

THE BLOCK–BLOCK BOOTSTRAP: IMPROVED ASYMPTOTIC REFINEMENTS

BY DONALD W. K. ANDREWS¹

The asymptotic refinements attributable to the block bootstrap for time series are not as large as those of the nonparametric iid bootstrap or the parametric bootstrap. One reason is that the independence between the blocks in the block bootstrap sample does not mimic the dependence structure of the original sample. This is the join-point problem.

In this paper, we propose a method of solving this problem. The idea is not to alter the block bootstrap. Instead, we alter the original sample statistics to which the block bootstrap is applied. We introduce *block* statistics that possess join-point features that are similar to those of the block bootstrap versions of these statistics. We refer to the application of the block bootstrap to block statistics as the block–block bootstrap. The asymptotic refinements of the block–block bootstrap are shown to be greater than those obtained with the block bootstrap and close to those obtained with the nonparametric iid bootstrap and parametric bootstrap.

KEYWORDS: Asymptotics, block bootstrap, block statistics, Edgeworth expansion, extremum estimator, generalized method of moments estimator, maximum likelihood estimator, t statistic, test of over-identifying restrictions.

1. INTRODUCTION

THE PRINCIPAL THEORETICAL ATTRIBUTE of bootstrap procedures is the asymptotic refinements they provide. That is, when properly applied, bootstrap tests have errors in null rejection probabilities that are of a smaller order of magnitude as the sample size, N , goes to infinity than those of standard asymptotic tests based on the delta method. Similarly, bootstrap confidence intervals (CI's) have coverage probability errors of a smaller order of magnitude than those of standard asymptotic CI's based on the delta method.

This paper is concerned with the magnitude of the asymptotic refinements of the block bootstrap for time series. These asymptotic refinements are not as large as those of the nonparametric iid bootstrap or the parametric bootstrap. For example, for iid observations, the error in rejection probability (ERP) of a one-sided bootstrap t test based on the nonparametric iid bootstrap is $O(N^{-1})$; e.g., see Hall (1992). In contrast, for stationary strong mixing observations, the ERP of a one-sided bootstrap t test based on nonoverlapping or overlapping blocks is $O(N^{-1/2-\xi})$ for $0 < \xi < 1/4$, where ξ depends on the block length; see Andrews (2002a), hereafter denoted A2002, and Zvingelis (2003). For the parametric bootstrap, the ERP of a one-sided bootstrap t test is essentially the same as that for the nonparametric iid bootstrap. This holds for iid

¹The author thanks a coeditor and two referees for helpful comments. The author gratefully acknowledges the research support of the National Science Foundation via Grant Numbers SBR-9730277 and SES-0001706.

observations as well as for stationary strong mixing Markov observations; see Andrews (2004).

There are two reasons why the asymptotic refinements of the block bootstrap are less than those of the nonparametric iid bootstrap. The first is that the independence between the blocks in the block bootstrap sample does not mimic the dependence structure of the original sample. This is the *join-point* problem. The second reason is that the use of blocks of length greater than one increases the variability of various moments calculated under the block bootstrap distribution in comparison to their variability under the nonparametric iid bootstrap distribution. The reason is that the variability is determined by the amount of averaging that occurs over the blocks and longer blocks yield fewer blocks and, hence, fewer terms in the averages.

In this paper, we propose a method of solving the join-point problem. We do not alter the block bootstrap, because there does not seem to be a way to avoid its join-point feature. Rather, we alter the original sample statistics to which the block bootstrap is applied. We introduce *block* statistics (for the original sample) that have join-point features that resemble those of the block bootstrap versions of these statistics. We call the application of the block bootstrap to block statistics the *block–block* bootstrap.

The asymptotic refinements obtained by the block–block bootstrap are shown to be greater than those obtained by the standard block bootstrap. In fact, the block length can be chosen such that the magnitude of the asymptotic refinements of the block–block bootstrap is arbitrarily close to that obtained in the iid context using the nonparametric iid bootstrap. In practice, however, one would not expect the block–block bootstrap to perform as well as the nonparametric iid bootstrap for iid data. But, the asymptotic results suggest that it should outperform the block bootstrap in terms of ERP's and CI coverage probabilities.

A block statistic is constructed by taking a statistic that depends on one or more sample averages and replacing the sample averages by averages with some summands deleted. Let ℓ denote the block length to be used by the block bootstrap. We take ℓ such that $\ell \propto N^\gamma$ for some $0 < \gamma < 1$. The join points of the block bootstrap sample are $\ell + 1, 2\ell + 1, \dots, (b - 1)\ell$, where b is the number of blocks and $N = b\ell$. We delete the $\lceil \pi\ell \rceil$ summands before each of the join points, where $\lceil \pi\ell \rceil$ denotes the smallest integer greater than or equal to $\pi\ell$, $\pi \in (0, 1)$, and $\pi = \pi_N \rightarrow 0$ and $\pi\ell - C \log(N) \rightarrow \infty$ as $N \rightarrow \infty$ for all constants $0 < C < \infty$. For example, $\pi \propto N^{-\delta\gamma}$ satisfies these conditions for any $0 < \delta < 1$. Note that π is the fraction of observations that are deleted from each block and from the whole sample.

For example, consider an estimator that minimizes a sample average of summands that depend on the observations and an unknown parameter θ , such as a quasi-maximum likelihood or least squares estimator. The corresponding block estimator minimizes the same sample average but with the summands described above deleted. A block t statistic for θ is based on a block estimator of θ normalized by a block standard deviation estimator.

Consider a sample average that appears in the definition of a block statistic. The last nonzero summand in one block is separated from the first summand in the next block by $\lceil \pi \ell \rceil$ time periods, where $\lceil \pi \ell \rceil \rightarrow \infty$ as $N \rightarrow \infty$. In consequence, for an asymptotically weakly dependent time series, such as a strong mixing process, the blocks are asymptotically independent. On the other hand, the blocks that appear in the bootstrap version of the block statistic are independent by construction.

Independence of the bootstrap blocks mimics the asymptotic independence of the original sample blocks sufficiently well that the join-point problem is solved. That is, join points do not affect the magnitude of the asymptotic refinements of the block–block bootstrap. See Section 2 of Andrews (2002b) for a detailed discussion of why this is true. Also, the join-point correction factors introduced in Hall and Horowitz (1996) and employed in A2002 for use with the block bootstrap are not needed with the block–block bootstrap.² Furthermore, in the case of an m -dependent process, the block length can be finite with the block–block bootstrap, whereas it must diverge to infinity with the standard block bootstrap.

Although the block–block bootstrap solves the join-point problem, the block–block bootstrap yields moments that are more variable than moments under the nonparametric iid bootstrap distribution—just as the standard block bootstrap does. In consequence, the asymptotic refinements obtained by the block–block bootstrap still depend on the block length. In particular, they are decreasing in the block length. In this paper, we show that the ERP of a one-sided bootstrap t test using the block–block bootstrap is $O(N^{-1/2-\xi})$ for all $\xi < 1/2 - \gamma$. In consequence, if γ is taken close to zero, the ERP is close to $O(N^{-1})$, which is the ERP of a one-sided nonparametric iid bootstrap t test.

In practice, one has to use a block length ℓ and a deletion fraction π that are large enough to accommodate the dependence in the data. Hence, one cannot just take γ arbitrarily close to zero. Thus, the above asymptotic result does not imply that one would expect the block–block bootstrap to work as well as the nonparametric iid bootstrap does with iid data. However, it does suggest that the block–block bootstrap should have smaller ERP's when $\gamma < 1/4$ than does the block bootstrap.

Block statistics have the same first-order asymptotic efficiency as the standard statistics upon which they are based, because $\pi \rightarrow 0$ as $N \rightarrow \infty$. Hence, block–block bootstrap tests have the same asymptotic local power as standard asymptotic tests and as block bootstrap tests. Nevertheless, block statistics sacrifice some higher-order asymptotic efficiency and finite sample efficiency

²These correction factors alleviate, but do not solve, the join-point problem for the block bootstrap applied to standard statistics. They allow the block bootstrap to attain ERP's for two-sided t tests of magnitude $O(N^{-1-\xi})$ for $\xi < 1/4$, but these are still noticeably larger than those attained by the nonparametric iid bootstrap. Correction factors are not needed for one-sided t tests to yield asymptotic refinements.

because some observations are deleted. This is a drawback of the use of the block–block bootstrap.

A second drawback of the block–block bootstrap is that it requires the specification of the deletion fraction π , as well as the block length ℓ . It may be possible to use higher-order expansions to determine suitable choices of π and ℓ . This is beyond the scope of the present paper. However, we do suggest a data-dependent method for choosing π and ℓ based on a nested bootstrap-type procedure and assess its performance using simulations.

This paper presents some Monte Carlo results that are designed to assess the finite sample performance of the block–block bootstrap. A dynamic regression model with regressors given by a constant, a lagged dependent variable, and three autoregressive variables is considered. Two-sided CI's for the coefficient on the lagged dependent variable are analyzed.

Standard delta method CI's are found to perform very poorly. For example, across the six cases considered with the coefficient on the lagged dependent variable ranging from .8 to .95 and sample size fifty, a nominal 95% delta method CI has average coverage probability of .766. Block and block–block bootstrap CI's are found to outperform the delta method CI by a noticeable margin. For example, the usual nominal 95% symmetric nonoverlapping block bootstrap CI with block length $\ell = 8$ has average coverage probability of .914 across the six cases. The block–block bootstrap is found to improve upon the block bootstrap in terms of coverage probability. With deletion fraction $\pi = .125$ and block length $\ell = 8$, its average coverage probability is .928 across the six cases. Using the data-dependent method for selecting π and ℓ , the average coverage probability of the block–block bootstrap is .933 or .923 depending upon the version employed.

On the other hand, the block bootstrap yields longer CI's than the delta method and the block–block bootstrap yields longer CI's than the block bootstrap. The average length of the delta method CI's over the six cases to which we refer above is .63. In contrast, the average lengths for the block bootstrap with $\ell = 8$, the block–block bootstrap with $\pi = .125$ and $\ell = 8$, and the block–block bootstrap with the two versions of the data-dependent method of choosing π and ℓ are 1.03, 1.15, 1.23, and 1.19, respectively. All of the CI's are shorter than they need to be in order to achieve coverage probabilities of .95. But, part of the increase in the lengths of the block–block bootstrap CI's over the delta method and block bootstrap CI's is due to the blocking of the original sample estimator.

