligned}
\zeta^*_N = \max_{j=1,\ldots,k}\Bigg\{ &\|(Q^*_{N,j-1})^{-1}\| \cdot \sum_{r=1}^{L_\theta}\left\|\frac{\partial^3}{\partial\theta_r\,\partial\theta\,\partial\theta'} J^*_N(\theta^+_{N,j-1,r})/2\right\| \\
&+ \|(Q^*_{N,j-1})^{-1}\| \cdot 1_{GN} N^c \left\|\frac{\partial^2}{\partial\theta\,\partial\theta'} J^*_N(\theta^*_{N,j-1}) - Q^*_{N,j-1}\right\| + 1\Bigg\}.
\end{aligned}$$

As in Robinson (1988, Pf. of Thm. 5), 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

$$(9.3)\qquad C(\zeta^*_N)^\phi \|\theta^*_{N,0} - \theta^*_N\|^{2^{k-j}} N^{-jc} 1^j_{GN} \quad \text{for} \quad j = 0,\ldots,k,$$

where ϕ is a positive integer and $0^0 = 1$ by definition. To see this, consider the solution in terms of x_0 of the equation $x_k = x^2_{k-1} + \lambda x_{k-1}$. Collect all terms in powers of λ that are multiplied by the smallest number of x_0 terms. Note that the upper bound in (9.3) for each of the three NR cases is $C(\zeta^*_N)^\phi \|\theta^*_{N,0} - \theta^*_N\|^{2^k}$.

Using (9.3) and $\theta^*_{N,0} = \hat\theta_N$, an upper bound on the right-hand side of (9.2) is

$$(9.4)\qquad C(\zeta^*_N)^\phi \max_{j=0,\ldots,k}(\lambda^*_N)^{2^{k-j}} N^{-(2^{k-j}c+jc)} 1^j_{GN} \le C(\zeta^*_N)^\phi \max_{j=0,\ldots,k}(\lambda^*_N)^{2^{k-j}}\mu_N, \quad \text{where}$$

$$\lambda^*_N = N^c\|\hat\theta_N - \theta^*_N\|.$$

Lemma 9 implies that for all $\varepsilon > 0$ $\lim_{N\to\infty} N^a P(P^*(\lambda^*_N > \varepsilon) > N^{-a}) = 0$ (using the fact that the result of Lemma 9 holds for all $c \in [0, 1/2)$). In addition, by Lemmas 11 and 17 and Assumption 3(b), there exists a finite constant K such that $\lim_{N\to\infty} N^a P(P^*(\zeta^*_N > K) > N^{-a}) = 0$. Combining these results with (9.2) and (9.4) gives

$$\begin{aligned}
(9.5)\qquad &\lim_{N\to\infty} N^a P\big(P^*\big((\|\theta^*_{N,k} - \theta^*_N\| > \mu_N) \cap \omega^*_N\big) > N^{-a}\big) \\
&\le \lim_{N\to\infty} N^a P\big(P^*\big(C(\zeta^*_N)^\phi \lambda^*_N > 1\big) > N^{-a}\big) = 0.
\end{aligned}$$

This result and Lemma 17(a) combine to yield the first result of part (a) of the Theorem for the one-step GMM estimator.

For the minimum ρ estimator, the result of Lemma 3 implies that $\hat{\theta}_N$ is in the interior of Θ, $N^{-1}\sum_{i=1}^{N} g(X_i, \hat{\theta}_N) = 0$, and $\hat{\theta}_N$ minimizes $J_N(\theta)$ (defined with an arbitrary positive definite weight matrix Ω) over $\theta \in \Theta$ with probability $1 - o(N^{-a})$. In consequence, the proof for the one-step GMM estimator also covers the minimum ρ estimator.

The proof of the first result of part (a) for the case where θ_N^* is the bootstrap two-step GMM estimator is similar to that given above with $J_N^*(\theta)$ replaced by $J_N^*(\theta, \tilde{\theta}_N^*)$ or $J_N^*(\theta, \tilde{\theta}_{N,k_1}^*)$ in the appropriate place and with reference to Lemma 9 replaced by reference to Lemma 10. However, two additional terms arise on the right-hand side of (9.1) because $J_N^*(\theta, \tilde{\theta}_N^*) \neq J_N^*(\theta, \tilde{\theta}_{N,k_1}^*)$. These terms are

$$M_{1,N}^* = \left(\frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta_{N,k-1}^*, \tilde{\theta}_N^*) - \frac{\partial^2}{\partial\theta\,\partial\theta'} J_N^*(\theta_{N,k-1}^*, \tilde{\theta}_{N,k_1}^*) \right)(\theta_N^* - \theta_{N,k-1}^*) \quad \text{and} \tag{9.6}$$

$$M_{2,N}^* = \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*, \tilde{\theta}_N^*) - \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*, \tilde{\theta}_{N,k_1}^*).$$

These terms can be shown to satisfy

$$\lim_{N\to\infty} N^a P\big(P^*\big(\|M_{j,N}^*\| > \mu_N\big) > N^{-a}\big) = 0 \quad \text{for} \quad j = 1, 2. \tag{9.7}$$

In consequence, the result of part (a) of the Theorem holds for the bootstrap two-step GMM estimator.

To prove (9.7), we first show that

$$\lim_{N\to\infty} N^a P\big(P^*\big(\big\|\Omega_N^*(\tilde{\theta}_{N,k_1}^*)^{-1} - \Omega_N^*(\tilde{\theta}_N^*)^{-1}\big\| > \mu_N\big) > N^{-a}\big) = 0 \tag{9.8}$$

using Lemma 12 with $m_N^*(\theta) = \Omega_N^*(\theta)^{-1}$, $\theta_N^* = \tilde{\theta}_N^*$, $\theta_{N,k}^* = \tilde{\theta}_{N,k_1}^*$, and $\vartheta_N = \mu_N$. The conditions of Lemma 12 are verified using the result of part (a) of the Theorem for the bootstrap one-step GMM estimator, the assumption that $k_1 \geq k$, and Lemma 9. The proof of (9.7) also uses the first and second results of Lemma 11 with $\hat{\theta}_N^* = \theta_{N,k-1}^*$, where the condition on $\hat{\theta}_N^*$ holds by applying the proof of part (a) of the Theorem for the k-step bootstrap two-step GMM estimator recursively for $k = 1, 2, \ldots$. The proof of (9.7) also uses $\lim_{N\to\infty} P(P^*(\|\theta_N^* - \theta_{N,k-1}^*\| > K) > N^{-a}) = 0$ for some $0 < K < \infty$, which holds by applying the current proof recursively because $K \geq \mu_N$ for N large.

For the second result of part (a), when $\gamma > 0$, we use (9.64) of the proof of Lemma 15(b) (which guarantees that $\tau_{N,r}$ is well behaved, utilizes the condition $d_1 \geq 2a + 2$, and holds for all $0 < \gamma < 1$). Let $\widetilde{T}_N$ and $\widetilde{T}_{N,k}$ denote T_N^* and $T_{N,k}^*$, respectively, with the correction factor $\tau_{N,r}$ deleted. Let $\sigma_{k,r}^*$ and σ_r^* denote $\sigma_N^*(\theta_{N,k}^*)_{rr}$ and $\sigma_N^*(\theta_N^*)_{rr}$ respectively. For $\gamma \geq 0$, we use the following:

$$\begin{aligned} \big|\widetilde{T}_{N,k} - \widetilde{T}_N\big| \leq{}& N^{1/2}\|\theta_{N,k}^* - \theta_N^*\|/(\sigma_{k,r}^*)^{1/2} \\ &+ N^{1/2}\|\theta_N^* - \hat{\theta}_N\| \cdot \big|(\sigma_{k,r}^*)^{1/2} - (\sigma_r^*)^{1/2}\big|/(\sigma_{k,r}^*\sigma_r^*)^{1/2}. \end{aligned} \tag{9.9}$$

By (9.9) and the result concerning $\tau_{N,r}$, the second result of part (a) is implied by the first result of part (a) plus the following: There exists a $K < \infty$ and a $\delta > 0$ such that

$$\lim_{N\to\infty} N^a P(P^*(|(\sigma_{k,r}^*)^{1/2} - (\sigma_r^*)^{1/2}| > \mu_N) > N^{-a}) = 0, \tag{9.10}$$

$$\lim_{N\to\infty} N^a P(P^*(\|\theta_N^* - \hat{\theta}_N\| > K) > N^{-a}) = 0, \tag{9.11}$$

$$\lim_{N\to\infty} N^a P(P^*(\sigma_{k,r}^* < \delta) > N^{-a}) = 0, \quad \text{and} \tag{9.12}$$

$$\lim_{N\to\infty} N^a P(P^*(\sigma_r^* < \delta) > N^{-a}) = 0. \tag{9.13}$$

Equation (9.11) holds by Lemma 9 or 10. Equations (9.12) and (9.13) hold by Lemmas 8–11 and the first result of part (a) of the Theorem.

