Econometrica, Vol. 64, No. 4 (July, 1996), 891–916

BOOTSTRAP CRITICAL VALUES FOR TESTS BASED ON GENERALIZED-METHOD-OF-MOMENTS ESTIMATORS

BY PETER HALL AND JOEL L. HOROWITZ[1]

Monte Carlo experiments have shown that tests based on generalized-method-of-moments estimators often have true levels that differ greatly from their nominal levels when asymptotic critical values are used. This paper gives conditions under which the bootstrap provides asymptotic refinements to the critical values of t tests and the test of overidentifying restrictions. Particular attention is given to the case of dependent data. It is shown that with such data, the bootstrap must sample blocks of data and that the formulae for the bootstrap versions of test statistics differ from the formulae that apply with the original data. The results of Monte Carlo experiments on the numerical performance of the bootstrap show that it usually reduces the errors in level that occur when critical values based on first-order asymptotic theory are used. The bootstrap also provides an indication of the accuracy of critical values obtained from first-order asymptotic theory.

KEYWORDS: Block bootstrap, asymptotic refinement, Edgeworth expansion, dependent data.

1. INTRODUCTION

THE GENERALIZED METHOD of moments (GMM) estimates parameters that are identified through moment conditions stating that the mean of a vector-valued random function of the parameters is zero. Under regularity conditions, GMM estimators are $n^{1/2}$-consistent and asymptotically normal (Hansen (1982)). However, Monte Carlo experiments have revealed that first-order asymptotic theory often provides poor approximations to the distributions of test statistics obtained from GMM estimators. It is not unusual for the true and nominal levels of the test of overidentifying restrictions and of t tests of hypotheses about parameters to differ greatly from one another when asymptotic critical values are used (Kocherlakota (1990), Tauchen (1986)).

This paper investigates the ability of the bootstrap to provide improved critical values for the test of overidentifying restrictions (henceforth the J test) and t tests based on GMM. The bootstrap amounts to treating the estimation data as if they were the population and carrying out a Monte Carlo experiment in which pseudo data are generated by randomly sampling the estimation data (bootstrap sampling). The distributions of test statistics are estimated by their empirical distributions under bootstrap sampling. When the estimation data are an independent random sample from some distribution, the bootstrap often

[1] We thank Narayana Kocherlakota, Whitney Newey, Gene Savin, and Kai Wang for helpful discussion and comments. Part of this research reported was carried out while Joel L. Horowitz was a visiting fellow at the Centre for Mathematics and Its Applications, Australian National University. The research of Joel L. Horowitz was supported in part by NSF Grants DMS-9208820 and SBR-9307677.

provides critical values for test statistics that are more accurate than those provided by asymptotic theory (Beran (1988), Hall (1986a, 1992), Singh (1981)). In some cases, the use of bootstrap critical values instead of asymptotic ones produces spectacular reductions in the finite-sample errors in the levels of test statistics (Horowitz (1994)).

The situation is more complicated when the data are dependent and (as in GMM) one does not have a structural model that reduces the data-generation process to a transformation of independent random variables (e.g., an AR model as in Bose (1988)). The bootstrap sample must be drawn in a way that suitably captures the dependence of the data-generation process, but this dependence cannot be replicated exactly in the bootstrap sample. As a result, correct bootstrap versions of the J and t statistics cannot be obtained by applying the familiar formulae for these statistics to the bootstrap sample. To obtain asymptotic refinements, it is necessary to develop special bootstrap versions of the statistics. These must have the same distributions as the sample versions through $O_p(n^{-1})$.

A further source of complication is establishing the existence of the Edgeworth expansions that are used to show that the bootstrap provides improvements over first-order asymptotic theory. Proving that Edgeworth expansions exist with dependent data requires satisfying regularity conditions that are stronger and more complex than those required with independent data.

When the data are dependent, the bootstrap can be implemented by dividing the sample into blocks and sampling the blocks independently with replacement. The blocks, whose lengths increase with increasing size of the estimation data set, may be nonoverlapping (Carlstein (1986)) or overlapping (Künsch (1989)). Hall (1985) also suggested these approaches in the context of spatial data. Lahiri (1992) showed that with dependent data and Künsch's blocking scheme, the bootstrap provides second-order asymptotic refinements to the distributions of normalized functions of sample moments and, if the data are m-dependent, studentized sample moments. These results are not applicable to GMM test statistics, however, for several reasons. First, the J and t statistics are not exact functions of sample moments, although they can be approximated by functions of sample moments, as will be discussed below. Second, the J and t statistics are studentized, but an assumption of m dependence is undesirable in many economic settings. Third, asymptotic refinements for J and symmetrical t tests are of size $O_p(n^{-1})$ (third order), not $O_p(n^{-1/2})$ (second order) as considered by Lahiri (1992).

This paper gives conditions under which the bootstrap provides asymptotic refinements to the critical values of J and symmetrical t tests in GMM estimation with dependent data.[2] It will be clear from the discussion that the results also apply to independent data, coverage probabilities of symmetrical

[2] Brown and Newey (1992) give conditions under which the bootstrap provides asymptotically valid critical values for GMM test statistics with independent data. Brown and Newey do not treat either dependent data or the possibility of obtaining asymptotic refinements.

confidence intervals, and, with obvious modifications, one-sided and equal-tailed t tests and confidence intervals. We use the block bootstrap with nonoverlapping blocks. We conjecture that similar results can be obtained with Künsch's blocking scheme, which is more difficult to analyze owing to its use of overlapping blocks.

The paper also reports the results of a small-scale Monte Carlo investigation of the numerical performance of the bootstrap for the J and symmetrical t tests. For the models and sample sizes that were investigated, the bootstrap usually reduces but does not eliminate the finite-sample distortions of level that occur when asymptotic critical values are used. The bootstrap also provides an indication of the accuracy of critical values obtained from first-order asymptotic theory. The Monte Carlo results reported here are quite limited. Further investigation of the numerical performance of the bootstrap for GMM is needed but is beyond the scope of this paper. Here, our main objective is to establish the theoretical basis for using the bootstrap to obtain asymptotic refinements of critical values for GMM test statistics with dependent data. Because special formulae are needed for the bootstrap versions of the test statistics, the development of theory must precede Monte Carlo or other investigations of numerical performance. The results presented here provide the theoretical foundation for further numerical as well as analytical investigation of the use of the bootstrap for GMM estimators with dependent data.

The remainder of the paper is organized as follows. Section 2 gives a heuristic explanation of why the bootstrap yields asymptotic refinements to the critical values of test statistics when the data are independent. This material is not new but provides intuition for the conceptually similar but technically more complicated analysis that is required with dependent data. Section 3 presents the sample and bootstrap versions of the J and t statistics, gives our assumptions, and provides theorems stating the theoretical results. Section 4 presents the results of the Monte Carlo investigation. Concluding comments are presented in Section 5. All proofs are in the Appendix.

2. WHY THE BOOTSTRAP PROVIDES ASYMPTOTIC REFINEMENTS

This section gives a heuristic explanation of why the bootstrap provides asymptotic refinements to the critical values of test statistics. It is assumed in this section that the data are a random sample of size n from a probability distribution whose cumulative distribution function (CDF) is F. The empirical distribution function based on the sample is denoted by F_n. To minimize the length and complexity of the discussion we consider only the symmetrical t test.

Let θ be a parameter and T_n be the t statistic for testing H_0: $\theta = \theta_0$. Suppose the exact finite-sample CDF of T_n is $H_n(z, F) \equiv P(T_n \leq z)$. The symmetrical t test rejects H_0 at the α level if $|T_n| > z_\alpha$, where the critical value, z_α, satisfies $H_n(z_\alpha, F) - H_n(-z_\alpha, F) = 1 - \alpha$.

Under regularity conditions, H_n has an asymptotic expansion of the form

$$H_n(z,F) = \Phi(z) + n^{-1/2}p_1(z,F) + n^{-1}p_2(z,F) + o(n^{-1}) \tag{2.1}$$

uniformly over z, where Φ is the standard normal CDF, p_1 and p_2 are functionals of (z,F), $p_1(z,F)$ is an even function of z for each F, $p_2(z,F)$ is an odd function of z, and $p_2(z,F_n) \to p_2(z,F)$ almost surely as $n \to \infty$ uniformly over z. It follows from (2.1) and the symmetry of Φ, p_1, and p_2 that for any $z > 0$

$$P(|T_n| > z) = 2\left[1 - \Phi(z) - n^{-1}p_2(z,F)\right] + o(n^{-1}). \tag{2.2}$$

Now consider the bootstrap. With independent data, the bootstrap samples a population whose CDF is F_n. Let $\hat{\theta}_n$ be a consistent estimate of θ. The bootstrap analog of H_0 is $H_0^*: \theta = \hat{\theta}_n$. Let T_n^* be the t statistic for testing H_0^* under bootstrap sampling. The CDF of T_n^* conditional on the estimation data is $H_n^*(z,F_n)$. Under regularity conditions

$$H_n^*(z,F_n) = \Phi(z) + n^{-1/2}p_1(z,F_n) + n^{-1}p_2(z,F_n) + o_p(n^{-1}) \tag{2.3}$$

uniformly over z. The leading terms of (2.1) and (2.3) are identical. Therefore, because p_1 is even and $p_2(z,F_n) \to p_2(z,F)$ almost surely, the distributions of $|T_n|$ and $|T_n^*|$ are identical through $O_p(n^{-1})$. This is the source of the bootstrap's ability to provide asymptotic refinements.

To obtain the refinement for the critical value of the symmetrical t test, let z_α^* denote the α-level critical value for testing H_0^*.[3] This can be computed with arbitrary accuracy by repeated bootstrap sampling. It follows from (2.3) and the almost sure convergence of $p_2(z_\alpha^*, F_n) - p_2(z_\alpha^*, F)$ to zero that

$$2\left[1 - \Phi(z_\alpha^*) - n^{-1}p_2(z_\alpha^*, F)\right] = \alpha + o_p(n^{-1}). \tag{2.4}$$

Combining (2.2) and (2.4) yields $z_\alpha^* = z_\alpha + o_p(n^{-1})$. Therefore,

$$P(|T_n| > z_\alpha^*) = \alpha + o(n^{-1}).$$

Thus, with the bootstrap critical value z_α^*, the level of the symmetrical t test is correct through $O(n^{-1})$. In contrast, first-order asymptotic theory ignores all but the leading term in (2.1). Therefore, when critical values based on first-order asymptotic theory are used, the error in the level of the symmetrical t test is $O(n^{-1})$.

In the next section, the foregoing ideas are formalized and extended to dependent data and the J test. The technical details of the formal treatment are complex, but the essential ideas are the same as those just outlined.

[3] Since F_n is the CDF of a discrete random variable, the exact α-level critical value of T_n^* may not exist. However, as discussed by Hall (1992, Appendix I), the error made by ignoring discreteness is an exponentially decreasing function of n and, therefore, can be ignored.