In sum, the Monte Carlo results illustrate that the block–block bootstrap improves the finite sample coverage probabilities of the block bootstrap in the dynamic regression models that are considered. The results also show that any of the bootstrap methods considered outperforms the delta method by a substantial margin.

We now discuss some alternative bootstraps for time series to the block bootstrap. One alternative is the parametric bootstrap for Markov time series. See Andrews (2004) and Andrews and Lieberman (2004) for analyses of

the higher-order improvements of this bootstrap. An obvious restriction of the parametric bootstrap is that it requires the existence of a parametric model or, at least, a conditional parametric model given some covariates.

Another alternative bootstrap procedure for Markov processes, that does not require a parametric model, is the Markov conditional bootstrap (MCB). Under some conditions, the MCB yields asymptotic refinements that exceed those of the block bootstrap; see Horowitz (2003). The MCB utilizes a nonparametric density estimator of the Markov transition density. This density has dimension equal to the product of the dimension of the observed data vector and the order of the Markov process plus one. For example, for a bivariate time series and a first-order Markov process, a four-dimensional density needs to be estimated. Since nonparametric density estimators are subject to the curse of dimensionality, they are reliable only when the density has dimension less than or equal to three or, perhaps, four. In consequence, the range of application of the MCB is restricted to cases in which it suffices to estimate a low-dimensional density.

The tapered bootstrap of Paparoditis and Politis (2001, 2002) (PP) is another alternative to the block bootstrap. The tapered bootstrap is a variant of the block bootstrap in which the observations near the ends of the bootstrap blocks are down-weighted. PP show that the tapered bootstrap is asymptotically correct to first order and that it reduces the asymptotic bias of the bootstrap variance estimator. PP do not address the issue of asymptotic refinements of the tapered bootstrap.

When the standard block bootstrap is applied to block statistics, as it is in this paper, the resulting bootstrap is a tapered bootstrap in which the tapering function is rectangular. Hence, the bootstrap procedure considered here is related to the tapered bootstrap of PP.³ However, the key to obtaining the improved asymptotic refinements of the block–block bootstrap over the block bootstrap is that *both* the original sample statistic and the block bootstrap down-weight observations near the end of the blocks. This is not considered in PP and it differentiates the approach taken in this paper from that of PP.

The discussion above indicates that the available alternatives to the block bootstrap for time series are useful, but are either only applicable in restrictive contexts or are not known to produce asymptotic refinements. In consequence, the problem addressed in this paper of how to increase the asymptotic refinements of the block bootstrap remains an important problem.

The results of this paper apply using the same assumptions and for the same cases as considered in A2002. In particular, two types of block bootstrap are considered—the nonoverlapping block bootstrap, introduced by Carlstein

³Furthermore, one could consider block statistics that are defined using a smooth tapering function. The block bootstrap applied to such statistics would be a tapered bootstrap. It is likely that the block bootstrap applied to such statistics would yield asymptotic refinements akin to those obtained in this paper.

(1986), and the overlapping block bootstrap, introduced by Künsch (1989). The results apply to extremum estimators, including quasi-maximum likelihood, least squares, and generalized method of moment (GMM) estimators. The results cover t statistics, Wald statistics, and J statistics based on the extremum estimators. One-sided, symmetric two-sided, and equal-tailed two-sided t tests and CI's are covered by the results. Tests of over-identifying restrictions are covered.

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 A2002 and Hall and Horowitz (1996).

Some papers in the literature that do not impose the uncorrelatedness restriction beyond κ lags are Götze and Künsch (1996), Lahiri (1996), and Inoue and Shintani (2000). However, if the uncorrelatedness restriction does not hold and one employs a heteroskedasticity and autocorrelation consistent covariance matrix estimator, then the asymptotic refinements of the block bootstrap are smaller than otherwise and they depend on the choice of the smoothing parameter. The use of block statistics also may prove to have advantages in such cases. This is left to further research.

The proofs of the results in this paper make extensive use of the results of A2002. That paper, in turn, relies heavily on the methods used by Hall and Horowitz (1996), Bhattacharya and Ghosh (1978), Chandra and Ghosh (1979), Götze and Hipp (1983, 1994), and Bhattacharya (1987).

The paper A2002 considers the k -step block bootstrap as well as the standard block bootstrap. The asymptotic refinements established in this paper for the block-block bootstrap also hold for the k -step block bootstrap applied to block statistics provided the condition in A2002 on the magnitude of k is satisfied.

The remainder of the paper is organized as follows: Section 2 defines the block extremum estimators. Section 3 defines the overlapping and nonoverlapping block-block bootstraps. Section 4 states the assumptions. Section 5 establishes the asymptotic refinements of the block-block bootstrap. Section 6 presents a data-dependent method of choosing π and ℓ . Section 7 reports some Monte Carlo results. An Appendix contains proofs of the results.

2. BLOCK EXTREMUM ESTIMATORS AND TESTS

In this section, we define the block statistics that are considered in the paper. As much as possible, we use the same notation as A2002 and Hall and Horowitz (1996). The observations are $\{X_i : i = 1, \dots, n\}$, where $X_i \in R^{L_x}$. The observations are assumed to be from a (strictly) stationary ergodic sequence of random vectors. We consider block versions of extremum estimators of an unknown parameter $\theta \in \Theta \subset R^{L_\theta}$. The estimators we consider are either GMM estimators or estimators that minimize a sample average, which we call “minimum ρ estimators.” Examples of minimum ρ estimators are maximum likelihood (ML), least squares (LS), and regression M estimators.

The GMM estimators that we consider 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 $L_g \geq L_\theta$. The minimum ρ estimators that we consider minimize a sample average of terms $\rho(X_i, \theta)$, where $\rho(\cdot, \cdot)$ is a known real function. Minimum ρ estimators can be written as GMM estimators with $g(X_i, \theta) = (\partial/\partial\theta)\rho(X_i, \theta)$.

We assume that the true moment vectors $\{g(X_i, \theta_0) : i \geq 1\}$ (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 A2002 and Hall and Horowitz (1996). For this reason, we let

$$(2.1) \quad \tilde{X}_i = (X'_i, X'_{i+1}, \dots, X'_{i+\kappa})' \quad \text{for } i = 1, \dots, n - \kappa.$$

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

$$(2.2) \quad \chi_N = \{\tilde{X}_i : i = 1, \dots, N\},$$

where $N = [(n - \kappa)/\ell]\ell$ for block bootstraps with block length ℓ and $[\cdot]$ denotes the integer part of \cdot . Thus, as in A2002, Hall and Horowitz (1996), and Götze and Künsch (1996), some observations \tilde{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 .⁴

⁴For convenience, we state that limits are as $N \rightarrow \infty$ below, although, strictly speaking, they are limits as $n \rightarrow \infty$.

Block statistics are based on sample averages of functions with certain summands deleted. The fraction of observations deleted is π , where π satisfies the conditions stated in the Introduction. As above, $\tau = 1 - \pi$ and ℓ is the block length. Given a function such as $g(X_i, \theta)$, we let $g_\pi(X_i, \theta)$ denote the function that is zero if the time subscript i corresponds to an observation that is one of the $\lceil \pi\ell \rceil$ observations before a join point and is $g(X_i, \theta)$ otherwise. Thus,

$$(2.3) \quad g_\pi(X_i, \theta) = \begin{cases} g(X_i, \theta) & \text{if } i \in [(j-1)\ell + 1, j\ell - \lceil \pi\ell \rceil] \\ & \text{for some } j = 1, \dots, b, \\ 0 & \text{otherwise.} \end{cases}$$

We consider two forms of block GMM estimator. The first is a one-step block 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 block GMM estimator that utilizes an asymptotically optimal weight matrix. It relies on a one-step block GMM estimator to define its weight matrix.

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

$$(2.4) \quad \min_{\theta \in \Theta} J_{N,\pi}(\theta) = \left((N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \theta) \right)' \Omega \left((N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \theta) \right).$$

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

$$(2.5) \quad \min_{\theta \in \Theta} J_{N,\pi}(\theta, \tilde{\theta}_N) = \left((N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \theta) \right)' \Omega_{N,\pi}(\tilde{\theta}_N) \\ \times \left((N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \theta) \right), \quad \text{where}$$

$$\Omega_{N,\pi}(\theta) = \overline{W}_{N,\pi}^{-1}(\theta),$$

$$\overline{W}_{N,\pi}(\theta) = (N\tau)^{-1} \sum_{i=1}^N \left(g_\pi(X_i, \theta)g_\pi(X_i, \theta)' + \sum_{j=1}^{\kappa} H_\pi(X_i, X_{i+j}, \theta) \right),$$

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

and $\tilde{\theta}_N$ solves (2.4). By definition, $H_\pi(X_i, X_{i+j}, \theta)$ equals zero or not depending on the value of i , not $i + j$.

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

$$(2.6) \quad \min_{\theta \in \Theta} (N\tau)^{-1} \sum_{i=1}^N \rho_{\pi}(X_i, \theta),$$

where $\rho_{\pi}(X_i, \theta)$ is defined analogously to $g_{\pi}(X_i, \theta)$ in (2.3) with $g(X_i, \theta)$ replaced by $\rho(X_i, \theta)$. For this estimator, we let $g_{\pi}(X_i, \theta)$ denote $(\partial/\partial\theta)\rho_{\pi}(X_i, \theta)$. Except for consistency properties, the block minimum ρ estimator can be analyzed simultaneously with the block GMM estimators. The reason is that with probability that goes to one (at an appropriate rate) the solution $\widehat{\theta}_N$ to the minimization problem (2.6) 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 block GMM criterion function.

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

$$(2.7) \quad \sigma = \begin{cases} (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1} \\ \quad \times \Omega D(D'\Omega D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4),} \\ (D'\Omega_0 D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5),} \\ D^{-1}\Omega_0^{-1}D^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.6),} \end{cases} \quad \text{where}$$

$$\Omega_0 = \lim_{N \rightarrow \infty} (E\overline{W}_{N,\pi}(\theta_0))^{-1} \quad \text{and} \quad D = E \frac{\partial}{\partial\theta'} g(X_i, \theta_0).$$

By stationarity, Ω_0 does not depend on π .

