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

By

Donald W. K. Andrews

January 2001

COWLES FOUNDATION DISCUSSION PAPER NO. 1230R

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

YALE UNIVERSITY

Box 208281

New Haven, Connecticut 06520-8281

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

Higher-order Improvements of a Computationally Attractive k-step Bootstrap for Extremum Estimators

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

March 1998
Revised: January 2001

Abstract

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

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

Keywords: Asymptotics, block bootstrap, Edgeworth expansion, extremum estimator, Gauss-Newton, generalized method of moments estimator, k-step bootstrap, maximum likelihood estimator, Newton-Raphson, parametric bootstrap, t statistic, test of over-identifying restrictions.

JEL Classification Numbers: C12, C13, C15.

1 Introduction

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

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

The results cover the nonparametric iid bootstrap, overlapping and non-overlapping block bootstraps for time series, and the parametric bootstrap for iid and time series data. The results apply to extremum estimators, such as GMM and ML estimators. 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 also employed in Hall and Horowitz (1996) (denoted HH hereafter).

This paper also provides a number of new results concerning the higher-order properties of different types of standard (i.e., non-k-step) bootstraps for extremum estimators. These results extend results of HH.

The results of the paper are as follows. First, the results of the paper establish that the standard bootstrap, as well as the k-step bootstrap, reduces the error in test rejection probability and confidence interval coverage probability (relative to standard first-order asymptotics) by at least $N^{-\xi}$ for some $\xi>0$. (In contrast, the results of HH show that the error in test rejection probability is reduced from $O(N^{-1})$ to $o(N^{-1})$ for symmetric two-sided t tests and J tests.) When the data are dependent and the block bootstrap is employed, the value of ξ depends on the block length parameter γ, where the block length ℓ satisfies $\ell \propto N^{\gamma}$. For the block length $\ell \propto N^{1/4}$ that maximizes the upper bound on ξ, ξ is bounded above by 1/4. For equal-tailed confidence intervals and one-sided confidence intervals and tests, this bound on ξ is sharp. Improvements of magnitude $\xi \geq 1/4$ are not possible. The source of this bound is the combination of (i) the bias of the block bootstrap in estimating population moments that appear in Edgeworth coefficients and the behavior of the correction factor that is needed to adjust the block bootstrap, both of which require $\xi < \gamma$, and (ii) the variability of the block bootstrap Edgeworth coefficients, which approximate the corresponding non-bootstrap Edgeworth coefficients sufficiently well only if $\xi + \gamma < 1/2$.

When the data are iid and a nonparametric or parametric bootstrap is employed or when the data are dependent and a parametric bootstrap is employed, our results yield improvements of magnitude $N^{-\xi}$ for $\xi < 1/2$. For two-sided t tests and symmetric percentile t confidence intervals and iid data, these results can be improved to $N^{-\xi}$ for $\xi < 1$ via the argument of Hall (1988). Also, it may be possible to improve the results for restricted parametric bootstrap tests, as suggested by Davidson and MacKinnon (1999b).

Note the difference between the results for the improvements due to the block bootstrap for time series, viz. $N^{-\xi}$ for all $\xi < 1/4$, and the results for the other bootstraps, viz., $N^{-\xi}$ for $\xi < 1/2$. The block bootstrap is only half as effective as the other bootstraps.

Second, the results given here allow for a great deal of flexibility in the choice of the block length parameter in the case of time series data. Specifically, if the block length is $\ell \propto N^{\gamma}$, where N is the sample size, then we just require $\gamma < 1/2$. Thus, the length $\ell \propto N^{1/4}$, which maximizes the upper bound on ξ, is covered. (The results of HH are restrictive in this dimension and do not cover this block length.)

Third, the results given here apply to both the overlapping and non-overlapping block bootstraps. HH's results apply to the non-overlapping block bootstrap, whereas much of the literature on the block bootstrap focuses on overlapping blocks, e.g., see Künsch (1989), Lahiri (1992, 1996), and Götze and Künsch (1996). The overlapping block bootstrap is slightly more efficient asymptotically for estimating the distribution function of a t statistic than the non-overlapping block bootstrap, see Hall, Horowitz, and Jing (1995). The overlapping block bootstrap also is more efficient asymptotically for estimating the variance of an estimator than the non-overlapping block bootstrap, see Lahiri (1999). Given these superior asymptotic properties of the overlapping block bootstrap, it is desirable to have results for extremum estimators that cover it.

Fourth, the results of the paper apply to restricted and unrestricted parametric bootstraps for likelihood-based statistics in κ-th order Markov processes. The restricted parametric bootstrap is applicable to hypothesis tests. It utilizes the restricted ML estimator, which satisfies the null hypothesis, to generate the bootstrap sample. The unrestricted parametric bootstrap is applicable to tests and confidence intervals. It utilizes the unrestricted ML estimator to generate the bootstrap sample. Surprisingly, there do not appear to be results in the literature that cover parametric bootstraps in this general nonlinear estimation case (although Davidson and MacKinnon (1999b) provide a general discussion of the higher-order improvements of the parametric bootstrap).

Fifth, the results apply to Wald tests of nonlinear restrictions and corresponding

Wald-based confidence regions, as well as to the t and J statistics that are considered in HH.

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

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

Two papers in the literature concerning the overlapping block bootstrap are Götze and Künsch (1996) and Lahiri (1996). They consider statistics that are smooth functions of sample averages and regression parameter t statistics respectively. They allow the asymptotic variances of the statistics of interest to depend on an infinite number of correlations, which is less restrictive than the assumption employed here. On the other hand, they obtain accuracy of confidence interval coverage probabilities only up to $o(N^{-1/2})$, whereas we obtain higher-order accuracy, as outlined above. Two recent papers that consider the block bootstrap for econometric models are Zvingelis (2001) and Inoue and Shintani (2000).

Andrews (2001a) provides results on the higher-order equivalence of non-bootstrap k-step estimators and corresponding extremum estimators.

The remainder of the paper is organized as follows: Section 2 defines the extremum estimators, the overlapping and non-overlapping block bootstraps, and the k-step block bootstrap. It also states the assumptions. Section 3 provides the results for the higher-order asymptotic equivalence and the higher-order improvements of the k-step and standard block bootstraps. Section 4 deals with the unrestricted and restricted parametric bootstraps. An Appendix contains proofs of the results.

2 Extremum Statistics and the k-step Bootstrap

2.1 Definition of Extremum Statistics

As much as possible, we use the same notation as HH. We consider extremum estimators that are either GMM estimators or estimators that minimize a sample average. We call the latter "minimum ρ estimators," because the sample average is taken to be $N^{-1}\sum_{i=1}^{N}\rho(X_i,\theta)$, where $X_i \in R^{L_x}$ is a random vector, $\theta \in \Theta \subset R^{L_\theta}$ is an unknown parameter, and $\rho(\cdot,\cdot)$ is a known real function. Maximum likelihood (ML) estimators, least squares (LS), and regression M estimators are examples of minimum

ρ estimators. GMM estimators are based on the moment conditions $Eg(X_i, \theta_0) = 0$, where $g(\cdot, \cdot)$ is a known L_g-valued function, X_i is as above, $\theta_0 \in \Theta \subset R^{L_\theta}$ is the true unknown parameter, and $L_g \geq L_\theta$.

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

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

$$\widetilde{X}_i = (X_i', X_{i+1}', ..., X_{i+\kappa}')' \text{ for } i = 1, ..., n - \kappa. \tag{2.1}$$

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

$$\chi_N = \{\widetilde{X}_i : i = 1, ..., N\}, \tag{2.2}$$

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

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

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

$$\min_{\theta \in \Theta} J_N(\theta) = \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta) \right)' \Omega \left(N^{-1} \sum_{i=1}^{N} g(X_i, \theta) \right). \tag{2.3}$$

The two-step GMM estimator which, for economy of notation, we also denote by

$\widehat{\theta}_N$, solves

$$\begin{aligned}
\min_{\theta\in\Theta} J_N(\theta,\widetilde{\theta}_N) &= \left(N^{-1}\sum_{i=1}^N g(X_i,\theta)\right)' \Omega_N(\widetilde{\theta}_N)\left(N^{-1}\sum_{i=1}^N g(X_i,\theta)\right), \text{ where}\\
\Omega_N(\theta) &= \overline{W}_N^{-1}(\theta),\\
\overline{W}_N(\theta) &= N^{-1}\sum_{i=1}^N\left(g(X_i,\theta)g(X_i,\theta)' + \sum_{j=1}^{\kappa} H(X_i,X_{i+j},\theta)\right),\\
H(X_i,X_{i+j},\theta) &= g(X_i,\theta)g(X_{i+j},\theta)' + g(X_{i+j},\theta)g(X_i,\theta)', \qquad (2.4)
\end{aligned}$$

and $\widetilde{\theta}_N$ solves (2.3).

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

$$\min_{\theta\in\Theta} N^{-1}\sum_{i=1}^N \rho(X_i,\theta). \qquad (2.5)$$

For this estimator, we let $g(X_i,\theta)$ denote $(\partial/\partial\theta)\rho(X_i,\theta)$. Except for consistency properties, the minimum ρ estimator can be analyzed simultaneously with the GMM estimators. The reason is that with probability that goes to one (at an appropriate rate) the solution $\widehat{\theta}_N$ to the minimization problem (2.5) is an interior solution and, hence, is also a solution to the problem of minimizing a quadratic form in the first-order conditions from this problem with weight matrix given by the identity matrix, which is just the one-step GMM criterion function.

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

$$\begin{aligned}
\sigma &= \begin{cases} (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.3)}\\ (D'\Omega_0 D)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4)}\\ D^{-1}\Omega_0^{-1}D^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5), where}\end{cases}\\
\Omega_0 &= (E\overline{W}_N(\theta_0))^{-1} \text{ and } D = E\frac{\partial}{\partial\theta'}g(X_i,\theta_0). \qquad (2.6)
\end{aligned}$$

A consistent estimator of σ is

$$\begin{aligned}
\sigma_N &= \begin{cases} (D_N'\Omega D_N)^{-1}D_N'\Omega\Omega_N^{-1}(\widehat{\theta}_N)\Omega D_N(D_N'\Omega D_N)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.3)}\\ (D_N'\Omega_N(\widehat{\theta}_N)D_N)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4)}\\ D_N^{-1}\Omega_N^{-1}(\widehat{\theta}_N)D_N^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5), where}\end{cases}\\
D_N &= N^{-1}\sum_{i=1}^N\frac{\partial}{\partial\theta'}g(X_i,\widehat{\theta}_N). \qquad (2.7)
\end{aligned}$$

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

$$T_N = N^{1/2}(\widehat{\theta}_{N,r}-\theta_{0,r})/(\sigma_N)_{rr}^{1/2}. \qquad (2.8)$$

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

$$\mathcal{W}_N = N\eta(\widehat{\theta}_N)' \left(\frac{\partial}{\partial \theta'}\eta(\widehat{\theta}_N)\sigma_N(\frac{\partial}{\partial \theta'}\eta(\widehat{\theta}_N))' \right)^{-1} \eta(\widehat{\theta}_N). \tag{2.9}$$

The J statistic for testing over-identifying restrictions is

$$\begin{aligned} J_N &= K_N(\widehat{\theta}_N)'K_N(\widehat{\theta}_N), \text{ where} \\ K_N(\theta) &= \Omega_N^{1/2}(\theta)N^{-1/2}\sum_{i=1}^{N} g(X_i,\theta) \end{aligned} \tag{2.10}$$

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

2.2 The Nonparametric Block Bootstrap

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

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

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

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

$$\begin{aligned} \min_{\theta\in\Theta} J_N^*(\theta) &= \left(N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)\right)' \Omega \left(N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)\right), \text{ where} \\ g^*(X_i^*,\theta) &= g(X_i^*,\theta) - E^*g(X_i^*,\widehat{\theta}_N), \end{aligned} \tag{2.11}$$

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

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N} E^* g(X_i^*,\theta) &= N^{-1}\sum_{i=1}^{N} g(X_i,\theta) \text{ and} \\
N^{-1}\sum_{i=1}^{N} E^* g(X_i^*,\theta) &= (N-\ell+1)^{-1}\sum_{i=1}^{N} w(i,\ell,N) g(X_i,\theta), \text{ where} \\
w(i,\ell,N) &= \begin{cases} i/\ell & \text{if } i\in[1,\ell-1] \\ 1 & \text{if } i\in[\ell,N-\ell+1] \\ (N-i+1)/\ell & \text{if } i\in[N-\ell+2,N] \end{cases} . \qquad (2.12)
\end{aligned}$$

The sample moment conditions $N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)$ in (2.11) are recentered (by subtracting off $E^*g(X_i^*,\widehat{\theta}_N)$) to ensure that the bootstrap moments $E^*N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)$ equal zero when $\theta=\widehat{\theta}_N$, which mimics the population moments $Eg(X_i,\theta)$, which equal zero when $\theta=\theta_0$.

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

$$\begin{aligned}
\min_{\theta\in\Theta} J_N^*(\theta,\widetilde{\theta}_N^*) &= \left(N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)\right)' \Omega_N^*(\widetilde{\theta}_N^*)\left(N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta)\right), \text{ where} \\
\Omega_N^*(\theta) &= \overline{W}_N^*(\theta)^{-1}, \\
\overline{W}_N^*(\theta) &= N^{-1}\sum_{i=1}^{N}\left(g^*(X_i^*,\theta)g^*(X_i^*,\theta)' + \sum_{j=1}^{\kappa} H^*(X_i^*,X_{i+j}^*,\theta)\right), \\
H^*(X_i^*,X_{i+j}^*,\theta) &= g^*(X_i^*,\theta)g^*(X_{i+j}^*,\theta)' + g^*(X_{i+j}^*,\theta)g^*(X_i^*,\theta)', \qquad (2.13)
\end{aligned}$$

$\widetilde{\theta}_N^*$ denotes the one-step bootstrap estimator that solves (2.11), and, with some abuse of notation, X_{i+j}^* denotes the $(j+1)$-st element of $\widetilde{X}_i^*$. (This abuse of notation can be avoided by writing X_i^* as $X_{i,0}^*$ and X_{i+j}^* as $X_{i,j}^*$, where $\widetilde{X}_i^*=(X_{i,0}^*,...,X_{i,\kappa}^*)'$ wherever they appear. To maintain consistency of notation with HH, we do not do so.)

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

$$\min_{\theta\in\Theta} N^{-1}\sum_{i=1}^{N}(\rho(X_i^*,\theta) - E^*g(X_i^*,\widehat{\theta}_N)'\theta), \qquad (2.14)$$

where $g(\cdot,\theta)=(\partial/\partial\theta)\rho(\cdot,\theta)$. The term $E^*g(X_i^*,\widehat{\theta}_N)'\theta$ properly recenters the minimum ρ bootstrap criterion function. It yields bootstrap population first-order conditions that equal zero at $\widehat{\theta}_N$, as desired. That is, $E^*(\partial/\partial\theta)(N^{-1}\sum_{i=1}^{N}(\rho(X_i^*,\theta) - E^*g(X_i^*,\widehat{\theta}_N)'\theta)) = E^*N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta) = 0$ when $\theta=\widehat{\theta}_N$. With this recentering, the first-order conditions for θ_N^* are $N^{-1}\sum_{i=1}^{N} g^*(X_i^*,\theta_N^*)=0$, rather than

$N^{-1}\sum_{i=1}^{N} g(X_i^*, \theta_N^*) = 0$, which means that θ_N^* minimizes the one-step GMM bootstrap criterion function $J_N^*(\theta)$ with $g(\cdot,\theta) = (\partial/\partial\theta)\rho(\cdot,\theta)$ and arbitrary positive definite weight matrix Ω. For the non-overlapping block bootstrap, $E^* g(X_i^*, \widehat{\theta}_N)$ $= N^{-1}\sum_{i=1}^{N} g(X_i, \widehat{\theta}_N) = 0$, where the second equality holds by the first-order conditions for $\widehat{\theta}_N$ because the dimensions of $g(\cdot,\cdot)$ and θ are equal, and the extra term in (2.14) disappears. (This is also true for the parametric bootstrap considered below.) For the overlapping block bootstrap, however, the first equality in the latter equation does not hold and the extra term is non-zero.

The bootstrap covariance matrix estimator is

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

$$\sigma_N^*(\theta) = \begin{cases} (D_N^*(\theta)'\Omega D_N^*(\theta))^{-1} D_N^*(\theta)\Omega\Omega_N^*(\theta)^{-1}\Omega D_N^*(\theta) \\ \quad \times(D_N^*(\theta)'\Omega D_N^*(\theta))^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.3)} \\ (D_N^*(\theta)'\Omega_N^*(\theta) D_N^*(\theta))^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4)} \\ D_N^*(\theta)^{-1}\Omega_N^*(\theta)^{-1} D_N^*(\theta)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5) and} \end{cases}$$

$$D_N^*(\theta) = N^{-1}\sum_{i=1}^{N} \frac{\partial}{\partial\theta'} g(X_i^*, \theta). \tag{2.15}$$

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

$$\begin{aligned} T_N^* &= \tau_{N,r} N^{1/2}((\theta_N^*)_r - \widehat{\theta}_{N,r})/\sigma_N^*(\theta_N^*)_{rr}^{1/2}, \\ \mathcal{W}_N^* &= H_N^*(\theta_N^*)' H_N^*(\theta_N^*), \text{ and} \\ J_N^* &= K_N^*(\theta_N^*)' K_N^*(\theta_N^*), \text{ where} \\ H_N^*(\theta) &= \Xi_N \left((\frac{\partial}{\partial\theta'}\eta(\theta))\sigma_N^*(\theta)(\frac{\partial}{\partial\theta'}\eta(\theta))' \right)^{-1/2} N^{1/2}(\eta(\theta) - \eta(\widehat{\theta}_N)), \\ K_N^*(\theta) &= (V_N^+)^{1/2}\Omega_N^*(\theta)^{1/2} N^{-1/2}\sum_{i=1}^{N} g^*(X_i^*, \theta), \end{aligned} \tag{2.16}$$

$(\theta_N^*)_r$ denotes the r-th element of θ_N^*,[3] $\sigma_N^*(\theta_N^*)_{rr}$ denotes the (r,r)–th element of $\sigma_N^*(\theta_N^*)$, and V_N^+ denotes the Moore-Penrose inverse of V_N. The correction factor $\tau_{N,r}$ is defined as follows:

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

$$\widetilde{\sigma}_N = \begin{cases} (D_N'\Omega D_N)^{-1} D_N'\Omega\widetilde{W}_N\Omega D_N(D_N'\Omega D_N)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.3)} \\ (D_N'\Omega_N(\widehat{\theta}_N)D_N)^{-1} D_N'\Omega_N(\widehat{\theta}_N)\widetilde{W}_N\Omega_N(\widehat{\theta}_N)D_N \\ \quad \times(D_N'\Omega_N(\widehat{\theta}_N)D_N)^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.4)} \\ D_N^{-1}\widetilde{W}_N D_N{}^{-1} & \text{if } \widehat{\theta}_N \text{ solves (2.5), and} \end{cases}$$

$$
\begin{aligned}
\widetilde{W}_N &= E^* N^{-1} \sum_{i=1}^{N} \sum_{j=1}^{N} g^*(X_i^*, \widehat{\theta}_N) g^*(X_j^*, \widehat{\theta}_N)' \\
&= \begin{cases} N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} g^*(X_{i\ell+j}, \widehat{\theta}_N) g^*(X_{i\ell+m}, \widehat{\theta}_N)' \\ \qquad\qquad \text{for non-overlapping blocks} \\ bN^{-1}(N-\ell+1)^{-1} \sum_{i=0}^{N-\ell} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} g^*(X_{i+j}, \widehat{\theta}_N) g^*(X_{i+m}, \widehat{\theta}_N)' \\ \qquad\qquad \text{for overlapping blocks.} \end{cases}
\end{aligned} \tag{2.17}
$$

The correction factor Ξ_N is defined to be

$$
\Xi_N = \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N) \widetilde{\sigma}_N \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N)\right)' \right)^{-1/2} \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N) \sigma_N \left(\frac{\partial}{\partial \theta'} \eta(\widehat{\theta}_N)\right)' \right)^{1/2}. \tag{2.18}
$$

The correction factor V_N is defined to be

$$
\begin{aligned}
V_N &= M_N \overline{W}_N^{-1/2}(\widehat{\theta}_N) \widetilde{W}_N \overline{W}_N^{-1/2}(\widehat{\theta}_N) M_N, \text{ where} \\
M_N &= I_{L_g} - \overline{W}_N^{-1/2}(\widehat{\theta}_N) D_N (D_N' \overline{W}_N^{-1}(\widehat{\theta}_N) D_N)^{-1} D_N' \overline{W}_N^{-1/2}(\widehat{\theta}_N).
\end{aligned} \tag{2.19}
$$

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 $z^*_{|T|,\alpha} = \inf\{z \in R : P^*(|T_N^*| \leq z) \geq 1-\alpha\}$ etc.)

The symmetric two-sided bootstrap t test of $H_0 : \theta_r = \theta_{0,r}$ versus $H_1 : \theta_r \neq \theta_{0,r}$ of significance level α rejects H_0 if

$$
|T_N| > z^*_{|T|,\alpha}. \tag{2.20}
$$

The equal-tailed two-sided bootstrap t test of significance level α for the same hypotheses rejects H_0 if

$$
T_N < z^*_{T,1-\alpha/2} \text{ or } T_N > z^*_{T,\alpha/2}. \tag{2.21}
$$

The one-sided bootstrap t test of $H_0 : \theta_r \leq \theta_{0,r}$ versus $H_1 : \theta_r > \theta_{0,r}$ of significance level α rejects H_0 if

$$
T_N > z^*_{T,\alpha}. \tag{2.22}
$$

The bootstrap Wald test of $H_0 : \eta(\theta_0) = 0$ versus $H_1 : \eta(\theta_0) \neq 0$ rejects the null hypothesis if

$$
\mathcal{W}_N > z^*_{\mathcal{W},\alpha}. \tag{2.23}
$$

The bootstrap J test of over-identifying restrictions of significance level α rejects the null if

$$
J_N > z^*_{J,\alpha}. \tag{2.24}
$$

Correspondingly, the symmetric two-sided bootstrap confidence interval for $\theta_{0,r}$ of confidence level $100(1-\alpha)\%$ is

$$
[\widehat{\theta}_{N,r} - z^*_{|T|,\alpha} (\sigma_N)^{1/2}_{rr} / N^{1/2},\ \widehat{\theta}_{N,r} + z^*_{|T|,\alpha} (\sigma_N)^{1/2}_{rr} / N^{1/2}]. \tag{2.25}
$$

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

$$[\widehat{\theta}_{N,r} - z^*_{T,\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2},\ \widehat{\theta}_{N,r} + z^*_{T,1-\alpha/2}(\sigma_N)^{1/2}_{rr}/N^{1/2}]. \tag{2.26}$$

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

$$[\widehat{\theta}_{N,r} - z^*_{T,\alpha}(\sigma_N)^{1/2}_{rr}/N^{1/2},\ \infty). \tag{2.27}$$

The Wald-based bootstrap confidence region for $\eta(\theta_0)$ of confidence level $100(1-\alpha)\%$ is

$$\{\eta \in R^{L_\eta} : N(\eta(\widehat{\theta}_N) - \eta)' \left(\frac{\partial}{\partial\theta'}\eta(\widehat{\theta}_N)\sigma_N(\frac{\partial}{\partial\theta'}\eta(\widehat{\theta}_N))' \right)^{-1} (\eta(\widehat{\theta}_N) - \eta) \leq z^*_{\mathcal{W},\alpha}\}. \tag{2.28}$$

For example, by taking $\eta(\theta)$ to equal a subvector, say θ_1, of θ, this yields a confidence region for θ_1.

2.3 The k-step Bootstrap

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

$$\theta^*_{N,j} = \theta^*_{N,j-1} - (Q^*_{N,j-1})^{-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j-1}) \text{ for } 1 \leq j \leq k \tag{2.29}$$

and $\theta^*_{N,0} = \widehat{\theta}_N$, where $\widehat{\theta}_N$ denotes the one-step GMM estimator. For two-step GMM and minimum ρ estimators, $\theta^*_{N,k}$ is defined in the same way with $(\partial/\partial\theta)J^*_N(\theta^*_{N,j-1})$ replaced by $(\partial/\partial\theta)J^*_N(\theta^*_{N,j-1}, \widetilde{\theta}^*_{N,k_1})$ and $N^{-1}\sum_{i=1}^N g^*(X^*_i, \theta^*_{N,j-1})$, respectively, and with $\widehat{\theta}_N$ denoting the two-step GMM estimator and the minimmum ρ estimator, respectively, where the derivative is taken with respect to the first argument of $J^*_N(\cdot,\cdot)$ and $\widetilde{\theta}^*_{N,k_1}$ denotes the k_1-step bootstrap one-step GMM estimator defined in (2.29). We assume that $k_1 \geq k$.

