On bootstrap coverage probability with dependent data

By Jānis J. Zvingelis*

Department of Finance, The University of Iowa, Iowa City, Iowa 52242, U.S.A.

Abstract

This paper establishes that the minimum error rates in coverage probabilities of one- and symmetric two-sided block bootstrap confidence intervals are of orders $\mathcal{O}(n^{-3/4})$ and $\mathcal{O}(n^{-4/3})$, respectively, for normalized and studentized smooth functions of sample averages. The block lengths that minimize the error in coverage probabilities of one- and symmetric two-sided block bootstrap confidence intervals are proportional to $n^{1/4}$ and $n^{1/3}$, respectively. Existing literature provides Monte Carlo evidence that such small improvement over the coverage precision of asymptotic confidence intervals is to be expected.

1. Introduction

The bootstrap, introduced by Efron (1979), is a statistical procedure for estimating the distribution of a given estimator. The distinguishing feature of the bootstrap is that it replaces the unknown population distribution of the data by an estimate of it, which is formed by resampling the original sample randomly with replacement.

*I thank Professor Joel Horowitz for his indispensable comments and continued support. All the errors are my own.
E-mail: janis-zvingelis@uiowa.edu.

When the observed data form a random sample, the bootstrap often provides more accurate critical values for the tests than asymptotic theory (e.g., Beran, 1988, Hall, 1986, 1992, Singh, 1981). Hall (1988) proves that the error in coverage probability made by symmetric two-sided confidence intervals, when bootstrap critical values are used in the IID case, is of order $\mathcal{O}(n^{-2})$. This amounts to sizable refinement[1] for the precision of asymptotic confidence intervals, since the errors made by one- and symmetric two-sided asymptotic confidence intervals are of orders $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(n^{-1})$, respectively. Monte Carlo experiments support the predictions of the theory, sometimes producing spectacular results (Horowitz, 1994).

In the case of dependent data the bootstrap procedure must be designed in a way that suitably captures the dependence structure of the original sample. Several different sampling procedures have been invented to tackle this task. Carlstein (1986) proposes to divide the original data set in non-overlapping blocks and then sample these blocks randomly with replacement. Künsch (1989) proceeds similarly, except he divides the original sample in overlapping blocks. Hall (1985) also suggested these techniques in the context of spatial data. Despite the blocking, the dependence structure of the original sample is not replicated exactly in the bootstrap sample. For example, if non-overlapping blocks are used, the observations from different blocks in the bootstrap sample are independent with respect to the probability measure induced by bootstrap sampling. Furthermore, observations from the same block are deterministically related. Lastly, the block bootstrap sample is non-stationary even if the original sample is stationary. This dependence structure is unlikely to be present in the original sample. As a result bootstrap performance deteriorates. Hall and Horowitz (1996) give conditions under which Carlstein's block bootstrap provides asymptotic refinements through $\mathcal{O}(n^{-1})$

[1] By "refinement through $O(n^{-r})$" we mean that the estimated parameter of interest is correct up to and including the term of order $O(n^{-r})$, and the estimation error is of size $o(n^{-r})$.

for coverage probabilities, when bootstrap critical values are used to construct symmetric two-sided confidence intervals for Generalized Method of Moments (GMM) estimators.

The random variables of interest in this paper are standardized and studentized smooth functions of sample moments of $\tilde{X}$ or sample moments of functions of $\tilde{X}$, where $\tilde{X}$ denotes the sample. For this broad class of random variables we have established the following result: the errors made in the coverage probabilities by one- and symmetric two-sided block bootstrap confidence intervals are of orders $\mathcal{O}(n^{-3/4})$ and $\mathcal{O}(n^{-4/3})$, respectively, when optimal block lengths are used. The optimal block lengths are equal to[2] $C_1 n^{1/4}$ and $C_2 n^{1/3}$ for one- and symmetric two-sided confidence intervals, respectively. Note, however, that the improvement from using the bootstrap over the asymptotics leaves much to be desired, especially in the symmetric two-sided confidence interval case. The lackluster performance of the block bootstrap, as demonstrated in the Monte Carlo experiments of Hall and Horowitz (1996) and Hansen (1999), among others, is consistent with the theory established in the present paper.

To achieve asymptotic refinement, the Edgeworth expansions[3] of the statistic of interest and its bootstrap equivalent have to have the same structure apart from replacing bootstrap cumulants with sample cumulants in the bootstrap expansion. Lahiri (1992) and Hall and Horowitz (1996) proposed "corrected" bootstrap estimators that achieve asymptotic refinement and partially account for the

[2]For convenience and simplicity we will employ C and/or C_i, $i = 1, 2, \ldots$ to denote some finite constants that depend on the specifics of the data generation process, but not on sample size, n. These constants may assume different values at each appearance.

[3]An *Edgeworth expansion* is an approximation to the distribution function of a random variable. Under certain assumptions Edgeworth expansion takes on the form of power series in n^{-r}, where the first term is the standard Normal distribution function and r depends on the type of a random variable. The power series form of an Edgeworth expansion makes it a convenient tool for determining the size of the error made by an estimator of a finite sample distribution function. See Hall (1992) for a detailed discussion on Edgeworth expansions.

change in the dependence structure in the bootstrap sample. The corrected versions of bootstrap test statistics are also used in this paper. The point of the correction factor is to make the exact variance of the leading term of the Taylor series expansion of the bootstrap test statistic equal to one and to do this without introducing new (bootstrap) stochastic terms that would affect the structure of the Edgeworth expansion.

An enlightening fact to note is that one does not need correction factors in one-sided confidence interval case to achieve asymptotic refinement through $\mathcal{O}(n^{-1/2})$ (see, for example, Lahiri, 1992, Davison and Hall, 1993, Götze and Künsch, 1996, and Lahiri, 1996). The reason for this is that the differences between the population and bootstrap variances of higher order terms of the Taylor series expansions of the random variable of interest are of order smaller than $\mathcal{O}(n^{-1/2})$.

The solution method used in this paper was introduced in Hall (1988). A crucial prerequisite for the usage of this technique is the existence of the Edgeworth expansions for the random variable of interest. Götze and Hipp (1983, 1994) and Götze and Künsch (1996) give regularity conditions under which the Edgeworth expansions exist for smooth functions of sample averages in the dependent data setting. However, the method of solution of this paper involves terms that are not smooth functions of sample moments. These terms are of the form $(1/b)^m \sum_{i=1}^{b}(\sum_{j=1}^{l} X_{ij}/l)^m$. At present time there are no Edgeworth expansion results in the literature that apply to the statistics of above type in the dependent data setting. Thus, the techniques used and results obtained in this paper are heuristic.

The paper is organized as follows: section 2 introduces the test statistics of interest, section 3 lays out the main theoretical results, and section 4 concludes. This is followed by the appendix containing the relevant mathematical derivations.

2. Test statistics

Notation will largely follow that used in Hall, et al (1995) and Hall and Horowitz (1996).

2.1. Sample

The test statistics of interest in this paper are either normalized or studentized smooth functions of sample moments of $\tilde{X}$ or sample moments of functions of $\tilde{X}$. Many test statistics and estimators are smooth functions of sample averages or can be approximated by such with negligible error. Test statistics based on GMM estimators constitute an example of the latter case (Hall and Horowitz, 1996, Proposition 1, 2).

Denote the data by $\mathcal{X} = (X_1, \ldots, X_{n_{full}})$, where $X_i \in \mathcal{R}^d$ is a $d \times 1$ random variable. Assume that $\{X_i\}$ is an ergodic stationary strongly mixing stochastic process and that $EX_iX_j' = 0$ if $|i-j| > k$ for some integer $k < \infty$.[4] Set $n = n_{full} - k$. Define the sample as $\tilde{\mathcal{X}} \equiv \{\tilde{X}_i : i = 1, \ldots, n\}$, where $\tilde{X}_i = \{X_i', \ldots, X_{i+k}'\}'$. We need to make this redefinition of the sample so that the consistent estimator of the asymptotic variance is a smooth function of sample moments; specifically, the problem lies in the cross-product components of the covariance estimator.