Equation (9.10) is implied by (9.12), (9.13), and

$$(9.14)\qquad \lim_{N\to\infty} N^a P(P^*(|\sigma^*_{k,r}-\sigma^*_r|>\mu_N)>N^{-a})=0$$

by applying the mean value theorem. Equation (9.14) is implied by

$$(9.15)\qquad \lim_{N\to\infty} N^a P(P^*(\|D^*_N(\theta^*_{N,k})-D^*_N(\theta^*_N)\|>\mu_N)>N^{-a})=0 \quad \text{and}$$
$$\lim_{N\to\infty} N^a P(P^*(\|\Omega^*_N(\theta^*_{N,k})^{-1}-\Omega^*_N(\theta^*_N)^{-1}\|>\mu_N)>N^{-a})=0.$$

These results hold by Lemma 12 with $\vartheta_N=\mu_N$ using the first result of part (a) of the Theorem and Assumption 3.

The third and fourth results of part (a) are established by analogous arguments; see Andrews (2001a) for details.

To establish part (b) of the Theorem, we apply Lemma 5(b) four times with $\vartheta_N=N^{1/2}\mu_N=o(N^{-a})$ (where the last equality uses the condition $2^k\geq 2a+2$ in the NR cases and $k\geq 2a+1$ in the GN case) and with (A^*_N,ξ^*_N) equal to $(N^{1/2}(\theta^*_N-\hat\theta_N), N^{1/2}(\theta^*_{N,k}-\theta^*_N))$, $(T^*_N, T^*_{N,k}-T^*_N)$, $(H^*_N(\theta^*_N), H^*_N(\theta^*_{N,k})-H^*_N(\theta^*_N))$, and $(K^*_N(\theta^*_N), K^*_N(\theta^*_{N,k})-K^*_N(\theta^*_N))$. In the third and fourth times Lemma 5(b) is applied, the result of Lemma 5(b) implies the third and fourth results of part (b) of the Theorem by a straightforward argument. The condition of Lemma 5(b) on ξ^*_N holds by part (a) of the Theorem for the first two applications of Lemma 5(b) and by analogous results given in Andrews (2001a) for the third and fourth applications. As required by Lemma 5(b), the random variables T^*_N, $H^*_N(\theta^*_N)$, and $K^*_N(\theta^*_N)$ have Edgeworth expansions with remainder $o(N^{-a})$ by Lemma 16(b) using the additional conditions on d_2, γ, and a in part (b). Lemma 16(b) does not state an Edgeworth expansion for $N^{1/2}(\theta^*_N-\hat\theta_N)$, but one can be obtained under the same assumptions and by the same argument as for T^*_N. *Q.E.D.*

9.2.2. *Proof of Theorem 2*

We establish the first result of part (a) of the Theorem first. By Theorem 1(b), Lemma 16(b), Lemma 14, and Lemma 16(a), respectively, each with $a=3/2$, we have:

$$(9.16)\qquad \lim_{N\to\infty} N^{3/2} P\left(\sup_{z\in R}|P^*(T^*_{N,k}|\leq z)-P^*(|T^*_N|\leq z)|>N^{-3/2}\right)=0,$$
$$\lim_{N\to\infty} N^{3/2} P\left(\sup_{z\in R}|P^*(|T^*_N|\leq z)-[1+N^{-1}\pi_2(\delta,\nu^*_{T,N,3/2})](\Phi(z)-\Phi(-z))|>N^{-3/2}\right)=0,$$
$$\lim_{N\to\infty} N^{3/2} P\left(\sup_{z\in R}|(\pi_2(\delta,\nu^*_{T,N,3/2})-\pi_2(\delta,\nu_{3/2}))(\Phi(z)-\Phi(-z))|>N^{-\xi}\right)=0,\quad \text{and}$$
$$\lim_{N\to\infty} N^{3/2}\sup_{z\in R}|P(|T_N|\leq z)-(1+N^{-1}\pi_2(\delta,\nu_{3/2}))(\Phi(z)-\Phi(-z))|=0,$$

using the evenness of $\pi_j(\delta,\nu^*_{T,N,3/2})\Phi(z)$ and $\pi_j(\delta,\nu_{3/2})\Phi(z)$ in z for $j=1,3$ in the second and fourth results respectively. The results of (9.16) combine to give

$$(9.17)\qquad \lim_{N\to\infty} N^{3/2} P\left(\sup_{z\in R}|P^*(|T^*_{N,k}|\leq z)-P(|T_N|\leq z)|>N^{-(1+\xi)}\right)=0.$$

Because $|T^*_{N,k}|$ has a discrete distribution, $P^*(|T^*_{N,k}|\leq z^*_{|T|,k,\alpha})$ might not equal $1-\alpha$ exactly. Nevertheless, the Edgeworth expansion for T^*_N with $a=3/2$ given in Lemma 16(b) combined with the equivalence of the higher-order asymptotic efficiency of T^*_N and $T^*_{N,k}$ given in Theorem 1(b) imply that

$$(9.18)\qquad \lim_{N\to\infty} N^{3/2} P(|P^*(|T^*_{N,k}|\leq z^*_{|T|,k,\alpha})-(1-\alpha)|>N^{-3/2}=0.$$

This holds because (i) there is a value $z^{**}_{|T|,\alpha}$ such that the Edgeworth expansion of $P^*(|T^*_N| < z)$, viz., $\pi_2(\delta, \nu^*_{T,N,3/2})(\Phi(z) - \Phi(-z))$, equals $1-\alpha$ at $z = z^{**}_{|T|,\alpha}$ by the continuity of $\Phi(z)$ and $\phi(z)$, (ii) by definition of $z^*_{|T|,\alpha}$, $|P^*(|T^*_{N,k}| \le z^*_{|T|,k,\alpha}) - (1-\alpha)| \le |P^*(|T^*_{N,k}| \le z^{**}_{|T|,\alpha}) - (1-\alpha)|$, (iii) the latter upper bound differs from $|P^*(|T^*_N| \le z^{**}_{|T|,\alpha}) - (1-\alpha)|$ by more than $N^{-3/2}$ with probability $o(N^{-3/2})$ by Theorem 1(b), (iv) the latter expression differs from $|\pi_2(\delta, \nu^*_{T,N,3/2})(\Phi(z^{**}_{|T|,\alpha}) - \Phi(-z^{**}_{|T|,\alpha})) - (1-\alpha)|$ by more than $N^{-3/2}$ with probability $o(N^{-3/2})$ by Lemma 16(b), and (v) the expression in (iv) equals 0 by definition of $z^{**}_{|T|,\alpha}$.

Taking $z = z^*_{|T|,k,\alpha}$ in (9.17) and combining it with (9.18) gives

$$\lim_{N\to\infty} N^{3/2} P(|1-\alpha - P(|T_N| \le z^*_{|T|,k,\alpha})| > N^{-(1+\xi)}) = 0. \tag{9.19}$$

The expression in the absolute value sign is nonrandom. Hence, for N large, $|1-\alpha - P(|T_N| \le z^*_{|T|,k,\alpha})| \le N^{-(1+\xi)}$, which establishes part (a).

The proof of the second result of part (a) is the same as for the first result but with $T^*_{N,k}$ and $z^*_{|T|,k,\alpha}$ replaced by T^*_N and $z^*_{|T|,\alpha}$ in (9.17)–(9.19). The proofs of the remaining results of part (a) are analogous, using the appropriate results from Theorem 1(b) and Lemma 16(a)–(b).

The proof of part (b) of the Theorem is quite similar to that of part (a). The main differences are that Theorem 1(b) and Lemma 16 are applied with $a = 1$ and the term involving $\pi_1(\delta, \nu^*_{J,N,1})\Phi(z)$ and $\pi_1(\delta, \nu_1)\Phi(z)$, which arises in the application of Lemma 16, does not cancel out because it no longer enters via $(\Phi(z) - \Phi(-z))$.