A consistent estimator of σ is

$$(2.8) \quad \sigma_{N,\pi} = \begin{cases} (D'_{N,\pi}\Omega D_{N,\pi})^{-1}D'_{N,\pi}\Omega\Omega_{N,\pi}^{-1}(\widehat{\theta}_N) \\ \quad \times \Omega D_{N,\pi}(D'_{N,\pi}\Omega D_{N,\pi})^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4),} \\ (D'_{N,\pi}\Omega_{N,\pi}(\widehat{\theta}_N)D_{N,\pi})^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5),} \\ D_{N,\pi}^{-1}\Omega_{N,\pi}^{-1}(\widehat{\theta}_N)D_{N,\pi}^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.6),} \end{cases} \quad \text{where}$$

$$D_{N,\pi} = (N\tau)^{-1} \sum_{i=1}^N \frac{\partial}{\partial\theta'} g_{\pi}(X_i, \widehat{\theta}_N).$$

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

$$(2.9) \quad T_N = (N\tau)^{1/2}(\widehat{\theta}_{N,r} - \theta_{0,r})/(\sigma_{N,\pi})_{rr}^{1/2}.$$

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 block Wald statistic for testing $H_0: \eta(\theta_0) = 0$ versus $H_1: \eta(\theta_0) \neq 0$ is

$$(2.10) \quad \mathcal{W}_N = (N\tau)\eta(\widehat{\theta}_N)' \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N) \sigma_{N,\pi} \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N) \right) \right)^{-1} \eta(\widehat{\theta}_N).$$

The block J statistic for testing over-identifying restrictions is

$$(2.11) \quad J_N = K_{N,\pi}(\widehat{\theta}_N)' K_{N,\pi}(\widehat{\theta}_N), \quad \text{where}$$

$$K_{N,\pi}(\theta) = \Omega_{N,\pi}^{1/2}(\theta) (N\tau)^{-1/2} \sum_{i=1}^N g_\pi(X_i, \theta)$$

and $\widehat{\theta}_N$ is the block two-step GMM estimator. Under H_0 , T_N has an asymptotic $N(0, 1)$ distribution. If $L_g > L_\theta$ and the over-identifying 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 block GMM estimator is used to define the block J statistic.)

3. THE BLOCK-BLOCK BOOTSTRAP

The observations to be bootstrapped are $\{\widetilde{X}_i: 1 \leq i \leq N\}$. As above, the block length ℓ satisfies $\ell \propto N^\gamma$ for some $0 < \gamma < 1$. (Note that one can take $\gamma = 0$ if the data are m -dependent.) We consider both nonoverlapping and overlapping block bootstraps. For the nonoverlapping block bootstrap, the first block is $\widetilde{X}_1, \dots, \widetilde{X}_\ell$, the second block is $\widetilde{X}_{\ell+1}, \dots, \widetilde{X}_{2\ell}$, etc. There are b different blocks, where $b\ell = N$. For the overlapping block bootstrap, the first block is $\widetilde{X}_1, \dots, \widetilde{X}_\ell$, the second block is $\widetilde{X}_2, \dots, \widetilde{X}_{\ell+1}$, etc. There are $N - \ell + 1$ different blocks.

The bootstrap is implemented by sampling b blocks randomly with replacement from either the b nonoverlapping or the $N - \ell + 1$ overlapping blocks. Let $\widetilde{X}_1^*, \dots, \widetilde{X}_N^*$ denote the bootstrap sample obtained from this sampling scheme.

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

$$(3.1) \quad \min_{\theta \in \Theta} J_{N,\pi}^*(\theta) = \left((N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta) \right)' \Omega$$

$$\times \left((N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta) \right), \quad \text{where}$$

$$g_\pi^*(X_i^*, \theta) = g_\pi(X_i^*, \theta) - E^* g_\pi(X_i^*, \widehat{\theta}_N),$$

X_i^* denotes the first element of \tilde{X}_i^* , E^* denotes expectation with respect to the distribution of the bootstrap sample conditional on the original sample, and $g_\pi(X_i^*, \theta)$ is defined as $g_\pi(X_i, \theta)$ is defined in (2.3) but with X_i^* in place of X_i . For the nonoverlapping and overlapping block bootstraps, respectively, we have:

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

The bootstrap sample moments $(N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta)$ in (3.1) are recentered (by subtracting off $E^* g_\pi(X_i^*, \hat{\theta}_N)$) to ensure that their expectation $E^*(N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta)$ equals zero when $\theta = \hat{\theta}_N$, which mimics the population moments $E g_\pi(X_i, \theta)$, 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 Hall and Horowitz (1996) and A2002.

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

$$\begin{aligned}
 (3.3) \quad & \min_{\theta \in \Theta} J_{N,\pi}^*(\theta, \tilde{\theta}_N^*) = \left((N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta) \right)' \Omega_{N,\pi}^*(\tilde{\theta}_N^*) \\
 & \times \left((N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta) \right), \quad \text{where} \\
 & \Omega_{N,\pi}^*(\theta) = \overline{W}_{N,\pi}^*(\theta)^{-1}, \\
 & \overline{W}_{N,\pi}^*(\theta) = (N\tau)^{-1} \sum_{i=1}^N \left(g_\pi^*(X_i^*, \theta) g_\pi^*(X_i^*, \theta)' \right. \\
 & \quad \left. + \sum_{j=1}^{\kappa} H_\pi^*(X_i^*, X_{i,i+j}^*, \theta) \right), \\
 & H_\pi^*(X_i^*, X_{i,i+j}^*, \theta) = g_\pi^*(X_i^*, \theta) g_\pi^*(X_{i,i+j}^*, \theta)' + g_\pi^*(X_{i,i+j}^*, \theta) g_\pi^*(X_i^*, \theta)',
 \end{aligned}$$

$\tilde{\theta}_N^*$ denotes the bootstrap one-step block GMM estimator that solves (3.1), and $X_{i,i+j}^*$ denotes the $(j + 1)$ st element of \tilde{X}_i^* for $j = 1, \dots, \kappa$.

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

$$(3.4) \quad \min_{\theta \in \Theta} (N\tau)^{-1} \sum_{i=1}^N (\rho_\pi(X_i^*, \theta) - E^* g_\pi(X_i^*, \hat{\theta}_N)' \theta),$$

where $g_\pi(\cdot, \theta) = (\partial/\partial\theta)\rho_\pi(\cdot, \theta)$. For the nonoverlapping block bootstrap, the term $(N\tau)^{-1} \sum_{i=1}^N E^* g_\pi(X_i^*, \hat{\theta}_N)' \theta$ is zero, because $(N\tau)^{-1} \sum_{i=1}^N E^* g_\pi(X_i^*, \hat{\theta}_N) = (N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \hat{\theta}_N) = 0$, where the second equality holds by the first-order conditions for $\hat{\theta}_N$ using the fact that the dimensions of $g_\pi(\cdot, \cdot)$ and θ are equal. For the overlapping block bootstrap, $(N\tau)^{-1} \sum_{i=1}^N E^* g_\pi(X_i^*, \hat{\theta}_N) \neq (N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i, \hat{\theta}_N) = 0$ and the extra term in (3.4) is nonzero. In this case, the term $(N\tau)^{-1} \sum_{i=1}^N E^* g_\pi(X_i^*, \hat{\theta}_N)' \theta$ properly recenters the block minimum ρ bootstrap criterion function. It yields bootstrap population first-order conditions that equal zero at $\hat{\theta}_N$, as desired. That is,

$$\begin{aligned} & E^*(\partial/\partial\theta) \left((N\tau)^{-1} \sum_{i=1}^N (\rho_\pi(X_i^*, \theta) - E^* g_\pi(X_i^*, \hat{\theta}_N)' \theta) \right) \\ &= E^*(N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta) = 0, \end{aligned}$$

when $\theta = \hat{\theta}_N$. With this recentring, the first-order conditions for θ_N^* are $(N\tau)^{-1} \sum_{i=1}^N g_\pi^*(X_i^*, \theta_N^*) = 0$, rather than $(N\tau)^{-1} \sum_{i=1}^N g_\pi(X_i^*, \theta_N^*) = 0$, which means that θ_N^* minimizes the one-step block GMM bootstrap criterion function $J_{N,\pi}^*(\theta)$ with $g_\pi(\cdot, \theta) = (\partial/\partial\theta)\rho_\pi(\cdot, \theta)$ and arbitrary positive definite weight matrix Ω .

The bootstrap block covariance matrix estimator is

$$(3.5) \quad \sigma_{N,\pi}^* = \sigma_{N,\pi}^*(\theta_N^*), \quad \text{where}$$

$$\sigma_{N,\pi}^*(\theta) = \begin{cases} \left(D_{N,\pi}^*(\theta)' \Omega D_{N,\pi}^*(\theta) \right)^{-1} D_{N,\pi}^*(\theta) \\ \quad \times \Omega \Omega_{N,\pi}^*(\theta)^{-1} \Omega D_{N,\pi}^*(\theta) \\ \quad \times \left(D_{N,\pi}^*(\theta)' \Omega D_{N,\pi}^*(\theta) \right)^{-1} & \text{if } \hat{\theta}_N \text{ solves (2.4),} \\ \left(D_{N,\pi}^*(\theta)' \Omega_{N,\pi}^*(\theta) D_{N,\pi}^*(\theta) \right)^{-1} & \text{if } \hat{\theta}_N \text{ solves (2.5),} \\ D_{N,\pi}^*(\theta)^{-1} \Omega_{N,\pi}^*(\theta)^{-1} \\ \quad \times D_{N,\pi}^*(\theta)^{-1} & \text{if } \hat{\theta}_N \text{ solves (2.6) and} \end{cases}$$

$$D_{N,\pi}^*(\theta) = (N\tau)^{-1} \sum_{i=1}^N \frac{\partial}{\partial\theta'} g_\pi(X_i^*, \theta).$$

The bootstrap block t , Wald, and J statistics are

$$\begin{aligned}
 (3.6) \quad T_N^* &= (N\tau)^{1/2}(\theta_{N,r}^* - \widehat{\theta}_{N,r})/\sigma_{N,\pi}^*(\theta_N^*)_{rr}^{1/2}, \\
 \mathcal{W}_N^* &= H_{N,\pi}^*(\theta_N^*)'H_{N,\pi}^*(\theta_N^*), \quad \text{and} \\
 J_N^* &= K_{N,\pi}^*(\theta_N^*)'K_{N,\pi}^*(\theta_N^*), \quad \text{where} \\
 H_{N,\pi}^*(\theta) &= \left(\left(\frac{\partial}{\partial \theta'} \eta(\theta) \right) \sigma_{N,\pi}^*(\theta) \left(\frac{\partial}{\partial \theta'} \eta(\theta) \right)' \right)^{-1/2} \\
 &\quad \times (N\tau)^{1/2}(\eta(\theta) - \eta(\widehat{\theta}_N)), \\
 K_{N,\pi}^*(\theta) &= \Omega_N^*(\theta)^{1/2}(N\tau)^{-1/2} \sum_{i=1}^N g_\pi^*(X_i^*, \theta),
 \end{aligned}$$

where $\theta_{N,r}^*$ denotes the r th element of θ_N^* and $\sigma_{N,\pi}^*(\theta_N^*)_{rr}$ denotes the (r, r) th element of $\sigma_{N,\pi}^*(\theta_N^*)$. Note that the bootstrap block t , Wald, and J statistics are *not* defined using correction factors, in contrast to the test statistics considered in Hall and Horowitz (1996) and A2002. Because of the block nature of the statistics, we do not have to correct for the fact that the bootstrap blocks are independent.