3. THE MAIN RESULTS

a. Estimators and Test Statistics

We consider GMM estimation based on the moment condition $Eg(X,\theta)=0$, where $X\in\Re^{L_x}$ is a $L_x\times 1$ random variable, g is a $L_g\times 1$ function, θ is a $L_\theta\times 1$ parameter whose true but unknown value is θ_0, and $L_g\geq L_\theta$. Denote the data by $\{X_i: i=1,\ldots,n\}$. We assume that $\{X_i\}$ is a stationary, ergodic stochastic process and that $Eg(X_i,\theta_0)g(X_j,\theta_0)'=0$ if $|i-j|>\kappa$ for some integer $\kappa<\infty$.[4] Set $N=n-\kappa$. Write the sample as $\chi\equiv\{\tilde{X}_i: i=1,\ldots,N\}$, where $\tilde{X}_i=\{X_i',\ldots,X_{i+\kappa}'\}'$. This makes the estimator of the asymptotically optimal weight matrix a sample moment, which enables us to use results of Götze and Hipp (1983) to prove the existence of Edgeworth expansions of the distributions of the J and t statistics.[5]

We consider two forms of the GMM estimator, one with a fixed weight matrix and one with an estimate of the asymptotically optimal weight matrix. In the first form, the estimator, $\hat{\theta}_N$, solves

$$\min_{\theta\in\Theta} J_N(\theta)\equiv\left[N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right]'\Omega\left[N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right], \tag{3.1}$$

where Θ is the parameter set and Ω is a $L_g\times L_g$, positive-semidefinite, symmetrical matrix of constants. In the second form, $\hat{\theta}_N$ solves

$$\min_{\theta\in\Theta} J_N(\theta,\tilde{\theta}_N)\equiv\left[N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right]'\Omega_N(\tilde{\theta}_N)\left[N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right], \tag{3.2}$$

where $\tilde{\theta}_N$ solves (3.1),

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

and $H(X_i,X_{i+j},\theta)=g(X_i,\theta)g(X_{i+j},\theta)'+g(X_{i+j},\theta)g(X_i,\theta)'$. Equation (3.2) can be iterated by replacing $\tilde{\theta}_N$ at each iteration with the solution to (3.2) obtained in the previous iteration. Although we do not treat multistep GMM

[4] See Hansen (1982) and Hansen and Singleton (1982) for examples of applications that satisfy this condition. The condition is not equivalent to assuming m-dependence because it does not restrict the covariances of products of components of $g(X_i,\theta_0)$ (e.g., ARCH). It is an open question whether the bootstrap delivers asymptotic refinements when $\{g(X_i,\theta_0)\}$ has an unknown covariance structure. The main problem is that when the covariance is unrestricted, the estimator of the GMM covariance matrix is not a function of sample moments and converges at a rate that is slower than $n^{-1/2}$ (Andrews (1991)). Existing theory of Edgeworth expansions with dependent data (Götze and Hipp (1983)) does not apply to such estimators.

[5] The cost of this notational device is that only $n-\kappa$ observations are used to estimate θ. In Lahiri (1992), where asymptotic refinements through $O(n^{-1/2})$ are obtained for the t test of a hypothesis about a population mean, this cost is avoided by making an approximation whose error is $O(n^{-7/12})$. This error is too large for the $O(n^{-1})$ refinements developed here.

estimators explicitly, it is not difficult to show that the bootstrap provides asymptotic refinements to critical values for tests based on these estimators.

To obtain the t and J statistics, let $\overline{W}_N(\theta) = \Omega_N(\theta)^{-1}$ and $\Omega_0 = [E\overline{W}_N(\theta_0)]^{-1}$. Define the $L_g \times L_\theta$ matrices $D = E[\partial g(X, \theta_0)/\partial\theta]$ and

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

Define

$$\sigma = (D'\Omega D)^{-1} D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} \tag{3.3}$$

if $\hat{\theta}_n$ solves (3.1) and

$$\sigma = (D'\Omega_0 D)^{-1} \tag{3.4}$$

if $\hat{\theta}_N$ solves (3.2). Let σ_N be the consistent estimator of σ that is obtained by replacing D and Ω_0 in (3.3) and (3.4) by D_N and $\Omega_N(\hat{\theta}_N)$. In addition, let $(\sigma_N)_{rr}$ be the (r,r) component of σ_N, and let θ_r and $\hat{\theta}_{Nr}$ be the rth components of θ and $\hat{\theta}_N$, respectively.

The t statistic for testing $H_0\colon \theta_r = \theta_{0r}$ is

$$T_{Nr} = N^{1/2}\left(\hat{\theta}_{Nr} - \theta_{0r}\right)/[(\sigma_N)_{rr}]^{1/2}.$$

The J test statistic can be written in the form $J_N = K_N(\hat{\theta}_N)'K_N(\hat{\theta}_N)$, where

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

Under $H_0\colon \theta_r = \theta_{0r}$, T_{Nr} is asymptotically distributed as $N(0,1)$. If $L_g > L_\theta$ and the overidentifying restrictions hold, J_N is asymptotically χ^2 distributed with $L_g - L_\theta$ degrees of freedom. In subsection 3d, we obtain higher-order approximations to the distributions of J_N and T_{Nr}.

b. *The Bootstrap Versions of J_N and T_{Nr}*

The bootstrap treats the estimation data as if they were the population and, by repeatedly sampling the data and computing appropriate versions of J_N and T_{Nr} from the resulting bootstrap samples, develops bootstrap empirical distributions of these statistics. The bootstrap estimates of the α-level critical values of the J and symmetrical t tests are the $1-\alpha$ quantiles of the empirical distributions of the bootstrap versions of J_N and $|T_{Nr}|$.

To obtain asymptotic refinements, bootstrap sampling must take account of the dependence of $\{X_i\}$ in the population data generation process. As was discussed in Section 1, we resample the data in nonoverlapping blocks. Let b be the number of blocks, l the length of each block, and $bl = N$. Both b and l may depend on N. Block 1 consists of $\tilde{X}_1, \ldots, \tilde{X}_l$, block 2 of $\tilde{X}_{l+1}, \ldots, \tilde{X}_{2l}$, etc. The bootstrap is implemented by sampling b blocks randomly with replacement to form a bootstrap sample.

Let $\{\tilde{X}_i^*\} = \{(X_i^{*\prime}, \ldots, X_{i+\kappa}^{*\prime})' : i = 1, \ldots, N\}$ denote the resulting bootstrap sample. Let E^* denote the expectation relative to the distribution of the bootstrap sample conditional on the original sample, χ. Define $g^*(x, \theta) = g(x, \theta) - E^*g(X, \hat{\theta}_N)$. The bootstrap version of the moment condition $Eg(X, \theta) = 0$ is $E^*g^*(X, \theta) = 0$. The bootstrap version is recentered relative to the population version because, except in special cases, there is no θ such that $E^*g(X, \theta) = 0$ when $L_g > L_\theta$. Without recentering, the bootstrap would implement a moment condition that does not hold in the population the bootstrap samples. Consequently, the bootstrap and sample versions of J_N and T_{Nr} would have different asymptotic distributions.

With recentering, the bootstrap GMM estimator, θ_N^*, solves

$$\min_{\theta \in \Theta} J_N^*(\theta) = \left[N^{-1} \sum_{i=1}^{N} g^*(X_i^*, \theta)\right]' \Omega \left[N^{-1} \sum_{i=1}^{N} g^*(X_i^*, \theta)\right] \tag{3.5}$$

if Ω is fixed, and

$$\min_{\theta \in \Theta} J_N^*\left(\theta, \tilde{\theta}_N^*\right) = \left[N^{-1} \sum_{i=1}^{N} g^*(X_i^*, \theta)\right]' \Omega_N^*\left(\tilde{\theta}_N^*\right) \left[N^{-1} \sum_{i=1}^{N} g^*(X_i^*, \theta)\right], \tag{3.6}$$

if the two-step estimator is used, where $\tilde{\theta}_N^*$ solves (3.5),

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

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

We now obtain the bootstrap versions of J_N and T_{Nr}. The bootstrap analog of D_N is

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

Define $W_N^* = \Omega_N^*(\theta_N^*)^{-1}$. Set $\overline{\Omega}_N^* = \Omega$ if $\hat{\theta}_N$ solves (3.1) and $\overline{\Omega}_N^* = (W_N^*)^{-1}$ if $\hat{\theta}_N$ solves (3.2). $\overline{\Omega}_N^*$ is the bootstrap analog of Ω if $\hat{\theta}_N$ solves (3.1) and of $\Omega_N(\hat{\theta}_N)$ if $\hat{\theta}_N$ solves (3.2). The bootstrap analog of σ_N is

$$\sigma_N^* = \left(D_N^{*\prime}\overline{\Omega}_N^* D_N^*\right)^{-1} D_N^{*\prime}\overline{\Omega}_N^* W_N^* \overline{\Omega}_N^* D_N^* \left(D_N^{*\prime}\overline{\Omega}_N^* D_N^*\right)^{-1}.$$

Let $(\sigma_N^*)_{rr}$ denote the (r, r) element of σ_N^*. Then a bootstrap version of T_{Nr} can be obtained by replacing all quantities in the formula for T_{Nr} with their bootstrap analogs. This yields

$$\tilde{T}_{Nr} = N^{1/2}\left(\theta_{Nr}^* - \hat{\theta}_{Nr}\right)/[(\sigma_N^*)_{rr}]^{1/2}.$$

A bootstrap version of $K_N(\hat{\theta}_N)$ can be obtained by replacing all quantities in the formula for $K_N(\hat{\theta}_N)$ with their bootstrap analogs. This yields

$$\tilde{K}_N(\theta_N^*) \equiv \Omega_N^*(\theta_N^*)^{1/2} N^{-1/2} \sum_{i=1}^{N} g^*(X_i^*, \theta_N^*).$$

Unfortunately, the asymptotic distributions of $\tilde{T}_{Nr}$ and $\tilde{K}_{Nr}(\theta_N^*)$ conditional on χ are not the same to $O(N^{-1})$ as the asymptotic distributions of T_{Nr} and $K_{Nr}(\hat{\theta}_N)$. To see why, define

$$\begin{aligned}\tilde{W}_N &= N^{-1}E^* \sum_{i=1}^{N} \sum_{j=1}^{N} g^*\left(X_i^*, \hat{\theta}_N\right) g^*\left(X_j^*, \hat{\theta}_N\right), \\ &= N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{l} \sum_{k=1}^{l} g^*\left(X_{il+j}, \hat{\theta}_N\right) g^*\left(X_{il+k}, \hat{\theta}_N\right)'.\end{aligned}$$

Set $\overline{\Omega}_N = \Omega$ if $\hat{\theta}_N$ solves (3.1) and $\overline{\Omega}_N = \overline{W}_N(\hat{\theta}_N)^{-1}$ if θ_N solves (3.2). Define

$$\bar{\sigma}_N = \left(D_N' \overline{\Omega}_N D_N\right)^{-1} D_N' \overline{\Omega}_N \overline{W}_N\left(\hat{\theta}_N\right) \overline{\Omega}_N D_N \left(D_N' \overline{\Omega}_N D_N\right)^{-1},$$

and

$$\tilde{\sigma}_N = \left(D_N' \overline{\Omega}_N D_N\right)^{-1} D_N' \overline{\Omega}_N \tilde{W}_N \overline{\Omega}_N D_N \left(D_N' \overline{\Omega}_N D_N\right)^{-1}.$$

$\overline{W}_N(\hat{\theta}_N)$ is the bootstrap analog of Ω_0^{-1}, and $\bar{\sigma}_N$ is the bootstrap analog of σ.