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

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

$$D^*_N(\theta^*_{N,j-1}) = N^{-1}\sum_{i=1}^N \frac{\partial}{\partial\theta'}g(X^*_i, \theta^*_{N,j-1}). \tag{2.30}$$

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

The *default* NR choice of $Q^*_{N,j-1}$, denoted $Q^{*,D}_{N,j-1}$, equals $Q^{*,NR}_{N,j-1}$ if $Q^{*,NR}_{N,j-1}$ leads to an estimator $\theta^*_{N,j}$ via (2.29) for which $J^*_N(\theta^*_{N,j}) \leq J^*_N(\theta^*_{N,j-1})$ for the one-step GMM estimator, but equals some other matrix otherwise. In practice, one wants this other matrix to be such that $J^*_N(\theta^*_{N,j}) < J^*_N(\theta^*_{N,j-1})$ (but the theoretical results do not require this). For example, one might use the matrix $(1/\varepsilon)I_{L_\theta}$ for some small $\varepsilon > 0$. (See Ortega and Rheinboldt (1970, Theorem 8.2.1) for a result that indicates that such a choice will decrease the criterion function.) For the two-step GMM and minimum ρ estimators, $J^*_N(\cdot)$ above is replaced by $J^*_N(\cdot, \widetilde{\theta}^*_{N,k_1})$ and $\rho^*_N(\cdot)$ respectively.

The *line-search* NR choice of $Q^*_{N,j-1}$, denoted $Q^{*,LS}_{N,j-1}$, uses a scaled version of the NR matrix $Q^{*,NR}_{N,j-1}$ that optimizes the step length. Specifically, let A be a finite subset of $(0,1]$ of step lengths that includes 1. One computes $\theta^*_{N,j}$ via (2.29) for $Q^*_{N,j-1} = (1/\alpha)Q^{*,NR}_{N,j-1}$ for each $\alpha \in A$. One takes $Q^{*,LS}_{N,j-1}$ to be the matrix $(1/\alpha)Q^{*,NR}_{N,j-1}$ that minimizes $J^*_N(\theta^*_{N,j})$ over all $\alpha \in A$ for the one-step GMM estimator. (If the minimizing of value of α is not unique, one takes the largest minimizing value of α in A.) For the two-step GMM and the minimum ρ estimators, one replaces $J^*_N(\theta^*_{N,j})$ by $J^*_N(\theta^*_{N,j}, \widetilde{\theta}^*_{N,k_1})$ and $\rho^*_N(\theta^*_{N,j})$ respectively.

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

$$Q^{*,GN}_{N,j-1} = \begin{cases} 2D^{*\prime}_{N,j-1}\Omega D^*_{N,j-1} & \text{for the one-step GMM estimator} \\ 2D^{*\prime}_{N,j-1}\Omega^*_N(\widetilde{\theta}^*_{N,k_1})D^*_{N,j-1} & \text{for the two-step GMM estimator} \\ D^*_{N,j-1} & \text{for the minimum } \rho \text{ estimator,} \end{cases} \quad (2.31)$$

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

$$\begin{aligned} D^*_{N,j-1} &= N^{-1}\sum_{i=1}^N \Delta(\widetilde{X}^*_i, \theta^*_{N,j-1}) \in R^{L_g \times L_\theta} \text{ and} \\ E\Delta(\widetilde{X}_i, \theta_0) &= E\frac{\partial}{\partial\theta'}g(X_i,\theta_0). \end{aligned} \quad (2.32)$$

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

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

$$g(X_i,\theta) = U(X_i,\theta)L(Z_i,\theta) \text{ and } E(U(X_i,\theta_0)|Z_i) = 0 \text{ a.s.,} \quad (2.33)$$

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

$$\frac{\partial}{\partial \theta'} g(X_i, \theta) = L(Z_i, \theta) \frac{\partial}{\partial \theta'} U(X_i, \theta) + U(X_i, \theta) \frac{\partial}{\partial \theta'} L(Z_i, \theta). \tag{2.34}$$

The GN choice of $Q^*_{N,j-1}$ omits the second summand of the bootstrap version of $(\partial/\partial\theta')g(X_i, \theta)$ in $D^*_{N,j-1}$ because $EU(X_i, \theta_0)(\partial/\partial\theta')L(Z_i, \theta_0) = 0$. That is, $Q^{*,GN}_{N,j-1}$ is as in (2.31) and (2.32) with

$$\Delta(\widetilde{X}^*_i, \theta) = L(Z^*_i, \theta) \frac{\partial}{\partial \theta'} U(X^*_i, \theta). \tag{2.35}$$

For an example of a GN matrix for a minimum ρ estimator, consider the least squares (LS) estimator of a nonlinear regression model:

$$\begin{aligned}
Y_i &= q(Z_i, \theta_0) + U_i \text{ for } i = 1, ..., n, \\
\rho(X_i, \theta) &= (Y_i - q(Z_i, \theta))^2/2, \\
g(X_i, \theta) &= -(Y_i - q(Z_i, \theta)) \frac{\partial}{\partial \theta} q(Z_i, \theta), \text{ and} \\
\frac{\partial}{\partial \theta'} g(X_i, \theta) &= \frac{\partial}{\partial \theta} q(Z_i, \theta) \frac{\partial}{\partial \theta'} q(Z_i, \theta) + (Y_i - q(Z_i, \theta)) \frac{\partial^2}{\partial \theta \partial \theta'} q(Z_i, \theta),
\end{aligned} \tag{2.36}$$

where Y_i is a scalar dependent variable, Z_i is a vector of regressor variables, U_i is an unobserved scalar error with $E(U_i|Z_i) = 0$ a.s., and $q(\cdot, \cdot)$ is a known real function that is twice differentiable in its second argument. The GN matrix $Q^{*,GN}_{N,j-1}$ omits the second summand of the bootstrap version of $(\partial/\partial\theta')g(X_i, \theta)$, because $E((Y_i - q(Z_i, \theta_0))\ (\partial^2/\partial\theta\partial\theta')q(Z_i, \theta_0) = 0$. That is, $Q^{*,GN}_{N,j-1}$ is as in (2.31) (for minimum ρ estimators) and (2.32) with

$$\Delta(\widetilde{X}^*_i, \theta) = \frac{\partial}{\partial \theta} q(Z^*_i, \theta) \frac{\partial}{\partial \theta'} q(Z^*_i, \theta). \tag{2.37}$$

A second example of a GN matrix $Q^{*,GN}_{N,j-1}$ for a minimum ρ estimator is the sample outer-product estimator of the bootstrap information matrix in a ML scenario. Suppose that $\rho_N(\theta)$ is a normalized negative log likelihood function and $g(X_i, \theta) = (\partial/\partial\theta)\rho(X_i, \theta)$ is the negative score (or conditional score) function for the X_i-th observation. By the information matrix equality,

$$E \frac{\partial}{\partial \theta'} g(X_i, \theta_0) = Eg(X_i, \theta_0) g(X_i, \theta_0)' \tag{2.38}$$

when the model is correctly specified. In this case, the NR matrix $Q^{*,NR}_{N,j-1}$ is the bootstrap version of the sample analogue of the expectation on the left-hand side of (2.38): $Q^{*,NR}_{N,j-1} = N^{-1} \sum_{i=1}^{N} (\partial/\partial\theta') g(X^*_i, \theta^*_{N,j-1})$. The GN matrix $Q^{*,GN}_{N,j-1}$ is the bootstrap version of the sample analogue of the expectation on the right-hand side of (2.38). Thus, $Q^{*,GN}_{N,j-1}$ is as in (2.31) (for minimum ρ estimators) and (2.32) with

$$\Delta(\widetilde{X}^*_i, \theta) = g(X^*_i, \theta) g(X^*_i, \theta)'. \tag{2.39}$$

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

For GMM estimators that have the same number of moment conditions as the dimension of θ, such as ML estimators defined via the likelihood equations, $\theta^*_{N,k}$ is the same whether defined using Ω or $\Omega^*_N(\widetilde{\theta}^*_{N,k_1})$ (because the moment conditions $N^{-1}\sum_{i=1}^N g^*(X_i^*, \theta)$ have an exact zero with probability that goes to one at an appropriate rate as $N \to \infty$).

We define the k-step bootstrap t statistic, $T^*_{N,k}$, Wald statistic, $\mathcal{W}^*_{N,k}$, and J statistic, $J^*_{N,k}$, as in (2.16) but with $(\theta^*_N)_r$ and θ^*_N replaced by $\theta^*_{N,k,r}$ and $\theta^*_{N,k}$, respectively, where $\theta^*_{N,k,r}$ denotes the r-th element of $\theta^*_{N,k}$. Let $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, $z^*_{\mathcal{W},k,\alpha}$, and $z^*_{J,k,\alpha}$ denote the $1-\alpha$ quantiles of $|T^*_{N,k}|$, $T^*_{N,k}$, $\mathcal{W}^*_{N,k}$, and $J^*_{N,k}$ respectively.

The k-step bootstrap versions of the bootstrap t tests, Wald test, J test, and confidence intervals and regions are given by (2.20)–(2.28) with $z^*_{|T|,\alpha}$, $z^*_{T,\alpha}$, $z^*_{\mathcal{W},\alpha}$, and $z^*_{J,\alpha}$ replaced by $z^*_{|T|,k,\alpha}$, $z^*_{T,k,\alpha}$, $z^*_{\mathcal{W},k,\alpha}$, and $z^*_{J,k,\alpha}$ respectively.

2.4 Assumptions

We now introduce the assumptions. They are similar to those of HH, but are more flexible, because we consider higher-order terms than are considered in HH.

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

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

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

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

Assumption 2. (a) Θ is compact and θ_0 is an interior point of Θ. (b) Either (i) $\widehat{\theta}_N$ minimizes $J_N(\theta)$ or $J_N(\theta, \widetilde{\theta}_N)$ over $\theta \in \Theta$; θ_0 is the unique solution in Θ to $Eg(X_1, \theta) = 0$; for some function $C_g(x)$, $||g(x, \theta_1) - g(x, \theta_2)|| \leq C_g(x)||\theta_1 - \theta_2||$ for all x in the support of X_1 and all $\theta_1, \theta_2 \in \Theta$; and $EC_g^{q_0}(X_1) < \infty$ and $E||g(X_1, \theta)||^{q_0} < \infty$ for all $\theta \in \Theta$ for some $q_0 \geq 2$ or (ii) $\widehat{\theta}_N$ minimizes $N^{-1}\sum_{i=1}^N \rho(X_i, \theta)$ over $\theta \in \Theta$ for some function $\rho(x, \theta)$ such that $(\partial/\partial\theta)\rho(x, \theta) = g(x, \theta)$ for all x in the support of X_1; θ_0 is the unique minimum of $E\rho(X_1, \theta)$ over $\theta \in \Theta$; and $E\sup_{\theta\in\Theta}||g(X_1, \theta)||^{q_0} < \infty$ and $E|\rho(X_1, \theta)|^{q_0} < \infty$ for all $\theta \in \Theta$ for some $q_0 \geq 2$.

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) $E||g(X_1, \theta_0)||^{q_1} < \infty$ for some $q_1 \geq 2$. (d) $g(x, \theta)$ 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 $d_2 \geq 0$. (e) 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 $j = 0, ..., d_2$. (f) $EC_{\partial f}^{q_2}(\widetilde{X}_1) < \infty$ and $E||(\partial^j/\partial\theta^j) f(\widetilde{X}_1, \theta_0)||^{q_2} \leq C_f < \infty$ for all $j = 0, ..., d_2$ for some constants $q_2 \geq 2$ and C_f. (g) $f(\widetilde{X}_1, \theta_0)$ is once differentiable with respect to $\widetilde{X}_1$ with uniformly continuous first derivative. (h) 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 $\delta < ||t|| < N^\zeta$,

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

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

Assumption 5. For some constants $0 \leq c_Q \leq \infty$ and $0 \leq a < \infty$ and all $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P(N^a P^*(||Q^*_{N,j-1} - \frac{\partial^2}{\partial\theta\partial\theta'} J^*_N(\theta^*_{N,j-1})|| > N^{-c_Q}\varepsilon) > \varepsilon) = 0 \text{ for } j = 1, ..., k$$

for one-step GMM estimators. For two-step GMM estimators, an analogous condition holds with $(\partial^2/\partial\theta\partial\theta')J^*_N(\theta^*_{N,j-1})$ replaced by $(\partial^2/\partial\theta\partial\theta')J^*_N(\theta^*_{N,j-1}, \widetilde{\theta}^*_{N,k_1})$. For minimum ρ estimators, an analogous condition holds with $(\partial^2/\partial\theta\partial\theta')J^*_N(\theta^*_{N,j-1})$ replaced by $D^*_N(\theta^*_{N,j-1})$.

The lower bounds on the number of derivatives d and moments q_0, q_1, q_2 in Assumptions 1–3 are minimal bounds. The results stated below specify more stringent lower bounds that vary depending upon the result.

Assumption 1 is the same as Assumption 1 of HH.[4] 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.

In the next section, we give sufficient conditions for Assumption 5.

3 Nonparametric Bootstrap Results

3.1 Higher-order Asymptotic Equivalence of the k-step and Standard Bootstraps

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

The following Theorem also shows that Assumption 5 holds with $c_Q = \infty$ for the NR, default NR, and line-search NR procedures under Assumptions 1-3 given sufficiently large values of $q_0 - q_2$. It shows that Assumption 5 holds with $c_Q = c \in (0, 1/2)$ for the GN procedure under Assumptions 1-3 given sufficiently large values of $q_0 - q_2$ and some conditions on the function $\Delta(\cdot,\cdot)$ of (2.32), where c depends on the magnitude of q_1, the number of finite moments of $g(X_i,\theta_0)$.

The stochastic difference between the k-step estimator and the extremum estimator is shown to be of magnitude $\mu_{N,k}$, where

$$\mu_{N,k} = \max_{j=0,\ldots,k} N^{-(2^{k-j}c+jc_Q)}, \tag{3.1}$$

where c_Q is as in Assumption 5 and $c \in (0,1/2)$ is a constant that depends on q_1. (If $c_Q = \infty$, then by definition $N^{-(2^{k-j}c+jc_Q)} = 0$ for $j > 0$ and $N^{-(2^{k-j}c+jc_Q)} = N^{-2^k c}$ for $j = 0$.) If Assumption 5 holds with $c_Q = \infty$, then $\mu_{N,k} = N^{-2^k c}$. If Assumption 5 holds with $0 \leq c_Q \leq c$, then $\mu_{N,k} = N^{-(c+kc_Q)}$ (because the $j = k$ term of $\mu_{N,k}$ is dominant). Thus, when $c_Q = \infty$, the distance decreases very quickly as k increases and when $0 \leq c_Q \leq c$, the distance decreases more slowly as k increases.

The conditions under which the following Theorem holds specify trade-offs between the number of steps k, the numbers of moments q_0, q_1, and q_2, the number of derivatives d $(= d_1 + d_2)$ of $g(x,\theta)$ with respect to θ, the closeness of $Q^*_{N,j-1}$ to the second derivative of the criterion function, as measured by c_Q, and the magnitude of the block length parameter γ. Examples of different combinations of these parameters under which all the conditions of the Theorem hold are given in Comment 5 following the Theorem.

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

Theorem 1 *Suppose Assumptions* 1–3 *and* 5 *hold with* $q_0 > 4a$, $q_1 > \max\{4a/(1-2c),\, 8a\}$, *and* $q_2 > 4a$ *for some* $c \in (0,1/2)$ *and* $a \geq 0$. *Assume* $\ell \propto N^\gamma$ *for some* $0 \leq \gamma \leq 1$ *and* k *is a positive integer.*
(a) *Then, for all* $\varepsilon > 0$,

$$P^*(||\theta^*_{N,k} - \theta^*_N|| > \mu_{N,k}\varepsilon) < N^{-a}\varepsilon,$$

except if $\{\chi_N : N \geq 1\}$ *are in a sequence of sets with probability* $o(N^{-a})$.
(b) *Suppose the following additional conditions hold when* $\gamma > 0$ *(for this part of the Theorem only):* $q_2 > 4a/(1-\gamma)$ *and* $d_1 \geq -1 + (a+\gamma)/c$. *Then, for any* $0 \leq \gamma < 1$ *and all* $\varepsilon > 0$,

$$P^*(|T^*_{N,k} - T^*_N| > N^{1/2}\mu_{N,k}\varepsilon) < N^{-a}\varepsilon,$$

$$P^*(|\mathcal{W}^*_{N,k} - \mathcal{W}^*_N| > N^{1/2}\mu_{N,k}\varepsilon) < N^{-a}\varepsilon, \text{ and}$$

$$P^*(|J^*_{N,k} - J^*_N| > N^{1/2}\mu_{N,k}\varepsilon) < N^{-a}\varepsilon,$$

except if $\{\chi_N : N \geq 1\}$ *are in a sequence of sets with probability* $o(N^{-a})$.
(c) *Suppose the following additional conditions hold (for this part of the Theorem only):* $\mu_{N,k} = O(N^{-(a+1/2)})$, $q_2 \geq 2a+3$, $q_2 > 6a/(1-2\gamma)$, $d_1 \geq (2a+1)/(2c)$,

$d_2 \geq -1 + (a+\gamma)/c$, $0 < \gamma < 1/2$ (*where* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent*), $2a$ *is an integer, and Assumption* 4 *holds. Then, for all* $\varepsilon > 0$,

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

$$\sup_{z \in R} \left| P^*(T^*_{N,k} \leq z) - P^*(T^*_N \leq z) \right| < N^{-a}\varepsilon \text{ under } H_0,$$

$$\sup_{z \in R} \left| P^*(\mathcal{W}^*_{N,k} \leq z) - P^*(\mathcal{W}^*_N \leq z) \right| < N^{-a}\varepsilon \text{ under } H_0,$$

$$\sup_{z \in R} \left| P^*(J^*_{N,k} \leq z) - P^*(J^*_N \leq z) \right| < N^{-a}\varepsilon \text{ under } H_0,$$

except if $\{\chi_N : N \geq 1\}$ *are in a sequence of sets with probability* $o(N^{-a})$.
(d) *Given the assumptions listed before part* (a), *Assumption* 5 *holds with* $c_Q = \infty$ *for the NR, default NR, and line-search NR choices of* $Q^*_{N,j-1}$ *for* $j = 1, ..., k$.
(e) *Given the assumptions listed before part* (a), *Assumption* 5 *holds with* $c_Q = c$ *for the GN choice of* $Q^*_{N,j-1}$ *for* $j = 1, ..., k$, *provided the function* $\Delta(\cdot,\cdot)$ *in* (2.32) *satisfies* $E(\Delta(\widetilde{X}_i, \theta_0) - (\partial/\partial\theta')g(X_i, \theta_0)) = 0$, $E||\Delta(\widetilde{X}_i, \theta_0) - (\partial/\partial\theta')g(X_i, \theta_0)||^{q_3} < \infty$ *for some* $q_3 > 4a/(1-2c)$ *and* $q_3 \geq 2$, *and* $E \sup_{\theta \in N_0} ||(\partial/\partial\theta_r)\Delta(\widetilde{X}_i, \theta)||^{q_4} < \infty$ *for all* $r = 1, ..., L_\theta$ *for some* $q_4 > 4a$ *and* $q_4 \geq 2$.

Comments: 1. Another way to express the result of part (a) of the Theorem (and analogously parts (b) and (c)) is: for all $\varepsilon > 0$,

$$\lim_{N \to \infty} N^a P(N^a P^*(||\theta^*_{N,k} - \theta^*_N|| > \mu_{N,k}\varepsilon) > \varepsilon) = 0. \tag{3.2}$$

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

3. If Assumption 5 holds with $c_Q = \infty$, as it does with the NR, default NR, and line-search NR k-step procedures by part (d) of the Theorem, then the condition $\mu_{N,k} = O(N^{-(a+1/2)})$ in part (c) of the Theorem holds if $2^k \geq (a+1/2)/c$. If Assumption 5 holds with $0 \leq c_Q \leq c$, as it does with the GN k-step procedure by part (e) of the Theorem, then the condition $\mu_{N,k} = O(N^{-(a+1/2)})$ in part (c) holds if $k \geq (a + 1/2 - c)/c_Q$. For example, if Assumptions 1-5 hold for all finite values of q_0, q_1, q_2, d_1, and d_2, then part (c) of the Theorem holds provided $2^k > 2a+1$ when $c_Q = \infty$ and when $k > 2a$ when $c_Q = c$ (because part (c) holds for all $c < 1/2$).

4. Because no conditions on d_2 are specified in parts (a) and (b) of the Theorem, one can take $d_2 = 0$ in these parts.

5. Examples of combinations of a, c, k, c_Q, q_0, q_1, q_2, d, d_1, d_2, and γ that satisfy the conditions of all parts of the Theorem are as follows: take $d_2 = 0$ and $d = d_1$ in parts (a) and (b). Take $d = d_1 + d_2$ with $d_2 > 0$ as specified below in part (c).

Suppose we consider $a = 1$, as in HH, and we take $c = 3/8$. Then, we need the number of steps $k \geq 2$ when $c_Q = \infty$ and $k \geq 3$ when $c_Q = c = 3/8$; the numbers of moments $q_0 > 4$, $q_1 > 16$, and $q_2 > 6/(1 - 2\gamma)$; the number of derivatives $d_1 \geq 4$ and, in part (c), $d_2 \geq 5/3 + 8\gamma/3$; and the magnitude of the block length parameter $\gamma < 1/2$. For independent data, we take the block length parameter $\gamma = 0$ and we need $q_2 > 6$ and, in part (c), $d_2 \geq 2$. For dependent data, with $\gamma = 1/4$, say, we need $q_2 > 12$ and, in part (c), $d_2 \geq 3$.

Alternatively, suppose we consider $a = 3/2$ and we take $c = 1/4$. Then, we need $k \geq 3$ when $c_Q = \infty$, $k \geq 7$ when $c_Q = c = 1/4$, $q_0 > 6$, $q_1 > 12$, $q_2 > 9/(1- 2\gamma)$, $d_1 \geq 8$, $d_2 \geq 5 + 4\gamma$ in part (c), and $\gamma < 1/2$. If we take $\gamma = 0$ or $1/4$, we need $q_2 > 9$ or 18, respectively, and, in part (c), $d_2 \geq 5$ or 6, respectively. As a second alternative, suppose we consider $a = 3/2$ and we take $c = 2/5$. Then, we need $k \geq 3$ when $c_Q = \infty$, $k \geq 4$ when $c_Q = c = 2/5$, $q_0 > 6$, $q_1 > 30$, $q_2 > 9/(1- 2\gamma)$, $d_1 \geq 5$, $d_2 \geq 11/4 + 5\gamma/2$ in part (c), and $\gamma < 1/2$.

Alternatively, suppose we consider $a = 2$ and we take $c = 5/16$. Then, we need $k \geq 3$ when $c_Q = \infty$, $k \geq 7$ when $c_Q = c = 5/16$, $q_0 > 8$, $q_1 > 21.34$, $q_2 > 12/(1-2\gamma)$, $d_1 \geq 8$, $d_2 \geq 27/5 + 16\gamma/5$ in part (c), and $\gamma < 1/2$. If we take $\gamma = 0$ or $1/4$, then we need $q_2 > 12$ or 24, respectively, and, in part (c), $d_2 \geq 6$ or 7, respectively. As a second alternative, suppose we consider $a = 2$ and we take $c = 5/12$. Then, we need $k \geq 3$ when $c_Q = \infty$, $k \geq 5$ when $c_Q = c = 5/12$, $q_0 > 8$, $q_1 > 48$, q_2 as above, $d_1 \geq 6$, $d_2 \geq 19/5 + 12\gamma/5$ in part (c), and $\gamma < 1/2$.

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

3.2 Higher-order Improvements of the k-step and Standard Block Bootstraps

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

The following Theorem shows that the k-step and standard symmetric two-sided bootstrap t, Wald, and J tests have rejection probabilities that are correct up to $o(N^{-(1+\xi)})$ for some $0 \leq \xi < 1/2 - \gamma$ (where the block length ℓ is proportional to N^{γ}). It shows that the k-step and standard bootstrap equal-tailed two-sided t and one-sided t tests have rejection probabilities that are correct up to $o(N^{-(1/2+\xi)})$ for some $0 \leq \xi < 1/2 - \gamma$. The coverage probabilities of the corresponding confidence intervals are correct to the same orders.

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

Theorem 2 (a) *Let Assumptions* 1–5 *hold with* $q_0 > 6$, $q_1 > \max\{6\ /(1 - 2c),\ 12\}$,

$q_2 > 9/(1-2\gamma-2\xi)$, $d_1 \geq 2/c$, $d_2 \geq -1+(3/2+\gamma+\xi)/c$, and $a = 3/2$ for some $c \in (0,1/2)$. Suppose $\mu_{N,k} = O(N^{-2})$ for a positive integer k, $0 \leq \xi < 1/2-\gamma$, and either (i) $\xi < \gamma$ and $0 < \gamma < 1/2$ or (ii) $\{X_i : i \geq 1\}$ are independent. Then, under $H_0 : \theta_r = \theta_{0,r}$,

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

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

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

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

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

(b) *Let Assumptions 1–5 hold with $q_0 > 4$, $q_1 > \max\{4/(1-2c), 8\}$, $q_2 > 6/(1-2\gamma-2\xi)$, $d_1 \geq 3/(2c)$, $d_2 \geq -1+(1+\gamma+\xi)/c$, and $a = 1$ for some $c \in (0,1/2)$. Suppose $\mu_{N,k} = O(N^{-3/2})$ for a positive integer k, $0 \leq \xi < 1/2-\gamma$, and either (i) $\xi < \gamma$ and $0 < \gamma < 1/2$ or (ii) $\{X_i : i \geq 1\}$ are independent. Then, under $H_0 : \theta_r = \theta_{0,r}$,*

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

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

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

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

Comments: 1. The errors in parts (a) and (b) of the Theorem when the critical values are based on standard first-order asymptotics (using the normal distribution or the chi-square distribution) are $O(N^{-1})$ and $O(N^{-1/2})$ respectively. Thus, the Theorem shows that the bootstrap critical values reduce the error in test rejection probability (and in confidence interval coverage probability) relative to first-order asymptotics by at least $N^{-\xi}$. When the data are independent, one takes $\gamma = 0$ and the results of the Theorem hold for all $\xi < 1/2$.