Let $z_{|T|,\alpha}^*$, $z_{T,\alpha}^*$, $z_{\mathcal{W},\alpha}^*$, and $z_{J,\alpha}^*$ denote the $1 - \alpha$ quantiles of $|T_N^*|$, T_N^* , \mathcal{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 $|P^*(|T_N^*| \leq z) - (1 - \alpha)|$ over $z \in R$. The precise definitions of $z_{T,\alpha}^*$, $z_{\mathcal{W},\alpha}^*$, and $z_{J,\alpha}^*$ are analogous.

Each of the following tests is of asymptotic significance level α . The symmetric two-sided block–block 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 block–block 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 block–block bootstrap t test of $H_0 : \theta_r \leq \theta_{0,r}$ versus $H_1 : \theta_r > \theta_{0,r}$ rejects H_0 if $T_N > z_{T,\alpha}^*$. The block–block bootstrap Wald test of $H_0 : \eta(\theta_0) = 0$ versus $H_1 : \eta(\theta_0) \neq 0$ rejects the null hypothesis if $\mathcal{W}_N > z_{\mathcal{W},\alpha}^*$. The block–block bootstrap J test of over-identifying 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 block–block bootstrap CI for $\theta_{0,r}$ is

$$\left[\widehat{\theta}_{N,r} - z_{|T|,\alpha}^*(\sigma_N)_{rr}^{1/2}/N^{1/2}, \widehat{\theta}_{N,r} + z_{|T|,\alpha}^*(\sigma_N)_{rr}^{1/2}/N^{1/2} \right].$$

The equal-tailed two-sided block–block bootstrap CI for $\theta_{0,r}$ is

$$\left[\widehat{\theta}_{N,r} - z_{T,\alpha/2}^*(\sigma_N)_{rr}^{1/2}/N^{1/2}, \widehat{\theta}_{N,r} + z_{T,1-\alpha/2}^*(\sigma_N)_{rr}^{1/2}/N^{1/2} \right].$$

The upper one-sided block-block bootstrap CI for $\theta_{0,r}$ is $[\widehat{\theta}_{N,r} - z_{T,\alpha}^*(\sigma_N)_{rr}^{1/2}/N^{1/2}, \infty)$. The block-block Wald-based bootstrap confidence region for $\eta(\theta_0)$ is

$$\left\{ \eta \in R^{L_\eta} : N(\eta(\widehat{\theta}_N) - \eta)' \left((\partial\eta(\widehat{\theta}_N)/\partial\theta)' \sigma_{N,\pi} (\partial\eta(\widehat{\theta}_N)/\partial\theta)' \right)^{-1} \times (\eta(\widehat{\theta}_N) - \eta) \leq z_{W,\alpha}^* \right\}.$$

4. ASSUMPTIONS

We now introduce the assumptions. They are essentially the same as those of A2002 and are similar to those of Hall and Horowitz (1996).

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 $j = 0, \dots, \kappa$, 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 block GMM, two-step block GMM, or block minimum ρ estimator.

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

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

ASSUMPTION 2: (a) Θ is compact and θ_0 is an interior point of Θ . (b) Either (i) $\widehat{\theta}_N$ minimizes $J_{N,\pi}(\theta)$ or $J_{N,\pi}(\theta, \theta_N)$ over $\theta \in \Theta$; θ_0 is the unique solution in Θ to $Eg(X_1, \theta) = 0$; for some function $C_g(x)$, $\|g(x, \theta_1) - g(x, \theta_2)\| \leq C_g(x) \times \|\theta_1 - \theta_2\|$ 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) $\widehat{\theta}_N$ minimizes $N^{-1} \sum_{i=1}^N \rho_\pi(X_i, \theta)$ 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 $0 \leq \kappa < \infty$. (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 $d_1 \geq 3$ and $d_2 \geq 0$. (d) There is a function $C_{\partial f}(\widetilde{X}_1)$ such that $\|(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta) - (\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)\| \leq C_{\partial f}(\widetilde{X}_1)\|\theta - \theta_0\|$ for all $\theta \in N_0$ for all $j = 0, \dots, d_2$. (e) $EC_{\partial f}^{q_2}(\widetilde{X}_1) < \infty$ and $E\|(\partial^j/\partial\theta^j)f(\widetilde{X}_1, \theta_0)\|^{q_2} \leq C_f < \infty$ for all $j = 0, \dots, d_2$ 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 \tilde{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 $L_\eta \leq L_\theta$.

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 $\delta < \|t\| < N^\zeta$,

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

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

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 $\{X_i : i \geq 1\}$ are iid. The moment conditions in Assumptions 2 and 3 are stronger than necessary, but lead to relatively simple results. See Andrews (2001) for a much more complicated set of assumptions, but with weaker moment conditions than those above, that are sufficient for the results given below.

5. ASYMPTOTIC REFINEMENTS OF THE BLOCK-BLOCK BOOTSTRAP

In this section, we show that the block-block bootstrap leads to greater asymptotic refinements in ERP's of tests and in CI coverage probabilities when compared to the block bootstrap, as well as in comparison to procedures based on first-order asymptotics.

The following theorem shows that the symmetric two-sided block-block bootstrap t , Wald, and J tests have ERP's of magnitude $o(N^{-(1+\xi)})$ for all $\xi < 1/2 - \gamma$ when the block length ℓ is chosen proportional to N^γ . It shows that the block-block bootstrap equal-tailed two-sided t and one-sided t tests have ERP's of magnitude $o(N^{-(1/2+\xi)})$ for all $\xi < 1/2 - \gamma$ when ℓ is chosen proportional to N^γ . The only restriction on γ is that $0 < \gamma < 1/2$. Hence, for γ close to zero, ξ is close to $1/2$. For m -dependent data, $\gamma = 0$ is permitted.

In contrast, with the block bootstrap, analogous results hold but with the additional restriction that $\xi < \gamma$. The latter restriction plus $\xi < 1/2 - \gamma$ imply that $\xi < 1/4$.

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

THEOREM 1: (a) Suppose Assumptions 1–4 hold with $d_1 \geq 5$ and $d_2 \geq 4$; $0 \leq \xi < 1/2 - \gamma$; $0 < \gamma < 1/2$; $\pi \in (0, 1)$; and $\pi \rightarrow 0$ and $\pi\ell - C \log(N) \rightarrow \infty$ as $N \rightarrow \infty$ for all constants $0 < C < \infty$. Then, under $H_0 : \theta_r = \theta_{0,r}$,

$$P(|T_N| > z_{|T|,\alpha}^*) = \alpha + o(N^{-(1+\xi)}).$$

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

$$P(\mathcal{W}_N > z_{\mathcal{W},\alpha}^*) = \alpha + o(N^{-(1+\xi)}).$$

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

$$P(J_N > z_{J,\alpha}^*) = \alpha + o(N^{-(1+\xi)}).$$

(b) Suppose Assumptions 1–4 hold with $d_1 \geq 4$ and $d_2 \geq 3$; $0 \leq \xi < 1/2 - \gamma$; $0 < \gamma < 1/2$; $\pi \in (0, 1)$; and $\pi \rightarrow 0$ and $\pi\ell - C \log(N) \rightarrow \infty$ as $N \rightarrow \infty$ for all $0 < C < \infty$. Then, under $H_0: \theta_r = \theta_{0,r}$,

$$P(T_N < z_{T,\alpha/2}^* \text{ or } T_N > z_{T,1-\alpha/2}^*) = \alpha + o(N^{-(1/2+\xi)}) \quad \text{and}$$

$$P(T_N > z_{T,\alpha}^*) = \alpha + o(N^{-(1/2+\xi)}).$$

(c) If the observations $\{X_i: i \geq 1\}$ are m -dependent for some integer $m < \infty$, then the results of parts (a) and (b) hold under the stated conditions, but with $\gamma = 0$ and with the restrictions on π replaced by $\limsup_{N \rightarrow \infty} \pi < 1$ and $\liminf_{N \rightarrow \infty} \lceil \pi\ell \rceil \geq m + \kappa$.

COMMENTS: 1. The errors in part (a) of the theorem when the critical values are based on standard first-order asymptotics (using the normal distribution or the chi-square distribution) are $O(N^{-1})$ for each of the three statistics. The errors in part (b) of the theorem are $O(N^{-1})$ for the equal-tailed t test and $O(N^{-1/2})$ for the one-sided t test. Thus, part (a) of the theorem and the one-sided t test result of part (b) show that the bootstrap critical values reduce the ERP (and the error in CI coverage probability) relative to first-order asymptotics by a factor of at least $N^{-\xi}$. The choice of γ close to zero maximizes ξ subject to the requirement of the theorem that $\xi < 1/2 - \gamma$. For such a choice of γ , the results of parts (a) and (b) hold for ξ close to $1/2$.

2. The equal-tailed t test result of part (b) of the theorem shows that standard first-order asymptotics yields smaller ERP than the block–block bootstrap ERP by $N^{1/2-\xi}$, where $\xi < 1/2 - \gamma$. This occurs because the standard equal-tailed t test is symmetric by the symmetry of the normal distribution and, in consequence, the $n^{-1/2}$ term in its Edgeworth expansion drops out. This does not occur with equal-tailed bootstrap tests or CI’s.