Next, we prove the second result of part (c). By Lemma 13 with $a = 2$, it suffices to establish the result with T_N and T^*_N replaced by $N^{1/2}G(S_N)$ and $N^{1/2}G(S^*_N)$ respectively. The second result of part (c) now can be established using methods developed for "smooth functions of sample averages," as in Hall (1988, 1992). Define $z_{|G|,\alpha}$ by $P(|N^{1/2}G(S_N)| \le z_{|G|,\alpha}) = 1-\alpha$ and let $\Delta = z_{|G|,\alpha} - z^*_{|T|,\alpha}$. The idea of the proof is to show that

$$\begin{aligned} &P(N^{-1/2}G(S_N) + \Delta \le z_{|G|,\alpha}) = 1 - \alpha/2 + N^{-3/2} r_1(z_{|G|,\alpha})\phi(z_{|G|,\alpha}) + O(N^{-2}) \quad \text{and} \\ &P(N^{1/2}G(S_N) - \Delta \le -z_{|G|,\alpha}) = \alpha/2 - N^{-3/2} r_1(-z_{|G|,\alpha})\phi(-z_{|G|,\alpha}) + O(N^{-2}), \end{aligned} \tag{9.20}$$

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

$$\begin{aligned} P(|T_N| \le z^*_{|T|,\alpha}) &= P(|N^{1/2}G(S_N)| \le z^*_{|T|,\alpha}) + O(N^{-2}) \\ &= 1 - \alpha + N^{-3/2} r_1(z_{|G|,\alpha})\phi(z_{|G|,\alpha}) + N^{-3/2} r_1(-z_{|G|,\alpha})\phi(-z_{|G|,\alpha}) + O(N^{-2}) \\ &= 1 - \alpha + O(N^{-2}), \end{aligned} \tag{9.21}$$

using the fact that $r_1(x)$ is an odd function and $\phi(\cdot)$ is an even function. The results of (9.20) are established by the same argument as used to prove (3.2) of Hall (1988), where his T corresponds to our $N^{1/2}G(S_N)$. (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 $N^{1/2}G(S_N)$ and $N^{1/2}G(S^*_N)$:

$$\begin{aligned} &\lim_{N\to\infty} N^2 \sup_{z\in R} \big| P(|N^{1/2}G(S_N)| \le z) \\ &\qquad - (1 + N^{-1}\pi_2(\delta, \nu_2) + N^{-2}\pi_4(\delta, \nu_2))(\Phi(z) - \Phi(-z)) \big| = 0 \quad \text{and} \\ &\lim_{N\to\infty} N^2 P\Big(\sup_{z\in R} \big| P^*(|N^{1/2}G(S^*_N)| \le z) \\ &\qquad - (1 + N^{-1}\pi_2(\delta, \tilde{\nu}_{N,2}) + N^{-2}\pi_4(\delta, \tilde{\nu}_{N,2}))(\Phi(z) - \Phi(-z)) \big| > N^{-2}\Big) = 0, \end{aligned} \tag{9.22}$$

which hold by Lemma 16 with $a = 2$, T_N and T^*_N replaced by $N^{1/2}G(S_N)$ and $N^{1/2}G(S^*_N)$, respectively, and $\nu^*_{T,N,2}$ replaced by $\tilde{\nu}_{N,2}$. The former replacements are valid by the proof of Lemma 16. The latter replacement holds because no correction factor is used in the iid case.

The proof of the first result of part (c) is the same as that of the second except that the reason that it suffices to establish the desired result with T_N and T^*_N replaced by $N^{1/2}G(S_N)$ and

$N^{1/2}G(S_N^*)$, respectively, is the second result of Theorem 1(b) with $a=2$ combined with Lemma 13 with $a=2$. *Q.E.D.*

9.3. *Proofs of Lemmas*

9.3.1 *Proof of Lemma 1*

A strong mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) gives $E\|\sum_{i=1}^N h(\widetilde{X}_i)\|^p < CN^{p/2}$ 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

$$\lim_{N\to\infty} \varepsilon^{-p} N^{a-p+pc} E\left\| \sum_{i=1}^N h(\widetilde{X}_i)\right\|^p \leq \lim_{N\to\infty} \varepsilon^{-p} C N^{a-p+pc+p/2} = 0. \tag{9.23}$$

Part (b) follows from part (a) applied to $h(\widetilde{X}_i) - Eh(\widetilde{X}_1)$ with $c=0$ and the triangle inequality. *Q.E.D.*

9.3.2. *Proof of Lemma 2*

The proof is the same as that of Lemma 2 of HH (which mimics a standard proof of a uniform law of large numbers) except that we apply Lemma 1 above with $c=0$ and $p=q_1$ rather than their Lemma 1. *Q.E.D.*

9.3.3. *Proof of Lemma 3*

First, we prove the result with N^{-c} replaced by arbitrary fixed $\varepsilon > 0$ for the minimum ρ estimator under Assumption 2(b)(ii). Let $\rho(\theta) = E_\rho(X_1, \theta)$ and $\rho_N(\theta) = N^{-1}\sum_{i=1}^N \rho(X_i, \theta)$. Given $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\theta - \theta_0\| > \varepsilon$ implies that $\rho(\theta) - \rho(\theta_0) \geq \delta > 0$. Thus,

$$\begin{aligned} N^a P(\|\hat{\theta}_N - \theta_0\| > \varepsilon) &\leq N^a P(\rho(\hat{\theta}_N) - \rho_N(\hat{\theta}_N) + \rho_N(\hat{\theta}_N) - \rho(\theta_0) > \delta) \\ &\leq N^a P(\rho(\hat{\theta}_N) - \rho_N(\hat{\theta}_N) + \rho_N(\theta_0) - \rho(\theta_0) > \delta) \\ &\leq N^a P\left(2\sup_{\theta\in\Theta} |\rho_N(\theta) - \rho(\theta)| > \delta\right) \to 0 \end{aligned} \tag{9.24}$$

using Lemma 2 with $j=2$. The corresponding proof for the one-step GMM estimator under Assumption 2(b)(i) is analogous with $\rho(\theta)$ and $\rho_N(\theta)$ replaced by $J(\theta) = Eg(X_1,\theta)'\Omega Eg(X_1,\theta)$ and $J_N(\theta)$ respectively.

Next, we prove the result as stated in the Lemma. For the minimum ρ estimator, the result of the Lemma for arbitrary $\varepsilon > 0$ implies that $\hat{\theta}_N$ is in the interior of Θ, $N^{-1}\sum_{i=1}^N g(X_i, \hat{\theta}_N) = 0$, and $\hat{\theta}_N$ minimizes not only $\rho_N(\theta)$ but $J_N(\theta)$ (defined with an arbitrary positive definite weight matrix Ω) over $\theta \in \Theta$ with probability $1 - o(N^{-a})$. In consequence, in the remainder of this proof, we can treat the minimum ρ estimator as a one-step GMM estimator.

For the one-step GMM estimator, $\hat{\theta}_N$ is in the interior of Θ and $(\partial/\partial\theta)J_N(\hat{\theta}_N) = 0$ with probability $1 - o(N^{-a})$. Hence, element by element mean value expansions of $(\partial/\partial\theta)J_N(\hat{\theta}_N)$ about θ_0 and rearrangement give

$$\hat{\theta}_N - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\theta_N^+)\right)^{-1} \frac{\partial}{\partial\theta} J_N(\theta_0) \tag{9.25}$$

with probability $1 - o(N^{-a})$, where θ_N^+ lies between $\hat{\theta}_N$ and θ_0 and may differ across rows. In consequence, the result of the Lemma follows from

$$\begin{aligned} &\lim_{N\to\infty} N^a P\left(\left\| \frac{\partial^2}{\partial\theta\partial\theta'} J_N(\theta_N^+) - \frac{\partial^2}{\partial\theta\partial\theta'} J_N(\theta_0)\right\| > \varepsilon\right) = 0, \\ &\lim_{N\to\infty} N^a P\left(\left\| \frac{\partial^2}{\partial\theta\,\partial\theta'} J_N(\theta_0) - 2D'\Omega D\right\| > \varepsilon\right) = 0, \end{aligned} \tag{9.26}$$

$$\lim_{N\to\infty} N^a P(\|D_N(\theta_0) - D\| > \varepsilon) = 0, \quad \text{and}$$

$$\lim_{N\to\infty} N^a P\left(\left\|N^{-1}\sum_{i=1}^{N} g(X_i, \theta_0)\right\| > N^{-c}\right) = 0.$$

The first result of (9.26) holds using the result of the present Lemma with N^{-c} replaced by $\varepsilon > 0$, Taylor expansions about θ_0, and multiple applications of Lemma 1(b) with $h(\widetilde{X}_i) = (\partial^j/\partial\theta^j)g(X_i, \theta_0)$ for $j = 0, \dots, 3$ or $h(\widetilde{X}_i) = C_g(X_i)$. The second result of (9.26) holds by multiple applications of Lemma 1(a) with $h(\widetilde{X}_i) = (\partial^j/\partial\theta^j)g(X_i, \theta_0) - E(\partial^j/\partial\theta^j)g(X_i, \theta_0)$ for $j = 0, 1, 2, c = 0$, and $p = q_2$ and standard manipulations. The third result holds by Lemma 1(a) with $h(\widetilde{X}_i)$ as in the proof of the second result with $j = 1$. The fourth result holds by Lemma 1(a) with $h(\widetilde{X}_i) = g(X_i, \theta_0)$, $c = c$, and $p = q_1$. *Q.E.D.*

9.3.4. *Proof of Lemma 4*

First, we show that $\lim_{N\to\infty} N^a P(\|\Omega_N(\tilde{\theta}_N) - \Omega_0\| > \varepsilon) = 0$. This follows from

$$\lim_{N\to\infty} N^a P(\|\Omega_N^{-1}(\tilde{\theta}_N) - \Omega_N^{-1}(\theta_0)\| > \varepsilon) = 0 \quad \text{and} \tag{9.27}$$

$$\lim_{N\to\infty} N^a P(\|\Omega_N^{-1}(\theta_0) - \Omega_0^{-1}\| > \varepsilon) = 0.$$

The first result of (9.27) holds by Lemma 3, mean value expansions, and multiple applications of Lemma 1(b) with $h(\widetilde{X}_i) = \sup_{\theta\in N_0} \|g(X_i, \theta)\| \cdot \|(\partial/\partial\theta')g(X_{i+j}, \theta)\|$ for $j = -\kappa, \dots, \kappa$. The second result of (9.27) holds by multiple applications of Lemma 1(a) with $h(\widetilde{X}_i) = g(X_i, \theta_0)g(X_{i+j}, \theta_0)' - Eg(X_i, \theta_0)g(X_{i+j}, \theta_0)'$ for $j = -\kappa, \dots, \kappa, c = 0$, and $p = q_1/2$.