By applying the familiar Taylor series expansions of asymptotic theory to the first-order conditions for minimizing $J_N^*(\theta)$ and $J_N^*(\theta, \tilde{\theta}_N^*)$, it may be shown that the variance of the asymptotic distribution of $N^{1/2}(\theta_N^* - \hat{\theta}_N)$ conditional on χ is $\tilde{\sigma}_N$, not $\bar{\sigma}_N$. Therefore, $\tilde{T}_{Nr}$ does not have a conditional asymptotic variance of 1. In contrast, the variance of the asymptotic distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$ is σ, so T_{Nr} has an asymptotic variance of 1. Similarly, a Taylor series expansion of $\tilde{K}_N(\theta_N^*)$ about $\theta_N^* = \hat{\theta}_N$ reveals that to $O(N^{-1})$ $\tilde{K}_N(\theta_N^*)'\tilde{K}_N(\theta_N^*)$ is not asymptotically chi-square distributed conditional on χ because the variance of the asymptotic distribution of $N^{-1/2}\sum_{i=1}^N g^*(X_i^*, \hat{\theta}_N)$ is $\tilde{W}_N$, not $\overline{W}_N(\hat{\theta}_N)$. In contrast, the variance of the asymptotic distribution of $N^{-1/2}\sum_{i=1}^N g(X_i, \theta_0)$ is Ω_0^{-1}. Because the asymptotic distributions $\tilde{T}_{Nr}$ and $\tilde{K}_N(\theta_N^*)$ conditional on χ are not the same to $O(N^{-1})$ as the asymptotic distributions of T_{Nr} and $K_N(\hat{\theta}_N)$, $\tilde{T}_{Nr}$ and $\tilde{K}_N(\theta_N^*)$ cannot be used to obtain asymptotic refinements to critical values for J_N and T_{Nr}.

The source of this difficulty is that $\tilde{W}_N$ contains terms of the form $E^*[g^*(X_i^*, \hat{\theta}_N)g^*(X_j^*, \hat{\theta}_N)']$ for all i and j in the same block, whereas Ω_0^{-1} contains terms of the form $E[g(X_i, \theta_0)g(X_j, \theta_0)']$ for all i and j such that $|i-j| \leq \kappa$. In other words, the block bootstrap does not replicate the dependence of the data generation process. This difficulty cannot be overcome by modifying σ_N^* and $\Omega_N^*(\theta_N^*)$ to include only cross terms that are in $\tilde{W}_N$. These terms would then be present in the Edgeworth expansions of the distributions of the bootstrap versions but not the sample versions of J_N and T_{Nr}.

We solve this problem by incorporating correction factors into $\tilde{T}_{Nr}$ and $\tilde{K}_N(\theta_N^*)$. These factors replace $\overline{W}_N$ with $\tilde{W}_N$ in the asymptotic distributions of T_{Nr} and J_N without affecting the higher-order terms of the Edgeworth expansions of the statistics. The correction factor for the bootstrap t statistic is

$$\tau_{Nr} = [(\bar{\sigma}_N)_{rr}/(\tilde{\sigma}_N)_{rr}]^{1/2}.$$

The correct bootstrap version of T_{Nr} is

$$T_{Nr}^* = \tau_{Nr} N^{1/2}\left(\theta_{Nr}^* - \hat{\theta}_{Nr}\right)/(\sigma_N^*)_{rr}.$$

The correction consists of multiplying $\tilde{T}_{Nr}$ by the factor needed to make the variance of its asymptotic distribution 1 conditional on χ.

The correction factor for the bootstrap version of $K_N(\hat{\theta}_N)$ is similarly motivated though algebraically more complex. To specify this factor, define

$$M_N = I_g - \overline{W}_N\left(\hat{\theta}_N\right)^{-1/2} D_N\left[D_N'\overline{W}_N\left(\hat{\theta}_N\right)^{-1} D_N\right]^{-1} D_N'\overline{W}_N\left(\hat{\theta}_N\right)^{-1/2}$$

and

$$V_N = M_N\overline{W}_N\left(\hat{\theta}_N\right)^{-1/2}\tilde{W}_N\overline{W}\left(\hat{\theta}_N\right)^{-1/2}M_N.$$

Let V_N^+ be the Moore-Penrose generalized inverse of V_N. The correct bootstrap version of $K_N(\hat{\theta}_N)$ is

$$K_N^*(\theta_N^*) = (V_N^+)^{1/2}\Omega_N^*(\theta_N^*)^{1/2}N^{-1/2}\sum_{i=1}^{N} g^*(X_i^*, \theta_N^*).$$

The correct bootstrap version of the J statistic is $J_N^* = K_N^*(\theta_N^*)'K_N^*(\theta_N^*)$.

We show in subsection 3d that bootstrap critical values for J_N^* and $|T_{Nr}^*|$ differ from exact finite-sample critical values for J_N and $|T_{Nr}|$ by terms of size $o_p(n^{-1})$.

c. *Regularity Conditions*

Our approach to obtaining asymptotic refinements from the bootstrap consists of three main steps. First, we show that J_N, T_{Nr}, and their bootstrap analogs can be approximated by functions of sample (or bootstrap sample) moments with an approximation error whose size is $o_p(n^{-1})$. We then use results of Bhattacharya (1987), Chandra and Ghosh (1979), and Götze and Hipp (1983) to obtain Edgeworth expansions through $O_p(n^{-1})$ of the distributions of the sample and bootstrap forms of J_N and T_{Nr}. Finally, we use the Edgeworth expansions to show that with bootstrap critical values, the levels of the J and symmetrical t tests are correct through $O_p(n^{-1})$.

To carry out the first step, we must insure that with probability $1 - o(n^{-1})$, $\hat{\theta}_N$ is in a neighborhood of θ_0 whose radius decreases at a sufficiently rapid rate as n increases. To do this we restrict the dependence among the components of X at large lags (Assumption 1 below) and the thicknesses of the tails of the distributions of $g(X, \theta_0)$ and its derivatives (Assumption 4). To carry out the second step, we must satisfy regularity conditions for the existence of Edgeworth expansions of functions of sample moments with dependent data (Götze and Hipp (1983)). These conditions require that $\{X_i: -\infty < i < \infty\}$ can be suitably approximated by an m-dependent process and that a strengthened form of the Cramér condition holds. Assumptions 1 and 6 accomplish this. We also assume that θ is identified, θ_0 is an interior point of a compact parameter set, X has at least one continuously distributed component, and $g(\cdot,\cdot)$ and its derivatives satisfy certain smoothness conditions (Assumptions 2, 3, and 5).

The assumptions are:

ASSUMPTION 1: *There is a sequence of iid vectors $\{\epsilon_i: i = -\infty,\ldots,\infty\}$ of dimension $L_\epsilon \geq L_x$ and a $L_x \times 1$ function h such that $X_i = h(\epsilon_i, \epsilon_{i-1}, \epsilon_{i-2},\ldots)$. There is a constant $d > 0$ such that for all $n = 1,2,\ldots$ and all $m > d^{-1}$*

$$\|h(\epsilon_n, \epsilon_{n-1}\epsilon_{n-2},\ldots) - h(\epsilon_n, \epsilon_{n-1}, \epsilon_{n-2},\ldots, \epsilon_{n-m}, 0, 0,\ldots)\| \leq d^{-1}\exp(-dm).$$

Assumption 1 implies, among other things, that $\{X_i\}$ is strongly mixing with a mixing parameter that is an exponentially decreasing function of the lag length. Assumption 1 is satisfied, for example, if h is a (possibly infinite-order) vector MA process with exponentially decreasing coefficients.

ASSUMPTION 2: *θ_0 is an interior point of the compact parameter set Θ and is the unique solution in Θ to the equation $Eg(X, \theta) = 0$.*

ASSUMPTION 3: *(a) $E\|g(X, \theta)\| < \infty$ for all $\theta \in \Theta$, where $\|\cdot\|$ is the Euclidean norm. (b) $Eg(X_1, \theta_0)g(X_{1+i}, \theta_0) = 0$ if $i > \kappa$ for some $\kappa < \infty$. (c) $E\sum_{i=0}^{\kappa}[g(X_1, \theta)g(X_{1+i}, \theta)' + g(X_{1+i}, \theta)g(X_1, \theta)']$ exists for all $\theta \in \Theta$. Its smallest eigenvalue is bounded away from zero uniformly over θ in an open sphere, N_0, centered on θ_0. (d) There is a function $C_g(x)$ such that $\|g(x, \theta_1) - g(x, \theta_2)\| \leq C_g(x)\|\theta_1 - \theta_2\|$ for any $\theta_1, \theta_2 \in \Theta$. (e) g is 4-times differentiable with respect to the components of θ everywhere in N_0. (f) Let $\bar{g}(x, \theta)$ be a vector containing the unique components of the derivatives of $g(x, \theta)$ through order 4 with respect to θ. There is a function $C^*(x)$ such that $\|\bar{g}(x, \theta_1) - \bar{g}(x, \theta_2)\| \leq C^*(x)\|\theta_1 - \theta_2\|$ for any $\theta_1, \theta_2 \in N_0$. (g) Let C denote C_g or C^*. Then $P[C(X) > z] = O(z^{-33})$ as $z \to \infty$.*

Let $f(\tilde{X}_1, \theta)$ be a vector containing the unique components of $g(X_1, \theta)$, $g(X_1, \theta)g(X_j, \theta)'$ $(1 \leq j \leq \kappa + 1)$, and their derivatives through order 4 with respect to the components of θ.

ASSUMPTION 4: *$f(\tilde{x}_1, \theta_0)$ is a Lipschitz continuous function of $\tilde{x}$. As $z \to \infty$, $P[\|f(\tilde{X}_1, \theta_0)\| > z] = O(z^{-33})$.*

ASSUMPTION 5: *X can be partitioned $(X^{(c)\prime}, X^{(d)\prime})'$, where $X^{(c)} \in \Re^c$ for some $c > 0$, the distributions of $X^{(c)}$ and $\partial g(X, \theta_0)/\partial\theta$ are absolutely continuous with respect to Lebesgue measure, and the distribution of $X^{(d)}$ is discrete. There need not be any discrete components of X, but there must be at least one continuous component.*

ASSUMPTION 6: *There exist $r > 0$ and $\delta > 0$ such that for all integers m satisfying $\delta^{-1} < m + 1 < N$ and all $t \in \Re^{\dim(f)}$ with $\|t\| > \delta$*

$$E\left|E\left\{\exp\left[it' \sum_{j=1}^{2m+1} f(\tilde{X}_j)\right]\middle|\epsilon_k : r < |m + 1 - k|\right\}\right| \leq \exp(-\delta).$$

Assumption 6 is a dependent-data version of the Cramér condition (Götze and Hipp (1983, 1992)). Götze and Hipp (1983, 1992) verify Assumption 6 for some simple models, but checking it is quite difficult in general. We know of no way to simplify it without imposing unacceptably severe restrictions on the data generation process. However, if the data are sampled independently from a probability distribution, Assumption 6 may be replaced by the standard Cramér condition of Edgeworth analysis (see, e.g., Bhattacharya and Ghosh (1978)).

d. The Results

This subsection presents theorems stating our theoretical results. The following additional notation is used. Let P denote the probability measure induced by the data-generation process and P^* the probability measure induced by bootstrap sampling conditional on the estimation data, χ. Let $f^*(\tilde{X}_i^*, \theta)$ be a vector containing the unique components of $g^*(\tilde{X}_i^*, \theta)$, $g^*(\tilde{X}_i^*, \theta)g^*(\tilde{X}_j^*, \theta)'$ $(i \leq j \leq i + \kappa)$, and their derivatives through order 4 with respect to the components of θ. Define $S_N = N^{-1}\Sigma_{i=1}^N f(\tilde{X}_i, \theta_0)$, $S_N^* = N^{-1}\Sigma_{i=1}^N f^*(\tilde{X}_i^*, \hat{\theta}_N)$, and $S = E(S_N)$. Set $\Psi_N = N^{1/2}(S_N - S)$ and $\Psi_N^* = N^{1/2}[S_N^* - E^*(S_N^*)]$. Let Ψ_{Nj} and Ψ_{Nj}^*, respectively, denote the jth components of Ψ_N and Ψ_N^*.