When the data are dependent, the choice of $\gamma = 1/4$ maximizes ξ subject to the requirements of the Theorem that $\xi < \gamma$ and $\xi + \gamma < 1/2$. For this choice of γ, the results of the Theorem hold for all $\xi < 1/4$.

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

2. For the NR, default NR, and line-search NR choices of $Q^*_{N,j-1}$, Assumption 5 automatically holds with $c_Q = \infty$, $a = 3/2$ in part (a), and $a = 1$ in part (b) by Theorem 1(d). For the GN choice of $Q^*_{N,j-1}$, Assumption 5 automatically holds with $c_Q = c$, $a = 3/2$ in part (a), and $a = 1$ in part (b) by Theorem 1(e).

The conditions on $\mu_{N,k}$ are only needed for the k-step bootstrap procedure, not the standard bootstrap. If Assumption 5 holds with $c_Q = \infty$, then the condition $\mu_{N,k} = O(N^{-2})$ in part (a) holds if $2^{k-1} \geq 1/c$ and the condition $\mu_{N,k} = O(N^{-3/2})$ in part (b) holds if $2^{k+1} \geq 3/c$. If Assumption 5 holds with $0 \leq c_Q \leq c$, then the condition $\mu_{N,k} = O(N^{-2})$ in part (a) holds if $k \geq (2-c)/c_Q$ and the condition $\mu_{N,k} = O(N^{-3/2})$ in part (b) holds if $k \geq (3/2 - c)/c_Q$.

3. In part (a) of the Theorem, taking $c = 1/4$, it suffices to have $k \geq 3$ when $c_Q = \infty$, $k \geq 7$ when $c_Q = c = 1/4$, $q_0 > 6$, $q_1 > 12$, $q_2 > 9/(1-2\xi)$ when $\gamma = 0$, $q_2 > 18/(1-4\xi)$ when $\gamma = 1/4$, $d_1 \geq 8$, and $d_2 \geq 5 + 4\gamma + 4\xi$ (e.g., $d_2 \geq 9$ suffices). Alternatively, taking $c = 2/5$, it suffices to have $k \geq 3$ when $c_Q = \infty$, $k \geq 4$ when $c_Q = c = 2/5$, $q_0 > 6$, $q_1 > 30$, q_2 as above, $d_1 \geq 5$, and $d_2 \geq 11/4 + 5\gamma/2 + 5\xi/2$ (e.g., $d_2 \geq 4$ suffices).

In part (b) of the Theorem, taking $c = 3/8$, it suffices to have $k \geq 2$ when $c_Q = \infty$, $k \geq 3$ when $c_Q = c = 3/8$, $q_0 > 4$, $q_1 > 16$, $q_2 > 6/(1-2\xi)$ when $\gamma = 0$, $q_2 > 12/(1-4\xi)$ when $\gamma = 1/4$, $d_1 \geq 4$, $d_2 \geq 5/3 + 8(\gamma+\xi)/3$ (e.g., $d_2 \geq 3$ suffices). Alternatively, taking $c = 1/4$, it suffices to have $k \geq 3$ when $c_Q = \infty$, $k \geq 5$ when $c_Q = c = 1/4$, $q_0 > 4$, $q_1 > 8$, q_2 as above, $d_1 \geq 6$, and $d_2 \geq 3 + 4\gamma + 4\xi$ (e.g., $d_2 \geq 5$ suffices).

4. When we take $\xi = 0$, the results of the Theorem for the standard bootstrap are comparable to those of HH. In this case, the conditions of Theorem 2 can be weakened, because we only require Theorem 1(c) to hold with $a = 1$ and $a = 1/2$ to prove parts (a) and (b), rather than $a = 3/2$ and $a = 1$, respectively, and we only require Edgeworth expansions through order $O(N^{-1})$ and $O(N^{-1/2})$ to prove parts (a) and (b), rather than $O(N^{-3/2})$ and $O(N^{-1})$. Accordingly, we require $\mu_{N,k} = O(N^{-3/2})$ in part (a) and $\mu_{N,k} = O(N^{-1})$ in part (b). Thus, when $\xi = 0$, part (a) holds provided $2^k \geq 3/(2c)$ when $c_Q = \infty$, $k \geq (3/2 - c)/c_Q$ when $c_Q \leq c$, $q_0 > 4$, $q_1 > \max\{4/(1-2c), 8\}$, $q_2 > 6/(1-2\gamma)$, $d_1 \geq 3/(2c)$, $d_2 \geq -1 + (1+\gamma)/c$, and $0 < \gamma < 1/2$ for some $c \in (0, 1/2)$ (where $\gamma = 0$ is permitted if $\{X_i : i \geq 1\}$ is independent). For example, taking $c = 1/4$, it suffices to have $k \geq 3$ when $c_Q = \infty$, $k \geq 5$ when $c_Q = c = 1/4$, $q_1 > 8$, $d_1 \geq 6$, and $d_2 \geq 3 + 4\gamma$. Alternatively, taking $c = 3/8$, it suffices to have $k \geq 2$ when $c_Q = \infty$, $k \geq 3$ when $c_Q = c = 3/8$, $q_1 > 16$, $d_1 \geq 4$, and $d_2 \geq 5/3 + 8\gamma/3$. If $\gamma = 0$ or $1/4$, it suffices to have $q_2 > 6$ or $q_2 > 12$.

In contrast, HH require $\min\{q_0, q_1, q_2\} \geq 32$, $d_1 \geq 4$, $d_2 \geq 0$, and $11/50 \leq \gamma \leq 12/50$. The latter condition precludes the choice of $\gamma = 1/4$, which maximizes the upper bound on ξ.

We note that Lemma 11 of HH does not appear to establish a strong enough result for the use it is given in the proof of their Theorem 3. A stronger property that suffices is that $\lim_{N\to\infty} NP(||\nu^*_{m(j)} - \nu_{m(j)}|| > \varepsilon) = 0$ rather $\lim_{N\to\infty} P(||\nu^*_{m(j)} - \nu_{m(j)}|| > \varepsilon) = 0$. Using the method of proof given here, it appears that HH would need more moment conditions than 32 and $d_2 > 0$ to establish the stronger result using the Lahiri moment inequality. Thus, the above comparison between the conditions used here and those used in HH should be treated with caution.

When we take $\xi = 0$, part (b) of the Theorem holds provided $2^k \geq 1/c$ when $c_Q = \infty$, $k \geq (1-c)/c_Q$ when $c_Q \leq c$, $q_0 > 2$, $q_1 > \max\{2/(1-2c), 4\}$, $q_2 \geq 4$,

$q_2 > 3/(1-2\gamma)$, $d_1 \geq 1/c$, $d_2 \geq -1 + (1/2+\gamma)/c$, and $0 < \gamma < 1/2$ for some $c \in (0, 1/2)$ (where $\gamma = 0$ is permitted if $\{X_i : i \geq 1\}$ is independent). For example, taking $c = 1/4$, it suffices to have $k \geq 2$ when $c_Q = \infty$, $k \geq 3$ when $c_Q = c = 1/4$, $q_1 > 4$, $d_1 \geq 4$, and $d_2 \geq 1 + 4\gamma$. Taking $\gamma = 0$ or $1/4$, it suffices to have $q_2 > 3$ or 6, respectively, and $d_2 \geq 1$ or 2, respectively.

HH do not provide results that correspond to those of part (b) of the Theorem. For the case of $\xi = 0$, however, such results for the standard bootstrap, but not the k-step bootstrap, can be obtained easily from the results given in their Appendix. For this, their results require $\min\{q_0, q_1, q_2\} \geq 32$, $d_1 \geq 4$, $d_2 \geq 0$, and $11/50 \leq \gamma \leq 12/50$. (The same caveat applies here as above.) The block length $\gamma = 1/4$, which maximizes the upper bound on ξ, is outside the range allowed by HH.

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

6. When the observations are iid, the result of Theorem 2(a) for $|T_N|$ can be improved to $\alpha + O(N^{-2})$ using Hall's (1988) argument for symmetric percentile t confidence intervals. In particular, the principal ingredients needed for his argument are Edgeworth expansions for T_N and T_N^* to suitable order, which are provided in Lemma 16(a) and (b), respectively, in the Appendix. Similar improvements for Wald and J statistics are likely to hold when the data are iid, although a somewhat different argument would be needed. Similar improvements to the results of Theorem 2(b) are not possible, because the symmetry property which makes the improvements possible for $|T_N|$ are not present.

7. For dependent data, the possibility of improving the result of Theorem 2(a) for $|T_N|$ using the symmetry argument of Hall (1988) is less clear. At best, this would lead to an error of $O(N^{-3/2})$, because the bootstrap moments and population moments differ by at least $N^{-1/4}$ in the dependent case, rather than $N^{-1/2}$ in the independent case. But, even an improvement of this magnitude is difficult to establish for the following reason. Hall's (1988) argument for the $O(N^{-2})$ error in the iid case relies on determining Edgeworth expansions of $T_N \pm \Delta$, where Δ denotes the difference between the exact critical value and the bootstrap critical value. This is done by establishing Cornish-Fisher expansions for these two critical values, approximating the difference of these two expansions by a linear combination of sample averages of the data, and utilizing the smooth function of sample averages approach to Edgeworth expansions to get the Edgeworth expansions for $T_N \pm \Delta$, see Hall (1988, Sec. 3; 1992, Sec. 5.3). This method relies on the fact that the coefficients of the Cornish-Fisher expansion of the bootstrap critical value depend on bootstrap moments that are sample averages. With dependent observations, the coefficients of the Cornish-Fisher expansion of the bootstrap critical value depend on bootstrap moments, but the bootstrap moments are not sample averages. The bootstrap moments depend on terms of the form $\sum_{j=1}^{\mathcal{N}_\ell}(\sum_{i \in b_j} h(\widetilde{X}_i))^s$, where s is an integer that depends on the order of the moment, $h(\cdot)$ is some function that depends on the criterion function or its derivatives, b_j is a set of indices for the j–th block, and $\mathcal{N}_\ell$ is the number of blocks. The number of indices in b_j increases with N. There are no Edgeworth expansion results in the literature that cover terms of the above type when the random variables $\{\widetilde{X}_i : i \geq 1\}$ are dependent. Thus, without new Edgeworth expansion results for statistics involving terms of the above type, one cannot prove results utilizing Hall's (1988) symmetry argument.

4 Parametric Bootstrap

In this section, we provide results for likelihood-based methods using *unrestricted* and *restricted parametric* bootstraps. The unrestricted parametric bootstrap utilizes the unrestricted ML estimator to generate bootstrap samples. It can be used for both bootstrap tests and confidence intervals. The restricted parametric bootstrap utilizes a restricted ML parameter estimator, which satisfies the null hypothesis of interest, to generate bootstrap samples. It is appropriate only for bootstrap tests. It typically outperforms the unrestricted parametric bootstrap for testing problems, because it minimizes the likelihood function over a smaller parameter space, see Davidson and MacKinnon (1999b).

We define the parametric model considered here in such a way that the notation and results of the previous sections can be utilized as much as possible. The main differences are that we consider a particular minimum ρ estimator, viz., the ML estimator, and we use a parametric bootstrap instead of a block bootstrap. We obtain higher-order improvements of the unrestricted and restricted parametric bootstraps that are the same whether or not the data are dependent. The improvements are of the same magnitude as were obtained in the previous section for the case of iid data.

We consider a correctly specified parametric model for a (strictly) stationary process $\{X_i : i = 1, ..., n\}$. Let $X_i = (Y_i', Z_i')'$, where Y_i is a vector of dependent (or response) variables and Z_i is a vector of "regressor" variables. The dependent random variables $\{Y_i : i = 1, ..., n\}$ form a κ-th order Markov process. The regressor variables $\{Z_i : i = 1, ..., n\}$ are strictly exogenous.

Assumption 6. **(a)** The parametric model specifies the density of Y_i given $(Z_i, X_{i-1}, X_{i-2}, ..., X_1)$ (with respect to some σ-finite measure μ) to be $d(\cdot|Z_i, X_{i-1}, X_{i-2}, ..., X_{i-\kappa}; \theta)$ for $i = \kappa + 1, ..., n$, and $d(\cdot|Z_i, X_{i-1}, X_{i-2}, ..., X_1; \theta)$ for $i = 1, ...\kappa$, for some integer $\kappa \geq 0$, where the densities for $i = 1, ..., \kappa$ yield a stationary start-up of the process and θ is a parameter in the parameter space $\Theta \subset R^{L_\theta}$.
(b) The true distribution of $\{Y_i : i = 1, ..., n\}$ is given by the parametric model with $\theta = \theta_0$.
(c) The variable Z_i is independent of $(Y_{i-1}, Y_{i-2}, ..., Y_1)$ for all $i = 2, ..., n$.
(d) The distribution of Z_i given $(Z_{i-1}, Z_{i-2}, ..., Z_1)$ does not depend on θ for all $i = 2, ..., n$.

Given Assumption 6(d), the contribution of the density of Z_i to the likelihood can be ignored when estimating θ. Given Assumption 6(c), the same regressors $\{Z_i : i = 1, ..., n\}$ can be used in each bootstrap sample.

We adopt the same notation as in the previous section and take $\widetilde{X}_i = (X_i', ..., X_{i+\kappa}')'$ for $i = 1, ..., N$, where $N = n - \kappa$. The part of the normalized negative of the log likelihood function that depends on θ is

$$\frac{1}{N}\sum_{i=1}^{N}\rho(\widetilde{X}_i, \theta), \text{ where } \rho(\widetilde{X}_i, \theta) = -\log d(Y_{i+\kappa}|Z_{i+\kappa}, X_{i+\kappa-1}, X_{i+\kappa-2}, ..., X_i; \theta).^5 \tag{4.1}$$

We consider the (unrestricted) ML estimator $\widehat{\theta}_N$, which is defined to be the estimator that minimizes $N^{-1}\sum_{i=1}^{N}\rho(\widetilde{X}_i, \theta)$ over Θ. Following the notation of the previous section, the moment conditions for $\widehat{\theta}_N$ are

$$g(\widetilde{X}_i, \theta) = -\frac{\partial}{\partial\theta}\log d(Y_{i+\kappa}|Z_{i+\kappa}, X_{i+\kappa-1}, X_{i+\kappa-2}, ..., X_i; \theta).^6 \tag{4.2}$$

The variance matrix estimator σ_N of $\widehat{\theta}_N$ can be defined in several ways, because D and Ω_0 are square matrices and the information matrix equality implies that D and Ω_0^{-1} are equal. In particular, one can use

$$\begin{aligned}\sigma_N &= D_N^{-1}(\widehat{\theta}_N)\Omega_N^{-1}(\widehat{\theta}_N)D_N^{-1}(\widehat{\theta}_N), \ \sigma_N = D_N^{-1}(\widehat{\theta}_N), \text{ or } \sigma_N = \Omega_N(\widehat{\theta}_N), \text{ where} \\ \Omega_N(\theta) &= \overline{W}_N^{-1}(\theta), \ \overline{W}_N(\theta) = N^{-1}\sum_{i=1}^{N} g(\widetilde{X}_i, \theta)g(\widetilde{X}_i, \theta)', \text{ and} \\ D_N(\theta) &= N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}g(\widetilde{X}_i, \theta).\end{aligned} \tag{4.3}$$

(The terms involving $g(\widetilde{X}_i, \theta)g(\widetilde{X}_{i+j}, \theta)'$ in the definition of $\overline{W}_N(\theta)$ in Section 2 are not needed here, because the score function evaluated at θ_0 is uncorrelated.) The t statistic, T_N, and the Wald statistic, $\mathcal{W}_N$, are defined exactly as in Section 2. The J–statistic is not relevant here, because there are no over-identifying restrictions.

The unrestricted parametric bootstrap sample $\{X_i^* : i = 1, ..., n\}$ is defined as follows. The bootstrap regressors are the same as the regressors in the original sample and the bootstrap dependent variables are generated recursively for $i = 1, ..., n$ using the parametric density evaluated at the unrestricted ML estimator $\widehat{\theta}_N$. That is, one takes $X_i^* = (Y_i^{*\prime}, Z_i')'$, where Y_i^* has density $d(\cdot|Z_i, X_{i-1}^*, X_{i-2}^*, ..., X_{i-\kappa_i}^*; \widehat{\theta}_N)$ for $i = 1, ..., n$, where $\kappa_i = \min\{\kappa, i+1\}$. The bootstrap observations $\widetilde{X}_i^*$ are defined to be $\widetilde{X}_i^* = (X_i^{*\prime}, ..., X_{i+\kappa}^{*\prime})'$ for $i = 1, ..., N$. Under Assumption 6, the conditional distribution of the bootstrap sample given $\widehat{\theta}_N$ is the same as the distribution of the original sample except that the true parameter is $\widehat{\theta}_N$ rather than θ_0.

The restricted parametric bootstrap employs a restricted ML estimator, denoted $\overline{\theta}_N$, that imposes the restrictions specified by the null hypothesis of interest. Consider the null hypothesis $H_0 : \eta(\theta_0) = 0$ considered above for the Wald statistic. (This covers the null hypothesis for the t statistic $H_0 : \theta_r = \theta_{0,r}$ by appropriate choice of $\eta(\cdot)$.) By definition, the restricted ML estimator $\overline{\theta}_N$ minimizes $N^{-1}\sum_{i=1}^{N}\rho(\widetilde{X}_i, \theta)$ over $\theta \in \Theta$ subject to the restrictions $\eta(\theta) = 0$. Under the assumptions employed here, the restricted estimator is $N^{1/2}$-consistent and asymptotically normal (with singular asymptotic covariance matrix), e.g., see Aitcheson and Silvey (1958) for the iid case and Andrews (1999) for the time series case.

The restricted parametric bootstrap sample, also denoted $\{X_i^* : i = 1, ..., n\}$, is defined in exactly the same way as the unrestricted bootstrap sample, but with $\widehat{\theta}_N$ replaced by the restricted ML estimator $\overline{\theta}_N$.

The bootstrap estimator θ_N^* is defined exactly as the original estimator $\widehat{\theta}_N$ is defined in (2.5) but with the original sample $\{\widetilde{X}_i : i = 1, ..., N\}$ replaced by the bootstrap sample $\{\widetilde{X}_i^* : i = 1, ..., N\}$.[7]

The k-step bootstrap estimator, $\theta_{N,k}^*$, is defined as in (2.29) for minimum ρ estimators with $g^*(X_i^*, \theta_{N,j-1}^*)$ replaced by $g(\widetilde{X}_i^*, \theta_{N,j-1}^*)$. The NR and GN matrices $Q_{N,j-1}^*$ are as in (2.30) and (2.31) respectively. The default NR and line-search NR matrices are as defined for minimum ρ estimators. The outer product form of the GN matrix is as defined in (2.31) (for minimum ρ estimators), (2.32), and (2.39). Alternatively, one can use a GN matrix $Q_{N,j-1}^*$ based on the *expected* information matrix:

$$Q_{N,j-1}^{*,GN2} = E_\theta^* \frac{\partial}{\partial \theta'} g(\widetilde{X}_i^*, \theta)\Big|_{\theta = \theta_{N,j-1}^*}, \tag{4.4}$$

where E_θ^* denotes expectation when the true parameter is θ. For the restricted parametric bootstrap, one replaces $\widehat{\theta}_N$ by $\overline{\theta}_N$. The matrix $Q_{N,j-1}^{*,GN2}$ satisfies Assumption 5 with $c_Q = c$ under the other conditions stated in part (a) of the Theorem below. The expected information matrix is often used in the statistical literature on (non-bootstrap) one-step and k-step estimators in likelihood scenarios, e.g., see Pfanzagl (1974).

The bootstrap covariance matrix estimators, σ_N^* and $\sigma_{N,k}^*$, are defined as σ_N is defined in (4.3), but with the bootstrap sample in place of the original sample and θ_N^* or $\theta_{N,k}^*$ in place of $\widehat{\theta}_N$.

For the restricted parametric bootstrap, the bootstrap t and Wald statistics, T_N^* and $\mathcal{W}_N^*$, are defined just as the original statistics are defined in (2.8) and (2.9), but with the original statistics $\widehat{\theta}_N$ and σ_N replaced by the bootstrap statistics θ_N^* and σ_N^*; and the k-step statistics $T_{N,k}^*$ and $\mathcal{W}_{N,k}^*$ are defined using the same formulae, but with θ_N^* and σ_N^* replaced by $\theta_{N,k}^*$ and $\sigma_{N,k}^*$.

For the unrestricted parametric bootstrap, just as for the block bootstrap, the bootstrap t and Wald statistics need to be such that their distributions mimic the null non-bootstrap distribution even when sample is generated by a parameter in the alternative hypothesis. This is done by centering the statistics at $\widehat{\theta}_N$ rather than θ_0. For the unrestricted parametric bootstrap, we define

$$\begin{aligned} T_N^* &= N^{1/2}((\theta_N^*)_r - \widehat{\theta}_{N,r})/(\sigma_N^*)_{rr}^{1/2} \text{ and} \\ \mathcal{W}_N^* &= H_N^*(\theta_N^*)' H_N^*(\theta_N^*), \text{ where} \\ H_N^*(\theta) &= \left((\frac{\partial}{\partial \theta'} \eta(\theta)) \sigma_N^*(\theta) \frac{\partial}{\partial \theta'} \eta(\theta)' \right)^{-1/2} N^{1/2}(\eta(\theta) - \eta(\widehat{\theta}_N)), \end{aligned} \tag{4.5}$$

$(\sigma_N^*)_{rr}$ denotes the (r, r)-th element of σ_N^*. For the unrestricted parametric bootstrap,

$T^*_{N,k}$ and $\mathcal{W}^*_{N,k}$ are defined as in (4.5), but with θ^*_N and σ^*_N replaced by $\theta^*_{N,k}$ and $\sigma^*_{N,k}$ respectively. Note that no correction factors are employed with either of the parametric bootstraps (because no blocking scheme is used).

For the parametric bootstraps to work properly, the regularity conditions of Section 2, i.e., Assumptions 1–4, need to hold not just when the true value is θ_0, but when the true value is θ_1 for any θ_1 in a neighborhood of θ_0 :

Assumption 7. Assumptions 1–3 hold with (i) θ_0 replaced by θ_1 throughout the assumptions when the observations are distributed according to the parametric model with true parameter θ_1 for any $\theta_1 \in N_0$, where N_0 is a neighborhood of θ_0, and the same functions $C_g(\cdot)$ and $C_{\partial f}(\cdot)$ and the same constants K, ξ, and C_f in Assumptions 1–3 apply for any true value $\theta_1 \in N_0$, (ii) X_i replaced with $\widetilde{X}_i$ in $\rho(X_i, \theta)$ and $g(X_i, \theta)$, (iii) $f(\widetilde{X}_i, \theta)$ defined to include only $g(\widetilde{X}_i, \theta)$, $g(\widetilde{X}_i, \theta)g(\widetilde{X}_i, \theta)'$, and their derivatives through order $d_1 \geq 3$ (and not $g(\widetilde{X}_i, \theta)g(\widetilde{X}_{i+j}, \theta)'$ for $j = 1, ..., \kappa$), and (iv) Assumption 3(a) deleted.

Assumption 8. Assumption 4 holds in the cases outlined in condition (i) of Assumption 7.

The function $f(\widetilde{X}_i, \theta)$ is defined without $g(\widetilde{X}_i, \theta)g(\widetilde{X}_{i+j}, \theta)'$ for $j = 1, ..., \kappa$ in Assumption 6, because the estimator $\Omega_N(\theta)$ does not contain these terms. Assumption 3(a) is not needed—it is redundant, because $g(\widetilde{X}_i, \theta_0)$ is the score function and the score function is necessarily uncorrelated.

The following Theorem provides analogues of Theorems 1 and 2 for the unrestricted and restricted parametric bootstraps.

Theorem 3 *The following results hold for both the unrestricted and restricted parametric bootstrap procedures.*
(a) *Suppose Assumptions* 5–7 *hold with* $q_0 > 4a$, $q_1 > \max\{4a/(1-2c), 8a\}$, *and* $q_2 > 4a$ *for some* $c \in (0, 1/2)$ *and* $a \geq 0$. *Then, the results of Theorem* 1(a), (d), *and* (e) *and the first two results of Theorem* 1(b) (*regarding* $T^*_{N,k}$ *and* $\mathcal{W}^*_{N,k}$) *hold.*
(b) *Suppose Assumptions* 5–8 *hold with* q_0, q_1, *and* q_2 *as in part* (a), $\mu_{N,k} = O(N^{-(a+1/2)})$, $d_1 > (2a+1)/(2c)$, $q_2 \geq 2a+3$, *and* $2a$ *is an integer. Then, the first three results of Theorem* 1(c) (*regarding* $\theta^*_{N,k}$, $T^*_{N,k}$, *and* $\mathcal{W}^*_{N,k}$) *hold.*
(c) *Suppose Assumptions* 5–8 *hold with* $q_0 > 6$, $q_1 > \max\{6/(1-2c), 6/(1-2\xi), 12\}$, $q_2 > 6$, *and* $d_1 > 2/c$ *for some* $c \in (0, 1/2)$. *Suppose* $\mu_{N,k} = O(N^{-2})$ *and* $0 \leq \xi < 1/2$. *Then, the first four results of Theorem* 2(a) (*regarding* $|T_N|$ *and* $\mathcal{W}_N$) *hold.*
(d) *Suppose Assumptions* 5–8 *hold with* $q_0 > 4$, $q_1 > \max\{4/(1-2c), 4/(1-2\xi), 8\}$, $q_2 \geq 5$, *and* $d_1 > 3/(2c)$ *for some* $c \in (0, 1/2)$. *Suppose* $\mu_{N,k} = O(N^{-3/2})$ *and* $0 \leq \xi < 1/2$. *Then, the results of Theorem* 2(b) *hold.*

Comments: 1. Part (a) of the Theorem shows that the stochastic difference between the k-step estimator and the ML estimator and the corresponding t-statistics converges to zero very quickly. It also gives sufficient conditions for Assumption 5. Part (b) establishes the equivalence of the higher-order efficiency of the k-step and

ML bootstrap estimators and the corresponding equivalence for the k-step and extremum t and Wald bootstrap statistics. Part (c) of the Theorem establishes higher-order improvements of the parametric bootstrap for a two-sided t test, a Wald test, a symmetric two-sided percentile t confidence interval, and a Wald-based confidence region. Part (d) does likewise for two-sided equal-tailed t tests and confidence intervals and one-sided t tests and confidence intervals.