Given the result of the previous paragraph, the proof of Lemma 4 is analogous to that of Lemma 3. *Q.E.D.*

9.3.5. *Proof of Lemma 5*

Consider part (a). Let ι denote a column L_A-vector of ones. Then, for all $z \in R^{L_A}$,

$$\begin{aligned} &N^a(P(A_N + \xi_N \le z) - P(A_N \le z)) \\ &\quad \le N^a(P(A_N + \xi_N \le z, \|\xi_N\| \le \vartheta_N) - P(A_N \le z)) + N^a P(\|\xi_N\| > \vartheta_N) \\ &\quad \le N^a(P(A_N \le z + \vartheta_N\iota) - P(A_N \le z)) + N^a P(\|\xi_N\| > \vartheta_N). \end{aligned} \tag{9.28}$$

The second term on the right-hand side converges in probability to zero by assumption. Now, consider the case where A_N has an Edgeworth expansion with remainder $o(N^{-a})$. The first term on the last line of (9.28) is less than or equal to N^a multiplied by

$$\left(1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\partial/\partial z)\right)\Phi_{\Sigma_N}(z + \vartheta_N\iota) - \left(1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\partial/\partial z)\right)\Phi_{\Sigma_N}(z) + o(N^{-a}). \tag{9.29}$$

This is $o(N^{-a})$ uniformly over $z \in R^{L_A}$ because the derivatives of $\Phi_{\Sigma_N}(z)$ of all orders are bounded over $z \in R^{L_A}$ given the assumptions on Σ_N and $\vartheta_N = o(N^{-a})$. Alternatively, in the case where $\{A_N : N \ge 1\}$ have uniformly bounded densities, the first term on the right-hand side of (9.28) is $o(1)$ because $\vartheta_N = o(N^{-a})$.

An analogous argument shows that $N^a(P(A_N \le z) - P(A_N + \xi_N \le z))$ is $o(1)$ uniformly over $z \in R^{L_A}$. This completes the proof of part (a).

The proof of part (b) is similar. For brevity, it is omitted. *Q.E.D.*

9.3.6. *Proof of Lemma 6*

First, we establish part (a). Define $\varGamma_N^* = N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*) - E^*h(\widetilde{X}_i^*)$. By Markov's inequality applied twice, we have

$$(9.30)\qquad N^a P(P^*(\|\varGamma_N^*\| > N^{-c}) > N^{-a}) \le N^a P(E^*\|\varGamma_N^*\|^p > N^{-a-cp}) \le N^{2a+cp} EE^*\|\varGamma_N^*\|^p.$$

Define $Y_{\ell i}^* = \ell^{-1}\sum_{j\in b_i^*} h(\widetilde{X}_j)$ and $Y_{\ell i} = \ell^{-1}\sum_{j\in b_i} h(\widetilde{X}_j)$ (where b_j^* and b_j are defined at the beginning of the Appendix). Then, $\varGamma_N^* = b^{-1}\sum_{i=1}^b (Y_{\ell i}^* - E^*Y_{\ell i}^*)$. By applying Burkholder's and Holder's inequality in a single step (e.g., see Hall and Heyde (1980, eqn. (3.67), p. 87)), we obtain

$$(9.31)\qquad E^*\|\varGamma_N^*\|^p = b^{-p}E^*\left\|\sum_{i=1}^b (Y_{\ell i}^* - E^*Y_{\ell i}^*)\right\|^p \le Cb^{-p/2}E^*\|Y_{\ell 1}^* - E^*Y_{\ell 1}^*\|^p \le Cb^{-p/2}E^*\|Y_{\ell 1}^*\|^p.$$

Now, for nonoverlapping blocks, we have

$$(9.32)\qquad EE^*\|Y_{\ell 1}^*\|^p = Eb^{-1}\sum_{i=1}^b \|Y_{\ell i}\|^p = E\|Y_{\ell 1}\|^p \le C\ell^{-p/2}$$

using Yokoyama's strong mixing moment inequality (see the proof of Lemma 1). For overlapping blocks, (9.32) holds, but with b replaced by $N-\ell+1$ after the first equality.

Combining (9.30), (9.31), and (9.32) gives

$$(9.33)\qquad N^a P(P^*(\|\varGamma_N^*\| > N^{-c}) > N^{-a}) \le CN^{2a+cp}b^{-p/2}\ell^{-p/2} = CN^{2a+cp-p/2} = o(1).$$

To establish part (b), note that the left-hand side in part (b) equals $\lim_{N\to\infty} N^a P(\|E^*N^{-1}\times \sum_{i=1}^N h(\widetilde{X}_i^*)\| > N^{-c})$, which we denote by *lhs*. For nonoverlapping blocks, then,

$$(9.34)\qquad lhs = \lim_{N\to\infty} N^a P\left(\left\|b^{-1}\sum_{i=1}^b Y_{\ell i}\right\| > N^{-c}\right) = \lim_{N\to\infty} N^a P\left(\left\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i)\right\| > N^{-c}\right) = 0$$

using Lemma 1. For overlapping blocks with $\ell \propto N^\gamma$ for $0 \le \gamma < 1$, we have

$$(9.35)\qquad lhs = \lim_{N\to\infty} N^a P\left(\left\|(N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1} Y_{\ell i}\right\| > N^{-c}\right) = \lim_{N\longrightarrow\infty} N^a P\left(\left\|(N-\ell+1)^{-1}\sum_{i=1}^N w(i,\ell,N)h(\widetilde{X}_i)\right\| > N^{-c}\right) = 0,$$

where $w(i,\ell,N)$ is defined in (4.2). Note that $|w(i,\ell,N)| \le 1$. The last equality of (9.35) holds by an argument analogous to that of Lemma 1 using Yokoyama's strong mixing moment inequality (which applies to nonstationary L^p-bounded random variables; see Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30)), using the fact that $\lim_{N\to\infty} N/(N-\ell+1) = 1$ when $0 \le \gamma < 1$. For overlapping blocks with $\gamma = 1$, we have $lhs = \lim_{N\to\infty} N^a P(\|N^{-1}\sum_{i=1}^N h(\widetilde{X}_i)\| > N^{-c}) = 0$.

Part (c) follows from parts (a) and (b). The first result of part (d) holds by using the triangle inequality $\|E^*N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*)\| \le \|E^*N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*) - Eh(\widetilde{X}_i)\| + \|Eh(\widetilde{X}_i)\|$ and applying part (b) to $h(\cdot) - Eh(\widetilde{X}_1)$ with c arbitrarily close to zero. The second result of part (d) is established analogously using part (c) in place of part (b). *Q.E.D.*

9.3.7. *Proof of Lemma 7*

The proof is the same as that of Lemma 8 of HH except that we apply Lemma 6 above with $c=0$ and $p=q_1$ rather than Lemma 7 of HH. *Q.E.D.*

9.3.8. *Proof of Lemma 8*

Define $\Omega_N^{**}(\theta)$ to equal $\Omega_N^*(\theta)$ with $g^*(X_i^*,\theta)$ $(=g(X_i^*,\theta)-E^*g(X_i^*,\hat{\theta}_N))$ replaced with $g(X_i^*,\theta)$. The result of the Lemma follows from

$$\lim_{N\to\infty} N^a P(P^*(\|\Omega_N^*(\hat{\theta}_N^*)^{-1}-\Omega_N^*(\theta_0)^{-1}\|>\varepsilon)>N^{-a})=0, \tag{9.36}$$

$$\lim_{N\to\infty} N^a P(P^*(\|\Omega_N^*(\theta_0)^{-1}-\Omega_N^{**}(\theta_0)^{-1}\|>\varepsilon)>N^{-a})=0, \tag{9.37}$$

$$\lim_{N\to\infty} N^a P(P^*(\|\Omega_N^{**}(\theta_0)^{-1}-E^*(\Omega_N^{**}(\theta_0)^{-1})\|>\varepsilon)>N^{-a})=0, \quad \text{and} \tag{9.38}$$

$$\lim_{N\to\infty} N^a P(P^*(\|E^*(\Omega_N^{**}(\theta_0)^{-1})-\Omega_0^{-1}\|>\varepsilon)>N^{-a})=0. \tag{9.39}$$