The following two theorems give conditions under which the distributions of J_N, T_{Nr}, J_N^*, and T_{Nr}^* have Edgeworth expansions through $O(n^{-1})$.

THEOREM 1: *Let Assumptions 1–6 hold. Let ν be a vector of moments of the form*

$$\lim_{N\to\infty} EN^{\alpha(m)} \prod_{k=1}^{m} \Psi_{Nj_k} \tag{3.7}$$

where $2 \leq m \leq 6$, $\alpha(m) = 0$ if m is even and $1/2$ if m is odd, and Ψ_{Nj_k} is the j_k component of the vector Ψ_N. Then

$$\sup_z \left|P(T_{Nr} \leq z) - \left[1 + \sum_{i=1}^{2} N^{-i/2}\pi_i(\delta, \nu)\right]\Phi(z)\right| = o(n^{-1}), \tag{3.8}$$

where $\delta = d/dz$, π_i $(i = 1,2)$ *is a polynomial in* δ *whose coefficients are continuous functions of* ν, $\pi_1(\delta, \nu)\Phi(z)$ *is an even function of* z, *and* $\pi_2(\delta, \nu)\Phi(z)$ *is an odd function of* z. *In addition*

$$(3.9) \qquad \sup_z \left| P(J_N < z) - \int_{-\infty}^{z} d\{[1 + N^{-1}\pi_J(\xi, \nu)] P(\chi_\lambda^2 \le \xi)\} \right| = o(n^{-1}),$$

where χ_λ^2 *is a random variable that has the chi-square distribution with* $\lambda \equiv L_g - L_\theta$ *degrees of freedom, and* $\pi_J(\xi, \nu)$ *is a polynomial function of* ξ *whose coefficients are continuous functions of* ν.

THEOREM 2: *Let Assumptions* 1–6 *hold. Let* ν_N^* *be a vector of bootstrap moments of the form*

$$(3.10) \qquad E^* \tau_{Nr}^{\beta} N^{\alpha(m)} \prod_{k=1}^{m} \Psi_{Nj_k}^*,$$

where $\beta = 0, 1,$ *or* 3, *depending on the moment*, $2 \le m \le 6$, $\alpha(m) = 0$ *if* m *is even and* $1/2$ *if* m *is odd, and* $\Psi_{Nj_k}^*$ *is the* j_k *component of* Ψ_N^*. *Assume that* $l = N^\gamma$, *where* $11/50 \le \gamma \le 12/50$. *Define* π_i $(i = 1,2)$, π_J, δ, *and* χ_λ^2 *as in Theorem* 1. *Then except, possibly, if* χ *is contained in a set of probability* $o(N^{-1})$

$$(3.11) \qquad \sup_z \left| P^*(T_{Nr}^* \le z) - \left[1 + \sum_{i=1}^{2} N^{-i/2}\pi_i(\delta, \nu_N^*)\right]\Phi(z) \right| = o(n^{-1}).$$

For bootstrap moments ν_N^* *that are obtained by setting* $\beta = 0$ *in* (3.10)

$$(3.12) \qquad \sup_z \left| P^*(J_N^* < z) - \int_{-\infty}^{z} d\{[1 + N^{-1}\pi_J(\xi, \nu_N^*)] P(\chi_\lambda^2 \le \xi)\} \right| = o(n^{-1}).$$

To prove these theorems, we first show that J_N, T_{Nr}, J_N^*, and T_{Nr}^* are functions of sample or bootstrap-sample moments up to asymptotically negligible remainder terms. This is done in Propositions 1 and 2 of the Appendix. We then use results of Götze and Hipp (1983), Bhattacharya (1987), and Chandra and Ghosh (1979) to establish the existence of the Edgeworth expansions.

Our final theorem gives conditions under which bootstrap critical values for J_N^* and T_{Nr}^* yield levels for the J and symmetrical t tests that are correct through $O(n^{-1})$. Let $z_{J\alpha}^*$ and $z_{T\alpha}^*$, respectively, denote the α-level critical values of the bootstrap versions of the J and symmetrical t tests.

THEOREM 3: *Let Assumptions* 1–6 *hold. Assume that* $l = N^\gamma$, *where* $11/50 \le \gamma \le 12/50$. *Under* H_0: $\theta_r = \theta_{0r}$

$$(3.13) \qquad P(|T_{Nr}| > z_{T\alpha}^*) = \alpha + o(n^{-1}).$$

If $L_g > L_\theta$ *and the overidentifying restrictions hold*,

$$(3.14) \qquad P(J_N > z_{J\alpha}^*) = \alpha + o(n^{-1}).$$

This theorem follows from Theorems 1 and 2 by showing that the quantities defined in (3.7) and (3.10) differ by $o_p(1)$ as $n \to \infty$.

Theorems 2 and 3 assume that $\{X_i\}$ may be a dependent sequence. If $\{X_i\}$ is known to be an independent sequence, the theorems hold with $l = 1$. No blocking is needed, and the correction factors for J_N^* and T_{Nr}^* may be ignored.

It is well known that first-order asymptotic theory provides an especially poor approximation to the finite-sample distributions of instrumental variables estimators when the correlation between the instruments and the variables being instrumented is low ($D \approx 0$ in our notation). See Hillier (1985), Nelson and Startz (1990a, b), and Phillips (1983). The bootstrap does not solve this problem. The higher-order terms of the Edgeworth expansions of J_N and T_{Nr} include powers of $(D'\Omega D)^{-1}$ and $(D'\Omega_0 D)^{-1}$ for the one- and two-step estimators, respectively. If $D \approx 0$, the higher-order terms may dominate the lower-order ones for a given sample size, in which case the bootstrap may provide little reduction in the errors of the levels of test statistics or even make the errors worse.

4. MONTE CARLO EXPERIMENTS

The Monte Carlo experiments provide numerical evidence on the magnitudes of the finite-sample errors in the levels of the J and symmetrical t tests with asymptotic and bootstrap critical values. Each experiment consisted of specifying a data-generation process and then carrying out the following steps:

1. Sample the data-generation process. Compute $\hat{\theta}_N$ using the two-step estimator. Also compute J_N and T_{Nr}. Determine whether the null hypotheses of the J and symmetrical t tests are rejected using asymptotic critical values.
2. Obtain bootstrap critical values for the J and t tests. Determine whether the null hypotheses of the tests are rejected using bootstrap critical values.
3. Estimate the levels of the J and symmetrical t tests using asymptotic and bootstrap critical values by repeating steps 2–3 1000 times.

Bootstrap critical values were based on 100 replications of the bootstrap sampling process. Use of larger numbers of replications did not change the results substantially. See Hall (1986b) for a theoretical explanation of the ability of the bootstrap to produce satisfactory results with few replications of the bootstrap sampling process. The experiments were carried out in GAUSS using GAUSS pseudo-random number generators.

The model used in the experiments is a simplified version of an asset-pricing model. It is defined by the moment conditions

$$E\{\exp[\mu - \theta(X+Z) + 3Z] - 1\} = 0$$

and

$$EZ\{\exp[\mu - \theta(X+Z) + 3Z] - 1\} = 0,$$

where θ is the estimated parameter, $\theta_0 = 3$, μ is a known normalization constant, and X and Z are scalars. X is sampled independently from $N(0, s^2)$ with $s = 0.2$ or 0.4. Z is independent of X, has a marginal $N(0, s^2)$ distribution, and is either sampled independently from this distribution or follows an AR(1) process with first-order serial correlation coefficient $r_z = 0$.

The experiments investigate the levels of the J test of the single overidentifying restriction and the symmetrical t test of $H_0: \theta = 3$. The sample sizes are $n = 50$ and $n = 100$. The number of blocks is $n/5$ or $n/10$ if $r_z > 0$ and n if $r_z = 0$.

The results of the experiments are shown in Table I. The empirical levels of the tests are much larger than nominal when asymptotic critical values are used. With bootstrap critical values, the errors in the levels of the tests are usually smaller than with asymptotic critical values. This is especially true of the nominal 0.10-level and 0.05-level t tests. The differences between the empirical and nominal levels of these tests are 40–90 percent less with bootstrap critical values than with asymptotic ones. However, the bootstrap does not completely remove the level distortions of either the J or t tests. Using either asymptotic or bootstrap critical values, the distortions are smaller with $s = 0.2$ than $s = 0.4$. The block lengths have little effect on the empirical levels of the tests with bootstrap critical values.

The Bootstrap as an Indicator of the Accuracy of Asymptotic Approximations

Phillips and Park (1988) argue that although Edgeworth expansions do not necessarily improve the numerical quality of asymptotic approximations to the levels of tests, they can nonetheless provide information on whether first-order approximations are accurate. Specificially, first-order approximations are likely

TABLE I

LEVELS OF T AND J TESTS FOR A NONLINEAR MODEL WITH ASYMPTOTIC AND BOOTSTRAP CRITICAL VALUES[a]

				t Test with Asymp. Crit. Values			t Test with Boot. Crit. Values			J Test with Asymp. Crit. Values			J Test with Boot. Crit. Values		
n	s	r_z	b	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01	0.10	0.05	0.01
50	0.2	0.0	50	0.153	0.092	0.042	0.104	0.063	0.031	0.124	0.088	0.022	0.123	0.082	0.032
50	0.2	0.75	5	0.194	0.144	0.073	0.150	0.104	0.064	0.143	0.093	0.052	0.138	0.096	0.058
		0.75	10				0.147	0.092	0.057				0.126	0.078	0.046
50	0.4	0.0	50	0.260	0.193	0.114	0.165	0.123	0.062	0.228	0.171	0.102	0.197	0.149	0.095
50	0.4	0.75	5	0.242	0.180	0.099	0.156	0.125	0.069	0.217	0.168	0.113	0.183	0.140	0.109
		0.75	10				0.182	0.131	0.086				0.198	0.164	0.122
100	0.2	0.0	100	0.155	0.090	0.024	0.122	0.067	0.023	0.119	0.070	0.026	0.116	0.068	0.027
100	0.2	0.75	10	0.182	0.110	0.050	0.133	0.080	0.036	0.139	0.090	0.032	0.138	0.093	0.045
		0.75	20				0.125	0.065	0.026				0.124	0.089	0.045
100	0.4	0.0	100	0.224	0.150	0.073	0.161	0.102	0.054	0.194	0.132	0.067	0.164	0.113	0.063
100	0.4	0.75	10	0.216	0.150	0.086	0.141	0.096	0.057	0.188	0.132	0.072	0.156	0.124	0.076
		0.75	20				0.149	0.110	0.064				0.172	0.126	0.083

[a] Underline indicates that the empirical level is not significantly different from the nominal level at the 0.01 level.

to be inaccurate if higher-order terms through $O(n^{-1})$ in the expansion of the distribution of a test statistic are large compared to the leading term. Since bootstrap corrections to critical values are based on Edgeworth expansions, it is possible that the bootstrap can provide an indication of the accuracy of first-order approximations without the tedious algebra associated with analytic Edgeworth expansions. In particular, large differences between bootstrap and asymptotic critical values may indicate that first-order theory is inaccurate.