[4] Andrews (1991) proposes a non-parametric covariance matrix estimator, but it is not a smooth function of sample moments and converges at a rate that is slower than $n^{-1/2}$. Existing theory on Edgeworth expansions with dependent data (Götze and Hipp (1983, 1994)) applies only to smooth functions of sampe moments. Note, also, that the above restriction is not equivalent to the assumption of m-dependence. Lastly, this assumption is not as restrictive as it might seem, because X_i, for example, can be equal to the GMM moment function. In that case the data can be dependent with an infinite lag, given that some assumptions are satisfied (see Hall and Horowitz, 1996).

2.2. Carlstein's blocking rule

Let b, l denote integers such that $n = bl$. Carlstein's rule divides the sample $\tilde{\mathcal{X}}$ in b disjoint blocks, where the k^{th} block is $\mathcal{B}_k = (\tilde{X}_{(k-1)l+1}, \ldots, \tilde{X}_{kl})$ for $1 \leq k \leq b$. According to the Carlstein's rule bootstrap sample $\tilde{\mathcal{X}}^*$ is formed by choosing b blocks randomly with replacement out of the set of blocks formed from the original sample and laying the chosen blocks side by side in the order that they are chosen. Bootstrap sample $\tilde{\mathcal{X}}^*$ then consists of $\{\tilde{X}_i^*\} = \{(X_i^{*\prime}, \ldots, X_{i+k}^{*\prime})' : i = 1, \ldots, n\}$.

2.3. Normalized statistic

Let us denote the random variable of interest by $U_N = (\hat{\theta} - \theta)/s$, where $\hat{\theta} = f(\bar{X})$, $\theta = f(E(X))$, $s = (V(\hat{\theta} - \theta))^{1/2}$, $V(\cdot)$ is an exact variance, and $f(\cdot) : \mathcal{R}^d \to \mathcal{R}$ is a smooth function of sample moments of $\tilde{\mathcal{X}}$ or sample moments of functions of $\tilde{\mathcal{X}}$.

Let U_N^* denote the bootstrap equivalent of U_N, where $U_N^* = (\hat{\theta}^* - \hat{\theta})/\tilde{s}$, $\hat{\theta}^* = f(\bar{X}^*)$, and $\bar{X}^* = n^{-1} \sum X_i^*$ is the resample mean. Define $\tilde{s} = (V'[\hat{\theta}^* - \hat{\theta}])^{1/2}$, where $V'[\hat{\theta}^* - \hat{\theta}] = E'(\hat{\theta}^* - E'[\hat{\theta}^*])^2$. Here $E'[\cdot]$ denotes the expectation induced by the bootstrap sampling, conditional on the sample, $\tilde{\mathcal{X}}$.

Next we define the Edgeworth expansions of U_N and U_N^*:

$$P(U_N < x) - \Phi(x) - n^{-1/2} p_1(x) - n^{-1} p_2(x) = o(n^{-1}),$$

where $p_1(z)$ and $p_2(z)$ are even and odd functions, respectively, both of the functions are polynomials with coefficients depending on cumulants[5] of U_N, and both are of order $\mathcal{O}(1)$.

$$P^*(U_N^* < x) - \Phi(x) - n^{-1/2} \hat{p}_1(x) - n^{-1} \hat{p}_2(x) = o(n^{-1}),$$

[5] Cumulants are defined as the coefficients of $\frac{1}{j!}(it)^j$ terms in a power series expansion of $\log \chi(t)$, where $\chi(t)$ is the characteristic function of a random variable and $\chi(t) = \exp(k_1 it + \frac{1}{2} k_2 (it)^2 + \ldots + \frac{1}{j!} k_j (it)^j + \ldots +)$.

except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-1})$. Here $\hat{p}_1(z)$ and $\hat{p}_2(z)$ are the same polynomials as above only the population cumulants of U_N are replaced by sample cumulants of U_N^*, and $P^*(\cdot)$ is a probability measure (conditional on the sample, $\tilde{\mathcal{X}}$) induced by the bootstrap sampling. Let k_i denote the ith cumulant of U_N. Then,

$$\begin{aligned} n^{-1/2}p_1(x) &= -k_1 - \frac{k_3}{6}(x^2-1) \\ n^{-1}p_2(x) &= -\frac{1}{2}k_1^2 x + \left(\frac{k_4}{24} + \frac{k_1 k_3}{6}\right)(3x - x^3) - \frac{k_3^2}{72}(x^5 - 10x^3 + 15x). \end{aligned}$$

First four cumulants of U_N have the following form (see Appendix):

$$\begin{aligned} k_1 &\equiv E(U_N) = \frac{k_{1,2}}{n^{1/2}} + \frac{k_{1,3}}{n^{3/2}} + \mathcal{O}(n^{-5/2}) \\ k_2 &\equiv E(U_N - E(U_N))^2 = 1 \\ k_3 &\equiv E(U_N - E(U_N))^3 = \frac{k_{3,1}}{n^{1/2}} + \frac{k_{3,2}}{n^{3/2}} + \mathcal{O}(n^{-5/2}) \\ k_4 &\equiv E(U_N - E(U_N))^4 - 3(V(U_N))^2 = \frac{k_{4,1}}{n} + \mathcal{O}(n^{-2}), \end{aligned}$$

where $k_{i,j}$'s are constants that do not depend on n and $E(U_N^2) = \mathcal{O}(1) + \mathcal{O}(n^{-1})$.

Define u_α as $P(U_N < u_\alpha) = \alpha$. Inverting the Edgeworth expansion produces Cornish-Fisher expansion:

$$u_\alpha - z_\alpha - n^{-1/2}p_{11}(z_\alpha) - n^{-1}p_{21}(z_\alpha) = o(n^{-1}),$$

where $0 < \varepsilon < 1/2$.[6]

Similarly, define $\hat{u}_\alpha$ as $P(U_N^* < \hat{u}_\alpha) = \alpha$.[7] Then for $\varepsilon > 0$,

$$\hat{u}_\alpha - z_\alpha - n^{-1/2}\hat{p}_{11}(z_\alpha) - n^{-1}\hat{p}_{21}(z_\alpha) = o(n^{-1}),$$

[6] In the notation $p_{ij}(\cdot)$ (and later $q_{ij}(\cdot)$), i denotes the term in the Cornish-Fisher expansion and j is equal to 1, if u_α is a percentile of a one-sided distribution, and 2, if it is a percentile of a two-sided distribution.

[7] A way to obtain an empirical estimate of $\hat{u}_\alpha$ is to carry out a Monte Carlo experiment that consists of resampling

except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-1})$. Here,

$$\begin{aligned} n^{-1/2}p_{11}(x) &= -n^{-1/2}p_1(x) \\ n^{-1}p_{21}(x) &= n^{-1/2}p_1(x)n^{-1/2}p_1^{\prime}(x) - \frac{1}{2}xn^{-1}p_1(x)^2 - n^{-1}p_2(x), \end{aligned} \quad (1)$$

with obvious modifications for $\hat{p}_{11}(x)$ and $\hat{p}_{21}(x)$.

Let us also introduce some notation for the two-sided distribution function of the normalized test statistic. Noting that $P(|U_N| < x) = P(U_N < x) - P(U_N < -x)$ and that $p_1(x)$ is an even polynomial, the Edgeworth expansions for $|U_N|$ and $|U_N^*|$ take on the following form:

$$P(|U_N| < x) - 2\Phi(x) + 1 - 2n^{-1}p_2(x) = o(n^{-2}),$$

$$P^*(|U_N^*| < x) - 2\Phi(x) + 1 - 2n^{-1}\hat{p}_2(x) = o(n^{-2}),$$

where the latter equality holds except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-2})$.