To establish (9.36), we take mean value expansions about θ_0, apply both parts of Lemma 6(d) with $h(\widetilde{X}_i)=\sup_{\theta\in N_0}\|g(X_i,\theta)\|\cdot\|(\partial/\partial\theta')g(X_{i+j},\theta)\|$ for $j=-\kappa,\ldots,\kappa$ and $p=q_2$, and use the assumption on $\hat{\theta}_N^*$. To establish (9.37), we use the Cauchy–Schwarz inequality and Lemma 6(b) and 6(c) with $h(\widetilde{X}_i)=g(X_i,\theta_0)$, $c\in(0,1/2)$, and $p=q_1$. To establish (9.38), we use Lemma 6(c) with $h(\widetilde{X}_i)=g(X_i,\theta_0)g(X_{i+j},\theta_0)'-Eg(X_1,\theta_0)g(X_{1+j},\theta_0)'$ for $j=-\kappa,\ldots,\kappa$, $c\in(0,1/2)$, and $p=q_1/2$. To establish (9.39), we use Lemma 6(b) with $h(\widetilde{X}_i)$, c, and p as immediately above. *Q.E.D.*

9.3.9. *Proof of Lemma 9*

First, we prove the result with N^{-c} replaced by arbitrary fixed $\varepsilon>0$ for the minimum ρ estimator under Assumption 2(b)(ii). Let $\rho_N(\theta)=N^{-1}\sum_{i=1}^N\rho(X_i,\theta)$, $\rho(\theta)=E\rho(X_1,\theta)$, and $\rho_N^*(\theta)=N^{-1}\sum_{i=1}^N(\rho(X_i^*,\theta)-E^*g(X_i^*,\hat{\theta}_N)'\theta)$. Consider the case of nonoverlapping blocks. Given $\varepsilon>0$, there exists a $\delta>0$ independent of N such that $\|\theta-\hat{\theta}_N\|>\varepsilon$ implies that $E^*\rho_N^*(\theta)-E^*\rho_N^*(\hat{\theta}_N)\geq\delta>0$ with probability $1-o(N^{-a})$ because (i) $E^*N^{-1}\sum_{i=1}^N\rho(X_i^*,\theta)=\rho_N(\theta)$ with probability $1-o(N^{-a})$, (ii) $E^*g(X_i^*,\hat{\theta}_N)=N^{-1}\sum_{i=1}^N g(X_i,\hat{\theta}_N)=0$ with probability $1-o(N^{-a})$ by the first-order conditions for $\hat{\theta}_N$ since the dimensions of $g(\cdot,\cdot)$ and θ are equal, (iii) $\lim_{N\to\infty}N^aP(\sup_{\theta\in\Theta}|\rho_N(\theta)-\rho_N(\hat{\theta}_N)-\rho(\theta)+\rho(\hat{\theta}_N)|>\lambda)=0$ for all $\lambda>0$ by Lemma 2, (iv) $\lim_{N\to\infty}N^aP(|\rho(\hat{\theta}_N)-\rho(\theta_0)|>\lambda)=0$ using Lemma 3, and (v) $\rho(\theta)$ is uniquely minimized at θ_0 and is continuous on Θ. Thus, we have

$$\begin{aligned}
&N^aP(P^*(\|\theta_N^*-\hat{\theta}_N\|>\varepsilon)>N^{-a})\\
&\quad\leq N^aP(P^*(E^*\rho_N^*(\theta_N^*)-\rho_N^*(\theta_N^*)+\rho_N^*(\theta_N^*)-E^*\rho_N^*(\hat{\theta}_N)>\delta)>N^{-a})\\
&\quad\leq N^aP(P^*(E^*\rho_N^*(\theta_N^*)-\rho_N^*(\theta_N^*)+\rho_N^*(\hat{\theta}_N)-E^*\rho_N^*(\hat{\theta}_N)>\delta)>N^{-a})\\
&\quad\leq N^aP\left(P^*\left(2\sup_{\theta\in\Theta}\left|\rho_N^*(\theta)-E^*\rho_N^*(\theta)\right|>\delta\right)>N^{-a}\right)\to 0,
\end{aligned} \tag{9.40}$$

using Lemma 7 with $j=2$.

For the case of overlapping blocks, (i) and (ii) of the previous paragraph do not hold. Instead, we have $E^*N^{-1}\sum_{i=1}^N\rho(X_i^*,\theta)=(N-\ell+1)^{-1}\sum_{i=1}^N w(i,\ell,N)\rho(X_i,\theta)$, where $w(i,\ell,N)$ is defined in (4.2). By the arguments used to prove Lemmas 2 and 6(b), (iii) holds with $\rho_N(\theta)$ replaced by $(N-\ell+1)^{-1}\sum_{i=1}^N w(i,\ell,N)\rho(X_i,\theta)$. In addition, some calculations using Lemmas 1 and 3 and a mean value expansion show that $\lim_{N\to\infty}N^aP(\|E^*g(X_i^*,\hat{\theta}_N)\|>\lambda)=0$ for all $\lambda>0$, because $E^*g(X_i^*,\hat{\theta}_N)=(N-\ell+1)^{-1}\sum_{i=1}^N(w(i,\ell,N)-(N-\ell+1)/N)g(X_i,\hat{\theta}_N)$. In consequence, the remainder of the proof above goes through unchanged with overlapping blocks.

The proof of the result of the Lemma for the one-step GMM estimator under Assumption 2(b)(i) when N^{-c} is replaced by $\varepsilon>0$ is analogous to that given above using Lemma 7 with $j=1$. The proof of the result of the Lemma as stated is analogous to that given in Lemma 3 with $J_N(\hat{\theta}_N)$ replaced by $J_N^*(\theta_N^*)$ using Lemmas 6 and 7 in place of Lemmas 1 and 2. *Q.E.D.*

9.3.10. *Proof of Lemma 10*

The proof is analogous to that of Lemma 9 using Lemma 8. *Q.E.D.*

9.3.11. *Proof of Lemma 11*

By Lemma 9 with N^{-c} replaced by arbitrary fixed $\varepsilon > 0$, $\tilde{\theta}_N^*$ satisfies the same condition as $\hat{\theta}_N^*$. In consequence, it suffices to show that the first result of the Lemma holds and that for all $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P(P^*(\|\Omega_N^*(\hat{\theta}_N^*) - \Omega_0\| > \varepsilon) > N^{-a}) = 0, \tag{9.41}$$

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N \frac{\partial^j}{\partial\theta^j} g^*(X_i^*, \hat{\theta}_N^*)\right\| > K\right) > N^{-a}\right) = 0 \tag{9.42}$$

for some $K < \infty$ for $j = 1, 2, 3$, and

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^N g^*(X_i^*, \hat{\theta}_N^*)\right\| > \varepsilon\right) > N^{-a}\right) = 0. \tag{9.43}$$

The first result of the Lemma, (9.42), and (9.43) hold by mean value expansions, multiple applications of Lemma 6, and the assumption on $\hat{\theta}_N^*$. Equation (9.41) holds by Lemma 8. *Q.E.D.*

9.3.12. *Proof of Lemma 12*

By a mean value expansion and the triangle inequality,

$$\begin{aligned} &\|m_N^*(\theta_{N,k}^*) - m_N^*(\theta_N^*)\| \\ &\quad \le \left(N^{-1}\sum_{i=1}^N \sup_{\theta\in N_0} \left\|(\partial/\partial\theta)m(\widetilde{X}_i^*, \theta)\right\| + E^* N^{-1}\sum_{i=1}^N \sup_{\theta\in N_0} \left\|(\partial/\partial\theta)m(\widetilde{X}_i^*, \theta)\right\|\right)\|\theta_{N,k}^* - \theta_N^*\|. \end{aligned} \tag{9.44}$$

Hence, the Lemma holds by the assumption on $\|\theta_{N,k}^* - \theta_N^*\|$ and Lemma 6(d) with $h(\widetilde{X}_i) = \sup_{\theta\in N_0}\|(\partial/\partial\theta)m(\widetilde{X}_i, \theta)\|$. *Q.E.D.*

9.3.13. *Proof of Lemma 13*

The proof of part (a) is analogous to that of Proposition 1 of HH except that we use Lemmas 1 and 3–5 above in place of their Lemmas 1 and 3–5, respectively, and we take the Taylor expansion through order d_1 rather than order 4. (For the Wald statistic, this requires that the function $\eta(\cdot)$ is d_1 times continuously differentiable.) The latter implies that the remainder term ζ_N from the Taylor expansion in the proof of Proposition 1 of HH (to which our Lemma 5 needs to be applied with $\xi_N = N^{1/2}\zeta_N$) satisfies $\|\zeta_N\| \le C\|\hat{\theta}_N - \theta_0\|^{d_1}$ with probability $1 - o(N^{-a})$. In consequence, $\lim_{N\to\infty} N^a P(\|\zeta_N\| > N^{-d_1 c}) = 0$ and for our Lemma 5 to apply with $\xi_N = N^{1/2}\zeta_N$, we need $d_1 c \ge a + 1/2$ for some $c \in [0, 1/2)$ and we need $2a$ to be an integer. The former holds if $d_1 \ge 2a + 2$, as is assumed. The latter holds by assumption.