To investigate this possibility, we computed the 25th, 50th, and 75th percentiles of the empirical distributions of the nominal 0.05-level bootstrap critical values of the J and symmetrical t tests in the Monte Carlo experiments. Figure 1 plots these percentiles against the empirical levels of the tests based on asymptotic critical values. The plots include horizontal lines indicating the asymptotic critical values.

Figure 1 shows that the differences between bootstrap and asymptotic critical values are increasing functions of the errors in the levels of the asymptotic tests. Thus, the bootstrap is informative about the accuracy of first-order asymptotic approximations despite its inability to fully correct the errors in level caused by these approximations. This is especially true for the symmetrical t test. The 25th percentiles of bootstrap critical values of this test exceed the asymptotic critical value whenever the level of the asymptotic test exceeds roughly 0.08. The bootstrap is less informative about the J test, reflecting the fact that the bootstrap provides smaller reductions in the errors in the level of the J test than of the t test. The 25th percentiles of the bootstrap distribution of the J test are roughly equal to the asymptotic critical value in most of the experiments, but the 50th percentile of the bootstrap critical value exceeds the asymptotic critical value whenever the level of the asymptotic J test exceeds roughly 0.10.

Of course, caution is needed in interpreting these results. Bootstrap critical values are random variables, and random sampling errors may cause bootstrap critical values to differ from asymptotic ones, even if first-order approximations are accurate. Conversely, downward fluctuations of bootstrap critical values may mask serious errors in level. In principle, these problems might be dealt with by using the bootstrap to construct confidence intervals for critical values based on $O(n^{-1})$ Edgeworth expansions of the distributions of the t and J statistics. We do not attempt this task here, however.

5. CONCLUSIONS

This paper has given conditions under which the bootstrap provides asymptotic refinements to the critical values of t and J tests based on GMM estimators. We have given particular attention to the case of dependent data and have shown that with such data, the formulae for the bootstrap versions of the t and J statistics are different from the formulae that apply with the original data. By giving correct formulae for bootstrap versions of the GMM J and symmetrical t statistics with dependent data, this paper provides the theoretical foundation for further research into the ability of the bootstrap to provide

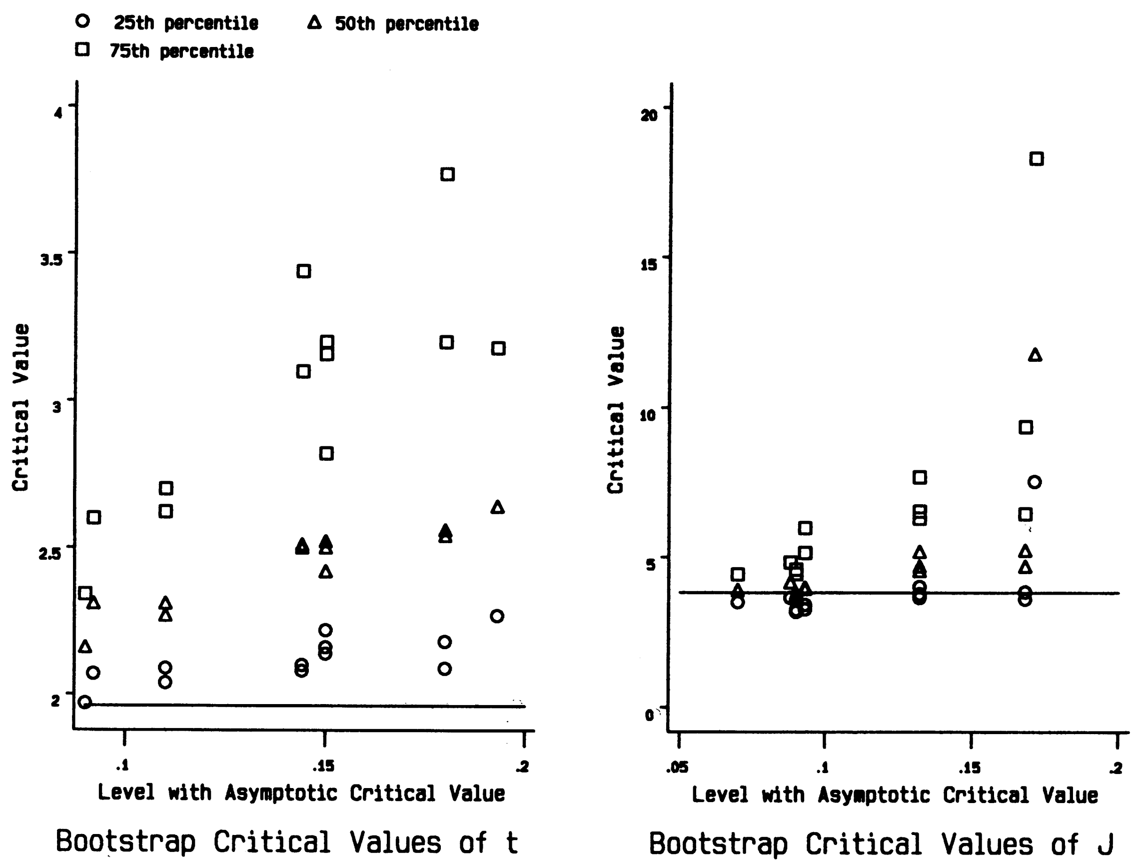

FIGURE 1—Percentiles of distributions of bootstrap critical values of t and J tests versus empirical levels based on asymptotic critical values.

improved finite-sample critical values for the J and t tests. This research should include further investigations of numerical performance, developing data-based methods for choosing block lengths in the block bootstrap for dependent data, and developing formal methods for using the bootstrap to detect serious inaccuracy in critical values obtained from first-order asymptotic approximations.

Centre for Mathematics and its Applications, Australian National University, Canberra, ACT 0200, Australia

and

Dept. of Economics, University of Iowa, Iowa City, IA 52242, U.S.A.

Manuscript received April, 1994; final revision received May, 1995.

APPENDIX

LEMMA 1: *Let $\{\xi_i: i = 1,2\ldots\}$ be a sequence of identically but not necessarily independently distributed random variables with $E(\xi) = 0$. Assume that: (a) As $z \to \infty$, $P(|\xi| > z) = O(z^{-33})$, and (b) ξ is strongly mixing with an exponentially decreasing mixing parameter. For $j = 1$ or 2, define*

$$R_{nj} \equiv n^{-1} \sum_{i=1}^{n} \xi_i^j.$$

For any ϵ such that $0 < \epsilon < 1/64$ and any $r > 0$

$$\lim_{n\to\infty} nP(|R_{n1}| > n^{-(2+\epsilon)/5}) = 0 \tag{A1}$$

and

$$\lim_{n\to\infty} nP(|R_{n2} - ER_{n2}| > r) = 0. \tag{A2}$$

PROOF: Only (A1) is proved; the proof of (A2) is similar. By (a)

$$\begin{aligned} nP\Big(\max_{1\le i\le n} |\xi_i| > n^{1/16}\Big) &= nP\left[\bigcup_{i=1}^{n} (|\xi_i| > n^{1/16})\right] \\ &\le n^2 P(|\xi_1| > n^{1/16}) \\ &\le n^{-1/16} = o(1) \end{aligned} \tag{A3}$$

as $n \to \infty$. Condition on $\max_{1\le i\le n} |\xi_i| \le n^{1/16}$. Then by Markov's inequality

$$\begin{aligned} P(|R_{n1}| > n^{-(2+\epsilon)/5}) &= P\left(\left|\sum_{i=1}^{n} \xi_i\right| > n^{3/5-\epsilon/5}\right) \\ &\le n^{-96/5+32\epsilon/5} E\left|\sum_{i=1}^{n} \xi_i\right|^{32}. \end{aligned}$$

By Lemma 5.1 of Lahiri (1992)

$$E\left|\sum_{i=1}^{n}\xi_i\right|^{32}\leq Cn^2[n^{16}m^{32}+n^{32}\alpha(m)],$$

where C is a constant, m is any number such that $1\leq m\leq Cn$, and $\alpha(m)$ is the mixing parameter of $\{\xi_i\}$ at lag m. Therefore

$$nP(|R_{n1}|>n^{-(2+\epsilon)/5})\leq C[n^{-1/5+32\epsilon/5}m^{32}+n^{79/5+32\epsilon/5}\alpha(m)].$$

Let $m=n^{\gamma}$, where $\gamma=1/160-2\epsilon/5$. Then $m<Cn$ for all sufficiently large n, and conditional on $\max_{1\leq i\leq n}|\xi_i|\leq n^{1/16}$,

$$nP(|R_{n1}|>n^{-(2+\epsilon)/5})\leq C[n^{-32\epsilon/5}+n^{79/5+64\epsilon/5}\alpha(n^{\gamma})] = o(1) \tag{A4}$$

as $n\to\infty$ by (b). The lemma follows by combining (A3) and (A4). *Q.E.D.*

LEMMA 2: *Let Assumptions* 1–4 *hold. Define* $G(x,\theta)=g(x,\theta)-Eg(X,\theta)$ *and* $\tilde H(\tilde X_i,\theta)=g(X_i,\theta)g(X_i,\theta)'+\sum_{j=1}^{\kappa}[g(X_i,\theta)g(X_{i+j},\theta')+g(X_{i+j},\theta)g(X_i,\theta)']$. *For any* $r>0$

$$\lim_{n\to\infty} nP\left[\sup_{\theta\in\Theta} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|>r\right]=0 \tag{A5}$$

and

$$\lim_{N\to\infty} NP\left[\sup_{\theta\in\Theta} N^{-1}\left\|\sum_{i=1}^{N}[\tilde H(\tilde X_i,\theta)-E\tilde H(\tilde X_1,\theta)]\right\|>r\right]=0, \tag{A6}$$

where $\|\cdot\|$ *in* (*A*6) *is the matrix norm.*

PROOF: Only (A5) is proved; the proof of (A6) is similar. Given $\epsilon>0$, divide Θ into subsets $\Theta_1,\Theta_2,\ldots,\Theta_{M(\epsilon)}$ such that $\|\theta_1-\theta_2\|<\epsilon$ whenever θ_1 and θ_2 are in the same subset. Let θ_i be a point in Θ_i for each $i=1,\ldots,M(\epsilon)$. Then

$$\sup_{\theta\in\Theta} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|=\max_j\sup_{\theta\in\Theta_j} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|.$$

Therefore

$$\begin{aligned}P^*&\equiv P\left[\sup_{\theta\in\Theta} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|>r\right]\\&-P\left\{\bigcup_{j=1}^{M(\epsilon)}\left[\sup_{\theta\in\Theta_j} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|>r\right]\right\}\\&\leq\sum_{j=1}^{M(\epsilon)}P\left[\sup_{\theta\in\Theta_j} n^{-1}\left\|\sum_{i=1}^{n}G(X_i,\theta)\right\|>r\right].\end{aligned}$$