Define $\xi = \frac{1}{2}(1 + \alpha)$, $P(|U_N| < w_\alpha) = \alpha$, and $n^{-1}p_{12}(\cdot) = -n^{-1}p_2(\cdot)$. Inverting the population Edgeworth expansion we obtain the following Cornish-Fisher expansion:

$$w_\alpha - z_\xi - n^{-1}p_{12}(z_\xi) = o(n^{-2}),$$

where $0 < \varepsilon < 1/2$. Equivalently, define $P(|U_N^*| < \hat{w}_\alpha) = \alpha$ and $n^{-1}\hat{p}_{12}(\cdot) = -n^{-1}\hat{p}_2(\cdot)$, where $\hat{p}_2(\cdot)$ is as $p_2(\cdot)$ with population moments replaced by their sample equivalents. Then

$$\hat{w}_\alpha - z_\xi - n^{-1}\hat{p}_{12}(z_\xi) = o(n^{-2}),$$

except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-2})$.

the original sample $\tilde{\mathcal{X}}$, calculating the bootstrap test statistic U_N^*, and forming the empirical distribution of U_N^* with the desired level of accuracy. The αth quantile of the empirical distribution of the bootstrap test statistic is the empirical estimate of $\hat{u}_\alpha$.

2.4. Studentized statistic

The random variable of interest here is $U_S = (\hat{\theta} - \theta)/\hat{s}$, where $\hat{s}^2$ is a consistent estimate of s^2. The functional forms of s^2 and $\hat{s}^2$ are:[8]

$$s^2 \sim \sum_{i=1}^{d} C_i^2 \cdot V(\bar{X}_i),$$

where $V(\bar{X}_i) = \gamma(0)/n + (2/n)\sum_{j=1}^{k}\gamma(j)\cdot(1 - n^{-1}j)$, $\gamma(j)$ is the jth autocovariance of X, k is the highest lag for non-zero covariance, $f(\cdot) : \mathcal{R}^d \rightarrow \mathcal{R}$, and C_i's are constants that depend on function $f(\cdot)$, but not on n. Also, $\bar{X}_i$ is a sample mean of the ith argument of the function $f(\cdot)$. A consistent estimator of s^2 is given by:

$$\hat{s}^2 \sim \sum_{i=1}^{d} C_i^2 \cdot \hat{V}(\bar{X}_i),$$

where $\hat{V}(\bar{X}_i) = n^{-2}\sum_{j=1}^{n}(X_{ij} - \bar{X}_i)^2 + (2/n)\sum_{m=1}^{k}(1 - n^{-1}m)\sum_{j=1}^{n-m}(X_{ij} - \bar{X}_i)(X_{i,j+m} - \bar{X}_i)/n$ and X_{ij} is the jth element of the sample from the ith argument .

The corrected bootstrap test statistic is $U_S^* = (\hat{s}/\tilde{s})\cdot(\hat{\theta}^* - \hat{\theta})/\hat{s}^*$, where $\hat{s}^{*2}$ is the bootstrap equivalent of $\hat{s}^2$:

$$\hat{s}^{*2} \sim \sum_{i=1}^{d} C_i^2 \cdot \hat{V}(\bar{X}_i^*).$$

Here $\hat{V}(\bar{X}_i^*) = n^{-2}\sum_{j=1}^{n}(X_{ij}^* - \bar{X}_i^*)^2 + (2/n)\sum_{m=1}^{k}(1 - n^{-1}m)\sum_{j=1}^{n-m}(X_{ij}^* - \bar{X}_i^*)(X_{i,j+m}^* - \bar{X}_i^*)/n$, X_{ij}^* is the jth observation of the ith argument of $f(\cdot)$ in the block bootstrap sample, and $\bar{X}_i^*$ is a sample mean of the block bootstrap sample for the ith argument. The exact bootstrap variance of $\hat{\theta}^* - \hat{\theta}$ is

[8] The following variances are expressed in terms of being asymptotically equivalent to something, because, in general, the function $f(\cdot)$ in the random variable of interest, U_S, is not a linear function. To be able to evaluate the variance of the random variable of interest, we have to linearize $f(\cdot)$ using Taylor's theorem.

denoted by $\tilde{s}^2$:

$$\begin{aligned}\tilde{s}^2 &\sim \sum_{i=1}^{d} C_i^2 \cdot V'(\bar{X}_i^* - \bar{X}_i) \\ &= \sum_{i=1}^{d} C_i^2 \frac{1}{b} \sum_{j=1}^{b} \frac{(\bar{X}_{ij} - \bar{X}_i)^2}{b} \\ &= \sum_{i=1}^{d} C_i^2 \frac{1}{n^2} \sum_{j=1}^{b} \sum_{k_1=1}^{l} \sum_{k_2=1}^{l} (X_{ijk_1} - \bar{X})(X_{ijk_2} - \bar{X}),\end{aligned}$$

where $V'(\cdot)$ is the variance induced by block bootstrap sampling, $\bar{X}_{ij}$ is the sample mean of the jth block of the ith argument, X_{ijk_m} is the k_mth observation in the jth block of the ith argument.

Note that the Taylor series expansions of U_S and U_S^* have the following forms:[9]

$$\begin{aligned}U_S &= U_N \times \left[1 - \frac{\hat{s}^2 - s^2}{2s^2} + \frac{3}{8}\left(\frac{\hat{s}^2 - s^2}{s^2}\right)^2 + o_p(n^{-1})\right] \qquad (2) \\ U_S^* &= U_N^* \times \left[1 - \frac{\hat{s}^{*2} - \hat{s}^2}{2\hat{s}^2} + \frac{3}{8}\left(\frac{\hat{s}^{*2} - \hat{s}^2}{\hat{s}^2}\right)^2 + o_p(b^{-1})\right],\end{aligned}$$

where the error in the second expansion holds conditional on the sample $\mathcal{X}$, D_i is a partial derivative with respect to the ith element of function $f(\cdot)$, μ_i is the population mean of the ith random variable in the vector X. The exact variances of U_N and U_N^* in the above two equations are equal to one. Furthermore, first four cumulants of U_S have the same expansions and rates as the cumulants of U_N above with an exception of the second cumulant. The second cumulant of U_S is equal to $1 + \mathcal{O}(n^{-1})$. With this change in the variance, the Edgeworth expansion, say, for U_S is:

$$P(U_S < x) - \Phi(x) - n^{-1/2} q_1(x)\phi(x) - n^{-1} q_2(x)\phi(x) = o(n^{-1}),$$

where

$$n^{-1/2} q_1(x) = -k_1' - \frac{k_3'}{6}(x^2 - 1)$$

[9] The expansion of U_S, of course, is a theoretical construct, since in the studentized case it is assumed that we do not know s^2.

$$n^{-1}q_2(x) = \left(-\frac{k'_{2,2}}{2} - \frac{k_1'^2}{2}\right)x + \left(\frac{k'_4}{24} + \frac{k'_1k'_3}{6}\right)(3x - x^3) - \frac{k_3'^2}{72}(x^5 - 10x^3 + 15x), \quad (3)$$

k'_i is the ith population cumulant of U_S, and $k'_2 = 1 + k'_{2,2}/n + o(n^{-1})$. The functional forms of the first two components of one-sided Cornish-Fisher expansion are defined as:

$$n^{-1/2}q_{11}(x) = -n^{-1/2}q_1(x)$$
$$n^{-1}q_{21}(x) = n^{-1/2}q_1(x)n^{-1/2}q'_1(x) - \frac{1}{2}xn^{-1}q_1(x)^2 - n^{-1}q_2(x). \quad (4)$$

The functional forms of $n^{-1/2}\hat{q}_{11}(\cdot)$ and $n^{-1}\hat{q}_{21}(\cdot)$ are the same as those of $n^{-1/2}q_{11}(\cdot)$ and $n^{-1}q_{21}(\cdot)$, respectively, with population moments of U_S replaced by the sample cumulants of U_S^*. Also, note that the following equality holds for the first polynomial, $n^{-1}q_{12}(\cdot)$, in the Cornish-Fisher expansion of the αth quantile of the two-sided population distribution function of the studentized test statistic: $n^{-1}q_{12}(\cdot) = -n^{-1}q_2(\cdot)$, with the obvious equivalent for the bootstrap case.

Note the difference between $n^{-1}p_2(\cdot)$ (introduced earlier) and $n^{-1}q_2(\cdot)$. Although the functional forms of the polynomials in the Edgeworth and Cornish-Fisher expansions for the standardized and the studentized statistics are the same (as functions of cumulants), some cancellations happen in the normalized case, when we replace the second cumulant with its expansion. In the normalized case the second cumulant is exactly equal to one, whereas it is equal to $1 + \mathcal{O}(n^{-1})$ in the studentized case.