The proof of part (b) is analogous to that of Proposition 2 of HH except that we use Lemmas 3–9 above in place of their Lemmas 3–9 and we take the Taylor expansion through order d_1 rather than order 4. *Q.E.D.*

9.3.14. *Proof of Lemma 14*

It is shown in Andrews (2001a) that the least favorable value of m for the bootstrap moment $N^{\alpha(m)} E^* \prod_{\mu=1}^{m} \Psi^*_{N,j_\mu}$ is three. Here we just consider this case. We show that

$$(9.45)\qquad A_1 = \lim_{N\to\infty} N^a P\left(\left| N^{1/2} E^* \prod_{\mu=1}^{3} \Psi^*_{N,j_\mu} - N^{1/2} E \prod_{\mu=1}^{3} \Psi_{N,j_\mu} \right| > 6N^{-\xi} \right) = 0 \quad \text{and}$$

$$A_2 = \lim_{N\to\infty} N^{\xi} \left| N^{1/2} E \prod_{\mu=1}^{3} \Psi_{N,j_\mu} - \lim_{N\to\infty} N^{1/2} E \prod_{\mu=1}^{3} \Psi_{N,j_\mu} \right| = 0.$$

For notational simplicity, suppose $j_\mu = 1$ for $\mu = 1, 2, 3$. Let $f_i = f_1(\widetilde{X}_i, \theta_0) - Ef_1(\widetilde{X}_i, \theta_0)$, where $f_1(\widetilde{X}_i, \theta_0)$ denotes the first element of $f(\widetilde{X}_i, \theta_0)$, and let $f_i^* = f_1(\widetilde{X}_i, \hat{\theta}_N) - E^* f_1(\widetilde{X}_i^*, \hat{\theta}_N)$. Let $Y_j = \sum_{i\in b_j} f_i$, $\widetilde{Y}_j = \sum_{i\in b_j} f_i^*$, and $Y_j^* = \sum_{i\in b_j^*} f_i^*$ (where b_j, b_j^*, and $\mathcal{N}_\ell$ are defined at the beginning of the Appendix). Then, $\Psi_{N,j_\mu} = N^{-1/2}\sum_{i=1}^{N} f_i$ and $\Psi^*_{N,j_\mu} = N^{-1/2}\sum_{j=1}^{b} Y_j^*$. We have

$$(9.46)\qquad N^{1/2} E^* \prod_{\mu=1}^{3} \Psi^*_{N,j_\mu} = N^{-1} \sum_{j_1=1}^{b}\sum_{j_2=1}^{b}\sum_{j_3=1}^{b} E^* Y^*_{j_1} Y^*_{j_2} Y^*_{j_3}$$

$$= N^{-1} b E^* Y_1^{*3} = N^{-1} b \mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} \widetilde{Y}_j^3.$$

A lengthy proof in Andrews (2001a) shows that

$$(9.47)\qquad \lim_{N\to\infty} N^a P\left(\left| N^{-1} b \mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} (\widetilde{Y}_j^3 - Y_j^3) \right| > N^{-\xi} \right) = 0.$$

The proof uses a Taylor expansion of order d_2 of $\widetilde{Y}_j - Y_j$, which depends on $\hat{\theta}_N$, about θ_0. To ensure that the remainder term is asymptotically negligible requires $d_2 \geq -1 + (a+\gamma+\xi)/c$, which is implied by the assumed condition that $d_2 \geq 2a+1$ because $\gamma + \xi < 1/2$ and we can take c $(< 1/2)$ arbitrarily close to 1/2.

Using (9.46) and (9.47), we obtain

$$(9.48)\qquad A_1 \leq B_1 + B_2, \quad \text{where}$$

$$B_1 = \lim_{N\to\infty} N^a P(|N^{-1} b \mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} (Y_j^3 - EY_1^3)| > N^{-\xi}) \quad \text{and}$$

$$B_2 = \lim_{N\to\infty} 1\left(\left| N^{-1} b E Y_1^3 - N^{1/2} E \prod_{\mu=1}^{3} \Psi_{N,j_\mu} \right| > N^{-\xi} \right).$$

For nonoverlapping blocks, a strong mixing moment inequality of Yokoyama and Doukhan (see Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30)) gives: for any $\delta > 0$, there exists a constant $C < \infty$ such that

$$(9.49)\qquad E\left| \sum_{i=1}^{b} (Y_i^r - EY_1^r) \right|^s \leq Cb^{s/2} (E|Y_1^r - EY_1^r|^{s+\delta})^{s/(s+\delta)} \leq Cb^{s/2} (E|Y_1^r|^{s+\delta})^{s/(s+\delta)},$$

for $r > 0$ and $s \geq 2$. (This result uses the fact that $\alpha_\ell(i) \leq \alpha(i)$ for all $\ell \geq 1$, where $\alpha(i)$ denotes the ith strong mixing number of $\{\widetilde{X}_i : i \geq 1\}$ and $\alpha_\ell(i)$ denotes the ith strong mixing number of $\{Y_i : i \leq b\}$.) In turn, the same moment inequality gives

$$(9.50)\qquad E|Y_1^r|^{s+\delta} = E\left| \sum_{j=1}^{\ell} f_j \right|^{r(s+\delta)} \leq C\ell^{r(s+\delta)/2} (E|f_1|^{r(s+\delta)+\delta})^{r(s+\delta)/(r(s+\delta)+\delta)}.$$

Combining these two inequalities gives

$$E\left|b\mathcal{N}_\ell^{-1}\sum_{i=1}^{\mathcal{N}_\ell}(Y_i^r - EY_1^r)\right|^s \leq Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}}, \tag{9.51}$$

where $c_{r,s} = sr(s+\delta)/[(r(s+\delta)+\delta)(s+\delta)]$.

Next, for overlapping blocks, we have

$$\begin{aligned}\left(E\left|\sum_{i=1}^{N-\ell+1}(Y_i^r - EY_1^r)\right|^s\right)^{1/s} &= \left(E\left|\sum_{u=1}^{\ell}\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)\right|^s\right)^{1/s} \\ &\leq \sum_{u=1}^{\ell}\left(E\left|\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)\right|^s\right)^{1/s} \\ &\leq \sum_{i=1}^{\ell}(Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}})^{1/s}\end{aligned} \tag{9.52}$$

where $b(u) = \max\{j : j \leq b-1, u+\ell j \leq N-\ell+1\}$, using Minkowski's inequality and (9.51), where the latter applies because $\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)$ is a sum of terms based on nonoverlapping blocks. Equation (9.52) gives

$$\begin{aligned}E\left|b\mathcal{N}_\ell^{-1}\sum_{i=1}^{\mathcal{N}_\ell}(Y_i^r - EY_1^r)\right|^s &\leq \frac{b^s\ell^s}{(N-\ell+1)^s}Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}} \\ &= Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}},\end{aligned} \tag{9.53}$$

using the fact that $(b\ell)/(N-\ell+1) = 1+o(1)$ if $\gamma < 1$.

Now we show that $B_1 = 0$. By Markov's inequality and (9.51) or (9.53) with $r = 3$, we have

$$\begin{aligned}B_1 &\leq C\lim_{N\to\infty} N^{a+p(\xi-1)}E\left|b\mathcal{N}_\ell^{-1}\sum_{j_2=1}^{\mathcal{N}_\ell}(Y_{j_2}^3 - EY_1^3)\right|^p \\ &\leq C\lim_{N\to\infty} N^{a+p(\xi-1)}b^{p/2}\ell^{3p/2}(E|f_1|^{3p+4\delta})^{c_{3,p}}.\end{aligned} \tag{9.54}$$

Given (9.54), $B_1 = 0$ provided s is such that $s > a/(1/2-\gamma-\xi)$ and $E|f_1|^{3s+4\delta} < \infty$. The latter holds for all $0 < s < \infty$ by Assumption 3(e) and the former holds for s sufficiently large because $\gamma+\xi < 1/2$.

For B_2, we have

$$\begin{aligned}N^{1/2}E\prod_{\mu=1}^{3}\Psi_{N,j_u} &= N^{-1}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}Ef_{i_1}f_{i_2}f_{i_3} \\ &= \sum_{i_1=-N+1}^{N-1}\sum_{i_2=-N+1}^{N-1}\omega(i_1+i_2,N)Ef_0f_{i_1}f_{i_2},\end{aligned} \tag{9.55}$$

where $\omega(i,N) = 1-i/N$. In addition, we have

$$N^{-1}bEY_1^3 = \ell^{-1}E\left(\sum_{i=1}^{\ell}f_i\right)^3 = \sum_{i_1=-\ell+1}^{\ell-1}\sum_{i_2=-\ell+1}^{\ell-1}\omega(i_1+i_2,\ell)Ef_0f_{i_1}f_{i_2}. \tag{9.56}$$

The quantity B_2 equals zero if the difference between the right-hand sides of (9.55) and (9.56) multiplied by N^ξ has limit equal to zero. The latter holds by a strong mixing covariance inequality, viz., $Ef_0f_{i_1} \leq C\alpha^r(i_1)$ for some $r > 0$ (where $\alpha(i_1)$ denotes the i_1th strong mixing number); e.g., see Doukhan (1995, Thm. 3, p. 9); the fact that the strong mixing numbers decline exponentially fast by Assumption 1, $N^\xi \propto \ell^{\xi/\gamma}$, and either (i) $\xi < \gamma$ and $0 < \gamma < 1$ or (ii) $\{X_i : i \geq 1\}$ are independent. The latter condition appears because $\lim_{N\to\infty}N^\xi\sum_{i_1=-\ell+1}^{\ell-1}\sum_{i_2=-\ell+1}^{\ell-1}((i_1+i_2)/\ell))Ef_0f_{i_1}f_{i_2} = \lim_{N\to\infty}N^\xi\ell^{-1}\sum_{i_1=-\infty}^{\infty}\sum_{i_2=-\infty}^{\infty}(i_1+i_2)Ef_0f_{i_1}f_{i_2} = 0$ if either $\xi < \gamma$ or $\{X_i : i \geq 1\}$ are independent.