On the other hand, if interest is in an equal-tailed test or CI, it is reasonable to be more concerned about the individual probabilities of a CI missing to the left (right) or of falsely rejecting the null in favor of larger (smaller) alternatives, rather than the overall error probabilities. For these one-sided error probabilities, the bootstrap obtains the same asymptotic refinements as for the one-sided t test (or one-sided CI) mentioned above.

3. When the data are m -dependent, part (c) of the theorem shows that one does not need the block length, ℓ , to diverge to infinity as $N \rightarrow \infty$ or the number of observations deleted per block, $\lceil \pi\ell \rceil$, to diverge to infinity as $N \rightarrow \infty$.

What is needed is that the number of observations deleted per block, $\lceil \pi \ell \rceil$, be greater than or equal to $m + \kappa$ for N large. This suffices, because the block statistics are based on sample averages, which are sums of independent blocks provided $\lceil \pi \ell \rceil \geq m + \kappa$, which is exactly mimicked by the independence of the bootstrap blocks.

In contrast, when the block bootstrap is applied to nonblock statistics and the observations are m -dependent, the length of the blocks needs to diverge to infinity as $N \rightarrow \infty$.

4. The reason that symmetric two-sided block–block bootstrap 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 $|T_N|$, W_N , and J_N are zero by a symmetry property. See Hall (1992), Hall and Horowitz (1996), or A2002 for details.

5. The possibility of improving the result of Theorem 1(a) for $|T_N|$ when the data are dependent via the symmetry argument of Hall (1988), which applies with iid data, is unclear; see the discussion in A2002.

6. DATA-DEPENDENT CHOICE OF ℓ AND π

In this section, we briefly discuss a data-dependent method for choosing ℓ and π . The method is a type of nested bootstrap. We do not present any theoretical results concerning its asymptotic behavior, but we investigate its finite sample properties in the next section.

The method is as follows. First, one specifies an approximate model for the observations and a grid of (ℓ, π) values. Next, one simulates the approximate model a large number of times. For each simulation, one computes the block–block bootstrap CI or test of interest for each (ℓ, π) combination in the grid (which requires a second level of simulation). One then selects the (ℓ, π) combination that optimizes a chosen criterion function computed using the simulated block–block bootstrap CI's or tests. For example, one could select the (ℓ, π) combination that minimizes the empirical coverage probability error based on simulated CI's or the empirical ERP based on simulated tests. Given the selected (ℓ, π) combination, one computes the block–block bootstrap CI or test based on the original sample.

The object in choosing the approximate model is to capture the amount of dependence in the time series because this is what determines a good choice of (ℓ, π) . The approximate model need not be correct asymptotically. For example, with the dynamic regression model considered in the simulations below, we take the approximate model to be the dynamic regression model with parameters equal to the estimated values (based on the original sample), with errors taken to be iid with distribution equal to the empirical distribution of the residuals from the original sample, and with the exogenous variables set

equal to their original sample values.⁵ In addition, in the simulations below, we consider the approximate model given by the block bootstrap with a relatively large block length. Other possibilities for approximate models (that we do not consider below) include the use of univariate autoregressive processes or a vector autoregressive process, possibly with orders determined by an information criterion.

The grid of (ℓ, π) values can be selected from a sequence of grids such that all sequences of (ℓ, π) values taken from the grids satisfy the properties specified above that $\pi \rightarrow 0$ and $\pi\ell - C \log(N) \rightarrow \infty$ as $N \rightarrow \infty$ for all constants $0 < C < \infty$. With the grid selected in this way, any chosen (ℓ, π) combination is compatible with the requirements for the asymptotic refinements established above. Hence, if the asymptotic refinements hold uniformly over the sequences from the grids, then the data-dependent method of selecting (ℓ, π) will yield the asymptotic refinements established above. However, we do not establish that the appropriate form of uniformity actually holds.

The criterion we consider in the simulation results below for selecting the best (ℓ, π) value is the absolute deviation of the empirical coverage probability (based on the simulations from the approximate model) from the desired coverage probability of a block–block bootstrap CI. The empirical coverage probability is calculated with the true value being the estimated value from the original sample. Thus, the empirical coverage probability for a given (ℓ, π) combination is the fraction of times out of all the simulations from the approximate model that the block–block bootstrap CI with this (ℓ, π) combination includes the estimated value from the original sample.

The data-dependent method for selecting (ℓ, π) can be time consuming because it is a nested bootstrap procedure. If need be, the amount of time can be reduced by using a smaller number of block–block bootstrap repetitions when selecting (ℓ, π) than when computing the final block–block bootstrap CI based on the selected (ℓ, π) combination. For example, in the Monte Carlo experiment below, we use 300 simulations of the approximate model, 399 block–block bootstrap simulations for each approximate model simulation and each (ℓ, π) combination, and 999 block–block bootstrap simulations for the final block–block bootstrap CI given the selected (ℓ, π) combination. In practice, it would be desirable to use more than 300 and 399 simulations in the selection of the (ℓ, π) value. But, for the Monte Carlo simulation results, the nested procedure must be repeated a large number of times—we computed 3,000 Monte Carlo repetitions—so it was not feasible to do so in the Monte Carlo results.

⁵This is a residual-based bootstrap. It is not appropriate to use this bootstrap to construct the final bootstrap CI unless one knows that the exogenous variables are strictly exogenous, rather than just weakly exogenous. But, it is appropriate to use the residual-based bootstrap as an approximate model for determining (ℓ, π) .

7. MONTE CARLO SIMULATIONS

In this section, we describe some Monte Carlo simulation results that are designed to assess the coverage probability accuracy of block–block bootstrap CI's.

7.1. *Experimental Design*

We consider a dynamic linear regression model estimated by LS:

$$\begin{aligned}
 (7.1) \quad Y_i &= \theta_{0,1} + Y_{i-1}\theta_{0,2} + \sum_{j=3}^5 Z_{i,j}\theta_{0,j} + U_i \\
 &= Z_i'\theta_0 + U_i \quad \text{for } i = 1, \dots, N, \quad \text{where} \\
 Z_i &= (1, Y_{i-1}, Z_{i,3}, Z_{i,4}, Z_{i,5})', \\
 \theta_0 &= (\theta_{0,1}, \dots, \theta_{0,5})', \\
 Z_{i,j} &= Z_{i-1,j}\rho_Z + V_{i,j} \quad \text{for } j = 3, 4, 5, \\
 X_i &= (Y_i, Z_i)', \quad \text{and} \\
 g(X_i, \theta) &= (Y_i - Z_i'\theta)Z_i.
 \end{aligned}$$

Five regressors are in the model. One is a constant; one is a lagged dependent variable; and the other three are first-order autoregressive (AR(1)) regressors with the same AR(1) parameter ρ_Z . The innovations, $V_{i,j}$, for the AR(1) regressors are iid across i and j with mean zero and variance one and are independent of the errors U_i . The regressor innovations and the errors are taken to have the same distribution. We consider four different distributions: standard normal, t -5 (rescaled to have variance one), half-normal (recentered and rescaled to have mean zero and variance one), and uniform on $[-\sqrt{12}, \sqrt{12}]$ (which has mean zero and variance one). The latter three distributions were chosen because the t -5 has thick tails, the half-normal is asymmetric but has the same kurtosis as the normal, and the uniform has thin tails. The initial observations used to start up the AR(1) regressors are taken to have the same distribution as the innovations, but are scaled to yield variance stationary processes. The moment vectors $g(X_i, \theta_0)$ are uncorrelated. In terms of the notation introduced above, $\kappa = 0$, $n = N$, and $\tilde{X}_i = X_i$.

The parameters $\theta_{0,1}$, $\theta_{0,3}$, $\theta_{0,4}$, $\theta_{0,5}$ are taken to be zero. Three combinations of $(\theta_{0,2}, \rho_Z)$ are considered: $(.9, .8)$, $(.95, .95)$, and $(.8, .7)$. Two sample sizes N are considered: 50 and 100.

We consider CI's for the parameter $\theta_{0,2}$ on the lagged dependent variable. The CI's are based on a t statistic that employs the LS estimator of $\theta_{0,2}$ coupled

with a heteroskedasticity consistent standard error estimator:

$$\begin{aligned}
 (7.2) \quad T_N &= \frac{N^{1/2}(\widehat{\theta}_{N,2} - \theta_{0,2})}{(\widehat{\sigma}_N)_{22}^{1/2}}, \\
 \widehat{\theta}_N &= \left(\sum_{i=1}^N Z_i Z_i' \right)^{-1} \sum_{i=1}^N Z_i Y_i, \\
 \widehat{\sigma}_N &= \left(N^{-1} \sum_{i=1}^N Z_i Z_i' \right)^{-1} N^{-1} \sum_{i=1}^N \widehat{U}_i^2 Z_i Z_i' \left(N^{-1} \sum_{i=1}^N Z_i Z_i' \right)^{-1}, \quad \text{and} \\
 \widehat{U}_i &= Y_i - Z_i' \widehat{\theta}_N.
 \end{aligned}$$

We compare standard two-sided delta method CI's to symmetric two-sided block-block bootstrap CI's. Nonoverlapping block-block bootstrap CI's are considered. (See Andrews (2002b) for some Monte Carlo results for equal-tailed two-sided and overlapping symmetric block-block bootstrap confidence intervals.) The delta method CI is given by $[\widehat{\theta}_{N,2} - z_{\alpha/2}(\widehat{\sigma}_N)_{22}^{1/2}/N^{1/2}, \widehat{\theta}_{N,2} + z_{\alpha/2}(\widehat{\sigma}_N)_{22}^{1/2}/N^{1/2}]$, where $z_{\alpha/2}$ denotes the $1 - \alpha/2$ quantile of the standard normal distribution. The bootstrap CI's are defined in Section 3 above.

The bootstrap CI's are based on blocks of length $\ell = 4, 6, 8,$ or 10 with the number of observations "skipped" in each block (*Skip*) in the computation of the block statistics equal to $0, 1, 2,$ or 3 with the restriction that $\pi = \text{Skip}/\ell \leq .35$. When *Skip* = 0 , the block-block bootstrap reduces to the standard block bootstrap. For brevity, we do not report results for all (ℓ, Skip) combinations.