3. Main results

In this paper we have established the optimal bootstrap block lengths by minimizing the error in the coverage probabilities of one- and symmetric two-sided block bootstrap confidence intervals of normalized and studentized smooth functions of sample averages. Solution methods to the problems involving normalized and studentized statistics are very similar. Section 3.1 deals with the normalized statistic, while the details of the solution to the case of the studentized statistic are discussed in section 3.2. Algebraic details of the important calculations can be found in the Appendix.

3.1. Normalized statistic

3.1.1. One-sided confidence intervals

Here we find the block length l that satisfies the following expression:

$$l^* = \arg\min_{l \in \mathcal{L}} \left| P(U_N < \hat{u}_\alpha) - \alpha \right|,$$

where $\mathcal{L}$ is the set of block lengths that are no larger than n and that go to infinity as the sample size n goes to infinity.

Intuitively, the above probability should equal α plus some terms that disappear asymptotically and are functions of l. The goal, therefore, is to find these approximating terms. We start out by expanding the objective function from the above minimization problem:

$$\begin{aligned} P(U_N < \hat{u}_\alpha) &= P\Big[U_N - n^{-1/2}(\hat{p}_{11}(z_\alpha) - p_{11}(z_\alpha)) - n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha)) \\ &\leq \sum_{j=1}^{2} n^{-j/2} p_{j1}(z_\alpha) + z_\alpha + r_N \Big], \end{aligned}$$

where $r_N = o(n^{-1})$, except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-1})$. Let's denote $n^{-1/2}\Delta_N \equiv n^{-1/2}(\hat{p}_{11}(z_\alpha) - p_{11}(z_\alpha))$, $S_N \equiv U_N - n^{-1/2}(\hat{p}_{11}(z_\alpha) - p_{11}(z_\alpha)) - n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha))$, and $p_{ij}(\cdot)$'s are as defined in equation 1. By the application of the Delta method (see Appendix):

$$P(U_N < \hat{u}_\alpha) = P(S_N < x) + o(n^{-1}),$$

where $x = \sum_{j=1}^{2} n^{-j/2} p_{j1}(z_\alpha) + z_\alpha$.

Now the objective is to develop the first four cumulants of S_N as functions of cumulants of U_N. Then, using cumulants of S_N, derive an Edgeworth expansion of S_N as an Edgeworth expansion[10] of U_N plus some error terms. Lastly, evaluate the resulting expression at $x = \sum_{j=1}^{2} n^{-j/2} p_{j1}(z_\alpha) + z_\alpha$.

Denote the cumulants of S_N by k_i^S. Then (see Appendix for details):

$$\begin{aligned}
k_1^S &= k_1 - n^{-1/2}E(\Delta_N) + o(n^{-1}) \\
k_2^S &= k_2 - 2n^{-1/2}E(U_N\Delta_N) + o(n^{-1}) \\
k_3^S &= k_3 - 3n^{-1/2}E(U_N^2\Delta_N) + 3n^{-1/2}E(U_N^2)E(\Delta_N) + o(n^{-1}) \\
k_4^S &= k_4 - 4n^{-1/2}E(U_N^3\Delta_N) + 12n^{-1/2}E(U_N^2)E(U_N\Delta_N) + o(n^{-1}),
\end{aligned}$$

where we have used the result that $U_N = \mathcal{O}_p(1)$, $n^{-1/2}\Delta_N = \mathcal{O}_p(A_1^{1/2})$ and $n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha)) = \mathcal{O}_p(A_2^{1/2})$ (see Appendix), and $A_1 = C_1 n^{-1} l^{-2} + C_2 n^{-2} l^2$ and $A_2 = C_3 n^{-2} l^{-2} + C_4 n^{-3} l^3$. The rates of A_1 and A_2 follow from Hall, et al (1995). Next, substitute these cumulants in the Edgeworth expansion of S_N. The resulting equation is:

$$\begin{aligned}
P(U_N \le \hat{u}_\alpha) &= P(S_N \le x) + o(n^{-1}) \\
&= P(U_N \le x) + n^{-1/2}E(\Delta_N)\phi(x) + n^{-1/2}E(U_N\Delta_N)x\phi(x) \\
&+ \left(\frac{1}{2}n^{-1/2}E(U_N^2\Delta_N) - \frac{1}{2}n^{-1/2}E(U_N^2)E(\Delta_N)\right)(x^2-1)\phi(x) \\
&+ \left(\frac{1}{2}n^{-1/2}E(U_N^2)E(U_N\Delta_N) - \frac{1}{6}n^{-1/2}E(U_N^3\Delta_N)\right)(3x - x^3)\phi(x) + o(n^{-1}).
\end{aligned}$$

Evaluating the above equation at $x = \sum_{j=1}^{2} n^{-j/2} p_{j1}(z_\alpha) + z_\alpha$ and noting that $P(U_N \le x) = \alpha + \mathcal{O}(n^{-1})$ does not depend on the block length, l, gives us the following objective function:

$$n^{-1/2}\sum_{i=0}^{3} E(U_N^i\Delta_N)C_i + n^{-1/2}E(U_N^2)E(\Delta_N)C_4 + n^{-1/2}E(U_N^2)E(U_N\Delta_N)C_5 + o(n^{-1}),$$

[10] In this paper we have not derived the regularity conditions under which this expansion exists.

where $n^{-1/2}E(U_N^i\Delta_N) \sim n^{-1/2}E(U_N^2)E(U_N^i\Delta_N)$, $\{i = 0, 1\}$. Thus, we are left with four terms: $n^{-1/2}E(U_N^i\Delta_N)$, $\{i = 0, \ldots, 3\}$. Appendix shows that these terms have the following orders:

$$\begin{aligned} n^{-1/2}E(\Delta_N) &= \mathcal{O}(n^{-3/2}l^{3/2}) + \mathcal{O}(n^{-1/2}l^{-1}) \\ n^{-1/2}E(U_N\Delta_N) &= \mathcal{O}(n^{-1}l) + \mathcal{O}(n^{-1}l^{-1}) \\ n^{-1/2}E(U_N^2\Delta_N) &= \mathcal{O}(n^{-3/2}l^{3/2}) + \mathcal{O}(n^{-1/2}l^{-1}) \\ n^{-1/2}E(U_N^3\Delta_N) &= \mathcal{O}(n^{-1}l^{1/2}) + \mathcal{O}(n^{-1}l^{-1}). \end{aligned}$$

Therefore the error in the bootstrap coverage probability of a one-sided block bootstrap confidence interval is: $\mathcal{O}(n^{-1}l) + \mathcal{O}(n^{-1/2}l^{-1})$. The block length, l, that minimizes this quantity is proportional to $n^{1/4}$. Furthermore, the size of the coverage error is $\mathcal{O}(n^{-3/4})$, when block lengths proportional to $n^{1/4}$ are used.

3.1.2. Symmetric two-sided confidence intervals

The solution methods for one- and symmetric two-sided confidence interval cases are very similar. Again, we are looking for the block length, l, that satisfies the following equation:

$$l^* = \min_{l \in \mathcal{L}} |P(|U_N| < \hat{w}_\alpha) - \alpha| .$$

Note that $\hat{w}_\alpha - w_\alpha = n^{-1}\Delta_N^A + o(n^{-2})$, except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-2})$, where $n^{-1}\Delta_N^A = n^{-1}(\hat{p}_{12}(z_\xi) - p_{12}(z_\xi))$ and $n^{-1}p_{12}(\cdot) = -n^{-1}p_2(\cdot)$. One can show (see Appendix) that $n^{-1}\Delta_N^A = \mathcal{O}_p(A_2^{1/2})$, where $A_2 = C_1 n^{-2}l^{-2} + C_2 n^{-3}l^3$ and the rate of A_2 follows from Hall, et al (1995).