We conclude that $B_2 = 0$. An analogous argument gives $A_2 = 0$. *Q.E.D.*

9.3.15. *Proof of Lemma 15*

First, we prove part (a) for the nonoverlapping block bootstrap. Let g_i denote $g(X_i, \theta_0)$. It is sufficient to establish the following results:

$$\lim_{N\to\infty} N^a P\left(\left\| \widetilde{W}_N - N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} g_{i\ell+j}g'_{i\ell+m}\right\| > N^{-\xi}\right) = 0. \tag{9.57}$$

$$\lim_{N\to\infty} N^a P\left(\left\| \overline{W}_N(\hat{\theta}_N) - N^{-1}\sum_{i=1}^{N}\left[g_i g'_1 + \sum_{j=1}^{\kappa}\{g_i g'_{i+j} + g_{i+j}g'_i\}\right]\right\| > N^{-\xi}\right) = 0. \tag{9.58}$$

$$\lim_{N\to\infty} N^a P\left(\left\| N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}[g_{i\ell+j}g'_{i\ell+m} - Eg_{i\ell+j}g'_{i\ell+m}]\right\| > N^{-\xi}\right) = 0. \tag{9.59}$$

$$\lim_{N\to\infty} N^a P\left(\left\| N^{-1}\sum_{i=1}^{N}\left[g_i g'_i - Eg_i g'_i + \sum_{j=1}^{\kappa}\{g_i g'_{i+j} - Eg_i g'_{i+j} + g_{i+j}g'_i - Eg_{i+j}g'_i\}\right]\right\| > N^{-\xi}\right) = 0. \tag{9.60}$$

$$\lim_{N\to\infty} N^{-1+\xi}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} Eg_{i\ell+j}g'_{i\ell+m} - N^{-1+\xi}\sum_{i=1}^{N}\left[Eg_i g'_i + \sum_{j=1}^{\kappa}\{E_{g_i g'_{i+j}} + Eg_{i+j}g'_i\}\right] = 0. \tag{9.61}$$

The proof of (9.57) is rather lengthy. It is given in Andrews (2001a). It involves taking a Taylor expansion of order d of $\widetilde{W}_N$, which depends on $\hat{\theta}_N$, about θ_0; showing that the first term in the expansion equals the second term in the norm in (9.57) plus an asymptotically negligible term; and showing that the higher order terms are asymptotically negligible. The proof of (9.58) is similar to that of (9.57), but simpler. Equation (9.59) is established using Markov's inequality, the Yokoyama-Doukhan strong mixing moment inequality of (9.49), Minkowski's inequality, and the assumptions that $E\|f_j\|^{q_2} < \infty$ for all $q_2 < \infty$ and $\xi + \gamma/2 < 1/2$; see Andrews (2001a) for details. Equation (9.60) holds by Lemma 1(a) with $c = \xi$ and $p > 2a/(1-2\xi)$ because $E\|f_j\|^p < \infty$ for all $p < \infty$ by Assumption 3(e) and $\xi < 1/2$.

Equation (9.61) holds only if $\xi < \gamma$ when $\sum_{j=1}^{\kappa} j(Eg_1g'_{1+j} + Eg_{1+j}g'_1) \neq 0$. To see this, for simplicity, suppose g_i is a scalar. Then, the left-hand side of (9.61) equals

$$\begin{aligned}
&\lim_{N\to\infty} N^{-1+\xi b}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} Eg_j g_m - N^{\xi}\left[Eg_1^2 + 2\sum_{j=1}^{\kappa} Eg_1 g_{1+j}\right] \\
&= \lim_{N\to\infty} N^{-\gamma+\xi}\left[\ell Eg_1^2 + 2\sum_{j=1}^{\kappa}(\ell - j)Eg_1 g_{1+j}\right] - N^{\xi}\left[Eg_1^2 + 2\sum_{j=1}^{\kappa} Eg_1 g_{1+j}\right] \\
&= -2\lim_{N\to\infty} N^{-\gamma+\xi}\sum_{j=1}^{\kappa} jEg_1 g_{1+j} = 0.
\end{aligned} \tag{9.62}$$

Next, we prove part (a) for the overlapping block bootstrap. The desired result follows from (9.57)–(9.61) with $N^{-1}\sum_{i=0}^{b-1}$ replaced by $N^{-1}b(N-\ell+1)^{-1}\sum_{i=0}^{N-\ell}$ and $g_{i\ell+j}g'_{i\ell+m}$ replaced by $g_{i+j}g'_{i+m}$ in (9.57), (9.59), and (9.61). Equations (9.58) and (9.60) have already been established. The proofs of (9.57) and (9.59) are similar to those for nonoverlapping blocks; see Andrews (2001a).

The analogue of (9.61) is established as follows. Some calculations show that

$$\begin{aligned}
&N^{-1}b(N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} Eg_{i+j}g'_{i+m} \\
&= (N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1}\left[v_N(i,0,\ell)Eg_i g'_i + \sum_{j=1}^{\kappa} v_N(i,j,\ell)(Eg_i g'_{i+j} + Eg_{i+j}g'_i)\right], \quad \text{where}
\end{aligned} \tag{9.63}$$

$$v_N(i,j,\ell) = \begin{cases} 1 - j/\ell & \text{for } \ell \le i \le N-\ell+1, \\ 1 - (\ell - i - j)/\ell & \text{for } 1 \le i < \ell, \\ 1 - (\ell + i - N + 1 - j)/\ell & \text{for } N-\ell+1 < i \le N. \end{cases}$$

In consequence, the analogue of (9.61) holds provided $\lim_{N\to\infty} N^{\xi}\sum_{j=1}^{\kappa}(j/\ell)\ (Eg_1g_{1+j}'+Eg_{1+j}g_1')=0$, which requires either $\xi<\gamma$ or $\sum_{j=1}^{\kappa} j(Eg_1g_{1+j}'+Eg_{1+j}g_1')=0$.

Next, the result of part (b) for t statistics follows from Lemma 14 and

$$\lim_{N\to\infty} N^{a}P(|\tau_{N,r}-1|>N^{-\xi})=0. \tag{9.64}$$

Equation (9.64) follows from part (a) of the present Lemma, $\lim_{N\to\infty} N^{a}P(\|\tilde{\sigma}_N-\sigma\|>\varepsilon)=0$ for all $\varepsilon>0$, which holds by (9.26) and (9.27), and the fact that σ is positive definite by Assumption 3(b). The proof for the $\mathscr{W}_N$ statistic is analogous with $N^{\xi}|\tau_{N,r}-1|$ replaced by $N^{\xi}\|\Xi_N-I_{L_\eta}\|$. The proof for the J statistic uses

$$\begin{aligned}&\lim_{N\to\infty} N^{a}P(\|M_N-M_0\|>N^{-\xi})=0,\quad \lim_{N\to\infty} N^{a}P(\|V_N-M_0\|>N^{-\xi})=0,\\ &\lim_{N\to\infty} N^{a}P(\|V_N^{+}-M_0\|>N^{-\xi})=0,\quad \text{and}\\ &\lim_{N\to\infty} N^{a}P(\|(V_N^{+})^{1/2}-M_0\|>N^{-\xi})=0.\end{aligned} \tag{9.65}$$

The first result of (9.65) follows from (9.26) and (9.27), the second result follows from the first result and part (a) of the Lemma, the third result follows from the second result, $M_0=M_0^{+}$ (because M_0 is a projection matrix), and the fact that $rk(V_N)=rk(M_0)=L_g-L_\theta$ with probability $1-o(N^{-a})$; see Andrews (1987, Theorem 2); and the fourth result follows from the third result and $M_0^{1/2}=M_0$. Now, part (b) of the Lemma for the J statistic follows from the fourth result of (9.65), Lemma 14, and the properties of the function $\lambda_J(\cdot,\cdot)$. *Q.E.D.*

9.3.16. *Proof of Lemma 16*

We use the same method as HH use in the proof of their Theorems 1 and 2. Because their description is very brief, we describe the method in a little more detail than HH do. Given Lemma 13, for the results of parts (a) and (b) regarding T_N and T_N^*, it suffices to show that $N^{1/2}G(S_N)$ and $N^{1/2}_{\tau_{N,r}}G(S_N^*)$ of Lemma 13 possess Edgeworth expansions with remainder $o(N^{-a})$. For the case of $N^{1/2}G(S_N)$, this follows by applying Theorem 3.1 of Bhattacharya (1987) with his integer parameter s satisfying $(s-2)/2=a$ for a given in the present Lemma (with $2a$ being an integer) and with the normalized sample average $N^{1/2}(\overline{X}-\mu)$ of the underlying random variables in his theorem satisfying an Edgeworth expansion not because they are iid and satisfy his condition (A_4), but because they are asymptotically weakly dependent and satisfy the conditions of Theorem 1.1 of Götze and Hipp (1994). The latter theorem is a special case of Corollary 2.9 of Götze and Hipp (1983). Conditions (2)–(4) of Götze and Hipp (1994) hold by Assumptions 1, 3(e), and 4. Conditions (A_1)–(A_3) of Bhattacharya (1987) hold by Assumption 3(e), the fact that $G(\cdot)$ is infinitely differentiable, and Assumption 3(b) respectively.