We also compute results for the data-dependent selection of (ℓ, Skip) for the block-block bootstrap using the method described in the preceding section. We consider both the dynamic regression approximate model and the nonoverlapping block bootstrap approximate model with block length equal to ten. We select (ℓ, Skip) from the subset of values listed above with *Skip* equal to one or two and also from the subset with *Skip* equal to one, two, or three.

The number of simulation repetitions used is 3,000 for each case considered. This yields simulation standard errors of (approximately) .0040 for the simulated coverage probabilities of nominal 95% CI's. For a single data-dependent choice of (ℓ, Skip) , each simulation repetition took 2.68 minutes using a 2,000 MHz Pentium III computer.

7.2. Simulation Results

Table I reports the simulation results. All results are for nominal 95% CI's. Case (a) is the base case, which has $(\theta_{0,2}, \rho_Z) = (.9, .8)$, standard normal $N(0, 1)$ distributions for the errors and regressor innovations, and sample $N = 50$. Variations on the base case are reported in cases (b)–(g). Cases

TABLE I
 COVERAGE PROBABILITIES AND AVERAGE LENGTHS OF NOMINAL 95% TWO-SIDED DELTA
 METHOD AND SYMMETRIC TWO-SIDED BLOCK AND BLOCK-BLOCK BOOTSTRAP
 CONFIDENCE INTERVALS^a

Confidence Interval Method		CI	CI	Fraction		Fraction				
		Cov	Avg	No. Skipped		Block Lengths				
		Prob	Length	1	2	4	6	8	10	
<i>Case (a): N(0, 1) Dist., $(\theta_{0,2}, \rho_Z) = (.9, .8), N = 50$</i>										
Delta Method:		.759	.61	—	—	—	—	—	—	—
Fixed ($\ell, Skip$):	(8, 0)	.920	.99	.00	.00	.00	.00	1.00	.00	.00
	(8, 1)	.934	1.10	1.00	.00	.00	.00	1.00	.00	.00
	(8, 2)	.949	1.25	.00	1.00	.00	.00	1.00	.00	.00
Data-depend ($\ell, Skip$):	Dyn Reg	.934	1.16	.53	.47	.19	.41	.27	.13	.13
	Blck Boot	.925	1.14	.56	.44	.14	.30	.28	.29	.29
<i>Case (b): N(0, 1) Dist., $(\theta_{0,2}, \rho_Z) = (.95, .95), N = 50$</i>										
Delta Method:		.701	.64	—	—	—	—	—	—	—
Fixed ($\ell, Skip$):	(8, 0)	.895	1.11	.00	.00	.00	.00	1.00	.00	.00
	(8, 1)	.909	1.23	1.00	.00	.00	.00	1.00	.00	.00
	(8, 2)	.927	1.41	.00	1.00	.00	.00	1.00	.00	.00
Data-depend ($\ell, Skip$):	Dyn Reg	.913	1.35	.47	.53	.18	.44	.27	.11	.11
	Blck Boot	.917	1.29	.52	.48	.14	.29	.27	.30	.30
<i>Case (c): N(0, 1) Dist., $(\theta_{0,2}, \rho_Z) = (.8, .7), N = 50$</i>										
Delta Method:		.830	.72	—	—	—	—	—	—	—
Fixed ($\ell, Skip$):	(8, 0)	.932	1.12	.00	.00	.00	.00	1.00	.00	.00
	(8, 1)	.944	1.24	1.00	.00	.00	.00	1.00	.00	.00
	(8, 2)	.948	1.42	.00	1.00	.00	.00	1.00	.00	.00
Data-depend ($\ell, Skip$):	Dyn Reg	.945	1.33	.58	.42	.18	.37	.30	.16	.16
	Blck Boot	.936	1.29	.56	.44	.14	.26	.26	.34	.34
<i>Case (d): t-5 Dist., $(\theta_{0,2}, \rho_Z) = (.9, .8), N = 50$</i>										
Delta Method:		.782	.62	—	—	—	—	—	—	—
Fixed ($\ell, Skip$):	(8, 0)	.923	1.00	.00	.00	.00	.00	1.00	.00	.00
	(8, 1)	.936	1.12	1.00	.00	.00	.00	1.00	.00	.00
	(8, 2)	.946	1.28	.00	1.00	.00	.00	1.00	.00	.00
Data-depend ($\ell, Skip$):	Dyn Reg	.943	1.18	.55	.45	.18	.37	.28	.17	.17
	Blck Boot	.931	1.16	.56	.44	.13	.27	.27	.33	.33
<i>Case (e): Half-Normal Dist., $(\theta_{0,2}, \rho_Z) = (.9, .8), N = 50$</i>										
Delta Method:		.764	.60	—	—	—	—	—	—	—
Fixed ($\ell, Skip$):	(8, 0)	.915	1.00	.00	.00	.00	.00	1.00	.00	.00
	(8, 1)	.926	1.11	1.00	.00	.00	.00	1.00	.00	.00
	(8, 2)	.939	1.26	.00	1.00	.00	.00	1.00	.00	.00
Data-depend ($\ell, Skip$):	Dyn Reg	.934	1.18	.52	.48	.20	.40	.27	.14	.14
	Blck Boot	.911	1.12	.56	.44	.12	.27	.26	.34	.34

TABLE I—Continued

Confidence Interval Method	CI Cov Prob	CI Avg Length	Fraction No. Skipped		Fraction Block Lengths				
			1	2	4	6	8	10	
			Case (f): Uniform Dist., $(\theta_{0,2}, \rho_Z) = (.9, .8), N = 50$						
Delta Method:		.759	.58	—	—	—	—	—	—
Fixed $(\ell, Skip)$:	(8, 0)	.901	.98	.00	.00	.00	.00	1.00	.00
	(8, 1)	.918	1.09	1.00	.00	.00	.00	1.00	.00
	(8, 2)	.932	1.23	.00	1.00	.00	.00	1.00	.00
Data-depend $(\ell, Skip)$:	Dyn Reg	.927	1.18	.52	.48	.16	.39	.28	.17
	Bck Boot	.918	1.14	.56	.44	.14	.31	.25	.30
Case (g): $N(0, 1)$ Dist., $(\theta_{0,2}, \rho_Z) = (.9, .8), N = 100$									
Delta Method:		.854	.36	—	—	—	—	—	—
Fixed $(\ell, Skip)$:	(8, 0)	.939	.49	.00	.00	.00	.00	1.00	.00
	(8, 1)	.943	.53	1.00	.00	.00	.00	1.00	.00
	(8, 2)	.953	.58	.00	1.00	.00	.00	1.00	.00
Data-depend $(\ell, Skip)$:	Dyn Reg	.940	.55	.52	.48	.12	.32	.30	.26
	Bck Boot	.941	.54	.57	.43	.14	.23	.26	.38

^aResults for Fixed $(\ell, Skip)$ with $Skip = 0$ are for the standard block bootstrap. Results for Fixed $(\ell, Skip)$ with $Skip = 1$ or 2 and for Data-depend $(\ell, Skip)$ with the Dyn Reg or Bck Boot approximate model are for the block-block bootstrap.

(b) and (c) have $(\theta_{0,2}, \rho_Z) = (.95, .95)$ and $(.8, .7)$, respectively, and otherwise are the same as in the base case. Cases (d), (e), and (f) differ from the base case in that the distributions of the errors and regressor innovations are t -5, half-normal, and uniform on $[-\sqrt{12}, \sqrt{12}]$, respectively. Case (g) differs from the base case in that $N = 100$.

In Table I we report results for fixed $(\ell, Skip)$ combinations of (8, 0), (8, 1), and (8, 2). For brevity, results for other combinations are not reported, but are commented on in the text. Table I also reports results for the data-dependent selection of $(\ell, Skip)$ using the dynamic regression approximate model, in the rows denoted “Dyn Reg,” and using the block bootstrap approximate model, in the rows denoted “Bck Boot.” For brevity, we only report results for the case where $Skip$ is allowed to equal one or two, but we comment on the results for the case where it may be one, two, or three in the text.

For each CI considered, Table I reports the CI coverage probabilities and CI average lengths based on the 3,000 Monte Carlo simulations. For the bootstrap CI’s, Table I also reports the fraction of $Skip$ values equal to one and two, denoted Fraction No. Skipped, and the fraction of block lengths equal 4, 6, 8, and 10, denoted Fraction Block Lengths, over the 3,000 Monte Carlo simulations. For the fixed $(\ell, Skip)$ results, these fractions are either zero or one. For the data-dependent $(\ell, Skip)$ results, these fractions are between zero and one.

The results of Table I show the following:

1. The coverage probabilities of the delta method CI's are poor. The coverage probabilities of the nominal 95% delta method CI are between .701 and .830 when $N = 50$ and are equal to .865 when $N = 100$.

2. The bootstrap CI's perform fairly well and, hence, outperform the delta method CI's by a wide margin. This is true regardless of the choice of the block length ℓ , the number of observations skipped, or the data-dependent method of choosing $(\ell, Skip)$. The nominal 95% bootstrap CI's reported in Table I have coverage probabilities between .895 and .949 when $N = 50$ and between .939 and .953 when $N = 100$.

3. The results for CI's based on fixed $(\ell, Skip)$ values with $\ell = 8$ show that the coverage probabilities increase and the coverage probability errors decrease as $Skip$ increases. For example, in the base case, the probabilities increase from .920 for $Skip = 0$, which is the standard block bootstrap, to .934 and .949 for $Skip = 1$ and 2, respectively. As expected, the average lengths of the CI's also increase as $Skip$ increases. For the base case, the average lengths are .99, 1.10, and 1.25 for $Skip = 0, 1,$ and 2, respectively. Hence, there is a trade-off between the CI coverage probability and the CI average length.

4. The results for fixed $(\ell, Skip)$ combinations with $\ell = 6$ (not reported in the table) are quite similar to those for $\ell = 8$. When $Skip = 0$, which corresponds to the standard block bootstrap, the length of the CI's with fixed $(\ell, Skip)$ combinations tends to increase with ℓ . When $Skip \geq 1$, the performance of the CI's in terms of coverage probabilities and average lengths is more closely related to the fraction of observations skipped, i.e., $\pi = Skip/\ell$, than to ℓ or $Skip$ individually.