Then:

$$P(|U_N| < \hat{w}_\alpha) = P(|U_N| < w_\alpha + n^{-1}\Delta_N^A + r_N^A)$$

$$= P(|U_N| < w_\alpha + n^{-1}\Delta_N^A) + o(n^{-2})$$

$$= P(U_N < w_\alpha + n^{-1}\Delta_N^A) - P(U_N < -w_\alpha - n^{-1}\Delta_N^A) + o(n^{-2}),$$

where $r_N^A = o(n^{-2})$, except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-2})$ and the second equality follows by the Delta method (see Appendix). The next task is to develop cumulants of $U_N - n^{-1}\Delta_N^A$ and $U_N + n^{-1}\Delta_N^A$ and substitute them in the Edgeworth expansion of $P(U_N - n^{-1}\Delta_N^A < w_\alpha) - P(U_N + n^{-1}\Delta_N^A < -w_\alpha)$. Following the steps of the solution method for the one-sided confidence interval case, it is straightforward to show that the relevant error terms are: $n^{-1}E(\Delta_N^A)$, $n^{-1}E(U_N^2\Delta_N^A)$, and $n^{-1}E(U_N\Delta_N^A)E(U_N)$, where $n^{-1}E(U_N\Delta_N^A)E(U_N) \sim n^{-3/2}E(U_N\Delta_N^A)$. The above terms are of the following orders (see Appendix for the methods used):

$$\begin{aligned} n^{-1}E(\Delta_N^A) &= \mathcal{O}(n^{-2}l^2) + \mathcal{O}(n^{-1}l^{-1}) \\ n^{-3/2}E(U_N\Delta_N^A) &= \mathcal{O}(n^{-2}l^{3/2}) + \mathcal{O}(n^{-2}l^{-1}) \\ n^{-1}E(U_N^2\Delta_N^A) &= \mathcal{O}(n^{-2}l^2) + \mathcal{O}(n^{-1}l^{-1}). \end{aligned}$$

Thus, the error in the coverage probability of a symmetric two-sided block bootstrap confidence interval is of order $\mathcal{O}(n^{-2}l^2)+\mathcal{O}(n^{-1}l^{-1})$. The block length, l, that minimizes this error is proportional to $n^{1/3}$. The error then is of size $\mathcal{O}(n^{-4/3})$.

3.2. Studentized statistic

It is intuitively clear that the error rates of the coverage probability in the studentized case should be the same as in the normalized case. The reason for this is that the Taylor series expansion of the studentized test statistic is equal to normalized test statistic plus some higher order error terms (see equation 2).

The solution method for the studentized statistic case is very similar to that of the normalized statistic. The derivation of the error terms is identical to the normalized statistic case for both, one- and symmetric two-sided confidence intervals. The dominant error terms are: $n^{-1/2}E(U_S^i\Delta_S)$, $\{i = 0, \ldots, 3\}$ for the one-sided case and $n^{-1}E(\Delta_S^A)$, $n^{-1}E(U_S^2\Delta_S^A)$, and $n^{-1}E(U_S\Delta_S^A)E(U_S)$ for the two-sided case, where $n^{-1/2}\Delta_S = n^{-1/2}(\hat{q}_{11}(z_\alpha) - q_{11}(z_\alpha))$ and $n^{-1}\Delta_S^A = n^{-1}(\hat{q}_{12}(z_\xi) - q_{12}(z_\xi))$ (see equation 4).

Let $k_i^{'}$ and $\hat{k}_i^{'}$ denote the population and bootstrap cumulants of U_S and U_S^*, respectively. Given the structure of the polynomials $q_1(\cdot)$ and $q_2(\cdot)$ in equations 3 we see that the following error terms have to be bounded; for one- sided case: $E[U_S^i \cdot (\hat{k}_1^{'} - k_1^{'})]$, $E[U_S^i \cdot (\hat{k}_3^{'} - k_3^{'})]$, $\{i = 0, \ldots, 3\}$, for symmetric two-sided case: $E[U_S^j \cdot (\hat{k}_2^{'} - k_2^{'})]$, $E[U_S^j \cdot (\hat{k}_1^{'2} - k_1^{'2})]$, $E[U_S^j \cdot (\hat{k}_4^{'} - k_4^{'})]$, $E[U_S^j \cdot (\hat{k}_1^{'}\hat{k}_3^{'} - k_1^{'}k_3^{'})]$, and $E[U_S^j \cdot (\hat{k}_3^{'2} - k_3^{'2})]$, $\{j = 0, \ldots, 2\}$.

Notice that the above terms are dominated by their normalized statistic equivalents. This is easy to see from equation 2, where we break down U_S and U_S^* in U_N and U_N^*, respectively, times something that is asymptotically equal to one. The only exception occurs in the case of the terms $E[U_S^j \cdot (\hat{k}_2^{'} - k_2^{'})]$, $\{j = 0, \ldots, 2\}$. In the normalized statistic case the exact variances of U_N and U_N^* are both equal to one. Thus, the leading terms of $k_2^{'}$ and $\hat{k}_2^{'}$ both cancel, and $E[U_S^j \cdot (\hat{k}_2^{'} - k_2^{'})]$ is dominated by the population and the bootstrap variances of the second brackets in equation 2. However, one can show (see Appendix) that $E[U_S^j \cdot (\hat{k}_2^{'} - k_2^{'})]$ terms are either equal to or dominated by $\mathcal{O}(n^{-1}l^{-1})$. It follows that the error rates in the coverage probabilities of one- and symmetric two- sided block bootstrap confidence intervals of studentized statistics are $\mathcal{O}(n^{-1}l) + \mathcal{O}(n^{-1/2}l^{-1})$ and $\mathcal{O}(n^{-2}l^2) + \mathcal{O}(n^{-1}l^{-1})$, respectively. Thus, the optimal block lengths and the coverage error rates are the same for both, studentized and normalized cases, where the former are proportional to $n^{1/4}$ and $n^{1/3}$ for one- and symmetric two-sided confidence intervals, respectively.

4. Conclusions

In this paper we have established that the minimum error rates in coverage probabilities of one- and symmetric two-sided block bootstrap confidence intervals are of orders $\mathcal{O}(n^{-3/4})$ and $\mathcal{O}(n^{-4/3})$, respectively, for normalized and studentized smooth functions of sample moments. These rates are attained, when the blocks for one- and symmetric two-sided block bootstrap confidence intervals are proportional to $n^{1/4}$ and $n^{1/3}$, respectively. The above rates in the coverage errors are consistent with Monte Carlo evidence of Hall and Horowitz (1996) and Hansen (1999). The reason for such slight refinement over the asymptotic case is clear: the blocking damages the dependence structure of the original sample, and even under the optimal blocking rate, the block bootstrap does not recover this dependence sufficiently well.

Bühlman (1997, 1998) has suggested a promising alternative to blocking called 'sieve bootstrap', which seems to work well under certain restrictions.

Appendix

Result 1 *Derivation of the probability bounds for* $n^{-1/2}\Delta_N$, $n^{-1}\Delta_N^A$, *and* $n^{-1}(\hat{p}_{21}(x) - p_{21}(x))$.

From Hall, et al (1995), we know that $n^{-1}E(\hat{p}_1(x)-p_1(x))^2 = \mathcal{O}(A_1)$, where $A_1 = C_1 n^{-1} l^{-2} + C_2 n^{-2} l^2$. Also, note that the probability rate of $n^{-1/2}(\hat{p}_1(x)-p_1(x))$ is the same as that of $n^{-1/2}(\hat{p}_{11}(x)-p_{11}(x))$, since $n^{-1/2}p_{11}(\cdot) = -n^{-1/2}p_1(\cdot)$ with the obvious modifications for $n^{-1/2}\hat{p}_{11}(\cdot)$. Then by Chebyshev's inequality:

$$P(A_1^{-1/2} n^{-1/2} |\Delta_N| > M_\varepsilon) \;\; < \;\; \frac{n^{-1} E(\Delta_N)^2}{A_1 M_\varepsilon^2} \equiv \varepsilon^*,$$

where $M_\varepsilon < \infty$ and ε^* can be made arbitrarily small. The latter statement is true, because $n^{-1}E(\Delta_N)^2/A_1 = \mathcal{O}(1)$. Thus, $A_1^{-1/2} n^{-1/2} \Delta_N = \mathcal{O}_p(1)$, i.e., it is bounded in probability.