For the case of $N^{1/2}\tau_{N,r}G(S_N^*)$, the result holds by an analogous argument as for $N^{1/2}G(S_N)$, but with Theorem 3.1 of Bhattacharya (1987) replaced by Theorem 3.3 of Bhattacharya (1987) and using Lemma 15(b) with $\xi=0$ to ensure that the coefficients $\nu_{T,N,a}^*$ are well behaved.

To obtain the remaining results of parts (a) and (b), we note that $N^{1/2}G(S_N)$ and $N^{1/2}\Xi_N G(S_N^*)$ (or $N^{1/2}(V_N^{+})^{1/2}G(S_N^*)$) of Lemma 13 possess multivariate Edgeworth expansions with remainder $o(N^{-a})$ when $G(\cdot)$ corresponds to $H_N(\hat{\theta}_N)$ or $K_N(\hat{\theta}_N)$, by the same argument as just given. Then, the results follow by applying Theorem 1 and Remark 2.2 of Chandra and Ghosh (1979) to obtain the given Edgeworth expansions of $H_N(\hat{\theta}_N)'H_N(\hat{\theta}_N)$, $K_N(\hat{\theta}_N)'K_N(\hat{\theta}_N)$, $H_N^*(\theta_N^*)'H_N^*(\theta_N^*)$, and $K_N^*(\theta_N^*)'K_N^*(\theta_N^*)$.

Part (c) follows from part (b) and Theorem 1(c). Note that the proof of Theorem 1(c) uses part (b), but not part (c), of the present Lemma in its proof. *Q.E.D.*

9.3.17. *Proof of Lemma 17*

The NR result of part (a) holds by definition of $Q_{N,j-1}^{*,NR}$. The default NR result of part (a) is established in Andrews (2001a) by showing that each step of the NR k-step estimator decreases the magnitude of the bootstrap criterion function except on a sequence of sets with asymptotically negligible probability. The line-search NR result of part (a) is established in Andrews (2001a) by showing that the optimal step length is one except on a sequence of sets with asymptotically negligible probability.

To establish part (b) of the Theorem, we use the second result of Lemma 11 and Lemma 8. Using these results, it suffices to show that

$$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^{N}\left(\Delta(\widetilde{X}_i^*, \theta_{N,j-1}^*) - \frac{\partial}{\partial\theta'}g(X_i^*, \theta_{N,j-1}^*)\right)\right\| > N^{-c}\right) > N^{-a}\right) = 0. \tag{9.66}$$

By mean value expansions about θ_0 and the triangle inequality, it suffices to show that

(i) $$\lim_{N\to\infty} N^a P\left(P^*\left(\left\|N^{-1}\sum_{i=1}^{N}\left(\Delta(\widetilde{X}_i^*, \theta_0) - \frac{\partial}{\partial\theta'}g(X_i^*, \theta_0)\right)\right\| > N^{-c}\right) > N^{-a}\right) = 0,$$

(ii) $$\lim_{N\to\infty} N^a P\left(P^*\left(N^{-1}\sum_{i=1}^{N}\sup_{\theta\in N_0, u\le L_\theta}\left\|\frac{\partial}{\partial\theta_u}\left(\Delta(\widetilde{X}_i^*, \theta_0) - \frac{\partial^2}{\partial\theta_u\partial\theta'}g(X_i^*, \theta)\right)\right\| > K\right) > N^{-a}\right) = 0,$$

(iii) $$\lim_{N\to\infty} N^a P\left(P^*\left(\|\theta_{N,j-1}^* - \theta_N^*\| > N^{-c}\right) > N^{-a}\right) = 0,$$

(iv) $$\lim_{N\to\infty} N^a P\left(P^*\left(\|\theta_N^* - \hat{\theta}_N\| > N^{-c}\right) > N^{-a}\right) = 0, \quad \text{and}$$

(v) $$\lim_{N\to\infty} N^a P\left(P^*\left(\|\hat{\theta}_N - \theta_0\| > N^{-c}\right) > N^{-a}\right) = 0$$

for $j = 1, \ldots, k$ and for some $K < \infty$. Condition (i) holds using Assumption 5 by Lemma 6(a) with $p = \min\{q_2, q_3\}$ for q_3 as in Assumption 5, (ii) holds using Assumptions 3 and 5 by Lemma 6(d) with $p = \min\{q_2, q_4\}$ for q_2 and q_4 as in Assumptions 3 and 5, (iv) holds by Lemma 9 or Lemma 10, (v) holds by Lemma 3 or 4, (iii) holds for $j = 1$ by Lemma 9 or Lemma 10 because $\theta_{N,0}^* - \hat{\theta}_N$, and (iii) holds for $j = 2, \ldots, k$ by recursively applying the first result of part (a) of Theorem 1 with $k = j-1$, which holds because the proof of Theorem 1(a) for $\theta_{N,j-1}^*$ only relies on the result of the present Lemma holding for $Q_{N,i}^{*,GN}$ for $i \le j-2$. *Q.E.D.*

REFERENCES

AITCHISON, J., AND S. D. SILVEY (1958): "Maximum Likelihood Estimation of Parameters Subject to Constraint," *Annals of Mathematical Statistics*, 29, 813–828.

ANDREWS, D. W. K. (1987): "Asymptotic Results for Generalized Wald Tests," *Econometric Theory*, 3, 348–358.

——— (2000): "Equivalence of the Higher-order Asymptotic Efficiency of k-step and Extremum Statistics," Cowles Foundation Discussion Paper No. 1269, Yale University. Available at http://cowles.econ.yale.edu.

——— (2001a): "Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators," Cowles Foundation Discussion Paper No. 1230R, Yale University. Available at http://cowles.econ.yale.edu.

——— (2001b): "Higher-order Improvements of the Parametric Bootstrap for Markov Processes," unpublished manuscript, Cowles Foundation, Yale University.

ANDREWS, D. W. K., AND M. BUCHINSKY (2000): "A Three-step Method for Choosing the Number of Bootstrap Repetitions," *Econometrica*, 68, 23–51.

BHATTACHARYA, R. N. (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.

BROWN, B. W., AND S. MAITAL (1981): "What Do Economists Know? An Empirical Study of Experts' Expectations," *Econometrica*, 49, 491–504.

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.

DAVIDSON, R., AND J. G. MACKINNON (1999): "Bootstrap Testing in Nonlinear Models," *International Economic Review*, 40, 487–508.

DENNIS, J. E., AND R. B. SCHNABEL (1983): *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, New Jersey: Prentice-Hall.

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.

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. (1985): "Resampling a Coverage Process," *Stochastic Process Applications*, 19, 259–269.

——— (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 C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. New York: Academic Press.

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

HALL, P., J. L. HOROWITZ, AND B.-Y. JING (1995): "On Blocking Rules for the Bootstrap," *Biometrika*, 82, 561–574.

HANSEN, L. P., AND R. J. HODRICK (1980): "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, 88, 829–853.

HANSEN, L. P., AND K. J. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

——— (1996): "Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors," *Journal of Business and Economic Statistics*, 14, 53–68.

INOUE, A., AND M. SHINTANI (2000): "Bootstraping GMM Estimators for Time Series," unpublished working paper, Department of Economics, Vanderbilt 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.

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.

——— (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.

MCCALLUM, B. T. (1979): "Topics Concerning the Formulation, Estimation, and Use of Macroeconometric Models with Rational Expectations," *American Statistical Association Proceedings of the Business and Economic Statistics Section*, 65–72.

Ortega, J. M., and W. C. Rheinboldt (1970): *Iterative Solution of Nonlinear Equations in Several Variables*. New York: Academic Press.

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. Prague: Charles University.

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

Rothenberg, T. J. (1984): "Approximating the Distributions of Econometric Estimators and Test Statistics," in *Handbook of Econometrics*, Vol. 2, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland, 881–935.

Shorack, G. R. (1982): "Bootstrapping Robust Regression," *Communications in Statistics–Theory and Methods*, 11, 961–972.

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

Zvingelis, Y. (2001): "On Bootstrap Coverage Probability with Dependent Data," forthcoming in *Computer-aided Econometrics*, ed. by D. Giles. New York: Marcel Dekker.