5. The results for the CI's with data-dependent choice of $(\ell, Skip)$ are quite good. They yield coverage probabilities between .911 and .945 across the seven cases considered. Not surprisingly, the worst performance is in case (b) in which $(\theta_{0,2}, \rho_Z) = (.95, .95)$ and the best performance is in case (g) in which $N = 100$. The data-dependent choices of $(\ell, Skip)$ perform similarly to the fixed $(\ell, Skip) = (8, 1)$ combination in terms of coverage probabilities and average lengths.

6. In most cases, the use of the dynamic regression approximate model yields better coverage probabilities and slightly longer CI's than the block bootstrap approximate model, but the differences are not large in most cases.

7. Using the data-dependent choices of $(\ell, Skip)$, one observation is skipped about half the time and two are skipped about half the time. The dynamic regression approximate model skips two observations slightly more often than the block bootstrap approximate model.

8. The dynamic regression approximate model selects $\ell = 6$ most frequently followed by $\ell = 8$. The variation across cases with $N = 50$ is not large. When $N = 100$, there is a shift toward longer blocks. The block bootstrap approximate model selects $\ell = 6, 8,$ and 10 with roughly the same frequency when $N = 50$. When $N = 100$, it selects $\ell = 10$ more often.

9. Results for data-dependent choice of (ℓ, Skip) when Skip is allowed to take on the values 1, 2, and 3 (not reported in the table) are similar in terms of coverage probabilities but somewhat worse in terms of CI average length across the cases considered, compared to the results reported in the table in which $\text{Skip} = 1$ or 2. $\text{Skip} = 3$ is chosen in a relatively small fraction of the cases. Hence, enlarging the choice of Skip values beyond 2 does not improve performance in the cases considered.

10. The effect of increasing and decreasing the amount of correlation, as shown in cases (b) and (c), respectively, is as expected for all CI's. Increasing the amount of correlation reduces the coverage probabilities and increases the coverage probability errors of all CI's. Decreasing the amount of correlation increases the coverage probabilities and decreases the coverage probability errors.

11. Shifting from normal to t -5 or half-normal distributions has little effect on CI coverage probabilities or average lengths except that the delta method coverage probability is somewhat higher with t -5 distributions (although still very low). This indicates robustness of the CI's to thick tails and asymmetry. Shifting from normal to uniform distributions reduces the coverage probabilities of the bootstrap CI's but does not change their average lengths. Hence, the bootstrap CI's exhibit some sensitivity to thin-tailed distributions.

12. The effect of increasing the sample size, as shown in case (g), is to increase the coverage probabilities and reduce the coverage probability errors for all CI's.

In sum, the Monte Carlo results show that all of the bootstrap CI's considered outperform the delta method CI's by a substantial margin. The results also show that the block–block bootstrap yields improved coverage probabilities in the cases considered compared to the standard block bootstrap. On the other hand, the block–block bootstrap yields longer CI's than the standard block bootstrap. Some of the increase in length is due to the fact that the block–block bootstrap coverage probabilities are higher and some is due to the loss in information attributable to the skipping of observations by the block estimators employed by the block–block bootstrap CI's. The data-dependent methods of selecting the (ℓ, Skip) combination are found to work fairly effectively in the cases considered. To conclude, there is some evidence that the theoretical advantages established in this paper for the block–block bootstrap are reflected in finite samples.

Cowles Foundation for Research in Economics, Yale University, P.O. Box 208281, New Haven, CT 06520-8281, U.S.A.; donald.andrews@yale.edu.

Manuscript received April, 2002; final revision received May, 2003.

APPENDIX: PROOFS

The proof of Theorem 1 holds by making some adjustments to the proof of Theorem 2 of A2002. The proof of Theorem 2 of A2002 relies on sixteen lemmas. These lemmas need to

be adjusted as follows. Lemma 1 needs to hold for triangular arrays of functions $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$, rather than a single function $h(\cdot)$, in order to apply the lemma with $h_{N,i}(X_i) = g_\pi(X_i, \theta_0)$, rather than $h(X_i) = g(X_i, \theta_0)$. This extension is easily achieved. It is stated as Lemma 1 below.

Given the new Lemma 1 (and the fact that $(N\tau)/N \rightarrow 1$ as $N \rightarrow \infty$ under the assumption that $\pi = 1 - \tau \rightarrow 0$), the proofs of Lemmas 2–13 and 16 of A2002 hold with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$ throughout without any significant changes in their proofs. Lemma 15 of A2002 is not needed when block statistics are considered because it involves the behavior of correction factors, which are not used with block statistics. Lemma 14 of A2002 needs to be changed. In particular, we need to show that it holds with the condition $\xi < \gamma$ deleted. Lemma 2 below gives the required result.

Given that Lemmas 2–14 and 16 of A2002 hold with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$, Theorem 1(a) and (b) hold by the proof of Theorem 2 of A2002. For the case of m -dependent observations $\{X_i : i \geq 1\}$ (covered in Theorem 1(c)), the only adjustment to the proof that is required is that the result of Lemma 14 of A2002 needs to hold with $\gamma = 0$. Lemma 2 below covers this case.

A.1. Lemmas

LEMMA 1: *Suppose Assumption 1 holds.*

(a) *Let $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$ be a triangular array of matrix-valued functions that satisfy $Eh_{N,i}(\tilde{X}_i) = 0$ for all i, N and $\sup_{i \leq N, N \geq 1} E\|h_{N,i}(\tilde{X}_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a/(1 - 2c)$ for some $c \in [0, 1/2)$ and $a \geq 0$. Then, for all $\varepsilon > 0$,*

$$\lim_{N \rightarrow \infty} N^a P\left(\left\|N^{-1} \sum_{i=1}^N h_{N,i}(\tilde{X}_i)\right\| > N^{-c} \varepsilon\right) = 0.$$

(b) *Let $\{h_{N,i}(\cdot) : i \leq N, N \geq 1\}$ be a triangular array of matrix-valued functions that satisfy $\sup_{i \leq N, N \geq 1} E\|h(\tilde{X}_i)\|^p < \infty$ for $p \geq 2$ and $p > 2a$ for some $a \geq 0$. Then, there exists a constant $K < \infty$ such that*

$$\lim_{N \rightarrow \infty} N^a P\left(\left\|N^{-1} \sum_{i=1}^N h_{N,i}(\tilde{X}_i)\right\| > K\right) = 0.$$

Asymptotic refinements of the block bootstrap depend on the differences between the Edgeworth expansions of the df's of T_N and T_N^* being small (and analogously for $(\mathcal{W}_N, \mathcal{W}_N^*)$ and (J_N, J_N^*)). Let $\nu_{N,\pi,a}$ denote a vector of population moments including those of $g_\pi(X_i, \theta_0)$ and some of its partial derivatives with respect to θ ; $\nu_{N,\pi,a}$ is defined precisely below. Let $\nu_{N,\pi,a}^*$ denote an analogous vector of bootstrap moments including those of $g_\pi^*(X_i^*, \theta_0)$ and some of its partial derivatives. Edgeworth expansions of the df's of T_N, \mathcal{W}_N , and J_N at a point y , with remainder of order $o(N^{-a})$, where $2a$ is an integer, depend on polynomials in y whose coefficients are polynomials in the elements of $\nu_{N,\pi,a}$. Analogously, Edgeworth expansions of the df's of T_N^*, \mathcal{W}_N^* , and J_N^* are the same as those of T_N, \mathcal{W}_N , and J_N , but with $\nu_{N,\pi,a}^*$ in place of $\nu_{N,\pi,a}$. In consequence, asymptotic refinements of the block bootstrap depend on the magnitude of the differences between $\nu_{N,\pi,a}^*$ and $\nu_{N,\pi,a}$. Lemma 2 shows that these differences are small asymptotically.

We now define $\nu_{N,\pi,a}$ and $\nu_{N,\pi,a}^*$ precisely. Let $f(\tilde{X}_i, \theta)$ be the vector-valued function defined at the beginning of Section 4. Let $f_\pi(\tilde{X}_i, \theta)$ be the function derived from $f(\tilde{X}_i, \theta)$ in the same way as $g_\pi(\tilde{X}_i, \theta)$ is derived from $g(\tilde{X}_i, \theta)$ in (2.3). Let $f_\pi^*(\tilde{X}_i^*, \theta)$ denote the vector containing the unique components of $g_\pi^*(X_i^*, \theta)$ and $g_\pi^*(X_i^*, \theta)g_\pi^*(X_{i+j}^*, \theta)$ for all $j = 0, \dots, \kappa$ and their derivatives with respect to θ through order d_1 . Let $S_{N,\pi} = (N\tau)^{-1} \sum_{i=1}^N f_\pi(\tilde{X}_i, \theta_0)$, $S_\pi = ES_{N,\pi}$, $S_{N,\pi}^* = (N\tau)^{-1} \sum_{i=1}^N f_\pi^*(\tilde{X}_i^*, \theta_N)$, and $S_\pi^* = E^*S_{N,\pi}^*$. Let $\Psi_{N,\pi} = (N\tau)^{1/2}(S_{N,\pi} - S_\pi)$ and $\Psi_{N,\pi}^* = (N\tau)^{1/2}(S_{N,\pi}^* - S_\pi^*)$. Let $\Psi_{N,\pi,j}$ and $\Psi_{N,\pi,j}^*$ denote the j th elements of $\Psi_{N,\pi}$ and $\Psi_{N,\pi}^*$,

respectively. Let $\nu_{N,\pi,a}$ and $\nu_{N,\pi,a}^*$ denote vectors of moments of the form $(N\tau)^{\alpha(m)} E \prod_{\mu=1}^m \Psi_{N,\pi,j_\mu}$ and $(N\tau)^{\alpha(m)} E^* \prod_{\mu=1}^m \Psi_{N,\pi,j_\mu}^*$, respectively, where $2 \leq m \leq 2a + 2$, $\alpha(m) = 0$ if m is even, and $\alpha(m) = 1/2$ if m is odd.

LEMMA 2: *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) $0 < \gamma < 1/2$ or (ii) the observations $\{X_i : i \geq 1\}$ are m -dependent for some integer $m < \infty$, $\gamma = 0$, $\pi \in (0, 1)$, and $\liminf_{N \rightarrow \infty} \lceil \pi \ell \rceil \geq m + \kappa$. Then,*

$$\lim_{N \rightarrow \infty} N^a P(\|\nu_{N,\pi,a}^* - \nu_{N,\pi,a}\| > (N\tau)^{-\xi}) = 0.$$

COMMENT: The condition $\xi < \gamma$, which is needed in Lemma 14 of A2002, is not needed in Lemma 2 because the moments considered are moments of block statistics. This is the key feature of block statistics that allows the block–block bootstrap to attain larger asymptotic refinements than the block bootstrap applied to standard statistics.