To find the probability bound for $n^{-1}\Delta_N^A$ we use the result from Hall, et al (1995): $n^{-2}E(\hat{p}_2(x) - p_2(x))^2 = \mathcal{O}(A_2)$, where $A_2 = C_1 n^{-2} l^{-2} + C_2 n^{-3} l^3$. Also, the probability rate of $n^{-1}(\hat{p}_2(x) - p_2(x))$ is the same as that of $n^{-1}(\hat{p}_{12}(x) - p_{12}(x))$, since $n^{-1}p_{12}(\cdot) = -n^{-1}p_2(\cdot)$ with the obvious modifications for $n^{-1}\hat{p}_{12}(\cdot)$. Then we follow the steps above to establish that $n^{-1}\Delta_N^A = \mathcal{O}_p(A_2^{1/2})$.

Lastly, to establish the probability bound of $n^{-1}(\hat{p}_{21}(x) - p_{21}(x))$, we note that the probability rate of $n^{-1}(\hat{p}_2(x) - p_2(x))$ is the same as that of $n^{-1}(\hat{p}_{21}(x) - p_{21}(x))$ (this is not hard to show), and then proceed as in the case above.

Result 2 *Derivation of the cumulants of U_N, S_N, and $U_N \pm n^{-1}\Delta_N^A$.*

The derivation of the cumulants of U_N depend on applying the Taylor series expansion to the random variable of interest. We know that

$$U_N = \frac{n^{1/2}(f(\bar{X}) - f(\mu))}{\sqrt{V\left[n^{1/2}(f(\bar{X}) - f(\mu))\right]}}.$$

Note that $V\left[n^{1/2}(f(\bar{X}) - f(\mu))\right] = \mathcal{O}(1)$. Then using the Taylor expansion with respect to $\bar{X}$ around μ:

$$\begin{aligned} n^{1/2}(f(\bar{X}) - f(\mu)) \;\; &= \;\; \sum_{i=1}^{d} (D_i f)(\mu) n^{1/2} (\bar{X}_i - \mu_i) \\ &+ \;\; \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (D_i D_j f)(\mu) n^{1/2} (\bar{X}_i - \mu_i)(\bar{X}_j - \mu_j) + o_p(n^{-1/2}), \end{aligned}$$

where the notation is as in equation 2. Then

$$E(U_N) = \frac{k_{1,2}}{n^{1/2}} + \frac{k_{1,3}}{n^{3/2}} + \mathcal{O}(n^{-5/2}),$$

where $k_{i,j}$ are constants that do not depend on n. Here we have used the following equalities from Hall, et al (1995):

$$\begin{aligned} E(\bar{X}_i - \mu_i)^2 &= \frac{\gamma(0)_i}{n} + \frac{2}{n}\sum_{j=1}^{k}(1-n^{-1}j)\gamma(j)_i && (5)\\ E(\bar{X}_i - \mu_i)^3 &= \frac{E(X_{i1}-\mu_i)^3}{n^2} + \frac{3}{n^2}\sum_{j_1=1}^{k}(1-n^{-1}j_1)E\Big((X_{i0}-\mu_i)(X_{ij_1}-\mu_i)^2 && (6)\\ &+ (X_{i0}-\mu_i)^2(X_{ij_1}-\mu_i)\Big)\\ &+ \frac{6}{n^2}\sum_{j_2,j_3\geq 1;}\sum_{j_2+j_3\leq k}(1-n^{-1}(j_2+j_3))E\left((X_{i0}-\mu_i)(X_{ij_2}-\mu_i)(X_{i,j_2+j_3}-\mu_i)\right), \end{aligned}$$

where X_{ij} is the jth observation of the ith element of the vector X, $\gamma(j)_i$ is the lag-j covariance of the ith element of the vector X. Also, we used the moment inequalities of Yokoyama (1980), Doukhan (1994) (Remark 2, p. 30), Hölder, and a consequence of Hölder and Burkholder inequalities (Hall and Heyde, 1980, eq. 3.67, p. 87). These tools will be used repeatedly, when we bound the error terms in the coverage probability (see Results 4 and 5).

To derive higher order cumulants, use the Taylor series expansion, taken to the appropriate power.

The method of derivation of cumulants of S_N and $U_N \pm n^{-1}\Delta_N^A$ is to derive them as sums of cumulants of U_N plus an error that is asymptotically equal to zero. Let's demonstrate this for the second cumulant of S_N:

$$\begin{aligned} k_2^S &= E(S_N)^2 - E^2(S_N)\\ &= E(U_N - n^{-1/2}\Delta_N - n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha)))^2\\ &- E^2(U_N - n^{-1/2}\Delta_N - n^{-1}(\hat{p}_{21}(z_\alpha) - p_{21}(z_\alpha))) \end{aligned}$$

$$= \quad k_2 - 2n^{-1/2}E(U_N\Delta_N) + n^{-1}E(\Delta_N^2) + 2n^{-1/2}E(U_N)E(\Delta_N) - n^{-1}E^2(\Delta_N) + o(n^{-1}).$$

Using this method it is straightforward to derive cumulants of higher orders.

Result 3 *Derivations involving the Delta method.*

Here we will demonstrate the derivation of equality $P(U_N < \hat{u}_\alpha) = P(S_N < x) + o(n^{-1})$. The derivation of other equalities involving applications of Delta method are similar.

$$P(U_N < \hat{u}_\alpha) \quad = \quad P(S_N < x + r_N),$$

where $r_N = o(n^{-1})$, except, possibly, if $\mathcal{Y}$ is contained in a set of probability $o(n^{-1})$. That is, $P(r_N \neq o(n^{-1})) = o(n^{-1})$. Therefore, as $n \to \infty$, $P(m \cdot |r_N| \geq \varepsilon,$ for some $m \geq n) = o(n^{-1})$, for all $\varepsilon > 0$. This is equivalent to $P(n \cdot |r_N| \geq \varepsilon) = o(n^{-1})$ for all $\varepsilon > 0$, as $n \to \infty$. Then,

$$\begin{aligned}
P(S_N < x + r_N) \quad &= \quad P\left(S_n < x + r_N, \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{\omega : m \cdot |r_N(\omega)| < \varepsilon\}\right) \\
&+ \quad P\left(S_N < x + r_N, \bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}\{\omega : m \cdot |r_N(\omega)| \geq \varepsilon\}\right) \\
&\leq \quad P\left(S_N < x + r_N, \bigcup_{n=1}^{\infty}\bigcap_{m=n}^{\infty}\{\omega : m \cdot |r_N(\omega)| < \varepsilon\}\right) \\
&+ \quad P\left(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}\{\omega : m \cdot |r_N(\omega)| \geq \varepsilon\}\right) \\
&\leq \quad \lim_{n\to\infty} P(S_N < x + r_N, m \cdot |r_N| \leq \varepsilon, \forall m \geq n) + P(\{\omega : r_N(\omega) \neq o(n^{-1})\}) \\
&\leq \quad P(S_N < x) + o(n^{-1}).
\end{aligned}$$

Similarly we can show that $P(S_N < x + r_N) \geq P(S_N < x) + o(n^{-1})$.

Result 4 *Bounding of* $n^{-1/2}E(U_N^i\Delta_N)$, $\{i = 0, \ldots, 3\}$.