2. The moment conditions in Theorem 3 are similar to, and the conditions on k are identical to, those of Theorems 1 and 2. The smoothness conditions are weaker in the present Theorem, because one can take $d_2 = 0$ throughout.

5 Appendix of Proofs

In the first subsection of this Appendix, we state Lemmas 1–16 that are used in the proofs of Theorems 1– 3. A number of these Lemmas are similar to Lemmas in HH, but in the Lemmas given here the rates of convergence to the limits are allowed to vary depending upon the assumptions. In the second subsection, we prove Theorems 1–3. In the third subsection, we prove Lemmas 1–16.

Throughout the Appendix, c and a denote constants that satisfy $c \in [0, 1/2)$ and $a \geq 0$. In Lemmas 13-16, $c \in (0, 1/2)$. In addition, C denotes a generic constant that may change from one equality or inequality to another.

5.1 Lemmas

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

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

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

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

Lemma 2 *Suppose Assumptions* 1–3 *hold with* $q_0 > 2a$. *Define* $G_1(X_i, \theta) = g(X_i, \theta) - Eg(X_1, \theta)$ *and* $G_2(X_i, \theta) = \rho(X_i, \theta) - E\rho(X_1, \theta)$. *Then, for all* $\varepsilon > 0$,

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

Lemma 3 *Let* $\widehat{\theta}_N$ *denote the one-step GMM or minimum* ρ *estimator. Suppose Assumptions* 1–3 *hold with* $q_0 > 2a$, $q_1 > 2a/(1-2c)$, *and* $q_2 > 2a$. *Then, for all* $\varepsilon > 0$,

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

Lemma 4 *Let* $\widetilde{\theta}_N$ *denote the one-step GMM estimator based on the weight matrix* Ω. *Let* $\widehat{\theta}_N$ *denote the two-step GMM estimator based on the weight matrix* $\Omega_N(\widetilde{\theta}_N)$. *Suppose Assumptions* 1–3 *hold with* $q_0 > 2a$, $q_1 > \max\{2a/(1-2c), 4a\}$, *and* $q_2 > 2a$. *Then, for all* $\varepsilon > 0$,

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

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

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

(b) *Let $\{A^*_N : N \geq 1\}$ be a sequence of $L_A \times 1$ bootstrap random vectors that possesses an Edgeworth expansion with coefficients of order $O(1)$ and remainder of order $o(N^{-a})$ that holds except if $\{\chi_N : N \geq 1\}$ is in a sequence of sets with probability $o(N^{-a})$. (That is, for all $\varepsilon > 0$, $\lim_{N\to\infty} N^a P(N^a \sup_{z\in R^{L_A}} |P^*(A^*_N \leq z)$ $-[1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi^*_i(\partial/\partial z)]\Phi(z)| > \varepsilon) = 0$, where $\pi^*_i(\delta)$ are polynomials in $\delta = \partial/\partial z$ whose coefficients, $C^*_{N,\ell}$, satisfy: for all $\rho > 0$, there exists $K_\rho < \infty$ such that $\lim_{N\to\infty} N^a P(P^*(|C^*_{N,\ell}| > K_\rho) > \rho) = 0$ for all ℓ and all $i = 1, ..., 2a$, $\Phi(z)$ is the distribution function of a $N(0,\Sigma)$ random variable, and Σ is some nonsingular matrix.) Let $\{\xi^*_N : N \geq 1\}$ be a sequence of $L_A \times 1$ random vectors with $\lim_{N\to\infty} N^a P(N^a P^*(||\xi^*_N|| > \vartheta_N\varepsilon) > \varepsilon) = 0$ for all $\varepsilon > 0$ for some constants $\vartheta_N = O(N^{-a})$. Then, for all $\varepsilon > 0$,*

$$\lim_{N\to\infty} N^a P(N^a \sup_{z\in R^{L_A}} |P^*(A^*_N + \xi^*_N \leq z) - P^*(A^*_N \leq z)| > \varepsilon) = 0.$$

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

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

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

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

(d) *for some $K_\varepsilon < \infty$, $\lim_{N\to\infty} N^a P\left(N^a P^*(||N^{-1}\sum_{i=1}^N E^*h(\widetilde{X}^*_i)|| > K_\varepsilon) > \varepsilon\right) = 0$ and $\lim_{N\to\infty} N^a P\left(N^a P^*(||N^{-1}\sum_{i=1}^N h(\widetilde{X}^*_i)|| > K_\varepsilon) > \varepsilon\right) = 0$ even if $Eh(X_i) \neq 0$ and p only satisfies $p > 4a$ and $p \geq 2$.*

Lemma 7 *Suppose Assumptions 1–3 hold with $q_0 > 4a$ and $\ell \propto N^\gamma$ for some $0 \leq \gamma \leq 1$. Define $G^*_1(X^*_i, \theta) = g(X^*_i, \theta) - E^*g(X^*_i, \theta)$ and $G^*_2(X^*_i, \theta) = \rho(X^*_i, \theta) - E^*\rho(X^*_i, \theta)$. Then, for all $\varepsilon > 0$,*

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

Lemma 8 *Suppose Assumptions 1–3 hold with* $q_1 > 8a$ *and* $q_2 > 4a$. *Suppose* $\ell \propto N^\gamma$ *for some* $0 \leq \gamma \leq 1$. *Let* $\widehat{\theta}^*_N$ *denote any bootstrap estimator that satisfies: For all* $\varepsilon > 0$, $\lim_{N\to\infty} N^a P(N^a P^*(||\widehat{\theta}^*_N - \theta_0|| > \varepsilon) > \varepsilon) = 0$. *Then, for all* $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P\left(N^a P^*(||\Omega^*_N(\widehat{\theta}^*_N) - \Omega_0|| > \varepsilon) > \varepsilon\right) = 0.$$

Lemma 9 *Let* $\widehat{\theta}_N$ *denote the one-step GMM or minimum* ρ *estimator. Let* θ^*_N *denote the corresponding bootstrap estimator. Suppose Assumptions 1–3 hold with* $q_0 > 4a$, $q_1 > 4a/(1-2c)$, *and* $q_2 > 4a$. *Suppose* $\ell \propto N^\gamma$ *for some* $0 \leq \gamma \leq 1$. *Then, for all* $\varepsilon > 0$,

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

Lemma 10 *Let* $\widetilde{\theta}_N$ *and* $\widetilde{\theta}^*_N$ *denote the one-step GMM estimator and its bootstrap analogue based on the weight matrix* Ω. *Let* $\widehat{\theta}_N$ *and* θ^*_N *denote the two-step GMM estimator and its bootstrap analogue based on the weight matrices* $\Omega_N(\widetilde{\theta}_N)$ *and* $\Omega^*_N(\widetilde{\theta}^*_N)$ *respectively. Suppose Assumptions 1–3 hold with* $q_0 > 4a$, $q_1 > \max\{4a/(1-2c), 8a\}$, *and* $q_2 > 4a$. *Suppose* $\ell \propto N^\gamma$ *for some* $0 \leq \gamma \leq 1$. *Then, for all* $\varepsilon > 0$,

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

Lemma 11 *Suppose Assumptions 1–3 hold with* $q_0 > 4a$, $q_1 > 8a$, *and* $q_2 > 4a$. *Suppose* $\ell \propto N^\gamma$ *for some* $0 \leq \gamma \leq 1$. *Let* $\widetilde{\theta}^*_N$ *denote the bootstrap one-step GMM estimator based on the weight matrix* Ω. *Let* $\widehat{\theta}^*_N$ *denote any bootstrap estimator that satisfies: For all* $\varepsilon > 0$, $\lim_{N\to\infty} N^a P(N^a\ P^*(||\widehat{\theta}^*_N - \theta_0|| > \varepsilon) > \varepsilon) = 0$. *Then, for all* $\varepsilon > 0$,

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

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

$$\lim_{N\to\infty} N^a P\left(N^a P^*(||\frac{\partial^3}{\partial\theta^3} J^*_N(\widehat{\theta}^*_N, \widetilde{\theta}^*_N)|| > K_\varepsilon) > \varepsilon\right) = 0 \text{ for some } K_\varepsilon < \infty,$$

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

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

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

$$\lim_{N\to\infty} N^a P(N^a P^*(||m^*_N(\theta^*_{N,k}) - m^*_N(\theta^*_N)|| > \vartheta_N\varepsilon) > \varepsilon) = 0.$$

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

Lemma 13 *Let Δ_N and Δ_N^* denote $N^{1/2}(\widehat{\theta}_N - \theta_0)$ and $N^{1/2}(\theta_N^* - \widehat{\theta}_N)$, or T_N and $\widetilde{T}_N$, or $H_N(\widehat{\theta}_N)$ and $\widetilde{H}_N(\theta_N^*)$, or $K_N(\widehat{\theta}_N)$ and $\widetilde{K}_N(\theta_N^*)$ (where the statistics may be defined using one-step GMM, two-step GMM, or minimum ρ estimators in each case except the last, in which case $\widehat{\theta}_N$ and θ_N^* are two-step GMM estimators). For each definition of Δ_N and Δ_N^*, there is an infinitely differentiable function $G(\cdot)$ with $G(S) = 0$ and $G(S^*) = 0$ such that the following results hold.*
(a) *Suppose Assumptions* 1–4 *hold with* $q_0 > 2a$, $q_1 > \max\{2a/(1-2c), 4a\}$, $q_2 > 2a$, $d_1 \geq (2a+1)/(2c)$, *and* $2a$ *equal to an integer. Then,*

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

(b) *Suppose Assumptions* 1–4 *hold with* $q_0 > 4a$, $q_1 > \max\{4a/(1-2c),\ 4a/(1-\gamma),\ 8a\}$, $q_2 > 4a$, $d_1 \geq (2a+1)/(2c)$, $2a$ *equal to an integer, and* $0 < \gamma < 1$ (*and* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent*). *Then, for all* $\varepsilon > 0$,

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

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

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

The Edgeworth expansion of T_N^* depends on $\pi_i(\delta, \nu_{T,N,a}^*)$, where $\nu_{T,N,a}^* = \tau_{N,r}^\beta\widetilde{\nu}_{N,a}$ and β is some positive integer that depends on the bootstrap moment $\widetilde{\nu}_{N,a}$ being considered. The Edgeworth expansions of $\mathcal{W}_N^*$ and J_N^* depend on $\pi_{\mathcal{W}i}(y, \nu_{\mathcal{W},N,a}^*)$

and $\pi_{Ji}(y, \nu^*_{J,N,a})$, respectively, where $\nu^*_{\mathcal{W},N,a} = \lambda_{\mathcal{W}}(\Xi_N, \widetilde{\nu}_{N,a})$, $\lambda_{\mathcal{W}}(\cdot,\cdot)$ is a function that is continuously differentiable at (I_{L_η}, ν_a), $\lambda_{\mathcal{W}}(I_{L_\eta}, \nu_a) = \nu_a$, $\nu^*_{J,N,a} = \lambda_J((V_N^+)^{1/2}, \widetilde{\nu}_{N,a})$, $\lambda_J(\cdot,\cdot)$ is a function that is continuously differentiable at (M_0, ν_a), and $\lambda_J(M_0, \nu_a) = \nu_a$. Here, M_0 is the projection matrix $I_{L_g} - \Omega_0^{1/2} D(D'\Omega_0 D)^{-1} D'\Omega_0^{1/2}$, which is the probability limit of the correction factor $(V_N^+)^{1/2}$. The functions $\lambda_{\mathcal{W}}(\cdot,\cdot)$ and $\lambda_J(\cdot,\cdot)$ are determined by the effect of the correction factors Ξ_N and $(V_N^+)^{1/2}$ on the Edgeworth expansions of the bootstrap Wald and J statistics respectively.

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

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

Lemma 14 *Suppose Assumptions* 1 *and* 3 *hold with* $q_0 > 2a$, $q_1 > 2a/(1-2c)$, $q_2 \geq 2a+2$, $q_2 > 6a/(1-2\gamma-2\xi)$, $d_2 \geq -1 + (a+\gamma+\xi)/c$, $0 \leq \xi < 1/2 - \gamma$, *and either (i)* $\xi < \gamma$ *and* $0 < \gamma < 1/2$ *or (ii)* $\{X_i : i \geq 1\}$ *are independent. Then, for all* $\varepsilon > 0$,

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

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

Lemma 15 *Suppose Assumptions* 1–3 *hold with* $q_0 > 2a$, $q_1 > 2a/(1-2c)$, $q_2 > 4a/(1-\gamma-2\xi)$, *and* $d = d_1 + d_2 \geq -1 + (a+\gamma+\xi)/c$. *Suppose* $0 \leq \xi < 1/2 - \gamma/2$ *and* $0 < \gamma < 1/2$ *(where* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent). Suppose either* $\xi < \gamma$ *or* $\sum_{j=1}^{\kappa} j(Eg_1 g'_{1+j} + Eg_{1+j} g'_1) = 0$, *where* $g_i = g(X_i, \theta_0)$.
(a) *Then,*

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

(b) *If, in addition,* $q_2 \geq 2a+2$, $q_2 > 6a/(1-2\gamma-2\xi)$, $d_2 \geq -1 + (a+\gamma+\xi)/c$, *and* $0 \leq \xi < 1/2 - \gamma$, *then*

$$\lim_{N\to\infty} N^a P(N^\xi ||\nu^*_{s,N,a} - \nu_a|| > \varepsilon) = 0 \text{ for all } \varepsilon > 0 \text{ for } s = T, \mathcal{W}, J.$$

Lemma 16 (a) *Suppose Assumptions* 1–4 *hold with* $q_0 > 2a$, $q_1 > \max\{2a/(1-2c), 4a\}$, $q_2 \geq 2a+3$, $d_1 \geq (2a+1)/(2c)$, *and* $2a$ *equal to an integer. Then,*

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

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

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

(b) *Suppose Assumptions* 1–4 *hold with* $q_0 > 4a$, $q_1 > \max\{4a/(1-2c), 8a\}$, $q_2 \geq 2a+3$, $q_2 > 6a/(1-2\gamma)$, $d_1 \geq (2a+1)/(2c)$, $d_2 \geq -1 + (a+\gamma)/c$, $2a$ *equal to an* *integer, and* $0 < \gamma < 1/2$ *(where* $\gamma = 0$ *is permitted if* $\{X_i : i \geq 1\}$ *are independent). Then, for all* $\varepsilon > 0$,

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

$$\lim_{N\to\infty} N^a P\Bigg(N^a \sup_{z\in R} |P^*(\mathcal{W}_N^* \leq z) - \int_{-\infty}^{z} d[1 + \sum_{i=1}^{[a]} N^{-i}\pi_{\mathcal{W}i}(y, \nu_{\mathcal{W}Na}^*)]P(\chi^2_{L_\eta} \leq y)| > \varepsilon\Bigg) = 0, \text{ and}$$

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

(c) *Under the assumptions given, the results of part* (b) *also hold with* T_N^*, $\mathcal{W}_N^*$, *and* J_N^* *replaced by* $T_{N,k}^*$, $\mathcal{W}_{N,k}^*$, *and* $J_{N,k}^*$, *provided* $\mu_{N,k} = O(N^{-(a+1/2)})$ *for* $\mu_{N,k}$ *as in* (3.1).

5.2 Proofs of Theorems

5.2.1 Proof of Theorem 1

We establish part (a) first. To start, suppose θ_N^* is the one-step GMM estimator. A Taylor expansion about $\theta_{N,k-1}^*$ gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta} J_N^*(\theta_N^*) \\
&= \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*) + \frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1}^*)(\theta_N^* - \theta_{N,k-1}^*) + R_{N,k}^* \\
&= \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*) + Q_{N,k-1}^*(\theta_{N,k}^* - \theta_{N,k-1}^*) + Q_{N,k-1}^*(\theta_N^* - \theta_{N,k}^*) \\
&\quad + (\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1}^*) - Q_{N,k-1}^*)(\theta_N^* - \theta_{N,k-1}^*) + R_{N,k}^* \\
&= Q_{N,k-1}^*(\theta_N^* - \theta_{N,k}^*) + (\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1}^*) - Q_{N,k-1}^*)(\theta_N^* - \theta_{N,k-1}^*) + R_{N,k}^*, \text{ where}
\end{aligned}$$

$$R_{N,k}^* = \left[(\theta_N^* - \theta_{N,k-1}^*)' \frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1,u}^+)(\theta_N^* - \theta_{N,k-1}^*)/2\right]_{L_\theta}, \tag{5.1}$$

$[\xi_u]_{L_\theta}$ denotes an L_θ vector whose u-th element is ξ_u, $\theta_{N,k-1,r}^+$ lies between θ_N^* and $\theta_{N,k-1}^*$, the first equality holds with P^*–probability $1 - o(N^{-a})$ on a set with P–probability $1 - o(N^{-a})$ by Lemma 9, and the fourth equality holds because

$(\partial/\partial\theta)J_N^*(\theta_{N,k-1}^*) + Q_{N,k-1}^* \ (\theta_{N,k}^* - \theta_{N,k-1}^*) = 0$ by the definition of $\theta_{N,k}^*$. Rearranging (5.1) yields

$$\begin{aligned}
&||\theta_{N,k}^* - \theta_N^*|| \\
&\leq ||(Q_{N,k-1}^*)^{-1}R_{N,k}^*|| + ||(Q_{N,k-1}^*)^{-1}(\frac{\partial^2}{\partial\theta\partial\theta'}J_N^*(\theta_{N,k-1}^*) - Q_{N,k-1}^*)(\theta_{N,k-1}^* - \theta_N^*)|| \\
&\leq \zeta_N^*(||\theta_{N,k-1}^* - \theta_N^*||^2 + N^{-c_Q}||\theta_{N,k-1}^* - \theta_N^*||), \text{ where}
\end{aligned}$$

$$\begin{aligned}
\zeta_N^* = \max_{j=1,\ldots,k}\{&||(Q_{N,j-1}^*)^{-1})|| \cdot \sum_{u=1}^{L_\theta}||\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J_N^*(\theta_{N,j-1,u}^+)/2|| \\
&+||(Q_{N,j-1}^*)^{-1})|| \cdot N^{c_Q}||\frac{\partial^2}{\partial\theta\partial\theta'}J_N^*(\theta_{N,j-1}^*) - Q_{N,j-1}^*|| + 1\} \qquad (5.2)
\end{aligned}$$

As in Robinson (1988, Pf. of Thm. 5), repeated substitution into the right-hand side of the inequality gives an upper bound that is a finite sum of terms with dominant terms of the form:

$$C(\zeta_N^*)^\phi||\theta_{N,0}^* - \theta_N^*||^{2^{k-j}}N^{-jc_Q} \text{ for } j = 0, ..., k, \qquad (5.3)$$

where ϕ is a positive integer. To see this, consider the solution in terms of x_0 of the equation $x_k = x_{k-1}^2 + \lambda x_{k-1}$. Collect all terms in powers of λ that are multiplied by the smallest number of x_0 terms.

Because $\theta_{N,0}^* = \widehat{\theta}_N$, an upper bound on the right–hand side of (5.2) is

$$C(\zeta_N^*)^\phi \max_{j=0,\ldots,k}(\lambda_N^*)^{2^{k-j}}N^{-(2^{k-j}c+jc_Q)}, \text{ where } \lambda_N^* = N^c||\widehat{\theta}_N - \theta_N^*||. \qquad (5.4)$$

By Lemma 9, for all $\varepsilon > 0$, $\lim_{N\to\infty} N^a \ P(N^aP^*(\lambda_N^* > \varepsilon) > \varepsilon) = 0$. In addition, by Lemma 11 and Assumptions 3(b) and 5, for all $\varepsilon > 0$, there exists a finite constant K_ε such that $\lim_{N\to\infty} N^aP\left(N^aP^*\left(\zeta_N^* > K_\varepsilon\right) > \varepsilon\right) = 0$. Combining these results with (5.2) and (5.4) gives: For all $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{N\to\infty} N^aP\left(N^aP^*\left(||\theta_{N,k}^* - \theta_N^*|| > \max_{j=0,\ldots,k} N^{-(2^{k-j}c+jc_Q)}\varepsilon\right) > \varepsilon\right) \\
&\leq \lim_{N\to\infty} N^aP\left(N^aP^*\left(C(\zeta_N^*)^\phi\lambda_N^* > \varepsilon\right) > \varepsilon\right) = 0. \qquad (5.5)
\end{aligned}$$

Hence, part (a) of the Theorem holds for the one-step GMM estimator.

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

The proof of part (a) for the case where θ_N^* is the bootstrap two-step GMM estimator is similar to that given above with $J_N^*(\theta)$ replaced by $J_N^*(\theta, \widetilde{\theta}_N^*)$ or $J_N^*(\theta, \widetilde{\theta}_{N,k_1}^*)$ in the appropriate place and with reference to Lemma 9 replaced by reference to

Lemma 10. However, two additional terms arise on the right-hand side of (5.1) because $J_N^*(\theta, \widetilde{\theta}_N^*) \neq J_N^*(\theta, \widetilde{\theta}_{N,k_1}^*)$. These terms are

$$M_{1,N}^* = (\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1}^*, \widetilde{\theta}_N^*) - \frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,k-1}^*, \widetilde{\theta}_{N,k_1}^*))(\theta_N^* - \theta_{N,k-1}^*) \text{ and}$$
$$M_{2,N}^* = \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*, \widetilde{\theta}_N^*) - \frac{\partial}{\partial\theta} J_N^*(\theta_{N,k-1}^*, \widetilde{\theta}_{N,k_1}^*). \tag{5.6}$$

These terms can be shown to satisfy

$$\lim_{N\to\infty} N^a P(P^*(||M_{j,N}^*|| > \mu_N) > N^{-a}) = 0 \text{ for } j = 1, 2. \tag{5.7}$$

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

To prove (5.7), we first show that

$$\lim_{N\to\infty} N^a P(P^*(||\Omega_N^*(\widetilde{\theta}_{N,k_1}^*)^{-1} - \Omega_N^*(\widetilde{\theta}_N^*)^{-1}|| > \mu_N) > N^{-a}) = 0 \tag{5.8}$$

using Lemma 12 with $m_N^*(\theta) = \Omega_N^*(\theta)^{-1}$, $\theta_N^* = \widetilde{\theta}_N^*$, $\theta_{N,k}^* = \widetilde{\theta}_{N,k_1}^*$, and $\vartheta_N = \mu_N$. The conditions of Lemma 12 are verified using the result of part (a) of the Theorem for the bootstrap one-step GMM estimator, the assumption that $k_1 \geq k$, and Lemma 9. The proof of (5.7) also uses the first and second results of Lemma 11 with $\widehat{\theta}_N^* = \theta_{N,k-1}^*$, where the condition on $\widehat{\theta}_N^*$ holds by applying the proof of part (a) of the Theorem for the k-step bootstrap two-step GMM estimator recursively for $k = 1, 2, \ldots$. The proof of (5.7) also uses $\lim_{N\to\infty} P(P^*(||\theta_N^* - \theta_{N,k-1}^*|| > K) > N^{-a}) = 0$ for some $0 < K < \infty$, which holds by applying the current proof recursively because $K \geq \mu_N$ for N large.

For the first result of part (b), when $\gamma > 0$, we use (5.103) of the proof of Lemma 15(b) with $\xi = 0$ (which guarantees that $\tau_{N,r}$ is well behaved and is responsible for the additional conditions imposed in part (b)). Let $\widetilde{T}_N$ and $\widetilde{T}_{N,k}$ denote T_N^* and $T_{N,k}^*$, respectively, with the correction factor $\tau_{N,r}$ deleted. Let $\sigma_{k,r}^*$ and σ_r^* denote $\sigma_N^*(\theta_{N,k}^*)_{rr}$ and $\sigma_N^*(\theta_N^*)_{rr}$ respectively. For $\gamma \geq 0$, we use the following:

$$\begin{aligned}|\widetilde{T}_{N,k} - \widetilde{T}_N| \leq\ & N^{1/2}||\theta_{N,k}^* - \theta_N^*||/(\sigma_{k,r}^*)^{1/2} \\ & + N^{1/2}||\theta_N^* - \widehat{\theta}_N|| \cdot |(\sigma_{k,r}^*)^{1/2} - (\sigma_r^*)^{1/2}|/(\sigma_{k,r}^*\sigma_r^*)^{1/2}.\end{aligned} \tag{5.9}$$

By (5.9) and the result concerning $\tau_{N,r}$, the first result of part (b) is implied by part (a) plus the following: For all $\varepsilon > 0$, there exists a $K_\varepsilon < \infty$ and a $\delta > 0$ such that

$$\lim_{N\to\infty} N^a P(N^a P^*(|(\sigma_{k,r}^*)^{1/2} - (\sigma_r^*)^{1/2}| > \mu_{N,k}\varepsilon) > \varepsilon) = 0, \tag{5.10}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(||\theta_N^* - \widehat{\theta}_N|| > K_\varepsilon) > \varepsilon) = 0, \tag{5.11}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(\sigma_{k,r}^* < \delta) > \varepsilon) = 0, \text{ and} \tag{5.12}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(\sigma_r^* < \delta) > \varepsilon) = 0. \tag{5.13}$$

Equation (5.11) holds by Lemma 9 or 10. Equations (5.12) and (5.13) hold by Lemmas 8–11 and part (a) of the Theorem.