For $\theta \in \Theta_j$

$$n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta)\right\| \le n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta_j)\right\| + n^{-1}\sum_{i=1}^{n}\|G(X_i,\theta)-G(X_i,\theta_j)\|$$
$$\le n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta_j)\right\| + (\epsilon/n)\sum_{i=1}^{n}[C_g(X_i)+EC_g(X)]$$
$$\le n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta_j)\right\| + (\epsilon/n)\left|\sum_{i=1}^{n}[C_g(X_i)-EC_g(X)]\right|$$
$$+2\epsilon C_g(X),$$

where $C_g(X)$ is as in Assumption 3. Choose ϵ so that $2\epsilon EC_g(X) < r/3$. Then

$$P\left[\sup_{\theta\in\Theta_j} n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta)\right\| > r\right] \le P\left[n^{-1}\left\|\sum_{i=1}^{n} G(X_i,\theta_j)\right\| > r/3\right]$$
$$+P\left\{n^{-1}\left|\sum_{i=1}^{n}[C_g(X_i)-EC_g(X)]\right| > r/(3\epsilon)\right\} = o(n^{-1})$$

by Lemma 1. *Q.E.D.*

LEMMA 3: *Let Assumptions* 1–4 *hold. Let* Ω *be a nonstochastic, symmetrical, positive-definite* $L_g \times L_g$ *matrix. Let* $\hat{\theta}_N$ *solve* (3.1). *For any* $r > 0$

(A7) $$\lim_{N\to\infty} NP(\|\hat{\theta}_N - \theta_0\| > rN^{-2/5}) = 0.$$

PROOF: It is first necessary to prove that

(A8) $$\lim_{N\to\infty} NP(\|\hat{\theta}_N - \theta_0\| > r) = 0.$$

To do this define

$$\Delta_N(\theta) \equiv N^{-1}\sum_{i=1}^{N}[g(X_i,\theta) - Eg(X,\theta)].$$

Then $J_N(\theta) = Eg(X,\theta)'\Omega Eg(X,\theta) + 2Eg(X,\theta)'\Omega\Delta_N(\theta) + \Delta_N(\theta)'\Omega\Delta_N(\theta)$. Given any $\epsilon > 0$, it follows from Lemma 2 that as $N \to \infty$

$$P\left[\sup_{\theta\in\Theta}\|\Delta_N(\theta)\| > \epsilon\right] = o(N^{-1}).$$

Therefore, given any $\delta > 0$, $|2Eg(X,\theta)'\Omega\Delta_N(\theta) + \Delta_N(\theta)'\Omega\Delta_N(\theta)| < \delta$ uniformly over $\theta \in \Theta$ with probability $1 - o(N^{-1})$. Define $M = \inf_\Gamma Eg(X,\theta)'\Omega Eg(X,\theta)$, where $\Gamma = \{\theta \in \Theta: \|\theta - \theta_0\| > r\}$. Set $\delta < M/2$. Then

(A9) $$J_N(\theta) - J_N(\theta_0) > M - 2\delta > 0$$

uniformly over Γ with probability $1 - o(N^{-1})$. By the definition of $\hat{\theta}_N, J_N(\hat{\theta}_N) \le J_N(\theta_0)$, which is inconsistent with (A9). Therefore, $P(\|\hat{\theta}_N - \theta_o\| > r) = o(N^{-1})$.

It now follows that $\partial J_N(\hat{\theta}_N)/\partial\theta = 0$ with probability $1 - o(N^{-1})$. Let $\delta_N = \hat{\theta}_N - \theta_0$, δ_{Ni} denote the ith component of δ_N, $J_{N\theta ij}(\theta) = \partial^3 J_N(\theta)/\partial\theta\,\partial\theta_i\theta_j$, and $J_{N\theta ijk}(\theta) = \partial^4 J_N(\theta)/\partial\theta\,\partial\theta_i\theta_j\theta_k$. Using the convention of summing over common subscripts, a Taylor series expansion of $\partial J_N(\hat{\theta}_N)/\partial\theta$ about $\hat{\theta}_N = \theta_0$ yields

(A10) $$\partial J_N(\theta_0)/\partial\theta + [\partial^2 J_N(\theta_0)/\partial\theta\,\partial\theta']\delta_N + (1/2)J_{N\theta ij}(\theta_0)\delta_{Ni}\delta_{Nj}$$
$$+(1/6)J_{N\theta ijk}(\theta_0)\delta_{Ni}\delta_{Nj}\delta_{Nk} + \zeta_N = 0$$

with probability $1-o(N^{-1})$, where $\bar{\theta}_N$ is between $\hat{\theta}_N$ and θ_0, and

$$\zeta_N = (1/6)[J_{N\theta ijk}(\bar{\theta}_N) - J_{N\theta ijk}(\theta_0)]\delta_{Ni}\delta_{Nj}\delta_{Nk}.$$

Let $V_N = [\partial^2 J_N(\theta_0)/\partial\theta\,\partial\theta']^{-1}$. V_N exists and $\|V_N\|$ is bounded with probability $1-o(N^{-1})$ by Lemma 1. Therefore, with probability $1-o(N^{-1})$

$$\text{(A11)}\quad (\hat{\theta}_N - \theta_0) = -V_N[\partial J_N(\theta_0)/\partial\theta + (1/2)J_{N\theta ij}(\theta_0)\delta_{Ni}\delta_{Nj} + (1/6)J_{N\theta ijk}(\theta_0)\delta_{Ni}\delta_{Nj}\delta_{Nk} + \zeta_N].$$

Further application of Lemma 1 shows that with probability $1-o(N^{-1})$, $\|\partial J_N(\theta_o)/\partial\theta\| < N^{-2/5-\epsilon}$ for sufficiently small $\epsilon > 0$, $\|J_{N\theta ij}(\theta_0)\|$ and $\|J_{N\theta ijk}(\theta_0)\|$ are bounded, and $\|\zeta_N\| < M\|\hat{\theta}_N - \theta_0\|^4$ for some $M < \infty$. Therefore, with probability $1-o(N^{-1})$, the norm of the right-hand side of (A11) is less than $rN^{-2/5}$ whenever $\|\hat{\theta}_N - \theta_0\| \le rN^{-2/5}$ and N is sufficiently large. Application of the Brouwer fixed point theorem to the right-hand side of (A11) establishes (A7). *Q.E.D.*

LEMMA 4: *Let Assumptions* 1–4 *hold. Let* $\tilde{\theta}_N$ *solve* (3.1) *and* $\hat{\theta}_N$ *solve* (3.2). *For any* $r > 0$

$$\text{(A12)}\quad \lim_{N\to\infty} NP(\|\hat{\theta}_N - \theta_0\| > rN^{-2/5}) = 0.$$

PROOF: By arguments similar to those used in proving (A8)

$$\lim_{N\to\infty} NP(\|\hat{\theta}_N - \theta_0\| > r) = 0.$$

To prove (A12), let $\partial_1 J_N/\partial\theta$ denote the gradient of J_N with respect to its first argument, and observe that $\partial_1 J_N(\hat{\theta}_N, \tilde{\theta}_N)/\partial\theta = 0$ with probability $1-o(N^{-1})$. Let $\delta_N = [(\hat{\theta}_N - \theta_0)', (\tilde{\theta}_N - \theta_0)']'$, and let δ_{Ni} be the ith component of δ_N. Let $\tilde{V}_N = [\partial_1^2 J_N(\theta_0, \theta_0)/\partial\theta\,\partial\theta']^{-1}$ whenever this quantity exists. $\tilde{V}_N$ exists and $\|\tilde{V}_N\|$ is bounded with probability $1-o(N^{-1})$ by Lemma 1. Using arguments similar to those used in proving (A7), it follows from Lemma 1 and a Taylor series expansion of $\partial_1 J_N(\hat{\theta}_N, \tilde{\theta}_N)/\partial\theta$ about $\delta_N = 0$ that with probability $1-o(N^{-1})$

$$\text{(A13)}\quad (\hat{\theta}_N - \theta_0) = -\tilde{V}_N[\partial_1 J_N(\theta_0, \theta_0)/\partial\theta + (1/2)Q_{Nij}\delta_{Ni}\delta_{Nj} + Q_{Nijk}\delta_{Ni}\delta_{Nj}\delta_{Nk} + \zeta_N],$$

where the Q_{Nij} and Q_{Nijk} are the second and third derivatives of $\partial_1 J_N(\cdot,\cdot)/\partial\theta$ with respect to both of its arguments evaluated at (θ_0, θ_0), $\|\tilde{V}_N\|$, $\|Q_{Nij}\|$, and $\|Q_{Nijk}\|$ are bounded, and $\|\zeta_N\| = O(\|\delta_N\|^4)$. By Lemma 3, the norm of the right-hand side of (A13) is less than $rN^{-2/5}$ with probability $1-o(N^{-1})$ whenever $\|\hat{\theta}_N - \theta_0\| \le rN^{-2/5}$. Application of the Brouwer fixed point theorem to the right-hand side of (A13) establishes (A12). *Q.E.D.*

LEMMA 5: *Let* $\{A_n\}$ *be a sequence of* $L_A \times 1$ *random variables with bounded Lebesgue densities. Let* $\{\xi_n\}$ *be a sequence of* $L_A \times 1$ *random variables such that* $P(\|\xi_n\| > rn^{-8/5}) = o(n^{-1})$ *as* $n \to \infty$ *for any* $r > 0$. *Then*

$$\lim_{n\to\infty} \sup_z n|P(A_n + n^{1/2}\xi_n < z) - P(A_n < z)| = 0,$$

where "<" applies to each component of the relevant vectors.

PROOF: Let e be a $L_A \times 1$ vector of ones. Then

$$\begin{aligned}
\text{(A14)}\quad & n[P(A_n + n^{1/2}\xi_n < z) - P(A_n < z)] \\
&= n[P(A_n + n^{1/2}\xi_n < z, \|\xi_n\| \le rn^{-8/5}) \\
&\quad - P(A_n < z) + P(A_n + n^{1/2}\xi_n < z, \|\xi_n\| > rn^{-8/5})] \\
&\le n[P(A_n < z + ern^{-11/10}) - P(A_n < z) + P(\|\xi_n\| > rn^{-8/5})] \\
\text{(A15)}\quad & \le Mn^{-1/10} + o(n^{-1})
\end{aligned}$$

uniformly over z for some $M < \infty$. It also follows from (A14) that

$$\text{(A16)}\qquad n[P(A_n + n^{1/2}\xi_n < z) - P(A_n < z)] \geq n[P(A_n < z - ern^{-11/10}) - P(A_n < z)] \geq -Mn^{-1/10} + o(n^{-1})$$

uniformly over z. The lemma follows by combining (A15) and (A16). *Q.E.D.*

Proposition 1: *Let Assumptions* 1–5 *hold*, $\hat{\theta}_N$ *solve either* (3.1) *or* (3.2), *and* Δ_N *denote either* T_{Nr} *or* (*if* $\hat{\theta}_N$ *solves* (3.2)) $K_N(\hat{\theta}_N)$. *For each definition of* Δ_N, *there is a function G that is continuously differentiable in a neighborhood of S such that*