(i)

$$
\begin{aligned}
n^{-1/2}E(\Delta_N) &= n^{-1/2}E(\hat{p}_1(x) - p_1(x)) \\
&= E(\hat{k}_1 - k_1) + C \cdot E(\hat{k}_3 - k_3).
\end{aligned}
$$

Start with $E(\hat{k}_1 - k_1)$ and define $\hat{\beta} \equiv E'(f(\bar{X}^*) - f(\bar{X}))$, $\beta \equiv E(f(\bar{X}) - f(\mu))$, and $\hat{k}_1 = \hat{\beta}/\tilde{s}$. Then

$$
\begin{aligned}
\hat{k}_1 &= \frac{\hat{\beta} - \beta + \beta}{s}\left(1 - \frac{\tilde{s}^2 - s^2}{2s^2} + \frac{3}{8}\left(\frac{\tilde{s}^2 - s^2}{s^2}\right)^2 + \dots\right) \\
&= k_1 + \frac{\hat{\beta} - \beta}{s} - \frac{\beta}{s}\frac{\tilde{s}^2 - s^2}{2s^2} - \frac{\hat{\beta} - \beta}{s}\frac{\tilde{s}^2 - s^2}{2s^2} + \frac{\beta}{s}\frac{3}{8}\left(\frac{\tilde{s}^2 - s^2}{s^2}\right)^2 + \mathcal{O}_p(n^{3/2}A_0),
\end{aligned}
$$

where $A_0 = C_1 n^{-2} l^{-2} + C_2 n^{-3} l$, $s^2 = \mathcal{O}(n^{-1})$, $\beta = \mathcal{O}(n^{-1})$, $\hat{\beta} - \beta = \mathcal{O}_p(A_0^{1/2})$, $\tilde{s}^2 - s^2 = \mathcal{O}_p(A_0^{1/2})$, $E(\hat{\beta} - \beta) \sim C_1 n^{-1} l^{-1} + C_2 n^{-2} l \sim E(\tilde{s}^2 - s^2)$, $E(\hat{\beta} - \beta)^2 \sim C_1 n^{-2} l^{-2} + C_2 n^{-3} l \sim E(\tilde{s}^2 - s^2)^2$. The last six bounds are from Hall, et al (1995). Also, note that $\hat{\beta} - \beta \sim (f''(\bar{X})/2) \cdot E'(\bar{X}_i^* - \bar{X}_i)^2 - (f''(\mu)/2) \cdot E(\bar{X}_i - \mu_i)^2$ and $\tilde{s}^2 - s^2 \sim (f'(\bar{X}))^2 E'(\bar{X}_i^* - \bar{X}_i)^2 - (f'(\mu))^2 E(\bar{X}_i - \mu_i)^2$. Then $E(\hat{k}_1 - k_1) \sim C_1 n^{-1/2} l^{-1} + C_2 n^{-3/2} l$.

Next bound $E(\hat{k}_3 - k_3)$:

$$
\begin{aligned}
\hat{k}_3 - k_3 &= E'\left(\frac{f(\bar{X}^*) - f(\bar{X})}{\sqrt{V'(f(\bar{X}^*))}} - E'\left(\frac{f(\bar{X}^*) - f(\bar{X})}{\sqrt{V'(f(\bar{X}^*))}}\right)\right)^3 \\
&- E\left(\frac{f(\bar{X}) - f(\mu)}{\sqrt{V(f(\bar{X}))}} - E\left(\frac{f(\bar{X}) - f(\mu)}{\sqrt{V(f(\bar{X}))}}\right)\right)^3 \\
&\sim C \cdot \left(\frac{E'(\bar{X}_i^* - \bar{X}_i)^3}{(V'(\bar{X}_i^* - \bar{X}_i))^{3/2}} - \frac{E(\bar{X}_i - \mu_i)^3}{(V(\bar{X}_i - \mu_i))^{3/2}}\right).
\end{aligned}
$$

Note that $\bar{X}$ is a vector random variable and $\bar{X}_i$ is a scalar random variable. Here we have used Taylor's theorem for vector-valued functions. By Hall, et al (1995):

$$
E\left(\frac{E'(\bar{X}_i^* - \bar{X}_i)^3}{(V'(\bar{X}_i^* - \bar{X}_i))^{3/2}} - \frac{E(\bar{X}_i - \mu_i)^3}{(V(\bar{X}_i - \mu_i))^{3/2}}\right) = \mathcal{O}(n^{-1/2} l^{-1}) + \mathcal{O}(n^{-3/2} l^{3/2}).
$$

Therefore $n^{-1/2}E(\Delta_N) \sim C_1 n^{-3/2} l^{3/2} + C_2 n^{-1/2} l^{-1}$.

(ii)

$$n^{-1/2}E(U_N\Delta_N) = E(U_N\hat{k}_1) - k_1^2 + C \cdot (E(U_N\hat{k}_3) - k_1 k_3).$$

Since $\hat{k}_1 = k_1 + (\hat{\beta} - \beta)/s - (\beta/s) \cdot ((\tilde{s}^2 - s^2)/2s^2) + \mathcal{O}_p(n^{3/2}A_0)$,

$$E(U_N\hat{k}_1) = k_1^2 + \frac{1}{s^2}E\left((f(\bar{X}) - f(\mu))(\hat{\beta} - \beta)\right) + E\left(U_N \frac{\beta}{s}\frac{\tilde{s}^2 - s^2}{2s^2}\right) + o(A_3),$$

where $A_3 = C_1 n^{-1} l + C_2 n^{-1} l^{-1}$. The rate of the error $o(A_3)$ stems from the following two considerations. First, the terms covered by the error are farther out in the Taylor series expansion of $\hat{k}_1$ than the terms left in the expansion, and therefore their rates are smaller than those of the terms left in the expansion. Second, the error of the term $n^{-1/2}E(U_N\Delta_N)$ turns out to be $\mathcal{O}(A_3)$. Next define $Y \equiv X - \mu$, i.e., Y is the demeaned random vector X. Some algebra:

$$\begin{aligned}
(f(\bar{X}) - f(\mu))(\hat{\beta} - \beta) &\sim f'(\mu)(\bar{X}_i - \mu_i)\left[\frac{f''(\mu)}{2b}\frac{\sum_{j=1}^{b}(X_{ij} - \bar{X}_i)^2}{b} - \frac{f''(\mu)}{2}E(\bar{X}_i - \mu_i)^2\right] \\
&= C \cdot \left(\bar{Y}_i \frac{1}{b}(\bar{Y}_i^{(2)} - \bar{Y}_i^2) - \bar{Y}_i \frac{1}{b}E(\bar{Y}_i^{(2)})\right) \\
&+ C \cdot \left(\bar{Y}_i \frac{1}{b}E(\bar{Y}_i^{(2)}) - \bar{Y}_i E(\bar{Y}_i^2)\right),
\end{aligned}$$

where $X_{ij} = (1/l)\sum_{k=1}^{l} X_{i,(j-1)\cdot l+k}$, $\bar{Y}_i^{(k)} = (1/b)\sum_{j=1}^{b}(X_{ij} - \mu_i)^k$, and $C = f'(\mu)f''(\mu)/2$.

$$\begin{aligned}
E\left[(f(\bar{X}) - f(\mu))(\hat{\beta} - \beta)\right] &\sim \frac{C}{b}E\left[\bar{Y}_i(\bar{Y}_i^{(2)} - \bar{Y}_i^2) - \bar{Y}_i E(Y_i^{(2)})\right] \\
&= \frac{C}{b}E\left[\bar{Y}_i(\bar{Y}_i^{(2)} - E(\bar{Y}_i^{(2)})) - \bar{Y}_i^3\right] \\
&= \mathcal{O}(n^{-2}) + \mathcal{O}(n^{-3}l),
\end{aligned}$$

where the second equality follows from Taylor's theorem and the last equality follows from the application of inequalities of Yokoyama (1980), Doukhan (1994) (Remark 2, p. 30), and equation 6. By noting that $E[U_N \cdot (\hat{\beta} - \beta)/s] \sim E[U_N \cdot (\beta/s) \cdot ((\tilde{s}^2 - s^2)/2s^2)]$, it follows then that $E(U_N \hat{k}_1) - k_1^2 = \mathcal{O}(n^{-1}) + \mathcal{O}(n^{-2}l)$.

Let us examine $E(U_N \hat{k}_3) - k_1 k_3$. From Hall, et al (1995), $\hat{k}_3 = k_3 + (l^{1/2}/n^{1/2})k_3^l - k_3 + \mathcal{O}_p(A_4^{1/2})$, where $A_4 = C_1 n^{-1} l^{-2} + C_2 n^{-2} l^2$ and k_3^l is the third cumulant for a sample with l observations. Then

$$\begin{aligned} E(U_N \hat{k}_3) - k_1 k_3 &\sim k_1 \left(\frac{l^{1/2}}{n^{1/2}} k_3^l - k_3 \right) + E\left(U_N R_N\right) \\ &= \frac{k_1}{n^{1/2}} \cdot \mathcal{O}(l^{-1}) + \mathcal{O}(n^{-1}l) \\ &= \mathcal{O}(n^{-1}l^{-1}) + \mathcal{O}(n^{-1}l), \end{aligned}$$

where $R_N = \mathcal{O}(n^{-1/2}l^2)\left(\bar{Y}_i^{(3)} - E(\bar{Y}_i^{(3)})\right) + \mathcal{O}(n^{-1/2}l)\bar{Y}_i + \mathcal{O}(n^{-1/2}l)\left(\bar{Y}_i^{(2)} - E(\bar{Y}_i^{(2)})\right)$. The first part of the second line of the above equation follows from Hall, et al (1995), and the second part of the second line follows from inequalities of Yokoyama (1980), Doukhan (1994) (Remark 2, p. 30), and equations 5 and 6.