Equation (5.10) is implied by (5.12), (5.13), and

$$\lim_{N\to\infty} N^a P(N^a P^*(|\sigma^*_{k,r} - \sigma^*_r| > \mu_{N,k}\varepsilon) > \varepsilon) = 0 \tag{5.14}$$

by applying the mean value theorem. Equation (5.14) is implied by

$$\lim_{N\to\infty} N^a P(N^a P^*(||D^*_N(\theta^*_{N,k}) - D^*_N(\theta^*_N)|| > \mu_{n,k}\varepsilon) > \varepsilon) = 0 \text{ and}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega^*_N(\theta^*_{N,k})^{-1} - \Omega^*_N(\theta^*_N)^{-1}|| > \mu_{n,k}\varepsilon) > \varepsilon) = 0. \tag{5.15}$$

These results hold by Lemma 12 with $\vartheta_N = \mu_{N,k}$ using part (a) of the Theorem and Assumption 3.

Let $\widetilde{K}_N(\theta)$ denote $K^*_N(\theta)$ with the correction factor $(V^+_N)^{1/2}$ deleted. The third result of part (b) is implied by

$$\lim_{N\to\infty} N^a P(P^*(||\widetilde{K}_N(\theta^*_{N,k}) - \widetilde{K}_N(\theta^*_N)|| > N^{1/2}\mu_{n,k}\varepsilon) > \varepsilon) = 0 \text{ and}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(||\widetilde{K}_N(\theta^*_N)|| > K_\varepsilon) > \varepsilon) = 0 \tag{5.16}$$

for all $\varepsilon > 0$ and some $K_\varepsilon < \infty$, and (5.105) of the proof of Lemma 15 with $\xi = 0$ (which guarantees that V^+_N is well behaved). The first result of (5.16) is implied by the second result of (5.15) and the following: For all $\varepsilon > 0$, there exists $K_\varepsilon < \infty$ such that

$$\lim_{N\to\infty} N^a P(N^a P^*(||N^{-1}\sum_{i=1}^N g^*(X^*_i, \theta^*_{N,k}) - N^{-1}\sum_{i=1}^N g^*(X^*_i, \theta^*_N)|| > \mu_{n,k}\varepsilon) > \varepsilon) = 0,$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||N^{-1/2}\sum_{i=1}^N g^*(X^*_i, \theta^*_N)|| > K_\varepsilon) > \varepsilon) = 0,$$
$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega^*_N(\theta^*_N)|| > K_\varepsilon) > \varepsilon) = 0, \text{ and}$$
$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega^*_N(\theta^*_{N,k})|| > K_\varepsilon) > \varepsilon) = 0. \tag{5.17}$$

The first result of (5.17) holds by Lemma 12 with $\vartheta_N = \mu_{n,k}$. The third and fourth results of (5.17) hold by Lemma 8 and part (a) of the Theorem. The second result of (5.17) holds by a mean value expansion about $\widehat{\theta}_N$, using the fact that $N^{-1/2}\sum_{i=1}^N g^*(X^*_i, \widehat{\theta}_N)$ equals zero for the non-overlapping bootstrap and is sufficiently close to zero for the overlapping bootstrap, and using Lemma 12 and Theorem 1(a) to handle the second term in the mean value expansion. The second result of (5.16) is implied by the second and third results of (5.17).

The second result of part (b) is proved analogously to the third result but with (5.105) replaced by (5.104).

To establish part (c) of the Theorem, we apply Lemma 5(b) four times with $\vartheta_N = N^{1/2}\mu_{N,k} = O(N^{-a})$ and with (A_N^*, ξ_N^*) equal to $(N^{1/2}(\theta_N^* - \widehat{\theta}_N), N^{1/2}(\theta_{N,k}^* - \theta_N^*))$, $(T_N^*,\ T_{N,k}^* - T_N^*)$, $(H_N^*(\theta_N^*),\ H_N^*(\theta_{N,k}^*)\ - H_N^*(\theta_N^*))$, and $(K_N^*(\theta_N^*),\ K_N^*(\theta_{N,k}^*)\ - K_N^*(\theta_N^*))$. In the third and fourth cases, the result of Lemma 5(b) implies the third and fourth results of part (c) of the Theorem by a straightforward argument. The condition of Lemma 5(b) on ξ_N^* holds by parts (a) and (b) of the Theorem for the first two applications of Lemma 5(b), by (5.16) for the fourth application, and by the analogue of (5.16) for the Wald statistic for the third application. As required by Lemma 5(b), the random variables T_N^*, $H_N^*(\theta_N^*)$, and $K_N^*(\theta_N^*)$ have Edgeworth expansions with remainder $o(N^{-a})$ by Lemma 16(b) using the additional conditions on q_2, d_1, d_2, γ, and a in part (c). Lemma 16(b) does not state an Edgeworth expansion for $N^{1/2}(\theta_N^* - \widehat{\theta}_N)$, but one can be obtained under the same assumptions and by the same argument as for T_N^*.

The NR result of part (d) holds by definition of $Q_{N,j-1}^{*,NR}$. We now establish the default NR result of part (d). Let $\theta_{N,j}^*$ denote the NR one-step GMM estimator for $j = 1, ..., k$. For the one-step GMM estimator, it suffices to show that

$$\lim_{N\to\infty} N^a P(N^a P^*(J_N^*(\theta_{N,j}^*) - J_N^*(\theta_{N,j-1}^*) > 0) > \varepsilon) = 0, \tag{5.18}$$

for all $\varepsilon > 0$, for all $j = 1, ..., k$, because this implies that $\lim_{N\to\infty} N^a P(N^a P^*(Q_{N,j-1}^{*,D} \neq Q_{N,j-1}^{*,NR}$ for some $j = 1, ..., k) > \varepsilon) = 0$. When $\theta_{N,j}^* \neq \theta_{N,j-1}^*$, Taylor expansion of $J_N^*(\theta_{N,j}^*)$ about $\theta_{N,j-1}^*$ gives

$$\begin{aligned}
& J_N^*(\theta_{N,j}^*) - J_N^*(\theta_{N,j-1}^*) \\
&= \frac{\partial}{\partial\theta'} J_N^*(\theta_{N,j-1}^*)\zeta_{N,j}^*\phi_{N,j}^* + \frac{1}{2}\zeta_{N,j}^{*\prime}\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^*)\zeta_{N,j}^*(\phi_{N,j}^*)^2 + \Gamma_{N,j}^*(\phi_{N,j}^*)^3 \\
&= -\frac{1}{2}\zeta_{N,j}^{*\prime}\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^*)\zeta_{N,j}^*(\phi_{N,j}^*)^2 + \Gamma_{N,j}^*(\phi_{N,j}^*)^3, \text{ where} \\
\Gamma_{N,j}^* &= \frac{1}{6}\sum_{u=1}^{L_\theta} \zeta_{N,j,u}^*\zeta_{N,j}^{*\prime}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^+)\zeta_{N,j}^*, \\
\zeta_{N,j}^* &= (\theta_{N,j}^* - \theta_{N,j-1}^*)/||\theta_{N,j}^* - \theta_{N,j-1}^*||,\ \phi_{N,j}^* = ||\theta_{N,j}^* - \theta_{N,j-1}^*||,
\end{aligned} \tag{5.19}$$

$\zeta_{N,j,u}^*$ denotes the u-th element of $\zeta_{N,j}^*$, and $\theta_{N,j-1}^+$ lies between $\theta_{N,j}^*$ and $\theta_{N,j-1}^*$. The second equality holds by the definition of $\theta_{N,j}^*$. Using (5.19), the left-hand side of (5.18) is less than or equal to

$$\lim_{N\to\infty} N^a P(N^a P^*(-\lambda_{\min}(\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^*))/2 + \Gamma_{N,j}^*\phi_{N,j}^* > 0) > \varepsilon). \tag{5.20}$$

The latter equals zero, because for $\delta = \lambda_{\min}(D'\Omega D)/2 > 0$,

$$\begin{aligned}
& \lim_{N\to\infty} N^a P(N^a P^*(\lambda_{\min}(\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^*)) < \delta) > \varepsilon) = 0, \\
& \lim_{N\to\infty} N^a P(N^a P^*(|\Gamma_{N,j}^*| > K_\varepsilon) > \varepsilon) = 0 \text{ for some } K_\varepsilon < \infty, \text{ and} \\
& \lim_{N\to\infty} N^a P(N^a P^*(\phi_{N,j}^* > \varepsilon) > \varepsilon) = 0,
\end{aligned} \tag{5.21}$$

for all $\varepsilon > 0$, where the first result holds by Lemma 11 and Assumption 3(b), the second holds by Lemma 11, and the third holds by part (a) of the Theorem with $k = j - 1$ and $k = j$ for the NR one-step GMM estimator. This completes the proof for the one-step GMM estimator. The proofs for the two-step and minimum ρ estimators are analogous.

We now establish the line-search NR result of part (d). We consider the one-step GMM estimator first. Let $\theta^*_{N,j}$ be the NR j-step estimator:

$$\begin{aligned}
\theta^*_{N,j} &= \theta^*_{N,j-1} - \psi^*_{N,j-1}\pi^*_{N,j-1}, \text{ where} \\
\psi^*_{N,j-1} &= ||(Q^{*,NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j-1})|| \text{ and } \pi^*_{N,j-1} = (Q^{*,NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j-1})/\psi^*_{N,j-1}.
\end{aligned} \tag{5.22}$$

Let

$$\theta^{*,\alpha}_{N,j} = \theta^*_{N,j-1} - \alpha(Q^{*,NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j-1}) = \theta^*_{N,j} + (1-\alpha)\psi^*_{N,j-1}\pi^*_{N,j-1}. \tag{5.23}$$

It suffices to show that

$$\lim_{N\to\infty} N^a P(N^a P^*(\inf_{\alpha\in A,\alpha\neq 1} J^*_N(\theta^{*,\alpha}_{N,j}) - J^*_N(\theta^*_{N,j}) < 0) > \varepsilon) = 0 \tag{5.24}$$

for all $j = 1, ..., k$, because this implies that $\lim_{N\to\infty} N^a P(N^a P^*(Q^{*,LS}_{N,j-1} \neq Q^{*,NR}_{N,j-1}$ for some $j = 1, ..., k) > \varepsilon) = 0$.

A Taylor expansion of $J^*_N(\theta^{*,\alpha}_{N,j})$ about $\theta^*_{N,j}$ gives

$$\begin{aligned}
& J^*_N(\theta^{*,\alpha}_{N,j}) - J^*_N(\theta^*_{N,j}) = (1-\alpha)\psi^*_{N,j-1}\pi^{*\prime}_{N,j-1}\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j}) \\
& +\frac{1}{2}(1-\alpha)^2(\psi^*_{N,j-1})^2\pi^{*\prime}_{N,j-1}\frac{\partial^2}{\partial\theta\partial\theta'}J^*_N(\theta^*_{N,j})\pi^*_{N,j-1} \\
& +\frac{1}{6}(1-\alpha)^3(\psi^*_{N,j-1})^3\sum_{u=1}^{L_\theta}\pi^*_{N,j-1,u}\pi^{*\prime}_{N,j-1}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J^*_N(\theta^+_{N,j})\pi^*_{N,j-1},
\end{aligned} \tag{5.25}$$

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

Taylor expansions of $(\partial/\partial\theta)J^*_N(\theta^*_{N,j})$ about $\theta^*_{N,j-1}$ give

$$\begin{aligned}
\frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j}) &= \frac{\partial}{\partial\theta}J^*_N(\theta^*_{N,j-1}) + \frac{\partial^2}{\partial\theta\partial\theta'}J^*_N(\theta^*_{N,j-1})(\theta^*_{N,j} - \theta^*_{N,j-1}) \\
& +\frac{1}{2}[(\theta^*_{N,j} - \theta^*_{N,j-1})'\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J^*_N(\theta^{++}_{N,j-1,u})(\theta^*_{N,j} - \theta^*_{N,j-1})]_{L_\theta} \\
&= 0 + \frac{1}{2}(\psi^*_{N,j-1})^2[\pi^{*\prime}_{N,j-1}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J^*_N(\theta^{++}_{N,j-1,u})\pi^*_{N,j-1}]_{L_\theta},
\end{aligned} \tag{5.26}$$

where $\theta^{++}_{N,j-1,u}$ lies between $\theta^*_{N,j}$ and $\theta^*_{N,j-1}$ and the second equality holds using the definition of $\theta^*_{N,j}$.

The following properties hold: For $\delta = \lambda_{\min}(D'\Omega D)/2 > 0$ and all $\varepsilon > 0$,

$$\begin{aligned}
&\lim_{N\to\infty} N^a P(N^a P^*(\lambda_{\min}(\frac{\partial^2}{\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^*)) < \delta) > \varepsilon) = 0, \\
&\lim_{N\to\infty} N^a P(N^a P^*(||[\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'} J_N^*(\theta_{N,j-1}^{++})]_{L_\theta}|| > K_\varepsilon) > \varepsilon) = 0 \text{ for some } K_\varepsilon < \infty, \\
&\lim_{N\to\infty} N^a P(N^a P^*(\psi_{N,j}^* > \varepsilon) > \varepsilon) = 0 \qquad (5.27)
\end{aligned}$$

for $j = 1, ..., k$, where the first result holds by Lemma 11 and Assumption 3(b), the second holds by Lemma 11, and the third holds by Lemma 11 and Assumption 3(b) to ensure that $(Q_{N,j}^{*,NR})^{-1}$ is well-behaved and by a mean value expansion of $(\partial/\partial\theta)J_N^*(\theta_{N,j-1}^*)$ about θ_N^* and application of part (a) of the Theorem with $k = j-1$ and Lemma 11. The second result of (5.27) also holds with $\theta_{N,j-1}^{++}$ replaced by $\theta_{N,j-1}^{+}$.

Substituting (5.26) into the right-hand side of (5.25) and dividing (5.25) by $(\psi_{N,j-1}^*)^2$ (when $\psi_{N,j-1}^* > 0$) yields the resultant first and third terms on the right-hand side of (5.25) to be asymptotically negligible and the second term to be strictly positive with appropriate probability asymptotically (uniformly over $\alpha \in A$ with $\alpha \neq 1$) using (5.27), which gives (5.24). This completes the proof for the one-step GMM estimator. The proofs for the two-step and minimum ρ estimators are analogous.

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

$$\lim_{N\to\infty} N^a P(N^a P^*(||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{X}_i^*, \theta_{N,j-1}^*) - \frac{\partial}{\partial\theta'} g(X_i^*, \theta_{N,j-1}^*))|| > N^{-c}\varepsilon) > \varepsilon) = 0 \qquad (5.28)$$

for all $\varepsilon > 0$. By mean value expansions about θ_0 and the triangle inequality,

$$\begin{aligned}
&||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{X}_i^*, \theta_{N,j-1}^*) - \frac{\partial}{\partial\theta'} g(X_i^*, \theta_{N,j-1}^*))|| \\
\leq\ &||N^{-1}\sum_{i=1}^N(\Delta(\widetilde{X}_i^*, \theta_0) - \frac{\partial}{\partial\theta'} g(X_i^*, \theta_0))|| \\
&+ N^{-1}\sum_{i=1}^N \sup_{\theta\in N_0, u\leq L_\theta} ||\frac{\partial}{\partial\theta_u}(\Delta(\widetilde{X}_i^*, \theta) - \frac{\partial^2}{\partial\theta_u\partial\theta'} g(X_i^*, \theta))|| \cdot ||\theta_{N,j-1}^* - \theta_0||. \qquad (5.29)
\end{aligned}$$

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

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

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

$> K_\varepsilon) > \varepsilon) = 0,$

(iii) $\lim_{N\to\infty} N^a P(N^a P^*(||\theta^*_{N,j-1} - \theta^*_N|| > N^{-c}\varepsilon) > \varepsilon) = 0,$

(iv) $\lim_{N\to\infty} N^a P(N^a P^*(||\theta^*_N - \widehat{\theta}_N|| > N^{-c}\varepsilon) > \varepsilon) = 0,$ and

(v) $\lim_{N\to\infty} N^a P(N^a P^*(||\widehat{\theta}_N - \theta_0|| > N^{-c}\varepsilon) > \varepsilon) = 0 \quad (5.30)$

for $j = 1, ..., k$, for all $\varepsilon > 0$, and for some $K_\varepsilon < \infty$. Condition (i) holds by Lemma 6(a) with $p = \min\{q_2, q_3\}$ and c as in the statement of the Theorem, (ii) holds by Lemma 6(d) with $p = \min\{q_2, q_4\}$, (iv) holds by Lemma 9 or Lemma 10, (v) holds by Lemma 3 or 4, (iii) holds for $j = 1$ by Lemma 9 or Lemma 10 because $\theta^*_{N,0} = \widehat{\theta}_N$, and (iii) holds for $j = 2, ..., k$ by recursively applying part (a) of the Theorem with $k = j - 1$, which holds without assuming Assumption 5 by the present proof that the result of Assumption 5 holds for $Q^{*,GN}_{N,i}$ for $i \leq j - 1$ under the assumptions. □

5.2.2 Proof of Theorem 2

We establish the first result of part (a) of the Theorem first. By Theorem 1(c), Lemma 16(b), Lemma 14, and Lemma 16(a), respectively, each with $a = 3/2$, we have: for all $\varepsilon > 0$,

$$\lim_{N\to\infty} N^{3/2} P(N^{3/2} \sup_{z\in R} \left|P^*(|T^*_{N,k}| \leq z) - P^*(|T^*_N| \leq z)\right| > \varepsilon) = 0,$$

$$\lim_{N\to\infty} N^{3/2} P(N^{3/2} \sup_{z\in R} |P^*(|T^*_N| \leq z) - [1 + N^{-1}\pi_2(\delta, \nu^*_{T,N,3/2})](\Phi(z) - \Phi(-z))| > \varepsilon) = 0,$$

$$\lim_{N\to\infty} N^{3/2} P(N^{\xi} \sup_{z\in R} |(\pi_2(\delta, \nu^*_{T,N,3/2}) - \pi_2(\delta, \nu_{3/2}))(\Phi(z) - \Phi(-z))| > \varepsilon) = 0, \text{ and}$$

$$\lim_{N\to\infty} N^{3/2} \sup_{z\in R} \left|P(|T_N| \leq z) - (1 + N^{-1}\pi_2(\delta, \nu_{3/2}))(\Phi(z) - \Phi(-z))\right| = 0, \quad (5.31)$$

using the evenness of $\pi_j(\delta, \nu^*_{T,N,3/2})\Phi(z)$ and $\pi_j(\delta, \nu_{3/2})\Phi(z)$ in z for $j = 1, 3$ in the second and fourth results respectively. The results of (5.31) combine to give

$$\lim_{N\to\infty} N^{3/2} P(N^{1+\xi} \sup_{z\in R} |P^*(|T^*_{N,k}| \leq z) - P(|T_N| \leq z)| > \varepsilon) = 0. \quad (5.32)$$

Let $F_{|T|}(\cdot)$ denote the distribution function of $|T_N|$. Equation (5.32) yields

$$\lim_{N\to\infty} N^{3/2} P(N^{1+\xi} |1 - \alpha - F_{|T|}(z^*_{|T|,k,\alpha})| > \varepsilon) = 0. \quad (5.33)$$

Using (5.33), we have

$$\begin{aligned} & P(|T_N| > z^*_{|T|,k,\alpha}) \\ = \; & P(F_{|T|}(|T_N|) > F_{|T|}(z^*_{|T|,k,\alpha}),\; N^{1+\xi}|1 - \alpha - F_{|T|}(z^*_{|T|,k,\alpha})| \leq \varepsilon) \\ & + P(F_{|T|}(|T_N|) > F_{|T|}(z^*_{|T|,k,\alpha}),\; N^{1+\xi}|1 - \alpha - F_{|T|}(z^*_{|T|,k,\alpha})| > \varepsilon) \quad (5.34) \\ \leq \; & P(F_{|T|}(|T_N|) > 1 - \alpha - \varepsilon/N^{1+\xi}) + o(N^{-3/2}) \\ \leq \; & \alpha + \varepsilon/N^{1+\xi} + o(N^{-3/2}), \end{aligned}$$

where the last inequality holds because $F_{|T|}(|T_N|)$ has a uniform $(0,1)$ distribution. (If $|T_N|$ is not absolutely continuous, then the Edgeworth expansion for $|T_N|$ in (5.31) is used to obtain the last inequality.) Equation (5.34) also holds with the inequalities reversed and "$-\varepsilon/N^{1+\xi}$" and "$+\varepsilon/N^{1+\xi}$" interchanged. This establishes the first result of part (a) of the Theorem.

The proof of the second result of part (a) is the same as for the first result but with $T^*_{N,k}$ and $z^*_{|T|,k,\alpha}$ replaced by T^*_N and $z^*_{|T|,\alpha}$ in (5.32)–(5.34). The proofs of the remaining results of part (a) are analogous, using the appropriate results from Theorem 1(c) and Lemma 16(a)-(b).

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

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

5.2.3 Proof of Theorem 3

The proofs of parts (a), (b), (c), and (d) of the Theorem are the same as those of Theorem 1(a), (b), (d), and (e), Theorem 1(c), Theorem 2(a), and Theorem 2(b), respectively, with the following changes. First, we use parametric bootstrap analogues of Lemmas 6–16, excluding Lemmas 10 and 15, which are not needed because the (unrestricted) ML estimator is a one-step GMM estimator and no correction factor is used with the parametric bootstraps. Lemmas 1–5 hold as stated because they do not involve the bootstrap. Second, the parts of the proofs that deal with the correction factor are deleted, because no correction factor is needed. Third, the parts of the proofs that deal with the J–statistic are deleted, because it is not considered.

It remains to establish that the Lemmas referred to above hold when the unrestricted and restricted parametric bootstraps are used in place of the block bootstrap. By the following argument, Lemma 6 holds under the conditions given (except that no condition on γ is needed) plus the conditions of Lemma 3 (which are needed to guarantee suitable behavior of the estimators $\widehat{\theta}_N$ and $\overline{\theta}_N$). Equation (5.46) of the Proof of Lemma 6 holds without change. Next, the Yokoyama-Doukhan strong mixing moment inequality (see the Proof of Lemma 1) applied to the unrestricted parametric bootstrap sample, conditional on the original sample, gives

$$E^*_{\widehat{\theta}_N}||N^{-1}\sum_{i=1}^{N}h(\widetilde{X}^*_i)-E^*_{\widehat{\theta}_N}h(\widetilde{X}^*_i)||^p\leq CN^{-p/2}, \tag{5.35}$$

where $E^*_{\widehat{\theta}_N}$ denotes expectation with respect to the unrestricted parametric bootstrap sample (which is generated using $\widehat{\theta}_N$). The inequality also holds for the restricted parametric bootstrap sample with $E^*_{\widehat{\theta}_N}$ replaced by $E^*_{\overline{\theta}_N}$. The constant C does not depend on $\widehat{\theta}_N$ or $\overline{\theta}_N$, because (i) Assumption 1 holds with the same constants K and ξ for all $\theta_1\in N_0$, (ii) Assumption 3 holds with the same functions $C_g(\cdot)$ and $C_{\partial f}(\cdot)$ and the same constant C_f for all $\theta_1\in N_0$, and (iii) $\widehat{\theta}_N\in N_0$ with probability $1-o(N^{-a})$ by Lemma 3 and $\overline{\theta}_N\in N_0$ with probability $1-o(N^{-a})$ by an analogous result. Given Lemma 6, the proofs of Lemmas 7–13 and 16(a) go through as stated under the same assumptions as for the block bootstrap.

Lemma 16(b) holds under the same conditions as stated for Lemma 16(a) plus Assumptions 6–8. The proof of Lemma 16(b) is the same as that of 16(a) conditional on $\widehat{\theta}_N$ or $\overline{\theta}_N$, because Assumptions 6–8 imply that conditional on $\widehat{\theta}_N$ or $\overline{\theta}_N$ the bootstrap sample has the same distribution as the original sample when the latter is generated with true value $\theta_1=\widehat{\theta}_N$ or $\theta_1=\overline{\theta}_N$.