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

Proof: The main problem to be solved is showing that $\hat{\theta}_N - \theta_0$ can be approximated by a function of sample moments. Accordingly, begin by letting $\delta_N = N^{1/2}(\hat{\theta}_N - \theta_0)$ with $\hat{\theta}_N$ a solution to (3.1). Let R_N be the column vector whose elements are the unique components of $\partial J_N(\theta_0)/\partial\theta$, $\partial^2 J_N(\theta_0)/\partial\theta\,\partial\theta'$, $J_{N\theta ij}(\theta_0)$, and $J_{N\theta ijk}(\theta_0)$ $(i, j, k = 1, \ldots, L_\theta)$. Let R denote almost sure limit of R_N as $N \to \infty$. Let e_N be the conformable vector $(\zeta_N', 0, \ldots, 0)'$, where ζ_N is defined as in (A10). Application of the implicit function theorem to (A10) yields the result that there is a function Λ such that $\Lambda(R) = 0$, Λ is continuously differentiable in a neighborhood of R, and

$$\text{(A17)}\qquad (\hat{\theta}_N - \theta_0) = \Lambda(R_N + e_N)$$

with probability $1 - o(N^{-1})$. Each component of R_N is a continuous function of S_N. Therefore, it follows from Lemma 1 that for any $r > 0$, $\|R_N - R\| \leq r$ with probability $1 - o(N^{-1})$. As in the proof of Lemma 3, $\|\zeta_N\| \leq M\,\|\hat{\theta}_N - \theta_0\|^4$ with probability $1 - o(N^{-1})$, so it follows from Lemma 3 that $\|e_N\| \leq rN^{-8/5}$ for any $r > 0$ with probability $1 - o(N^{-1})$. Therefore, for some $\tilde{M} < \infty$ and any $r > 0$

$$NP[\|(\hat{\theta}_N - \theta_0) - \Lambda(R_N)\| > rN^{-8/5}] \leq NP[(\tilde{M}\,\|e_N\| > rN^{-8/5})] = o(1)$$

as $N \to \infty$. By this result and Lemma 5

$$\text{(A18)}\qquad \lim_{N\to\infty} \sup_z N\,|P[N^{1/2}(\hat{\theta}_N - \theta_0) < z] - P[N^{1/2}\Lambda(R_N) < z]| = 0.$$

Now let $\delta_N = N^{1/2}(\hat{\theta}_N - \theta_0)$ with θ_N a solution to (3.2). Expand $\partial_1 J_N(\hat{\theta}_N, \tilde{\theta}_N)/\partial\theta$ in a Taylor series through order 3 about $\hat{\theta}_N = \tilde{\theta}_N = \theta_0$. Apply (A17) and the implicit function theorem to obtain

$$\text{(A19)}\qquad (\hat{\theta}_N - \theta_0) = \Lambda^*[S_N, \zeta_N, \Lambda(R_N + e_N)]$$

with probability $1 - o(N^{-1})$ for some Λ^*, where ζ_N is the remainder term of the Taylor series expansion, $\Lambda^*(S, 0, 0) = 0$ and Λ^* is continuously differentiable in a neighborhood of $(S, 0, 0)$. As in the proof of Lemma 4, $\|\zeta_N\| = O(\|\delta_N\|^4)$. Therefore, by Lemma 4, $\|\zeta_N\| < rN^{-8/5}$ with probability $1 - o(N^{-1})$ for any $r > 0$. It follows from (A19) and Lemma 5 that there is a function Λ such that an equation of the form (A18) holds for the two-step estimator $\hat{\theta}_N$.

The remaining forms of Δ_N are functions of $\hat{\theta}_N$ and, possibly, $\tilde{\theta}_N$. Taylor series expansions of these functions about $\hat{\theta}_N = \tilde{\theta}_N = \theta_0$ through order 3 yield results of the form $\Delta_N = N^{1/2}[\Lambda^{**}(S_N, \hat{\theta}_N - \theta_0, \tilde{\theta}_N - \theta_0) + \zeta_N]$, where ζ_N is the remainder term of the expansion, $\|\zeta_N\| = O(\|\delta_N\|^4)$, and Λ^{**} is continuously differentiable in a neighborhood of $(S, 0, 0)$. Since $\|\delta_N\| < rN^{-2/5}$ with probability $1 - o(N^{-1})$ for any $r > 0$, the result follows from Lemma 5. *Q.E.D.*

Lemma 6: *Let Assumptions* 1 *and* 2 *hold. Let* $b_\iota = \{1 + l(i-1), 2 + l(i-1), \ldots, li\}$. *Let h be a function such that* $Eh(X) = 0$ *and* $P(|h(X)| > z) = O(z^{-33})$ *as* $z \to \infty$. *For any* $c > 0$, $l \propto N^\gamma$, $11/50 \leq \gamma \leq 12/50$,

$$\lim_{N\to\infty} NP\left[b^{-1}\sum_{\iota=1}^{b}\left|l^{-1}\sum_{j\in b_\iota} h(X_j)\right|^{32} > cN^{-9/25}\right] = 0.$$

PROOF: Define $\tilde{c} = c^{1/32}$ and

$$Q_N = NP\left[b^{-1} \sum_{i=1}^{b} \left| l^{-1} \sum_{j \in b_i} h(X_j) \right|^{32} > cN^{-9/25} \right].$$

Some algebra and an application of Markov's inequality yield

$$Q_N \le NbP\left[\left| \sum_{j \in b_1} h(X_j) \right| > \tilde{c} l N^{-9/800} \right]$$

$$\text{(A20)} \qquad \le bN^{34/25} (\tilde{c} l)^{-32} E \left| \sum_{j \in b_1} h(X_j) \right|^{32}.$$

Moreover

$$\text{(A21)} \qquad NP\left(\max_{j \in b_1} |h(X_j)| > N^{1/25} \right) \le NlO(N^{-33/25}) = o(1).$$

Now condition on the event

$$\max_{j \in b_1} |h(X_j)| \le N^{1/25}.$$

Conditional on this event, it follows from Lemma 5.1 of Lahiri (1992) that

$$E \left| \sum_{j \in b_1} h(X_j) \right|^{32} \le CN^{32/35} [l^{16} m^{32} + l^{32} \alpha(m)],$$

where C is a constant, m is any number such that $1 \le m \le Cl$, and $\alpha(\cdot)$ is the strong mixing parameter. So by (A20)

$$\text{(A22)} \qquad NbP\left[\left| \sum_{j \in b_1} h(X_j) \right| > \tilde{c} l N^{-9/800} \right] \le C \tilde{c}^{-32} b N^{66/25} l^{-32} [l^{16} m^{32} + l^{32} \alpha(m)].$$

Let $m = l^{1/128}$. Then the right-hand side of (A22) is $o(1)$ as $N \to \infty$. The Lemma now follows from this result and (A21). *Q.E.D.*

LEMMA 7: *Let h be a function such that $Eh(X) = 0$, $P(|h(X)| > z) = O(z^{-33})$ as $z \to \infty$, and for each N and sample $\{X_i: i = 1, \ldots, N\}$, $N^{-1} \sum_{i=1}^{N} h(X_i) = 0$. For blocks b_j $(j = 1, \ldots, b)$ of length $l \propto N^{\gamma}$, $11/50 \le \gamma \le 12/50$, that are sampled randomly according to Carlstein's scheme, define*

$$R_i^* = l^{-1} \sum_{j \in b_j} h(X_j)$$

and

$$\bar{R}^* = b^{-1} \sum_{i=1}^{b} R_i^*.$$

For each $\delta > 0$

$$\lim_{N \to \infty} NP[NP^*(|\bar{R}^*| > N^{-9/25}) > \delta] = 0.$$

PROOF: By Markov's inequality

$$NP^*(|\bar{R}^*| > N^{-9/25}) = NP^*\left(\left| \sum_{i=1}^{b} R_i^* \right| > bN^{-9/25} \right)$$

$$\text{(A23)} \qquad \le N^{313/25} b^{-32} E^* \left| \sum_{i=1}^{b} R_i^* \right|^{32}.$$

By Burkholder's inequality (Hall and Heyde (1980, p. 23)) there is a C such that

$$\text{(A24)} \qquad E^*\left|\sum_{i=1}^{b} R_i^*\right|^{32} \leq CE^*\left[\sum_{i=1}^{b}(R_i^*)^2\right]^{16}.$$

By Hölder's inequality

$$\text{(A25)} \qquad \left[\sum_{i=1}^{b}(R_i^*)^2\right]^{16} \leq b^{15}\sum_{i=1}^{b}|R_i^*|^{32}.$$

Substituting (A25) into (A24) and the resulting equation into (A23) yields

$$NP^*(|\bar{R}^*| > N^{-9/25}) \leq CN^{313/25}b^{-16}E^*|R_1^*|^{32}.$$

Now apply Lemma 6. *Q.E.D.*

LEMMA 8: *Let Assumptions 1–4 hold. Let X^* be sampled using Carlstein's blocking scheme conditional on χ with block length $l \propto N^{\gamma}$, $11/50 \leq \gamma \leq 12/50$. For $\theta \in \Theta$ define $G(x,\theta) = g(x,\theta) - E^*g(X,\theta)$ and $\tilde{H}(\tilde{X}_i^*,\theta) = g(X_i^*,\theta)g(X_i^*,\theta)' + \Sigma_{j=1}^{\kappa}[g(X_i^*,\theta)g(X_{i+j}^*,\theta)' + g(X_{i+j}^*,\theta)g(X_i^*,\theta)']$. For any $\delta > 0$ and $r > 0$,*

$$\text{(A26)} \qquad \lim_{N\to\infty} NP\left\{NP^*\left[\sup_{\theta\in\Theta} N^{-1}\left\|\sum_{i=1}^{N} G(X_i^*,\theta)\right\| > r\right] > \delta\right\} = 0$$

and

$$\text{(A27)} \qquad \lim_{N\to\infty} NP\left\{NP^*\left[\sup_{\theta\in\Theta} N^{-1}\left\|\sum_{i=1}^{N}[H(\tilde{X}_i^*,\theta) - E^*H(\tilde{X}_i^*,\theta)]\right\| > r\right] > \delta\right\} = 0.$$

Analogous results hold for first through third derivatives of G and H with respect to θ.