A.2. Proofs of Lemmas

PROOF OF LEMMA 1: A strong mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) gives $E \|\sum_{i=1}^N h_{N,i}(\tilde{X}_i)\|^p < CN^{p/2}$ provided $p \geq 2$, where C does not depend on N . 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

$$(A.1) \quad \lim_{N \rightarrow \infty} \varepsilon^{-p} N^{a-p+pc} E \left\| \sum_{i=1}^N h_{N,i}(\tilde{X}_i) \right\|^p \leq \lim_{N \rightarrow \infty} \varepsilon^{-p} CN^{a-p+pc+p/2} = 0.$$

Part (b) follows from part (a) applied to $h_{N,i}(\tilde{X}_i) - Eh_{N,i}(\tilde{X}_i)$ with $c = 0$ and the triangle inequality. Q.E.D.

PROOF OF LEMMA 2: The proof of Lemma 14 of A2002 goes through with $g(X_i, \theta_0)$ replaced by $g_\pi(X_i, \theta_0)$ except for the proof that $B_2 = 0$.

More specifically, as in A2002, the least favorable value of m for the bootstrap moment $(N\tau)^{\alpha(m)} E^* \prod_{\mu=1}^m \Psi_{N,\pi,j_\mu}^*$ (in terms of its distance from the corresponding population moment) is three. Hence, we just consider this case. For notational simplicity, suppose $j_\mu = 1$ for $\mu = 1, 2, 3$. Thus, we need to show that

$$(A.2) \quad \lim_{N \rightarrow \infty} N^a P(|((N\tau)^{1/2} E^*(\Psi_{N,\pi,1}^*)^3 - (N\tau)^{1/2} E\Psi_{N,\pi,1}^3)| > (N\tau)^{-\xi}) = 0.$$

Let $f_{\pi,i} = f_{\pi,1}(\tilde{X}_i, \theta_0) - Ef_{\pi,1}(\tilde{X}_i, \theta_0)$, where $f_{\pi,1}(\tilde{X}_i, \theta_0)$ denotes the first element of $f_\pi(\tilde{X}_i, \theta_0)$. Let $b_1 = \{1, \dots, \ell\}$, $b_2 = \{\ell + 1, \dots, 2\ell\}, \dots, b_b = \{(b - 1)\ell + 1, \dots, b\ell\}$, where $N = b\ell$. Let $Y_{\pi,j} = \sum_{i \in b_j} f_{\pi,i}$. Then,

$$(A.3) \quad \Psi_{N,\pi,1} = (N\tau)^{-1/2} \sum_{i=1}^N f_{\pi,i} = (N\tau)^{-1/2} \sum_{j=1}^b Y_{\pi,j}.$$

By the arguments in the proof of Lemma 14 of A2002, provided $\xi < 1/2 - \gamma$,

$$(A.4) \quad \lim_{N \rightarrow \infty} N^a P(|((N\tau)^{1/2} E^*(\Psi_{N,\pi,1}^*)^3 - (N\tau)^{-1} bEY_{\pi,1}^3)| > (N\tau)^{-\xi}) = 0.$$

Hence, it suffices to show that

$$(A.5) \quad \limsup_{N \rightarrow \infty} (N\tau)^\xi |(N\tau)^{1/2} E\Psi_{N,\pi,1}^3 - (N\tau)^{-1} bEY_{\pi,1}^3| = 0.$$

(Equation (A.5) shows that B_2 of A2002 equals zero.)

Using (A.3), we have

$$(A.6) \quad (N\tau)^{1/2} E\Psi_{N,\pi,1}^3 = (N\tau)^{\xi-1} \sum_{j_1=1}^b \sum_{j_2=1}^b \sum_{j_3=1}^b EY_{j_1} Y_{j_2} Y_{j_3}.$$

Hence,

$$(A.7) \quad (N\tau)^\xi |(N\tau)^{1/2} E\Psi_{N,\pi,1}^3 - (N\tau)^{-1} b EY_{\pi,1}^3| = (N\tau)^{\xi-1} \sum_{j_1=1}^b \sum_{j_2=1}^b \sum_{\substack{j_3=1 \\ j_1 \neq j_2 \neq j_3}}^b EY_{j_1} Y_{j_2} Y_{j_3}.$$

If the observations $\{X_i : i \geq 1\}$ are m -dependent, then the observations $\{\tilde{X}_i : i \geq 1\}$ are $(m + \kappa)$ -dependent and Y_{j_1} and Y_{j_2} are independent for all $j_1 \neq j_2$ for N large because the number of deleted observations at the end of each block satisfies $\lceil \pi \ell \rceil \geq m + \kappa$ for N large. Since $EY_{j_1} = 0$ for all j_1 , the right-hand side of (A.7) equals zero in the m -dependent case.

Next, we consider the case where the observations are not necessarily m -dependent, but $0 < \gamma < 1/2$. By a strong mixing covariance inequality of Davydov (e.g., see Doukhan (1995, Theorem 3(1), p. 9)),

$$(A.8) \quad |EY_{j_1} Y_{j_2} Y_{j_3}| \leq 8 \|Y_{j_1}\|_p \cdot \|Y_{j_2} Y_{j_3}\|_q \alpha^r(\lceil \pi \ell \rceil - \kappa),$$

where $p, q, r \geq 1, 1/p + 1/q + 1/r = 1, \|\cdot\|_p$ denotes the L^p norm, and $\{\alpha(s) : s \geq 1\}$ are the strong mixing numbers of $\{X_i : i \geq 1\}$, which decline to zero exponentially fast by Assumption 1. This inequality holds because the summands in Y_{j_1} are separated from those in any blocks Y_{j_2} and Y_{j_3} by at least $\lceil \pi \ell \rceil - \kappa$ by the block feature of $f_{\pi,i}$. This is the key part of the proof.

Next, by Minkowski's inequality, $\|Y_{j_1}\|_p = \|Y_1\|_p \leq \ell \|f_{\pi,1}\|_p \leq \ell \tau C_1$ for some constant $C_1 < \infty$. By an application of the Cauchy-Schwarz inequality and the fact that $\|Y_2\|_{2q} = \|Y_3\|_{2q}$, we have $\|Y_{j_2} Y_{j_3}\|_q = \|Y_2 Y_3\|_q \leq \|Y_2\|_{2q}^2 \leq (\ell \tau C_1)^2$. Hence, (A.5) holds by (A.7) and (A.8) provided

$$(A.9) \quad (N\tau)^{\xi-1} b^3 (\ell \tau)^3 \alpha^r(\lceil \pi \ell \rceil - \kappa) = (N\tau)^{2+\xi} \alpha^r(\lceil \pi \ell \rceil - \kappa) \rightarrow 0 \quad \text{as } N \rightarrow \infty.$$

Since the α -mixing numbers, $\{\alpha(s) : s \geq 1\}$, decline to zero exponentially fast in s , (A.9) holds provided $\pi \ell - C \log(N) \rightarrow \infty$ for all $C < \infty$, as is assumed. Q.E.D.

REFERENCES

ANDREWS, D. W. K. (2001): "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>.
 ——— (2002a): "Higher-Order Improvements of a Computationally Attractive k -Step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119–162.
 ——— (2002b): "The Block-Block Bootstrap: Improved Asymptotic Refinements," Cowles Foundation Discussion Paper No. 1370, Yale University. Available at <http://cowles.econ.yale.edu>.
 ——— (2004): "Higher-Order Improvements of the Parametric Bootstrap for Markov Processes," in *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock. Cambridge, U.K.: Cambridge University Press, forthcoming.
 ANDREWS, D. W. K., AND O. LIEBERMAN (2004): "Higher-Order Improvements of the Parametric Bootstrap for Long-Memory Gaussian Processes," *Journal of Econometrics*, forthcoming. Available at <http://cowles.econ.yale.edu>.

- 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: John Wiley & Sons, 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," *Sankhyā*, Series A, 41, 22–47.
- DOUKHAN, P. (1995): *Mixing: Properties and Examples*. New York: Springer-Verlag.
- GÖTZE, F., AND C. HIPF (1983): "Asymptotic Expansions for Sums of Weakly Dependent Random Vectors," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64, 211–239.
- (1994): "Asymptotic Distribution of Statistics in Time Series," *Annals of Statistics*, 22, 2062–2088.
- GÖTZE, F., AND H. R. KÜNSCH (1996): "Second-Order Correctness of the Blockwise Bootstrap for Stationary Observations," *Annals of Statistics*, 24, 1914–1933.
- HALL, P. (1988): "On Symmetric Bootstrap Confidence Intervals," *Journal of the Royal Statistical Society, Series B*, 50, 35–45.
- (1992): *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.
- HALL, P., AND J. L. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moment Estimators," *Econometrica*, 64, 891–916.
- 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.
- HOROWITZ, J. L. (2003): "Bootstrap Methods for Markov Processes," *Econometrica*, 71, 1049–1082.
- INOUE, A., AND M. SHINTANI (2000): "Bootstrapping GMM Estimators for Time Series," unpublished working paper, Department of Economics, Vanderbilt University.
- KÜNSCH, H. R. (1989): "The Jackknife and the Bootstrap for General Stationary Observations," *Annals of Statistics*, 17, 1217–1241.
- LAHIRI, S. N. (1996): "On Edgeworth Expansion and Moving Block Bootstrap for Studentized M-estimators in Multiple Linear Regression Models," *Journal of Multivariate Analysis*, 56, 42–59.
- 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.
- PAPARODITIS, E., AND D. N. POLITIS (2001): "Tapered Block Bootstrap," *Biometrika*, 88, 1105–1119.
- (2002): "The Tapered Block Bootstrap for General Statistics from Stationary Sequences," *Econometrics Journal*, 5, 131–148.
- 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," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52, 45–57.
- ZVINGELIS, Y. (2003): "On Bootstrap Coverage Probability with Dependent Data," in *Computer-Aided Econometrics*, ed. by D. Giles. New York: Marcel Dekker.