The result of Lemma 14 holds for any $0\leq\xi<1/2$ provided (i) the conditions of Lemma 3 hold for $c=\xi$, which requires that $q_1>2a/(1-2\xi)$, and (ii) Assumption 3 holds with $q_2\geq 2a+3$. Lemma 14 is established by showing that for all $\theta_1\in\Theta$:

$$|N^{\alpha(m)}E_{\theta_1}\prod_{\mu=1}^{m}\Psi^*_{N,j_\mu}-N^{\alpha(m)}E_{\theta_0}\prod_{\mu=1}^{m}\Psi_{N,j_\mu}|\leq B_N||\theta_1-\theta_0||, \tag{5.36}$$

where $\limsup_{N\to\infty}B_N<\infty$ and E_θ denotes expectation taken when the true parameter value is θ. This is coupled with the result of Lemma 3 with $c=\xi$ for the unrestricted parametric bootstrap and an analogous result for $\overline{\theta}_N$ for the restricted parametric bootstrap. Under the assumptions, (5.36) holds provided

$$E_{\theta_1}\prod_{\mu=1}^{m}f_{j_\mu}(\widetilde{X}_i,\theta_1) \tag{5.37}$$

satisfies a Lipschitz condition in θ_1 at θ_0 for all $m\leq 2a+2$, where $f_{j_\mu}(\widetilde{X}_i,\theta_1)$ denotes the j_μ–th element of $f(\widetilde{X}_i,\theta_1)$. The triangle inequality, a mean-value expansion, and some calculations show that the latter holds if

$$E_{\theta_1}||C^j_{\partial f}(\widetilde{X}_i)f^{2a+3-j}_{j_\mu}(\widetilde{X}_i,\theta_1)||<\infty \text{ for all } j=0,...,2a+2, \tag{5.38}$$

$\theta_1\in N_0$, and all elements j_μ of $f(\widetilde{X}_i,\theta_1)$. This holds if $q_2\geq 2a+3$.

Hence, in parts (c) and (d) of the Theorem, we must have $q_2\geq 2a+3$ and $q_2>4a$ (where the latter is used in Lemma 9). Taking $a=3/2$ in part (c) and $a=1$ in part (d) gives conditions stated in the Theorem for q_2. In addition, we must have

$q_1 > 2a/(1-2\xi)$. No condition on d_2 is needed in the Theorem and, hence, $d_2 = 0$, because such a condition is only needed for Lemma 14 with the block bootstrap. No condition on q_2 of the form $q_2 > 6a/(1-2\gamma)$ is needed in part (b) of the Theorem, in contrast to Theorem 1(c), because this condition is needed in the latter case to control the behavior of the bootstrap moments via Lemma 14 with $\xi = 0$, whereas the bootstrap moments are well-behaved in the parametric bootstrap cases due to Lemma 3 and its analogue for $\overline{\theta}_N$. □

5.3 Proofs of Lemmas

5.3.1 Proof of Lemma 1

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

$$\lim_{N\to\infty} \varepsilon^{-p} N^{a-p+pc} E||\sum_{i=1}^N h(\widetilde{X}_i)||^p \leq \lim_{N\to\infty} \varepsilon^{-p} C N^{a-p+pc+p/2} = 0. \tag{5.39}$$

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

5.3.2 Proof of Lemma 2

The proof is the same as that of Lemma 2 of HH (which mimics a standard proof of a uniform law of large numbers) except that we apply Lemma 1 above with $c = 0$ and $p = q_0$ rather than their Lemma 1. □

5.3.3 Proof of Lemma 3

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

$$\begin{aligned} N^a P(||\widehat{\theta}_N - \theta_0|| > \varepsilon) &\leq N^a P(\rho(\widehat{\theta}_N) - \rho_N(\widehat{\theta}_N) + \rho_N(\widehat{\theta}_N) - \rho(\theta_0) > \delta) \\ &\leq N^a P(\rho(\widehat{\theta}_N) - \rho_N(\widehat{\theta}_N) + \rho_N(\theta_0) - \rho(\theta_0) > \delta) \\ &\leq N^a P(2\sup_{\theta\in\Theta} |\rho_N(\theta) - \rho(\theta)| > \delta) \to 0 \end{aligned} \tag{5.40}$$

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

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

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

$$\widehat{\theta}_N - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\theta_N^+)\right)^{-1}\frac{\partial}{\partial\theta}J_N(\theta_0) \tag{5.41}$$

with probability $1 - o(N^{-a})$, where θ_N^+ lies between $\widehat{\theta}_N$ and θ_0 and may differ across rows. In consequence, the result of the Lemma (with $c > 0$) follows from

$$\begin{aligned}
\lim_{N\to\infty} N^a P(\|\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\theta_N^+) - \frac{\partial^2}{\partial\theta\partial\theta'}J_N(\theta_0)\| > \varepsilon) &= 0,\\
\lim_{N\to\infty} N^a P(\|\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\theta_0) - 2D'\Omega D\| > \varepsilon) &= 0,\\
\lim_{N\to\infty} N^a P(\|D_N(\theta_0) - D\| > \varepsilon) &= 0, \text{ and}\\
\lim_{N\to\infty} N^a P(\|N^{-1}\sum_{i=1}^{N} g(X_i,\theta_0)\| > N^{-c}\varepsilon) &= 0.
\end{aligned} \tag{5.42}$$

The first result of (5.42) holds using the result of the present Lemma with $c = 0$, Taylor expansions about θ_0, and multiple applications of Lemma 1(b) with $h(\widetilde{X}_i) = (\partial^j/\partial\theta^j)g(X_i,\theta_0)$ for $j = 0, ..., 3$ or $h(\widetilde{X}_i) = C_g(X_i)$. The second result of (5.42) holds by multiple applications of Lemma 1(a) with $h(\widetilde{X}_i) = (\partial^j/\partial\theta^j)\, g(X_i,\theta_0) - E(\partial^j/\partial\theta^j)$ $g(X_i,\theta_0)$ for $j = 0, 1, 2$, $c = 0$, and $p = q_2 > 2a$ and standard manipulations. The third result holds by Lemma 1(a) with $h(\widetilde{X}_i)$ as in the proof of the second result with $j = 1$. The fourth result holds by Lemma 1(a) with $h(\widetilde{X}_i) = g(X_i,\theta_0)$, $c = c$, and $p = q_1 > 2a/(1 - 2c)$. □

5.3.4 Proof of Lemma 4

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

$$\begin{aligned}
&\lim_{N\to\infty} N^a P(\|\Omega_N^{-1}(\widetilde{\theta}_N) - \Omega_N^{-1}(\theta_0)\| > \varepsilon) = 0 \text{ and}\\
&\lim_{N\to\infty} N^a P(\|\Omega_N^{-1}(\theta_0) - \Omega_0^{-1}\| > \varepsilon) = 0.
\end{aligned} \tag{5.43}$$

The first result of (5.43) holds by Lemma 3, mean value expansions, and multiple applications of Lemma 1(b) with $h(\widetilde{X}_i) = \sup_{\theta\in N_0} \|g(X_i,\theta)\|\cdot\|(\partial/\partial\theta')g(X_{i+j},\theta)\|$ for $j = -\kappa, ..., \kappa$. The second result of (5.43) holds by multiple applications of Lemma 1(a) with $h(\widetilde{X}_i) = g(X_i,\theta_0)g(X_{i+j},\theta_0)' - Eg(X_i,\theta_0)g(X_{i+j},\theta_0)'$ for $j = -\kappa, ..., \kappa$, $c = 0$, and $p = q_1/2 > 2a$.

Given the result of the previous paragraph, the proof of Lemma 4 is analogous to that of Lemma 3. □

5.3.5 Proof of Lemma 5

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

$$\begin{aligned}
&N^a(P(A_N + \xi_N \leq z) - P(A_N \leq z)) \\
&\leq N^a(P(A_N + \xi_N \leq z,\ ||\xi_N|| \leq \vartheta_N \varepsilon) - P(A_N \leq z)) + N^a P(||\xi_N|| > \vartheta_N \varepsilon) \\
&\leq N^a(P(A_N \leq z + \vartheta_N \varepsilon \iota) - P(A_N \leq z)) + N^a P(||\xi_N|| > \vartheta_N \varepsilon) \qquad (5.44)
\end{aligned}$$

The second term on the right-hand side converges in probability to zero by assumption. Now, consider the case where A_N has an Edgeworth expansion with remainder $o(N^{-a})$. Then, the first term on the last line of (5.44) is less than or equal to N^a multiplied by

$$(1+\sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\partial/\partial z))\Phi(z+\vartheta_N\varepsilon\iota) - (1+\sum_{i=1}^{[2a]} N^{-i/2}\pi_i(\partial/\partial z))\Phi(z) + o(N^{-a}). \quad (5.45)$$

This can be made less than any given $\delta > 0$ for all $z \in R^{L_A}$ by taking ε sufficiently small because the derivatives of $\Phi(z)$ of all orders are bounded over $z \in R^{L_A}$ and $N^a \vartheta_N = O(1)$. Alternatively, in the case where $\{A_N : N \geq 1\}$ have uniformly bounded densities, the first term on the right-hand side of (5.44) is less than any given $\delta > 0$ by taking ε sufficiently small because $N^a \vartheta_N = O(1)$.

An analogous argument shows that $N^a(P(A_N \leq z) - P(A_N + \xi_N \leq z)) < \delta$ for all $z \in R^{L_A}$ and all N sufficiently large by taking ε small. This completes the proof of part (a).

The proof of part (b) is similar. For brevity, it is omitted. □

5.3.6 Proof of Lemma 6

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

$$\begin{aligned}
N^a P(N^a P^*(||\Gamma_N^*|| > N^{-c}\varepsilon) > \varepsilon) &\leq N^a P(E^*||\Gamma_N^*||^p > N^{-a-cp}\varepsilon^{p+1}) \\
&\leq N^{2a+cp} E E^* ||\Gamma_N^*||^p \varepsilon^{-p-1}. \qquad (5.46)
\end{aligned}$$

Define $Y_{\ell i}^* = \ell^{-1}\sum_{j \in b_i^*} h(\widetilde{X}_j)$ and $Y_{\ell i} = \ell^{-1}\sum_{j \in b_i} h(\widetilde{X}_j)$ (where b_j^* and b_j are defined just before the proof of Theorem 3). Then, $\Gamma_N^* = b^{-1}\sum_{i=1}^b (Y_{\ell i}^* - E^* Y_{\ell i}^*)$. By applying Burkholder's and Holder's inequality in a single step (e.g., see Hall and Heyde (1980, eqn. (3.67), p. 87)) , we obtain

$$\begin{aligned}
E^*||\Gamma_N^*||^p &= b^{-p} E^* || \sum_{i=1}^b (Y_{\ell i}^* - E^* Y_{\ell i}^*)||^p \leq C b^{-p/2} E^* ||Y_{\ell 1}^* - E^* Y_{\ell 1}^*||^p \\
&\leq C b^{-p/2} E^* ||Y_{\ell 1}^*||^p. \qquad (5.47)
\end{aligned}$$

Now, for non-overlapping blocks, we have

$$EE^*||Y_{\ell 1}^*||^p = Eb^{-1}\sum_{i=1}^{b}||Y_{\ell i}||^p = E||Y_{\ell 1}||^p \leq C\ell^{-p/2} \tag{5.48}$$

using Yokoyama's strong mixing moment inequality (see the proof of Lemma 1). For overlapping blocks, (5.48) holds, but with b replaced by $N-\ell+1$ after the first equality.

Combining (5.46), (5.47), and (5.48) gives

$$N^a P(N^a P^*(||\Gamma_N^*|| > N^{-c}\varepsilon) > \varepsilon) \leq CN^{2a+cp}b^{-p/2}\ell^{-p/2} = CN^{2a+cp-p/2} = o(1). \tag{5.49}$$

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

$$\begin{aligned} lhs &= \lim_{N\to\infty} N^a P(||b^{-1}\sum_{i=1}^{b} Y_{\ell i}|| > N^{-c}\varepsilon) \\ &= \lim_{N\to\infty} N^a P(||N^{-1}\sum_{i=1}^{N} h(\widetilde{X}_i)|| > N^{-c}\varepsilon) = 0 \end{aligned} \tag{5.50}$$

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

$$\begin{aligned} lhs &= \lim_{N\to\infty} N^a P(||(N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1} Y_{\ell i}|| > N^{-c}\varepsilon) \\ &= \lim_{N\to\infty} N^a P(||(N-\ell+1)^{-1}\sum_{i=1}^{N} w(i,\ell,N)h(\widetilde{X}_i)|| > N^{-c}\varepsilon) = 0, \end{aligned} \tag{5.51}$$

where $w(i,\ell,N)$ is defined in (2.12). Note that $|w(i,\ell,N)| \leq 1$. The last equality of (5.51) holds by an argument analogous to that of Lemma 1 using Yokoyama's strong mixing moment inequality (which applies to non-stationary L^p–bounded random variables, see Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30)), using the fact that $\lim_{N\to\infty} N/(N-\ell+1) = 1$ when $0 \leq \gamma < 1$. For overlapping blocks with $\gamma = 1$, we have $\ell hs = \lim_{N\to\infty} N^a P(||N^{-1}\sum_{i=1}^N h(\widetilde{X}_i)|| > N^{-c}\varepsilon) = 0$.

Part (c) follows from parts (a) and (b). The first result of part (d) holds by using the triangle inequality $||E^*N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*)|| \leq ||E^*N^{-1}\sum_{i=1}^N h(\widetilde{X}_i^*) - Eh(\widetilde{X}_i)|| + ||Eh(\widetilde{X}_i)||$ and applying part (b) with $c=0$ to $h(\cdot) - Eh(\widetilde{X}_1)$. The second result of part (d) is established analogously using part (c) in place of part (b). □

5.3.7 Proof of Lemma 7

The proof is the same as that of Lemma 8 of HH except that we apply Lemma 6 above with $c=0$ and $p=q_0$ rather than Lemma 7 of HH. □

5.3.8 Proof of Lemma 8

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

$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega_N^*(\widehat{\theta}_N^*)^{-1} - \Omega_N^*(\theta_0)^{-1}|| > \varepsilon) > \varepsilon) = 0, \quad (5.52)$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega_N^*(\theta_0)^{-1} - \Omega_N^{**}(\theta_0)^{-1}|| > \varepsilon) > \varepsilon) = 0, \quad (5.53)$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega_N^{**}(\theta_0)^{-1} - E^*(\Omega_N^{**}(\theta_0)^{-1})|| > \varepsilon) > \varepsilon) = 0, \text{ and} \quad (5.54)$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||E^*(\Omega_N^{**}(\theta_0)^{-1}) - \Omega_0^{-1}|| > \varepsilon) > \varepsilon) = 0. \quad (5.55)$$

To establish (5.52), we take mean value expansions about θ_0, apply both parts of Lemma 6(d) with $h(\widetilde{X}_i) = \sup_{\theta\in N_0} ||g(X_i,\theta)|| \cdot ||(\partial/\partial\theta')g(X_{i+j},\theta)||$ for $j = -\kappa, ..., \kappa$ and $p = q_2 > 4a$, and use the assumption on $\widehat{\theta}_N^*$. To establish (5.53), we use Lemma 6(b) and 6(c) with $h(\widetilde{X}_i) = g(X_i,\theta_0)$, $c = 0$, and $p = q_1 > 4a$. To establish (5.54), we use Lemma 6(c) with $h(\widetilde{X}_i) = g(X_i,\theta_0)g(X_{i+j},\theta_0)' - Eg(X_1,\theta_0)g(X_{1+j},\theta_0)'$ for $j = -\kappa, ..., \kappa$, $c = 0$, and $p = q_1/2 > 4a$. To establish (5.55), we use Lemma 6(b) with $h(\widetilde{X}_i)$, c, and p as immediately above. □

5.3.9 Proof of Lemma 9

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

$$\begin{aligned}
& N^aP(N^aP^*(||\theta_N^* - \widehat{\theta}_N|| > \varepsilon) > \varepsilon) \\
\leq\ & N^aP(N^aP^*(E^*\rho_N^*(\theta_N^*) - \rho_N^*(\theta_N^*) + \rho_N^*(\theta_N^*) - E^*\rho_N^*(\widehat{\theta}_N) > \delta) > \varepsilon) \\
\leq\ & N^aP(N^aP^*(E^*\rho_N^*(\theta_N^*) - \rho_N^*(\theta_N^*) + \rho_N^*(\widehat{\theta}_N) - E^*\rho_N^*(\widehat{\theta}_N) > \delta) > \varepsilon) \\
\leq\ & N^aP(N^aP^*(2\sup_{\theta\in\Theta} |\rho_N^*(\theta) - E^*\rho_N^*(\theta)| > \delta) > \varepsilon) \to 0, \quad (5.56)
\end{aligned}$$

using Lemma 7 with $j = 2$.

For the case of overlapping blocks, (i) and (ii) of the previous paragraph do not hold. Instead, we have $E^*N^{-1}\sum_{i=1}^N \rho(X_i^*,\theta) = (N-\ell+1)^{-1}\sum_{i=1}^N w(i,\ell,N)\rho(X_i,\theta)$, where $w(i,\ell,N)$ is defined in (2.12). By the arguments used to prove Lemmas 2 and 6(b), (iii) holds with $\rho_N(\theta)$ replaced by $(N-\ell+1)^{-1}\sum_{i=1}^N w(i,\ell,N)\rho(X_i,\theta)$. In addition, some calculations using Lemmas 1 and 3 and a mean value expansion show that $\lim_{N\to\infty} N^a P(||E^* g(X_i^*, \widehat{\theta}_N)|| > \lambda) = 0$ for all $\lambda > 0$, because $E^* g(X_i^*, \widehat{\theta}_N)$ $= (N - \ell + 1)^{-1} \sum_{i=1}^N (w(i, \ell, N) - (N - \ell + 1)/N)\, g(X_i, \widehat{\theta}_N)$. In consequence, the remainder of the proof above goes through unchanged with overlapping blocks.

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

5.3.10 Proof of Lemma 10

The proof is analogous to that of Lemma 9 using Lemma 8. □

5.3.11 Proof of Lemma 11

By Lemma 9 with $c = 0$, $\widetilde{\theta}_N^*$ satisfies the same condition as $\widehat{\theta}_N^*$. In consequence, it suffices to show that the first result of the Lemma holds and that for all $\varepsilon > 0$,

$$\lim_{N\to\infty} N^a P(N^a P^*(||\Omega_N^*(\widehat{\theta}_N^*) - \Omega_0|| > \varepsilon) > \varepsilon) = 0, \tag{5.57}$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||N^{-1} \sum_{i=1}^N \frac{\partial^j}{\partial \theta^j} g^*(X_i^*, \widehat{\theta}_N^*)|| > K_\varepsilon) > \varepsilon) = 0 \tag{5.58}$$

$$\text{for some } K_\varepsilon < \infty \text{ for } j = 1, 2, 3, \text{ and}$$

$$\lim_{N\to\infty} N^a P(N^a P^*(||N^{-1} \sum_{i=1}^N g^*(X_i^*, \widehat{\theta}_N^*)|| > \varepsilon) > \varepsilon) = 0. \tag{5.59}$$

The first result of the Lemma, (5.58), and (5.59) hold by mean value expansions, multiple applications of Lemma 6, and the assumption on $\widehat{\theta}_N^*$. Equation (5.57) holds by Lemma 8. □

5.3.12 Proof of Lemma 12

By a mean value expansion and the triangle inequality,

$$\begin{aligned} & ||m_N^*(\theta_{N,k}^*) - m_N^*(\theta_N^*)|| \\ \le\ & (N^{-1} \sum_{i=1}^N \sup_{\theta \in N_0} ||(\partial/\partial\theta) m(\widetilde{X}_i^*, \theta)|| + E^* N^{-1} \sum_{i=1}^N \sup_{\theta \in N_0} ||(\partial/\partial\theta) m(\widetilde{X}_i^*, \theta)||) \\ & \times ||\theta_{N,k}^* - \theta_N^*||. \end{aligned} \tag{5.60}$$

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

5.3.13 Proof of Lemma 13

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

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

5.3.14 Proof of Lemma 14

The condition $\xi < 1/2 - \gamma$ is used to ensure that the denominator of the lower bound on q_2 is positive. There exists $c \in (0, 1/2)$ such that $q_1 > 2a/(1-2c)$ and $d_2 \geq -1 + (a + \gamma + \xi)/c$ if, and only if, the same two conditions hold as strict inequalities. Hence, we can assume that the latter holds.

We show that

$$\begin{aligned} A_1 &= \lim_{N\to\infty} N^a P(N^{\xi+\alpha(m)}|E^* \prod_{\mu=1}^{m} \Psi^*_{N,j_\mu} - E \prod_{\mu=1}^{m} \Psi_{N,j_\mu}| > 6\varepsilon) = 0 \text{ and} \\ A_2 &= \lim_{N\to\infty} N^{\xi} |N^{\alpha(m)} E \prod_{\mu=1}^{m} \Psi_{N,j_\mu} - \lim_{N\to\infty} N^{\alpha(m)} E \prod_{\mu=1}^{m} \Psi_{N,j_\mu}| = 0 \qquad (5.61) \end{aligned}$$

for all $\varepsilon > 0$ for $m = 5$. The proof for $m = 5$ illustrates the proof for arbitrary m. In the proof, we specify m generically rather than as 5, so that the conditions needed for the arbitrary m case become clear.

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

$$\begin{aligned} E^* \prod_{\mu=1}^{m} \Psi^*_{N,j_\mu} &= N^{-m/2} \sum_{j_1=1}^{b}\sum_{j_2=1}^{b}\sum_{j_3=1}^{b}\sum_{j_4=1}^{b}\sum_{j_5=1}^{b} E^* Y^*_{j_1} Y^*_{j_2} Y^*_{j_3} Y^*_{j_4} Y^*_{j_5} \\ &= N^{-m/2}(bE^* Y_1^{*5} + 10(b^2 - b)E^* Y_1^{*2} E^* Y_1^{*3}) \\ &= N^{-m/2}\left(b\mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} \widetilde{Y}_j^5 + 10\frac{b^2-b}{\mathcal{N}_\ell^2} \sum_{j_1=1}^{\mathcal{N}_\ell} \widetilde{Y}_{j_1}^2 \sum_{j_2=1}^{\mathcal{N}_\ell} \widetilde{Y}_{j_2}^3 \right), \qquad (5.62) \end{aligned}$$

where $\alpha(m)-m/2=-2$ when $m=5$, the second equality holds because the bootstrap random variables Y_j^* are iid with mean zero, and the third equality holds by the definition of E^*.

Below we show that

$$\lim_{N\to\infty} N^a P\left(N^{\xi+\alpha(m)-m/2}\left|b\mathcal{N}_\ell^{-1}\sum_{j=1}^{\mathcal{N}_\ell}(\widetilde{Y}_j^5-Y_j^5)\right.\right.$$
$$\left.\left.+10\frac{b^2-b}{\mathcal{N}_\ell^2}\sum_{j_1=1}^{\mathcal{N}_\ell}\sum_{j_2=1}^{\mathcal{N}_\ell}(\widetilde{Y}_{j_1}^2\widetilde{Y}_{j_2}^3-Y_{j_1}^2Y_{j_2}^3)\right|>\varepsilon\right)=0. \tag{5.63}$$

Using this and (5.62), we obtain

$$A_1\le B_1+B_2+B_3+B_4+B_5+B_6,\ \text{where}$$

$$\begin{aligned}
B_1&=\lim_{N\to\infty}N^aP(10N^{\xi+\alpha(m)-m/2}|(b/\mathcal{N}_\ell)^2\textstyle\sum_{j_1=1}^{\mathcal{N}_\ell}(Y_{j_1}^2-EY_1^2)\sum_{j_2=1}^{\mathcal{N}_\ell}(Y_{j_2}^3-EY_1^3)|>\varepsilon),\\
B_2&=\lim_{N\to\infty}N^aP(10N^{\xi+\alpha(m)-m/2}|b\mathcal{N}_\ell^{-1}\textstyle\sum_{j=1}^{\mathcal{N}_\ell}(Y_j^3-EY_1^3)bEY_1^2|>\varepsilon),\\
B_3&=\lim_{N\to\infty}N^aP(10N^{\xi+\alpha(m)-m/2}|b\mathcal{N}_\ell^{-1}\textstyle\sum_{j=1}^{\mathcal{N}_\ell}(Y_j^2-EY_1^2)bEY_1^3|>\varepsilon),\\
B_4&=\lim_{N\to\infty}N^aP(N^{\xi+\alpha(m)-m/2}|b\mathcal{N}_\ell^{-1}\textstyle\sum_{j=1}^{\mathcal{N}_\ell}(Y_j^5-EY_1^5)|>\varepsilon),\\
B_5&=\lim_{N\to\infty}1(|N^{\xi+\alpha(m)-m/2}bEY_1^5|>\varepsilon),\ \text{and}\\
B_6&=\lim_{N\to\infty}1(|10N^{\xi+\alpha(m)-m/2}b^2EY_1^2EY_1^3-N^{\xi+\alpha(m)}E\textstyle\prod_{\mu=1}^5\Psi_{N,j_\mu})|>\varepsilon).
\end{aligned} \tag{5.64}$$

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

$$E|\sum_{i=1}^{b}(Y_i^r-EY_1^r)|^s\le Cb^{s/2}(E|Y_1^r-EY_1^r|^{s+\delta})^{s/(s+\delta)}\le Cb^{s/2}(E|Y_1^r|^{s+\delta})^{s/(s+\delta)}, \tag{5.65}$$

for $r>0$ and $s\ge 2$. (This result uses the fact that $\alpha_\ell(i)\le\alpha(i)$ for all $\ell\ge 1$, where $\alpha(i)$ denotes the i-th strong mixing number of $\{\widetilde{X}_i: i\ge 1\}$ and $\alpha_\ell(i)$ denotes the i-th strong mixing number of $\{Y_i: i\le b\}$.) Note that the moment inequality used here is stronger than the one used in the proof of Lemma 1, because the upper bound depends on moments of the random variables in the sum rather than just an unknown constant. Yokoyama (1980, (4.1)) proved this moment inequality for the case where s is an even integer. Doukhan's results extend it to arbitrary s using an interpolation lemma.