PROOF: Only (A26) is proved. The proof of (A27) is similar. To prove (A26), proceed as in the proof of Lemma 2 to obtain

$$P^*\left[\sup_{\theta\in\Theta_j} N^{-1}\left\|\sum_{i=1}^{N} G(X_i^*,\theta)\right\| > r_N\right] \leq P^*\left[N^{-1}\left\|\sum_{i=1}^{N} G(X_i^*,\theta_j)\right\| > r/3\right]$$
$$+ P^*\left\{N^{-1}\left|\sum_{i=1}^{N}[C_g(X_i^*) - E^*C_g(X)]\right| > r/(3\epsilon)\right\}$$

for any $\epsilon > 0$. Therefore

$$P\left\{NP^*\left[\sup_{\theta\in\Theta_j} N^{-1}\left\|\sum_{i=1}^{N} G(X_i^*,\theta)\right\| > r\right] > \delta\right\}$$
$$\leq P\left\{NP^*\left[N^{-1}\left\|\sum_{i=1}^{N} G(X_i^*,\theta_j)\right\| > r/3\right] > \delta/2\right\}$$
$$+ P\left(NP^*\left\{N^{-1}\left|\sum_{i=1}^{N}[C_g(X_i^*) - E^*C_g(X)]\right| > r/(3\epsilon)\right\} > \delta/2\right).$$

Now apply Lemma 7. *Q.E.D.*

LEMMA 9: *Let Assumptions* 1–4 *hold. Let X be sampled according to Carlstein's blocking scheme with block length* $l \propto N^{\gamma}$, $11/50 \leq \gamma \leq 12/50$. *Let* θ_N^* *solve* (3.5). *For any* $\delta > 0$, $r > 0$ *and all sufficiently small* η

$$\lim_{N\to\infty} NP[NP^*(\|\theta_N^* - \theta_N\| > rN^{-(1+\eta)/3}) > \delta] = 0.$$

PROOF: This is the bootstrap version of Lemma 3 and is proved using similar arguments. *Q.E.D.*

LEMMA 10: *Let Assumptions* 1–4 *hold. Let X be sampled according to Carlstein's blocking scheme with* $l \propto N^{\gamma}$, $11/50 \leq \gamma \leq 12/50$. *Let* θ_N^* *solve* (3.6), *where* $\tilde{\theta}_N^*$ *solves* (3.5). *For any* $\delta > 0$, $r > 0$ *and all sufficiently small* η,

$$\lim_{N\to\infty} NP[NP^*(\|\theta_N^* - \hat{\theta}_N\| > rN^{-(1+\eta)/3}) > \delta] = 0.$$

PROOF: This is the bootstrap version of Lemma 4 and is proved using similar arguments. *Q.E.D.*

PROPOSITION 2: *Let Assumptions* 1–5 *hold, and let* θ_N *and* θ_N^*, *respectively, solve either* (3.1) *and* (3.5) *or* (3.2) *and* (3.6). *Assume that* $l = N^{\gamma}$, *where* $11/50 \leq \gamma \leq 12/50$. *Let* Δ_N^* *denote either* T_{Nr}^* *or* (*if* $\hat{\theta}_N$ *and* θ_N^* *solve* (3.2) *and* (3.6)) $K_N^*(\theta_N^*)$. *Define G as in Theorem* 1, $\omega_N = \tau_{Nr}$ *if* $\Delta_N^* = T_{Nr}^*$, *and* $\omega_N = (V_N^+)^{1/2}$ *if* $\Delta_N^* = K_N^*(\theta_N^*)$. *Then* $\Delta_N^* = \omega_N N^{1/2} G(S_N^*) + o(N^{-1})$ *with* P^* *probability* $o(N^{-1})$ *except, possibly, if* χ *is in a set of P probability* $o(N^{-1})$.

PROOF: Let R_N^* be the column vector whose elements are the unique components of $\partial J_N^*(\hat{\theta}_N)/\partial\theta$, $\partial^2 J_N^*(\hat{\theta}_N)\partial\theta\,\partial\theta'$, $J_{N\theta ij}^*(\hat{\theta}_N)$, and $J_{N\theta ijk}^*(\hat{\theta}_N)$. Let e_N^* be a conformable column vector with zeros for all but its first L_θ elements. Define Λ as in (A17). Suppose that θ_N^* solves (3.5). By using Lemmas 4 and 9 and the bootstrap analog of (A11), it can be seen that $(\theta_N^* - \hat{\theta}_N) = \Lambda(R_N^* + e_N^*)$ with P^* probability $1 - o(N^{-1})$ except, possibly, if χ is in a set of P probability $o(N^{-1})$. Moreover, $\|e_N^*\| = o(N^{-1})$ with P^* probability $1 - o(N^{-1})$ except, possibly, if χ is in a set of P probability $o(N^{-1})$. Therefore, $(\theta_N^* - \hat{\theta}_N) = \Lambda(R_N^*) + o(N^{-1})$ with P^* probability $1 - o(N^{-1})$ except, possibly, if χ is in a set of P probability $o(N^{-1})$. As in the proof of Proposition 1, similar arguments apply if θ_N^* solves (3.6). The remainder of the proof of Proposition 2 proceeds as in the proof of Proposition 1. *Q.E.D.*

PROOF OF THEOREM 1: By Proposition 1, to prove part (a) it suffices to show that the function $G(S_N)$ that approximates T_{Nr} in Theorem 1 has an Edgeworth expansion of the form (3.8). This result follows Theorem (2.8) of Götze and Hipp (1983). Part (b) is proved by using Theorem (2.8) of Götze and Hipp (1983) to show that Ψ_N has a multivariate Edgeworth expansion and then using Theorem 1 and Remark (2.2) of Chandra and Ghosh (1979) to obtain (3.9) from the expansion of Ψ_N. *Q.E.D.*

PROOF OF THEOREM 2: Let $G(S_N)$ be the approximation to T_{Nr} in Proposition 1. By Proposition 2 and Lemma 5, to prove part (a), it suffices to prove that $\tau_{Nk} G(S_N^*)$ has an Edgeworth expansion of the form (3.11). This result follows from a proof identical to that of Theorem 3.1 of Bhattacharya (1987). To prove part (b), let $N^{1/2}G(S_N)$ be the approximation to $K_N(\theta_N)$ in Proposition 1. Use a proof identical to that of Theorem 3.1 of Bhattacharya (1987) to show that Ψ_N^* has an Edgeworth expansion through $o(N^{-1})$. Then use Theorem 1 and Remark (2.2) of Chandra and Ghosh (1979) to show that $K_N^*(\theta_N^*)'K_N^*(\theta_N^*) = G(S_N^*)'V_N^+ G(S_N^*)$ has an Edgeworth expansion of the form (3.12). Finally, use Lemma 5 to show that the CDFs of $K_N^*(\theta_N^*)'K_N^*(S_N^*)$ and J_N^* differ by $o(N^{-1})$ except, possibly, if χ is in a set of probability $o(N^{-1})$. *Q.E.D.*

LEMMA 11: *Define* $\{j\} = (j_1, \ldots, j_m)$. *Let* $\nu_{m\{j\}}$ *and* $\nu_{m\{j\}}^*$ *denote the quantities in* (3.7) *and* (3.10), *respectively. Let Assumptions* 1–4 *hold, and set* $l = o(N^{1/4})$ *as* $N \to \infty$. *Then for each integer m such that* $2 \leq m \leq 6$,

$$\operatorname*{plim}_{N\to\infty} \nu_{m\{j\}}^* = \nu_{m\{j\}}.$$

PROOF: Because blocks are sampled independently in the block bootstrap, it is not difficult to show that if $m > 3$, $\nu^*_{m\{j\}}$ can be written as a product of terms of the form $\nu_{r\{s\}}$ ($r = 2$ or 3) plus a remainder that is $o_p(1)$ as $n \to \infty$. Therefore, it suffices to prove the result for $m = 2$ or 3. Suppose, first, that $m = 3$. Define $f_r(\cdot,\cdot)$ as the rth component of $f(\cdot,\cdot)$, $\mu_{Nr}(\theta) = E^* f_r(X, \theta)$, and $\mu_r = Ef_r(X, \theta_0)$. Set $\{j\} = (r, s)$. Since the bootstrap sample consists of b independent blocks of length l, some algebra shows that $\nu^*_{2\{j\}} = A_N + B_N + o_p(1)$, where

$$A_N = N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{l} \sum_{k=1}^{l} [f_r(X_{il+j}, \theta_0) f_s(X_{il+k}, \theta_0) - \mu_r \mu_s]$$

and $B_N = l[\mu_{rN}(\theta_0)\mu_{sN}(\theta_0) - \mu_r \mu_s]$. It follows from the arguments of Hall and Horowitz (1993, pp. 15–16) that $B_N = o_p(1)$ and $A_N \to p_{\nu_{2\{j\}}}$.

Now let $m = 3$. Again using independence of the sampled blocks, as well as the convergence of A_N and B_N above,

$$\begin{aligned}\nu^*_{3\{j\}} &= N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{l} \sum_{k=1}^{l} \sum_{p=1}^{l} [f_r(X_{il+j}, \theta_N) - \mu_{Nr}(\theta_N)] \\ &\quad \cdot [f_s(X_{il+k}, \theta_N) - \mu_{Ns}(\theta_N)][f_t(X_{il+p}, \theta_N) - \mu_{Nt}(\theta_N)] + o_p(1) \\ &= N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{l} \sum_{k=1}^{l} \sum_{p=1}^{l} [f_r(X_{il+j}, \theta_0) - \mu_r] \\ &\quad \cdot [f_s(X_{il+k}, \theta_0) - \mu_s][f_t(X_{il+p}, \theta_0) - \mu_t] + o_p(1) \\ &= C_N + o_p(1).\end{aligned}$$

It is clear that $E(C_N) = \nu_{3\{j\}}$. Convergence to 0 of $\mathrm{var}(C_N)$ follows from convergence $E\nu^*_{m\{j\}}$ for even m. *Q.E.D.*

PROOF OF THEOREM 3: $\pi_1(\delta, \nu)\Phi(z)$ and $\pi_1(\delta, \nu^*)\Phi(z)$ are even functions of z. Therefore, by Theorems 1 and 2, Lemma 11, and $\tau_{Nr} \overset{p}{\to} 1$

(A28) $$P(|T_{Nr}| > z) - P^*(|T^*_{Nr}| > z) = o_p(N^{-1})$$

uniformly over z and

(A29) $$P(J_N > z) - P^*(J^*_N > z) = o_p(N^{-1})$$

where P^* denotes probability under bootstrap sampling conditional on χ. Equations (3.13) and (3.14) follow from (A28) and (A29) and an expected value argument. *Q.E.D.*

REFERENCES

ANDREWS, D. W. K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

BERAN, R. (1988): "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687–697.

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. Vilaplana, and Wolfgang Wertz. New York: John Wiley & Sons, pp. 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 W. K. NEWEY (1992): "Bootstrapping from GMM," Seminar Notes, July, 1992.

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

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.

GÖTZE, F., AND C. HIPP (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Zeitschrift für Warscheinlichkeitstheorie und verwandte Gebiete*, 64, 211–239.

GÖTZE, F., AND C. HIPP (1992): "On the Validity of Edgeworth Expansions," Preprint 92-045, Sonderforschungsbereich 343, Universisät Bielefeld, Bielefeld, Germany.

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

——— (1986a): "On the Bootstrap and Confidence Intervals," *Annals of Statistics*, 14, 1431–1452.

——— (1986b): "On the Number of Bootstrap Simulations Needed to Construct a Confidence Interval," *Annals of Statistics*, 14, 1453–1462.

——— (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 (1993): "Corrections and Blocking Rules for the Block Bootstrap with Dependent Data," Working Paper 93-11, Department of Economics, University of Iowa.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

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

HILLIER, G. H. (1985): "On the Joint and Marginal Densities of Instrumental Variables Estimators in a General Structural Equation," *Econometric Theory*, 1, 53–72.

HOROWITZ, J. L. (1994): "Bootstrap-Based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395–411.

KOCHERLAKOTA, N. R. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.

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," in *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard. New York: Wiley, pp. 183–214.

NELSON, C. R., AND R. STARTZ (1990a): "The Distribution of the Instrumental Variable Estimator and its t Ratio when the Instrument is a Poor One," *Journal of Business*, 63, 3125–3140.

——— (1990b): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

PHILLIPS, P. C. B. (1983): "Exact Small Sample Theory in the Simultaneous Equations Model," in *Handbook of Econometrics*, Vol. 1, ed. by Z. Griliches and M. D. Intriligator. Amsterdam: North-Holland Publishing Co., Ch. 8.

PHILLIPS, P. C. B., AND J. Y. PARK (1988): "On the Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065–1083.

SINGH, K. (1981): "On the Asymptotic Accuracy of Efron's Bootstrap," *Annals of Statistics*, 9, 1187–1195.

TAUCHEN, G. (1986): "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397–425.