Thus, $n^{-1/2}E(U_N \Delta_N) = \mathcal{O}(n^{-1}l) + \mathcal{O}(n^{-1}l^{-1})$. Following methods developed above we can establish that $n^{-1/2}E(U_N^2 \Delta_N) = \mathcal{O}(n^{-3/2}l^{3/2}) + \mathcal{O}(n^{-1/2}l^{-1})$ and $n^{-1/2}E(U_N^3 \Delta_N) = \mathcal{O}(n^{-1}l^{1/2}) + \mathcal{O}(n^{-1}l^{-1})$. Therefore the error in coverage probability of the one-sided confidence interval is equal to $\mathcal{O}(n^{-1/2}l^{-1}) + \mathcal{O}(n^{-1}l)$.

To obtain the rate of the error in coverage probability for the symmetric two-sided confidence interval, we have to bound the following terms: $n^{-1}E(\Delta_N^A)$, $n^{-3/2}E(U_N \Delta_N^A)$, and $n^{-1}E(U_N^2 \Delta_N^A)$. Applying the methods above we can show that $n^{-1}E(\Delta_N^A) = \mathcal{O}(n^{-2}l^2) + \mathcal{O}(n^{-1}l^{-1})$, $n^{-3/2}E(U_N \Delta_N^A) = \mathcal{O}(n^{-2}l) + \mathcal{O}(n^{-2}l^{-1})$, and $n^{-1}E(U_N^2 \Delta_N^A) = \mathcal{O}(n^{-2}l^2) + \mathcal{O}(n^{-1}l^{-1})$.

Result 5 *Bounding of* $E\left[U_S^j \cdot (V'(U_S^*) - V(U_S))\right]$, $\{j = 0, \ldots, 2\}$

$$E\left[V'(U_S^*) - V(U_S)\right] \sim C \cdot E\left\{E'\left(\left(\frac{f'(\bar{X})(\bar{X}_i^* - \bar{X}_i)}{\tilde{s}}\right)^2 \frac{\hat{s}^{*2} - \hat{s}^2}{2\hat{s}^2}\right)\right.$$

$$
\begin{aligned}
- & \quad E\left(\left(\frac{f'(\mu)(\bar{X}_i-\mu_i)}{s}\right)^2 \frac{\hat{s}^2-s^2}{2s^2}\right)\Bigg\} \\
\sim & \quad \mathcal{O}(n)E\left\{E'\left((\bar{X}_i^*-\bar{X}_i)^2(\bar{X}_i^{*(2)}-\bar{X}^{(2)})\right)-E\left((\bar{X}_i-\mu_i)^2(\bar{X}_i^{(2)}-E(X_i^2))\right)\right\} \\
+ & \quad \mathcal{O}(n)E\left\{E'\left((\bar{X}_i^*-\bar{X}_i)^2(\bar{X}_i^{*2}-\bar{X}_i^2)\right)-E\left((\bar{X}_i-\mu_i)^2(\bar{X}_i^2-\mu_i^2)\right)\right\} \\
\sim & \quad \mathcal{O}(n)\cdot E\left\{E'(\bar{X}_i^*-\bar{X}_i)^3-E(\bar{X}_i-\mu_i)^3\right\} \\
+ & \quad \mathcal{O}(n)\cdot E\left\{\frac{1}{b^3}E(\bar{X}_{ij}-\bar{X}_i)^4-E(\bar{X}_i-\mu_i)^4\right\} \\
\sim & \quad C\cdot E\left\{\frac{l^2}{n}\sum_{j=1}^{b}\frac{(\bar{X}_{ij}-\mu_i)^3}{b}-nE(\bar{X}_i-\mu_i)^3\right\} \\
= & \quad C\cdot\left(\frac{l^2}{n}E(\bar{X}_{ij}-\mu_i)^3-nE(\bar{X}_i-\mu_i)^3\right),
\end{aligned}
$$

where $E'\left(\bar{X}^*-\bar{X}\right)^3=(1/b^2)\sum_{i=1}^{b}(\bar{X}_i-\bar{X})^3/b$, $s^2=\mathcal{O}(n^{-1})$, and the rest of the notation is as in Result 4. Then by equation 6 we have that $\frac{l^2}{n}E(\bar{X}_{ij}-\mu_i)^3-nE(\bar{X}_i-\mu_i)^3=\mathcal{O}(n^{-1}l^{-1})$. Using above methodology it is straightforward to show that the terms $E\left[U_S\cdot(\hat{k}_2'-k_2')\right]$ and $E\left[U_S^2\cdot(\hat{k}_2'-k_2')\right]$ are dominated by the term $E\left[\hat{k}_2'-k_2\right]$.

References

[1] D.W.K. Andrews. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858, 1991.

[2] R. Beran. Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association*, 83:687–697, 1988.

[3] P. Bühlman. Sieve bootstrap for time series. *Bernoulli*, 3(2):123–148, 1997.

[4] P. Bühlman. Sieve bootstrap for smoothing in nonstationary time series. *The Annals of Statistics*, 26, 1:48–83, 1998.

[5] E. Carlstein. The use of subseries methods for estimating the variance of a general statistic from a stationary time series. *The Annals of Statistics*, 14:1171–1179, 1986.

[6] A.C. Davison and P. Hall. On studentizing and blocking methods for implementing the bootstrap with dependent data. *Australian Journal of Statistics*, 35:215–224, 1993.

[7] P. Doukhan. *Mixing: Properties and Examples.* Springer-Verlag, 1994.

[8] B. Efron. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 7:1–26, 1979.

[9] F. Götze and C. Hipp. Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 64:211–239, 1983.

[10] F. Götze and C. Hipp. Asymptotic distributions of statistics in time series. *The Annals of Statistics*, 22:2062–2088, 1994.

[11] F. Götze and H.R. Künsch. Second-order correctness of the blockwise bootstrap for stationary observations. *The Annals of Statistics*, 24:1914–1933, 1996.

[12] P. Hall. Resampling a coverage process. *Stochastic Processes and Applications*, 20:231–246, 1985.

[13] P. Hall. On the bootstrap and confidence intervals. *The Annals of Statistics*, 14:1431–1452, 1986.

[14] P. Hall. On symmetric bootstrap confidence intervals. *Journal of the Royal Statistical Society*, 50(1):35–45, 1988.

[15] P. Hall. *The Bootstrap and Edgeworth Expansion.* Springer-Verlag, 1992.

[16] P. Hall and C.C. Heyde. *Martingale Limit Theory and Its Application*. New York: Academic Press, 1980.

[17] P. Hall and J. L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64, 4:891–916, 1996.

[18] P. Hall, J.L. Horowitz, and Bing-Yi Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82,3:561–574, 1995.

[19] B. Hansen. Non-parametric dependent data bootstrap for conditional moment models. *Working paper*, 1999.

[20] J. L. Horowitz. Bootstrap-based critical values for the information matrix test. *Journal of Econometrics*, 61:395–411, 1994.

[21] H.R. Künsch. The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, 17:1217–1241, 1989.

[22] S.N. Lahiri. Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. In R. LePage and L. Billard, editors, *Exploring the Limits of Bootstrap*, pages 183–214. New York: Wiley, 1992.

[23] S.N. Lahiri. On Edgeworth expansion and moving block bootstrap for studentized M-estimators in multiple linear regression models. *Journal of Multivariate Analysis*, 56:42–59, 1996.

[24] D. Politis and J. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *The Annals of Statistics*, 22:2031–2050, 1994.

[25] K. Singh. On the asymptotic accuracy of Efron's bootstrap. *The Annals of Statistics*, 9:1187–1195, 1981.

[26] R. Yokoyama. Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52:45–57, 1980.