In turn, the same moment inequality gives

$$E|Y_1^r|^{s+\delta}=E|\sum_{j=1}^{\ell}f_j|^{r(s+\delta)}\le C\ell^{r(s+\delta)/2}(E|f_1|^{r(s+\delta)+\delta})^{r(s+\delta)/(r(s+\delta)+\delta)}. \tag{5.66}$$

Combining these two inequalities gives

$$E|b\mathcal{N}_\ell^{-1}\sum_{i=1}^{\mathcal{N}_\ell}(Y_i^r-EY_1^r)|^s\le Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}}, \tag{5.67}$$

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

Next, for overlapping blocks, we have

$$\begin{aligned}(E|\sum_{i=1}^{N-\ell+1}(Y_i^r - EY_1^r)|^s)^{1/s} &= (E|\sum_{u=1}^{\ell}\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)|^s)^{1/s} \\ \leq \sum_{u=1}^{\ell}(E|\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)|^s)^{1/s} &\leq \sum_{i=1}^{\ell}(Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}})^{1/s} \quad (5.68)\end{aligned}$$

where $b(u) = \max\{j : j \leq b-1,\ u+\ell j \leq N-\ell+1\}$, using Minkowski's inequality and (5.67), where the latter applies because $\sum_{i=0}^{b(u)}(Y_{i\ell+u}^r - EY_1^r)$ is a sum of terms based on non-overlapping blocks. Equation (5.68) gives

$$\begin{aligned}E|b\mathcal{N}_\ell^{-1}\sum_{i=1}^{\mathcal{N}_\ell}(Y_i^r - EY_1^r)|^s &\leq \frac{b^s\ell^s}{(N-\ell+1)^s}Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}} \\ &= Cb^{s/2}\ell^{rs/2}(E|f_1|^{rs+(r+1)\delta})^{c_{r,s}}, \quad (5.69)\end{aligned}$$

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

Now we show that $B_1 = 0$. For some small $\delta > 0$, we take p such that $mp+4\delta = q_2$. Then, by Markov's inequality, Holder's inequality, and (5.67) or (5.69) applied twice, we have

$$\begin{aligned}B_1 &\leq C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)}E|b\mathcal{N}_\ell^{-1}\sum_{j_1=1}^{\mathcal{N}_\ell}(Y_{j_1}^2 - EY_1^2)b\mathcal{N}_\ell^{-1}\sum_{j_2=1}^{\mathcal{N}_\ell}(Y_{j_2}^3 - EY_1^3)|^p \\ &\leq C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)}(E|b\mathcal{N}_\ell^{-1}\sum_{j_1=1}^{\mathcal{N}_\ell}(Y_{j_1}^2 - EY_1^2)|^{5p/2})^{2/5} \\ &\quad\times(E|b\mathcal{N}_\ell^{-1}\sum_{j_2=1}^{\mathcal{N}_\ell}(Y_{j_2}^3 - EY_1^3)|^{5p/3})^{3/5} \\ &\leq C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)}(b^{5p/4}\ell^{5p/2}(E|f_1|^{5p+3\delta})^{c_{2,5p/2}})^{2/5} \\ &\quad\times(b^{5p/6}\ell^{5p/2}(E|f_1|^{5p+4\delta})^{c_{3,5p/3}})^{3/5} \\ &\leq C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)}b^{2(p/2)}\ell^{5p/2}(E|f_1|^{5p+4\delta})^{3c_{3,5p/3}/5} \\ &= C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)}b^{[m/2]p/2}\ell^{mp/2}(E|f_1|^{mp+4\delta})^{3c_{3,5p/3}/5} \\ &= C\lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-(m-[m/2])(1-\gamma)/2)}(E|f_1|^{q_2})^{3c_{3,5p/2}/5}, \quad (5.70)\end{aligned}$$

where the first equality replaces quantities for the $m = 5$ case with the appropriate quantities for the generic m case. Note that $[m/2]$ denotes the number of sums over j that appear in the generic m case and equals 2 when $m = 5$.

Given (5.70), $B_1 = 0$ provided $p > a/\{-\xi - \alpha(m) + (m-[m/2])\ (1-\gamma)/2\}$ or, because $\delta > 0$ is arbitrarily small, provided

$$q_2 > ma/\{-\xi - \alpha(m) + (m-[m/2])(1-\gamma)/2\}. \quad (5.71)$$

For $m = 5$, this becomes $q_2 > 5a/(1-3\gamma/2-\xi)$, which holds because $q_2 > 6a/(1-2\gamma-2\xi)$. For arbitrary m, (5.71) can be analyzed separately for m even and m odd. In each case, the derivative with respect to m of the right-hand side is strictly negative, so the most restrictive condition occurs when $m = 2$ or $m = 3$. It is easy to see that it actually occurs when $m = 3$ and that it is the condition assumed in the Lemma, viz., $q_2 > 6a/(1-2\gamma-2\xi)$.

Next, we consider B_2. The number of times that the term (bEY_1^2) appears in B_2 in the arbitrary odd m case is $(m/2-\alpha(m)-1)$ (presuming that B_2 is defined to include one sum over terms of the form $Y_i^3 - EY_i^3$ and the remaining terms are of the form (bEY_1^2)), which equals 1 when $m = 5$. For some small $\delta > 0$, we take p such $3p + 4\delta = q_2$. Then, by Markov's inequality, (5.66), and either (5.67) or (5.69), we have

$$\begin{aligned}
B_2 &\leq C \lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)} E|b\mathcal{N}_\ell^{-1}\sum_{i=1}^{\mathcal{N}_\ell}(Y_j^3 - EY_1^3)|^p (bEY_1^2)^{(m/2-\alpha(m)-1)p} \\
&\leq C \lim_{N\to\infty} N^{a+p(\xi+\alpha(m)-m/2)} b^{p/2}\ell^{3p/2}(E|f_1|^{3p+4\delta})^{c_{3,p}} \\
&\quad \times (b\ell(Ef_1^{2+\delta})^{2/(2+\delta)})^{p(m/2-\alpha(m)-1)} \\
&= C \lim_{N\to\infty} N^{a+p(\xi+\gamma-1/2)}(E|f_1|^{3p+4\delta})^{c_{3,p}} \to 0, \qquad (5.72)
\end{aligned}$$

where the convergence to zero holds provided $p > a/(-\gamma-\xi+1/2)$ or, because δ is arbitrarily small, provided

$$q_2 > 3a/(-\gamma-\xi+1/2). \qquad (5.73)$$

The latter is equivalent to the assumption that $q_2 > 6a/(1-2\gamma-2\xi)$. Note that this condition does not depend on m. Hence, the same binding constraint on q_2 arises for all m odd. (It arises only for m odd, because terms of the form B_2 only arise when m is odd.)

By analogous arguments to those for B_1 and B_2, we can show that $B_3 = 0$ and $B_4 = 0$ provided $q_2 > 2a/(-\gamma-\xi+1/2)$ and $q_2 > ma/\{-\xi-\alpha(m)+(m-1)(1-\gamma)/2\}$ respectively. The former condition, which is independent of m, is implied by the assumption that $q_2 > 6a/(1-2\gamma-2\xi)$. The latter condition can be analyzed by the same method as used for the condition in (5.71). It is most restrictive when $m = 3$ and in this case is equivalent to the assumption of the Lemma on q_2.

Next, we show $B_5 = 0$. It suffices to show that $N^{\xi+\alpha(m)-m/2}b|EY_1^5|$ has limit zero. It is bounded by

$$\begin{aligned}
N^{\xi+\alpha(m)-m/2} bE|\sum_{i=1}^{\ell} f_i|^m &\leq N^{\xi+\alpha(m)-m/2} b\ell^{m/2}(E|f_i|^{m+\delta})^{m/(m+\delta)} \\
&= CN^{\xi+\alpha(m)-(m-2)(1-\gamma)/2} \qquad (5.74)
\end{aligned}$$

for $\delta > 0$, using (5.66). The right-hand side converges to zero as $N \to \infty$ provided $\xi + \alpha(m) - (m-2)(1-\gamma)/2 < 0$. The latter holds when $m \geq 4$ because $\xi < 1/2 - \gamma$. When $m = 2$ or 3, B_5 does not appear.

For B_6, we have

$$\begin{aligned}
& N^{\alpha(m)} E \prod_{\mu=1}^{m} \Psi_{N,j_\mu} \\
= \ & N^{\alpha(m)-m/2} \sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{i_3=1}^{N}\sum_{i_4=1}^{N}\sum_{i_5=1}^{N} E f_{i_1} f_{i_2} f_{i_3} f_{i_4} f_{i_5} \\
= \ & 10 N^{-1} \sum_{i_1=1}^{N}\sum_{i_2=1}^{N} E f_{i_1} f_{i_2} N^{-1} \sum_{i_3=1}^{N}\sum_{i_4=1}^{N}\sum_{i_5=1}^{N} E f_{i_3} f_{i_4} f_{i_5} + O(N^{-1}) \\
= \ & 10 \sum_{i_1=-N+1}^{N-1} \omega(i_1, N) E f_0 f_{i_1} \sum_{i_2=-N+1}^{N-1} \sum_{i_3=-N+1}^{N-1} \omega(i_2+i_3, N) E f_0 f_{i_2} f_{i_3} + O(N^{-1}) \\
\to \ & 10 \sum_{i_1=-\infty}^{\infty} E f_0 f_i \sum_{i_2=-\infty}^{\infty} \sum_{i_3=-\infty}^{\infty} E f_0 f_{i_2} f_{i_3} \text{ as } N \to \infty, \qquad (5.75)
\end{aligned}$$

where $\omega(i, N) = 1 - i/N$, the second equality holds by standard manipulations of stationary strong mixing random variables, and the third equality holds by change of variables.

In addition, we have

$$\begin{aligned}
& N^{\alpha(m)-m/2} b^2 E Y_1^2 E Y_1^3 = \ell^{-1} E(\sum_{i=1}^{\ell} f_i)^2 \ell^{-1} E(\sum_{i=1}^{\ell} f_i)^3 \\
= \ & \sum_{i_1=-\ell+1}^{\ell-1} \omega(i_1, \ell) E f_0 f_{i_1} \sum_{i_2=-\ell+1}^{\ell-1} \sum_{i_3=-\ell+1}^{\ell-1} \omega(i_2+i_3, \ell) E f_0 f_{i_2} f_{i_3}. \qquad (5.76)
\end{aligned}$$

by change of variables. Equations (5.75) and (5.76) imply that B_6 is zero if

$$\begin{aligned}
& \lim_{N\to\infty} N^{\xi} | \sum_{i_1=-\ell+1}^{\ell-1} \omega(i_1, \ell) E f_0 f_{i_1} \sum_{i_2=-\ell+1}^{\ell-1} \sum_{i_3=-\ell+1}^{\ell-1} \omega(i_2+i_3, \ell) E f_0 f_{i_2} f_{i_3} \\
- & \sum_{i_1=-N+1}^{N-1} \omega(i_1, N) E f_0 f_{i_1} \sum_{i_2=-N+1}^{N-1} \sum_{i_3=-N+1}^{N-1} \omega(i_2+i_3, N) E f_0 f_{i_2} f_{i_3} + O(N^{-1}) | = 0. \qquad (5.77)
\end{aligned}$$

The latter holds by a strong mixing covariance inequality, viz., $E f_0 f_{i_1} \le C \alpha^r(i_1)$ for some $r > 0$ (where $\alpha(i_1)$ denotes the i_1-th strong mixing number), e.g., see Doukhan (1995, Thm. 3, p.9), the fact that the strong mixing numbers decline exponentially fast by Assumption 1, $N^{\xi} \propto \ell^{\xi/\gamma}$, and either (i) $\xi < \gamma$ and $0 < \gamma < 1$ or (ii) $\{X_i : i \ge 1\}$ are independent. The latter condition appears because $\lim_{N\to\infty} N^{\xi} \sum_{i_1=-\ell+1}^{\ell-1} (i_1/\ell) E f_0 f_{i_1} \sum_{i_2=-\ell+1}^{\ell-1} \sum_{i_3=-\ell+1}^{\ell-1} \omega(i_2+i_3, \ell) E f_0 f_{i_2} f_{i_3} = \lim_{N\to\infty} N^{\xi} \ell^{-1} \sum_{i_1=-\infty}^{\infty} i_1 E f_0 f_{i_1} \sum_{i_2=-\infty}^{\infty} \sum_{i_3=-\infty}^{\infty} E f_0 f_{i_2} f_{i_3} = 0$ if either $\xi < \gamma$ or $\{X_i : i \ge 1\}$ are independent. We conclude that $B_6 = 0$.

An analogous argument to that given for (5.77) gives $A_2 = 0$.

When m is arbitrary, rather than equal to 5, more than six terms arise in (5.64). Nevertheless, the least favorable term (with respect to the assumptions required) is the case where m is odd and B_2 depends on $\sum_{j=1}^{b}(Y_j^3 - EY_1^3)\ (bEY_1^2)^{(m-3)/2}$. This case is considered above. Thus, the conditions given in the Theorem are sufficient for the stated result to hold for arbitrary $m \leq 2a + 2$.

When $\{X_i : i \geq 1\}$ are independent, then (5.75) holds with ∞ replaced by 0 and (5.76) and (5.77) hold with $\ell - 1$ and $N - 1$ replaced by 0.

It remains to show (5.63). It suffices to show that

$$\lim_{N\to\infty} N^a P(N^{\xi+\alpha(m)-m/2}|b\mathcal{N}_\ell^{-1}\sum_{j=1}^{\mathcal{N}_\ell}(\widetilde{Y}_j^r - Y_j^r)| > \varepsilon) = 0 \tag{5.78}$$

for $r = 2, 3, ..., m.$ Define

$$\lambda_j = \widetilde{Y}_j - Y_j = \sum_{i\in b_j}(f_i^* - f_i), \tag{5.79}$$

where f_i^* and f_i are defined above. We have

$$\begin{aligned}\sum_{i\in b_j} f_i^* &= \sum_{i\in b_j} f_1(\widetilde{X}_i, \widehat{\theta}_N) - E^*\sum_{i\in b_j} f_1(\widetilde{X}_i^*, \widehat{\theta}_N) = \sum_{i\in b_j} f_1(\widetilde{X}_i, \widehat{\theta}_N) - \ell\overline{f}_{1N}(\widehat{\theta}_N), \text{ and}\\ \overline{f}_{1N}(\widehat{\theta}_N) &= N^{-1}\sum_{i=1}^{N} f_1(\widetilde{X}_i, \widehat{\theta}_N).\end{aligned} \tag{5.80}$$

(The second equality for $\sum_{i\in b_j} f_i^*$ is only approximately true for overlapping blocks due to end effects. But, these effects are asymptotically negligible in the calculations below and can be ignored.)

Because $\widetilde{Y}_j^r = (Y_j + \lambda_j)^r = \sum_{k=0}^{r}\binom{r}{k}Y_j^k\lambda_j^{r-k}$, it suffices to show that

$$\lim_{N\to\infty} N^a P(N^{\xi+\alpha(m)-m/2}b|\mathcal{N}_\ell^{-1}\sum_{j=1}^{\mathcal{N}_\ell} Y_j^{k_1}\lambda_j^{k_2})| > \varepsilon) = 0, \tag{5.81}$$

where k_1 and k_2 are integers that satisfy $0 \leq k_1 \leq m-1$, $1 \leq k_2 \leq m$, and $k_1 + k_2 \leq m$.

For notational simplicity, suppose θ is a scalar. A Taylor expansion about θ_0 of order d_2 gives

$$\begin{aligned}\lambda_j &= \sum_{u=1}^{d_2}\frac{1}{u!}\sum_{i\in b_j}(\frac{\partial^u}{\partial\theta^u}f_1(\widetilde{X}_i, \theta_0) - E\frac{\partial^u}{\partial\theta^u}f_1(\widetilde{X}_i, \theta_0))(\widehat{\theta}_N - \theta_0)^u\\ &\quad -\sum_{u=0}^{d_2}\frac{1}{u!}\ell(\frac{\partial^u}{\partial\theta^u}\overline{f}_{1N}(\theta_0) - E\frac{\partial^u}{\partial\theta^u}\overline{f}_{1N}(\theta_0))(\widehat{\theta}_N - \theta_0)^u\\ &\quad +\frac{1}{d_2!}\sum_{i\in b_j}(\frac{\partial^{d_2}}{\partial\theta^{d_2}}f_1(\widetilde{X}_i, \theta_N^+) - \frac{\partial^{d_2}}{\partial\theta^{d_2}}f_1(\widetilde{X}_i, \theta_0))(\widehat{\theta}_N - \theta_0)^{d_2}\end{aligned}$$

$$-\frac{1}{d_2!}\ell(\frac{\partial^{d_2}}{\partial\theta^{d_2}}\overline{f}_{1N}(\theta_N^+) - \frac{\partial^{d_2}}{\partial\theta^{d_2}}\overline{f}_{1N}(\theta_0))(\widehat{\theta}_N - \theta_0)^{d_2}$$
$$= \lambda_{1j} + \lambda_{2N} + \lambda_{3j} + \lambda_{4N}, \quad (5.82)$$

where θ_N^+ lies between $\widehat{\theta}_N$ and θ_0 and $\lambda_{1j}, \lambda_{2N}, \lambda_{3j}$, and λ_{4N} are defined implicitly.

Equation (5.81) is implied by

$$\lim_{N\to\infty} N^a P(N^{\xi+\alpha(m)-m/2} b|\mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} Y_j^{k_1} \lambda_{1j}^{k_{21}} \lambda_{3j}^{k_{23}}| \cdot |\lambda_{2N}|^{k_{22}} \cdot |\lambda_{4N}|^{k_{24}} > \varepsilon) = 0 \quad (5.83)$$

for all integers $k_{2i} \geq 0$ for $i = 1, ..., 4$ for which $1 \leq k_2 = \sum_{i=1}^4 k_{2i} \leq m$. Let $\eta = k_{22} + k_{24}$. Equation (5.83) is implied by: for some $\xi_1 > \xi$,

$$\text{(i)} \lim_{N\to\infty} N^a P(N^{\xi-\xi_1\eta+\alpha(m)-m/2} b|\mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} Y_j^{k_1} \lambda_{1j}^{k_{21}} \lambda_{3j}^{k_{23}}| > \varepsilon) = 0,$$
$$\text{(ii)} \lim_{N\to\infty} N^a P(N^{\xi_1}|\lambda_{2N})| > \varepsilon) = 0, \text{ and}$$
$$\text{(iii)} \lim_{N\to\infty} N^a P(N^{\xi_1}|\lambda_{4N})| > \varepsilon) = 0. \quad (5.84)$$

First, we show (ii) holds. Because $q_2 > 6a/(1-2\gamma-2\xi)$, there exists $\xi_1 > \xi$ such that $q_2 > 6a/(1-2\gamma-2\xi_1)$. Result (ii) follows from $\lim_{N\to\infty} N^a P(N^{\xi_1}\ell|(\partial^u/\partial\theta^u)\overline{f}_{1N}(\theta_0) - E(\partial^u/\partial\theta^u)\overline{f}_{1N}(\theta_0)| > \varepsilon) = 0$ for all $u = 0, ..., d_2$ and $\lim_{N\to\infty} N^a P(|\widehat{\theta}_N - \theta_0| > \varepsilon) = 0$. The former holds by Lemma 1(a) with $c = \xi_1 + \gamma$ and $p = q_2$. The latter holds by Lemmas 3 and 4 with $c = 0$.

Result (iii) of (5.84) follows from $\lim_{N\to\infty} N^a P(N^{-1}\sum_{i=1}^N C_{\partial f}(\widetilde{X}_i) > \varepsilon) = 0$ and $\lim_{N\to\infty} N^a P(N^{\xi_1}\ell|\widehat{\theta}_N - \theta_0|^{d_2+1} > \varepsilon) = 0$. The former holds by Lemma 1(b) because $EC_{\partial f}^{2a+2}(\widetilde{X}_i) < \infty$. The latter holds by Lemmas 3 and 4 with c such that $q_1 > 2a/(1-2c)$ and $d_2 > -1 + (a+\gamma+\xi)/c$, because the latter implies that $d_2 \geq -1 + (a+\gamma+\xi_1)/c$ for some $\xi_1 > \xi$ and $c \geq (\gamma+\xi_1)/(d_2+1)$.

To show (i) of (5.84), we consider the cases $k_{23} = 0$ and $k_{23} > 0$ separately. First, suppose $k_{23} = 0$. It suffices to show that for all $u = 1, ..., d_2$ and some $c_2 \geq 0$,

$$\text{(a)} \lim_{N\to\infty} N^a P(N^{c_2}|\widehat{\theta}_N - \theta_0| > \varepsilon) = 0,$$
$$\text{(b)} \lim_{N\to\infty} N^a P(N^{\xi-\xi_1\eta+\alpha(m)-m/2-c_2k_{21}}|b\mathcal{N}_\ell^{-1}\sum_{j=1}^{\mathcal{N}_\ell}(Y_j^{k_1}R_j^{k_{21}} - EY_j^{k_1}R_j^{k_{21}})| > \varepsilon) = 0,$$
$$\text{(c)} \lim_{N\to\infty} N^{\xi-\xi_1\eta+\alpha(m)-m/2-c_2k_{21}} bEY_j^{k_1}R_j^{k_{21}} = 0, \text{ where}$$
$$R_j = \sum_{i\in b_j}(\frac{\partial^u}{\partial\theta^u}f_1(\widetilde{X}_i,\theta_0) - E\frac{\partial^u}{\partial\theta^u}f_1(\widetilde{X}_i,\theta_0)). \quad (5.85)$$

Result (a) holds by Lemmas 3 and 4 with c_2 defined below.

Result (b) holds by applying Markov's inequality and analogues of (5.67) and (5.69), as in the proofs above that $B_1 = 0$ and $B_2 = 0$, provided $a+p(\xi-\xi_1\eta+\alpha(m)-$

$(m-1)/2$ $-c_2 k_{21}$ $+(k_1+k_{21}-1)\gamma/2) < 0$, where p is such that $(k_1+k_{21})p+4\delta = q_2$ for some $\delta > 0$. We drop the term $-c_2 k_{21}$, which makes the condition more stringent. With k_{21} dropped, the left-hand side of the condition is decreasing in η, noting that $k_1 + k_{21} = m - \eta$. In consequence, the least favorable choice of η is zero and the condition becomes $p > a/\{-\xi - \alpha(m)$ $+(m-1)(1-\gamma)/2\}$. Because $q_2 = mp + 4\delta$ for $\delta > 0$ arbitrarily small, the latter condition is equivalent to $q_2 > ma/\{-\xi - \alpha(m)$ $+(m-1)(1-\gamma)/2\}$. This inequality is less restrictive than (5.71), so result (b) holds.

For result (c), we have $|EY_j^{k_1} R_j^{k_{21}}| \leq C\ell^{(k_1+k_{21})/2}$ by Holder's inequality and the Yokoyama-Doukhan moment inequality. Hence, result (c) holds provided $\xi - \xi_1\eta +$ $\alpha(m)$ $-m/2$ $-c_2 k_{21} + 1$ $-\gamma$ $+(k_1+k_{21})\gamma/2 < 0$. Given that $k_1 + k_{21} = m - \eta$, the least favorable choice of η for any fixed value of k_{21} is the minimum possible value. When $k_{21} > 0$, the minimum value is $\eta = 0$. In this case, the previous condition becomes $\xi + \alpha(m)$ $-(m-2)(1-\gamma)/2$ $-c_2 k_{21} < 0$. The least favorable value of k_{21} is one, so the condition becomes $\xi + \alpha(m)$ $-(m-2)(1-\gamma)/2$ $-c_2 < 0$. The left-hand side is decreasing in m for m odd and for m even. Hence, the least favorable value of m is either two or three. It is easily seen to be three and the requisite condition is $c_2 > \gamma/2 + \xi$, where c_2 must satisfy $q_1 > 2a/(1-2c_2)$ for Lemmas 3 and 4 to hold for result (a). Because $q_2 > 6a/(1-2\gamma-2\xi)$, there exists $\phi > 0$ sufficiently small that $q_2 > 6a/(1-2\gamma-2\xi-2\phi)$. Take $c_2 = \gamma/2 + \xi + \phi$. Then, $c_2 > \gamma/2 + \xi$ and $q_1 \geq q_2 > 2a/(1-2c_2)$. Next, when $k_{21} = 0$, the minimum value of η is $\eta = 1$, because $k_{21} = k_{23} = 0$ and $\sum_{j=1}^{4} k_{2j} \geq 1$. In this case, $k_1 = m - 1$ and the condition at the beginning of the paragraph becomes $\xi - \xi_1 + \alpha(m)$ $-(m-2)(1-\gamma)/2$ $-\gamma/2 < 0$. The left-hand side is decreasing in m for m odd and for m even. Hence, the least favorable value of m is either two or three. For $m = 2$, the condition is $\xi - \xi_1 - \gamma/2 < 0$ and for $m = 3$, the condition is $\xi - \xi_1 < 0$. Both conditions hold because $\xi_1 > \xi$ and $\gamma \geq 0$.

Next, we establish (i) of (5.84) when $k_{23} > 0$. The least favorable case is when $k_{23} = 1$ and $k_{21} = 0$, because λ_{3j} and λ_{1j} involve multiples of $\widehat{\theta}_N - \theta_0$, and when $\eta = 0$, because this maximizes $-\xi_1\eta$ and $k_1 = m - 1 - \eta$. It suffices to show that

$$\lim_{N\to\infty} N^a P(N^{\xi+\alpha(m)-m/2-c(d_2+1)} b|\mathcal{N}_\ell^{-1} \sum_{j=1}^{\mathcal{N}_\ell} Y_j^{m-1} \sum_{i\in b_j} C_{\partial f}(\widetilde{X}_i)| > \varepsilon) = 0 \text{ and}$$

$$\lim_{N\to\infty} N^a P(N^c|\widehat{\theta}_N - \theta_0| > \varepsilon) = 0. \tag{5.86}$$

The latter condition holds by Lemmas 3 and 4 with c as in the statement of Lemma 14. By Holder's inequality and the Yokoyama-Doukhan inequality,

$$\begin{aligned} E|Y_j|^{m-1} \sum_{i\in b_j} C_{\partial f}(\widetilde{X}_i) &\leq (E|Y_j|^m)^{(m-1)/m} (E|\sum_{i\in b_j} C_{\partial f}(\widetilde{X}_i)|^m)^{1/m} \\ &\leq C\ell^{(m-1)/2}\ell = CN^{(m+1)\gamma/2}. \end{aligned} \tag{5.87}$$

This and Markov's inequality imply that the first equation of (5.86) holds provided $a + \xi + \alpha(m) - m/2 - c(d_2+1) + 1 - \gamma + (m+1)\gamma/2 < 0$, $E||f(\widetilde{X}_i, \theta_0)||^m < \infty$, and $EC_{\partial f}^m(\widetilde{X}_i) < \infty$ for $m \leq 2a + 2$. The most restrictive case for the first of these three conditions is when $m = 3$. For $m = 3$, we need $d_2 > -1$ $+(a+\gamma+\xi)/c$, as is assumed, see the first paragraph of the proof. □

5.3.15 Proof of Lemma 15

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

$$\lim_{N\to\infty} N^a P(N^\xi ||\widetilde{W}_N - N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} g_{i\ell+j} g'_{i\ell+m}|| > \varepsilon) = 0. \tag{5.88}$$

$$\lim_{N\to\infty} N^a P(N^\xi ||\overline{W}_N(\widehat{\theta}_N) - N^{-1} \sum_{i=1}^{N} [g_i g'_i + \sum_{j=1}^{\kappa} \{g_i g'_{i+j} + g_{i+j} g'_i\}]|| > \varepsilon) = 0. \tag{5.89}$$

$$\lim_{N\to\infty} N^a P(N^\xi ||N^{-1} \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} [g_{i\ell+j} g'_{i\ell+m} - E g_{i\ell+j} g'_{i\ell+m}|| > \varepsilon) = 0. \tag{5.90}$$

$$\begin{aligned} &\lim_{N\to\infty} N^a P(N^\xi ||N^{-1} \sum_{i=1}^{N} [g_i g'_i - E g_i g'_i \\ &\qquad + \sum_{j=1}^{\kappa} \{g_i g'_{i+j} - E g_i g'_{i+j} + g_{i+j} g'_i - E g_{i+j} g'_i\}]|| > \varepsilon) = 0. \end{aligned} \tag{5.91}$$

$$\begin{aligned} &\lim_{N\to\infty} N^{-1+\xi} \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} E g_{i\ell+j} g'_{i\ell+m} - N^{-1+\xi} \sum_{i=1}^{N} [E g_i g'_i + \sum_{j=1}^{\kappa} \{E g_i g'_{i+j} + E g_{i+j} g'_i\}] \\ &= \ 0. \end{aligned} \tag{5.92}$$

Equation (5.91) holds by Lemma 1(a) with $c = \xi$ and $p = q_2$ because $q_2 > 2a/(1-2\xi)$ and $q_2 \geq 2$.

Equation (5.92) holds only if $\xi < \gamma$ when $\sum_{j=1}^{\kappa} j(E g_1 g'_{1+j} + E g_{1+j} g'_1) \neq 0$. To see this, for notational simplicity, suppose g_i is a scalar. Then, the left-hand side of (5.92) equals

$$\begin{aligned} &\lim_{N\to\infty} N^{-1+\xi} b \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} E g_j g_m - N^\xi [E g_1^2 + 2 \sum_{j=1}^{\kappa} E g_1 g_{1+j}] \\ &= \lim_{N\to\infty} N^{-\gamma+\xi} [\ell E g_1^2 + 2 \sum_{j=1}^{\kappa} (\ell - j) E g_1 g_{1+j}] - N^\xi [E g_1^2 + 2 \sum_{j=1}^{\kappa} E g_1 g_{1+j}] \\ &= -2 \lim_{N\to\infty} N^{-\gamma+\xi} \sum_{j=1}^{\kappa} j E g_1 g_{1+j} = 0. \end{aligned} \tag{5.93}$$

Let $A_N = \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} (g_{i\ell+j} g'_{i\ell+m} - E g_{i\ell+j} g'_{i\ell+m})$. Equation (5.90) holds by the following argument with $p = -\delta + q_2/2$ for some small $\delta > 0$:

$$N^a P(||N^{-1+\xi} A_N|| > \varepsilon) \ \leq \ N^{a+\xi p - p} E||A_N||^p / \varepsilon^p$$

$$
\begin{aligned}
&\leq CN^{a+\xi p-p}b^{p/2}(E||\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(g_j g_m' - Eg_j g_m')||^{p+\delta})^{p/(p+\delta)} \\
&\leq CN^{a+\xi p-p}b^{p/2}(2^{p+\delta}E||\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}g_j g_m'||^{p+\delta})^{p/(p+\delta)} \\
&= CN^{a+\xi p-p/2-\gamma p/2}(E||\sum_{j=1}^{\ell}g_j||^{2(p+\delta)})^{p/(p+\delta)} \\
&\leq CN^{a+\xi p-p/2-\gamma p/2}\ell^p \\
&= CN^{a-p(1-\gamma-2\xi)/2} = o(1) \qquad (5.94)
\end{aligned}
$$

where the four inequalities hold by Markov's inequality, the Yokoyama–Doukhan strong mixing moment inequality of (5.65), Minkowski's inequality, and the Yokoyama–Doukhan inequality using the assumption that $E||f_j||^{2(p+\delta)} = E||f_j||^{q_2} < \infty$, respectively, and the last equality holds because $p = -\delta + q_2/2 > 2a/(1-\gamma-2\xi)$ for some $\delta > 0$.

Next, we establish (5.88). For notational simplicity, suppose θ is a scalar. A Taylor expansion of order d about θ_0 gives

$$
\begin{aligned}
\widetilde{W}_N \;=\; & N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}[g^*(X_{i\ell+j},\theta_0)g^*(X_{i\ell+m},\theta_0)' \\
&+\sum_{s=1}^{d-1}(s!)^{-1}(\partial^s/\partial\theta^s)(g^*(X_{i\ell+j},\theta_0)g^*(X_{i\ell+m},\theta_0)')(\widehat{\theta}_N-\theta_0)^s \\
&+(d!)^{-1}(\partial^d/\partial\theta^d)(g^*(X_{i\ell+j},\theta_N^+)g^*(X_{i\ell+m},\theta_N^+)')(\widehat{\theta}_N-\theta_0)^d], \qquad (5.95)
\end{aligned}
$$

where θ_N^+ lies between $\widehat{\theta}_N$ and θ_0. The first two summands on the right-hand side of (5.95) are comprised of terms of the form $C(\widehat{\theta}_N-\theta_0)^{s_1+s_2}$ multiplied by

$$
N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(\partial^{s_1}/\partial\theta^{s_1})g^*(X_{i\ell+j},\theta_0)(\partial^{s_2}/\partial\theta^{s_2})g^*(X_{i\ell+m},\theta_0)' \qquad (5.96)
$$

for $0 \leq s_1+s_2 \leq d-1$. Let $g_i^c = (\partial^{s_c}/\partial\theta^{s_c})g(X_i,\theta_0)$ and $\overline{g}_N^c = N^{-1}\sum_{i=1}^{N}(\partial^{s_c}/\partial\theta^{s_c})$ $g(X_i,\theta_0)$ for $c = 1, 2$. The term in (5.96) can be rewritten as

$$
N^{-1}\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(g_{i\ell+j}^1-\overline{g}_N^1)(g_{i\ell+m}^2-\overline{g}_N^2)' = \sum_{i=1}^{5}K_{iN}, \text{ where}
$$

$$
\begin{aligned}
K_{1N} &= N^{-1}\textstyle\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(g_{i\ell+j}^1-Eg_{i\ell+j}^1)(g_{i\ell+m}^2-Eg_{i\ell+m}^2)' \\
&\quad -E(g_{i\ell+j}^1-Eg_{i\ell+j}^1)(g_{i\ell+m}^2-Eg_{i\ell+m}^2)', \\
K_{2N} &= \ell^{-1}\textstyle\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}E(g_j^1-Eg_j^1)(g_m^2-Eg_m^2)', \\
K_{3N} &= \ell(\overline{g}_N^1-Eg_1^1)(\overline{g}_N^2-Eg_1^2)', \qquad (5.97)\\
K_{4N} &= -(\overline{g}_N^1-Eg_1^1)N^{-1}\ell\textstyle\sum_{i=0}^{b-1}\sum_{m=1}^{\ell}(g_{i\ell+m}^2-Eg_{i\ell+m}^2)', \text{ and} \\
K_{5N} &= -N^{-1}\ell\textstyle\sum_{i=0}^{b-1}\sum_{j=1}^{\ell}(g_{i\ell+j}^1-Eg_{i\ell+j}^1)(\overline{g}_N^2-Eg_1^2)'.
\end{aligned}
$$

For the case where $s_1 = s_2 = 0$, $K_{1N} + K_{2N}$ equals the second term in the norm in (5.88) and, hence, these terms cancel in (5.88) and we only need to consider K_1 and K_2 when $s_1 + s_2 > 0$.

First, we consider K_{1N}. Let $f_i = g_i^1 - Eg_i^1$ and $h_i = g_i^2 - Eg_i^2$. Let $p = -\delta + q_2/2$ for some small $\delta > 0$. We have: for all $\varepsilon > 0$,

$$\begin{aligned}
& N^a P(N^\xi ||K_{1N}|| > \varepsilon) \\
&= N^a P(N^{-1+\xi} || \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} (f_{i\ell+j} h'_{i\ell+m} - E f_{i\ell+j} h'_{i\ell+m}) || > \varepsilon) \\
&\leq C N^{a+\xi p - p} E || \sum_{i=0}^{b-1} \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} (f_{i\ell+j} h'_{i\ell+m} - E f_{i\ell+j} h'_{i\ell+m}) ||^p \\
&\leq C N^{a+\xi p - p} b^{p/2} (E || \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} (f_j h'_m - E f_j h'_m) ||^{p+\delta})^{p/(p+\delta)} \\
&\leq C N^{a+\xi p - p/2 - \gamma p/2} (2^{p+\delta} E || \sum_{j=1}^{\ell} \sum_{m=1}^{\ell} f_j h'_m ||^{p+\delta})^{p/(p+\delta)} \\
&= C N^{a+\xi p - p/2 - \gamma p/2} (E || \sum_{j=1}^{\ell} f_j ||^{p+\delta} || \sum_{m=1}^{\ell} h_m ||^{p+\delta})^{p/(p+\delta)} \\
&\leq C N^{a+\xi p - p/2 - \gamma p/2} (E || \sum_{j=1}^{\ell} f_j ||^{2(p+\delta)} E || \sum_{m=1}^{\ell} h_m ||^{2(p+\delta)})^{p/(2p+2\delta)} \\
&\leq C N^{a+\xi p - p/2 - \gamma p/2} \ell^p = C N^{a - p(1/2 - \xi - \gamma/2)} = o(1) \qquad (5.98)
\end{aligned}$$

where the five inequalities hold by Markov's inequality, the Yokoyama–Doukhan strong mixing moment inequality as in (5.65), Minkowski's inequality, the Cauchy-Schwartz inequality, and the Yokoyama–Doukhan inequality using the assumption that $E||f_j||^{2(p+\delta)} = E||f_j||^{q_2} < \infty$, respectively, and the last equality holds because $p = -\delta + q_2/2 > 2a/(1 - 2\xi - \gamma)$ for some $\delta > 0$.

For the case where $s_1 + s_2 > 0$, $\limsup_{N\to\infty} K_{2N} < \infty$ by a strong mixing covariance inequality. Hence, $\lim_{N\to\infty} N^a P(||K_{2N}|| \cdot N^\xi ||\widehat{\theta}_N - \theta_0||^{s_1+s_2} > \varepsilon) = 0$ using Lemma 3 or Lemma 4 with $c = \xi$ because $q_2 > 2a/(1 - \gamma - 2\xi)$ implies $q_1 > 2a/(1 - 2\xi)$.

Next, $\lim_{N\to\infty} N^a P(N^\xi ||K_{3N}|| > \varepsilon) = 0$, because $\lim_{N\to\infty} N^a P(N^{\xi/2} \ell^{1/2} ||\overline{g}_N^j - Eg_i^j|| > \varepsilon) = 0$ for $j = 1, 2$ using Lemma 1(a) with $c = (\xi + \gamma)/2$ since $\xi + \gamma < 1$. We have $\lim_{N\to\infty} N^a P(N^\xi ||K_{jN}|| > \varepsilon) = 0$ for $j = 4, 5$ by the arguments used for K_{jN} for $j = 1, 3$. This establishes the desired result for the first two terms of (5.95).

Let $\widetilde{W}_{3N}$ denote the third term of (5.95). The desired result $\lim_{N\to\infty} N^a P(N^\xi ||\widetilde{W}_{3N}|| > \varepsilon) = 0$ follows from

$$\lim_{N\to\infty} N^a P(N^{c(d+1)} ||\widehat{\theta}_N - \theta_0||^{d+1} > \varepsilon) = 0 \text{ for all } \varepsilon > 0 \text{ and}$$

$$\lim_{N\to\infty} N^a P(N^{\xi-c(d+1)-1} \sum_{i=0}^{b-1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} C_{\partial f}(X_{i\ell+j})C_{\partial f}(X_{i\ell+m}) > K_\varepsilon)$$
$$\leq \lim_{N\to\infty} CN^{a+\xi-c(d+1)-1}\ell^2 bEC^2_{\partial f}(\widetilde{X}_i) < \infty, \quad (5.99)$$

where the first result holds by Lemma 3 or 4, the first inequality of the second result holds by Markov's inequality, and the second inequality of the second result holds because $\gamma \leq c(d+1) - a - \xi$. This completes the proof of (5.88).

The proof of (5.89) is similar to that of (5.88), but simpler. For brevity, we do not give the details.

Next, we prove part (a) for the overlapping block bootstrap. The desired result follows from (5.88)–(5.92) with $N^{-1}\sum_{i=0}^{b-1}$ replaced by $N^{-1}b(N-\ell+1)^{-1}\sum_{i=0}^{N-\ell}$ and $g_{i\ell+j}g'_{i\ell+m}$ replaced by $g_{i+j}g'_{i+m}$ in (5.88), (5.90), and (5.92). Equations (5.89) and (5.91) have already been established.

The analogue of (5.92) is established as follows. Some calculations show that

$$N^{-1}b(N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell} Eg_{i+j}g'_{i+m}$$
$$= (N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1}[v_N(i,0,\ell)Eg_ig'_i + \sum_{j=1}^{\kappa} v_N(i,j,\ell)(Eg_ig'_{i+j} + Eg_{i+j}g'_i)], \text{ where}$$
$$v(i,j,\ell) = \begin{cases} 1 - j/\ell & \text{for } \ell \leq i \leq N-\ell+1 \\ 1-(\ell-i-j)/\ell & \text{for } 1 \leq i < \ell \\ 1-(\ell+i-N+1-j)/\ell & \text{for } N-\ell+1 < i \leq N. \end{cases} \quad (5.100)$$

In consequence, the analogue of (5.92) holds provided $\lim_{N\to\infty} N^\xi \sum_{j=1}^{\kappa}(j/\ell)(Eg_1g'_{1+j} + Eg_{1+j}g'_1) = 0$, which requires either $\xi < \gamma$ or $\sum_{j=1}^{\kappa} j(Eg_1g'_{1+j} + Eg_{1+j}g'_1) = 0$.

The analogue of (5.90) for overlapping blocks is established as follows:

$$N^a P(N^\xi||N^{-1}b(N-\ell+1)^{-1}\sum_{i=1}^{N-\ell+1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(g_{i+j}g'_{i+m} - Eg_{i+j}g'_{i+m})|| > \varepsilon)$$
$$\leq CN^{a+\xi p-\gamma p-p}E||\sum_{i=1}^{N-\ell+1}\sum_{j=1}^{\ell}\sum_{m=1}^{\ell}(g_{i+j}g'_{i+m} - Eg_{i+j}g'_{i+m})||^p$$
$$= CN^{a+\xi p-\gamma p-p}E||\sum_{u=1}^{\ell}\sum_{i=0}^{b(u)}(\eta_{i\ell+u} - E\eta_1)||^p, \text{ where}$$
$$\eta_i = \sum_{j=1}^{\ell} g_{i-1+j}\sum_{m=1}^{\ell} g'_{i-1+m} \quad (5.101)$$

and $b(u)$ is defined just below (5.68), using Markov's inequality. Note that $\{\eta_{i\ell+u} : i = 0, ..., b(u)\}$ are non-overlapping blocks with strong mixing coefficients less than or equal to those of $\{X_i : i \geq 1\}$. Let $p = -\delta + q_2$ for some small $\delta > 0$. Next, we have

$$
\begin{aligned}
(E||\sum_{u=1}^{\ell}\sum_{i=0}^{b(u)}(\eta_{i\ell+u} - E\eta_1)||^p)^{1/p} &\leq \sum_{u=1}^{\ell}(E||\sum_{i=0}^{b(u)}(\eta_{i\ell+u} - E\eta_1)||^p)^{1/p} \\
\leq \sum_{u=1}^{\ell}(Cb^{p/2}(E||\eta_u - E\eta_1||^{p+\delta})^{p/(p+\delta)})^{1/p} &\leq C\ell b^{1/2}(2^{p+\delta}E||\eta_1||^{p+\delta})^{1/(p+\delta)} \\
= C\ell b^{1/2}(E||\sum_{j=1}^{\ell} g_j||^{2(p+\delta)})^{1/(p+\delta)} &\leq C\ell^2 b^{1/2}, \qquad (5.102)
\end{aligned}
$$

where the four inequalities hold by Minkowski's inequality, the Yokoyama–Doukhan strong mixing moment inequality as in (5.65) using the strong mixing property of $\{\eta_{i\ell+u} : i = 0, ..., b(u)\}$, Minkowski's inequality, and the Yokoyama–Doukhan inequality using the assumption that $E||g_i||^{q_2} < \infty$, respectively.

Using (5.102), the left-hand side of (5.101) is less than or equal to $CN^{a+\xi p-\gamma p-p}\ell^{2p}b^{p/2} = CN^{a-p(1/2-\xi-\gamma/2)} = o(1)$ because $p = -\delta + q_2 > 2a/(1-2\xi-\gamma)$ for some $\delta > 0$.

Lastly, we establish the analogue of (5.88) for overlapping blocks. We use the same method as for non-overlapping blocks. Equations (5.95)–(5.97) hold with $N^{-1}\sum_{i=0}^{b-1}$, $X_{i\ell+j}$, and $g_{i\ell+j}$ replaced by $N^{-1}b(N-\ell+1)^{-1}\sum_{i=0}^{N-\ell}$, X_{i+j}, and g_{i+j} respectively. The analogue of (5.98) is proved by an argument analogous to that used above to prove the analogue of (5.90). The treatment of the analogues of the terms $K_{2N}, ..., K_{5N}$ is the same as for non-overlapping blocks, but with sums of the form $\sum_{i=1}^{N-\ell+1}\zeta_i$ broken up into sums of the form $\sum_{u=1}^{\ell}\sum_{i=0}^{b(u)}\zeta_{i\ell+u}$ as in (5.101). This completes the proof of part (a).

Next, the result of part (b) for t statistics (i.e., $s = T$) follows from Lemma 14 and

$$\lim_{N\to\infty} N^a P(N^\xi||\tau_{N,r} - 1|| > \varepsilon) = 0 \text{ for all } \varepsilon > 0. \qquad (5.103)$$

Equation (5.103) follows from part (a) of the present Lemma, $\lim_{N\to\infty} N^a P(||\widetilde{\sigma}_N - \sigma|| > \varepsilon) = 0$ for all $\varepsilon > 0$, which holds by (5.42) and (5.43), and the fact that σ is positive definite by Assumption 3(b).

The result of part (b) of the Lemma for the $\mathcal{W}_N$ statistic follows from Lemma 14,

$$\lim_{N\to\infty} N^a P(N^\xi||\Xi_N - I_{L_\eta}|| > \varepsilon) = 0 \text{ for all } \varepsilon > 0, \qquad (5.104)$$

and the properties of the function $\lambda_{\mathcal{W}}(\cdot,\cdot)$. Equation (5.104) holds by the same argument as for (5.103).

Now, we establish the result of part (b) of the Lemma for the J statistic. We have

$$
\begin{aligned}
&\lim_{N\to\infty} N^a P(N^\xi||M_N - M_0|| > \varepsilon) = 0, \quad \lim_{N\to\infty} N^a P(N^\xi||V_N - M_0|| > \varepsilon) = 0, \\
&\lim_{N\to\infty} N^a P(N^\xi||V_N^+ - M_0|| > \varepsilon) = 0, \text{ and } \lim_{N\to\infty} N^a P(N^\xi||(V_N^+)^{1/2} - M_0|| > \varepsilon) = 0. \qquad (5.105)
\end{aligned}
$$

The first result of (5.105) follows from (5.42) and (5.43), the second result follows from the first result and part (a) of the Lemma, the third result follows from the second result, $M_0 = M_0^+$ (because M_0 is a projection matrix), and the fact that $rk(V_N) = rk(M_0) = L_g - L_\theta$ with probability $1 - o(N^{-a})$, see Andrews (1987, Theorem 2), and the fourth result follows from the third result and $M_0^{1/2} = M_0$. Now, part (b) of the Lemma for the J statistic follows from the fourth result of (5.105), Lemma 14, and the properties of the function $\lambda_J(\cdot, \cdot)$. □

5.3.16 Proof of Lemma 16

We use the same method as HH use in the proof of their Theorems 1 and 2. Because their description is very brief, we describe the method in a little more detail than HH do. Given Lemma 13, for the results of parts (a) and (b) regarding T_N and T_N^*, it suffices to show that $N^{1/2}G(S_N)$ and $N^{1/2}\tau_{N,r}G(S_N^*)$ of Lemma 13 possess Edgeworth expansions with remainder $o(N^{-a})$. For the case of $N^{1/2}G(S_N)$, this follows by applying Theorem 3.1 of Bhattacharya (1987) with his integer parameter s satisfying $(s-2)/2 = a$ for a given in the present Lemma (with $2a$ being an integer) and with the normalized sample average $N^{1/2}(\overline{X} - \mu)$ of the underlying random variables in his theorem satisfying an Edgeworth expansion not because they are iid and satisfy his condition (A_4), but because they are asymptotically weakly dependent and satisfy the conditions of Theorem 1.1 of Götze and Hipp (1994). The latter theorem is a special case of Corollary 2.9 of Götze and Hipp (1983). Conditions (2)–(4) of Götze and Hipp (1994) hold by Assumptions 1 and 4 and $q_2 \geq 2a + 3$. Conditions $(A_1) - (A_3)$ of Bhattacharya (1987) hold by $q_2 \geq 2a + 3$, the fact that $G(\cdot)$ is infinitely differentiable, and Assumption 3(b) respectively.

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

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

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

Footnotes

[1] The author is very grateful to Joel Horowitz for his helpful comments. The author thanks Carol Copeland for proofreading the manuscript. The author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277.

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

[3] The r-th element of θ_N^* is denoted $(\theta_N^*)_r$, rather than $\theta_{N,r}^*$, to distinguish it from the k-step bootstrap estimator, $\theta_{N,k}^*$ defined in Section 2.3.

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

[5] This specification of the log likelihood does not utilize the first κ observations except as conditioning variables in order to maintain the stationarity of the summands $\{\rho(\widetilde{X}_i, \theta) : i = 1, ..., N\}$.

[6] We have altered the notation slightly from the previous section in that we allow $\rho(\cdot, \theta)$ and $g(\cdot, \theta)$ to depend on $\widetilde{X}_i$ here, rather than just X_i. This is an innocuous difference.

[7] For both the unrestricted and restricted parametric bootstraps, θ_N^* is defined to minimize the bootstrap criterion function over $\theta \in \Theta$—not over $\theta \in \Theta$ subject to the restrictions $\eta(\theta) = 0$. The only place where the restricted ML estimator is used is in generating the bootstrap sample for the restricted parametric bootstrap.

References

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

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

————(1999): "Estimation When a Parameter Is on the Boundary of the Parameter Space," *Econometrica,* 67, 1341–1383.

————(2001): "Equivalence of the Higher-order Asymptotic Efficiency of k-step and Extremum Statistics," Cowles Foundation Discussion Paper, Yale University. Available at www.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: Wiley, 157–170.

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

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

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

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

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

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

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

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

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

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

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

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

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

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and Its Application.* New York: Academic Press.

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

Inoue, A. and M. Shintani (2000): "Bootstrapping GMM Estimators for Time Series," unpublished working paper, Department of Economics, Vanderbilt University.

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

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

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

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

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

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

Michel, R. (1974): "Results on Probabilities of Moderate Deviations," *Annals of Probability*, 2, 349–353.

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

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

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

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

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

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