On bootstrap coverage probability with dependent data

By Jānis J. Zvingelis*

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

Abstract

This paper establishes the optimal bootstrap block lengths for coverage probabilities when the bootstrap is applied to covariance stationary ergodic dependent data. It is shown that the block lengths that minimize the error in coverage probabilities of one- and two-sided block bootstrap confidence intervals of normalized and studentized smooth functions of sample averages are proportional to $n^{1/4}$. The minimum error rates in coverage probabilities of one- and two-sided block bootstrap confidence intervals are of order $\mathcal{O}(n^{-3/2})$ and $\mathcal{O}(n^{-5/4})$, respectively, for normalized and studentized statistics. This constitutes a refinement over the asymptotic confidence intervals.

1. Introduction

The bootstrap is a statistical procedure for estimating the distribution of an estimator. The distinguishing feature of the bootstrap is that it replaces the unknown population distribution of the data by an estimate of it. This estimate of the unknown population distribution is formed by resampling the original sample randomly with replacement. There are several different resampling procedures available depending on whether the data are dependent or independent. The block bootstrap is a bootstrap method applicable to stationary dependent data. In the blocking procedure the original data set is divided into blocks, and these blocks rather than the individual observations are sampled. The intuitive argument for the resampling of the blocks consists in trying to account for the dependence structure of the original data. For this reason the block length, l, has to increase as the sample size, n, increases. There are several different ways to implement the block bootstrap. The two most common methods were suggested in Hall (1985). Although introduced in Hall (1985), in the literature they are known as 'Carlstein's rule' (Carlstein, 1986) and 'Künsch's rule' (Künsch, 1989). Carlstein's rule consists of sampling non-overlapping blocks, and Künsch's rule samples overlapping blocks.[1]

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

[1]Next section gives details on Carlstein's block bootstrap rule.

This paper is partly based on results in Hall, et al (1995) and Hall and Horowitz (1996). The first paper derives, among other things, the block lengths that minimize the error made by bootstrap estimator of the value of one- and two-sided distribution functions, evaluated at a constant. Hall, et al (1995) show that the asymptotic formulae for the optimal block lengths are $l = C_1 n^{1/4}$ and $l = C_2 n^{1/5}$, respectively.[2] Hall and Horowitz (1996) give conditions under which Carlstein's block bootstrap provides asymptotic refinements[3] through $\mathcal{O}(n^{-1})$ for coverage probabilities, when bootstrap critical values are used for constructing symmetrical, two-tailed confidence intervals for Generalized Method of Moments (GMM) test statistics. In this paper we will use a slightly modified version of regularity conditions of Hall and Horowitz (1996).

Related literature also includes Lahiri (1992), who provides conditions under which asymptotic refinements through $\mathcal{O}(n^{-1/2})$ are obtained for estimating one-sided distribution functions of normalized statistics and for studentized statistics with m-dependent data. The m-dependence condition was not imposed in Hall and Horowitz (1996). Götze and Künsch (1996) apply Künsch's block bootstrap to estimating one-sided distribution functions. Among other things, they show that the block bootstrap provides asymptotic refinement through $\mathcal{O}(n^{-1/2})$ for estimating one-sided distribution functions of studentized statistics, provided the variance estimator is chosen appropriately. Lahiri (1996) gives a proof of a similar result for studentized M-estimators in multiple linear regression models.

In this paper we find the block lengths, l, that minimize the error in the values of one- and two-sided distribution functions of normalized and studentized smooth functions of sample averages evaluated at bootstrap critical values. This amounts to estimating the error in the coverage probabilities of one- and two-sided bootstrap confidence intervals of normalized and studentized smooth functions of sample averages. It turns out that the optimal block lengths are proportional to $n^{1/4}$ for one- and two-sided confidence intervals for both, normalized and studentized statistics. Furthermore, the errors made in the coverage probabilities by the one- and two-sided block bootstrap confidence intervals are $\mathcal{O}(n^{-3/4})$ and $\mathcal{O}(n^{-5/4})$, respectively, when optimal block lengths are used. This amounts to refinement over the asymptotic confidence intervals, since the errors made by one- and two-sided asymptotic confidence intervals in the dependent data setup are $\mathcal{O}(n^{-1/2})$ and $\mathcal{O}(n^{-1})$, respectively. Note, however, that the improvement from using the bootstrap over the asymptotics leaves much to be desired, especially in the two-sided confidence interval case. As a comparison, the errors in coverage probability made by one- and two-sided confidence intervals, when bootstrap critical values are used in an IID case, are $\mathcal{O}(n^{-1})$ and $\mathcal{O}(n^{-2})$, respectively.

[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]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 $o(n^{-r})$.

The random variable of interest here is a standardized and studentized smooth function of sample moments of $\mathcal{X}$ or sample moments of functions of $\mathcal{X}$. GMM estimators, for example, fall in this category, because they can be approximated by a smooth function of sample moments with a negligible error. In the case of normalized statistic the assumption is made that the appropriate variance is known. An important statistic that falls in this category is the Durbin-Watson test statistic for serial correlation. The second case of interest is a studentized random variable. Studentization introduces a new set of difficulties when we are applying bootstrap methods to dependent data. The reason is that the exact bootstrap variance of the demeaned random variable has a different functional form than its population equivalent. This is because 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. This dependence structure is unlikely to be present in the original sample.

To achieve asymptotic refinement, the Edgeworth expansions 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 change in the dependence structure in the bootstrap sample. The corrected versions of bootstrap test statistics are also used in this paper. If, instead of using a correction factor, we used a bootstrap equivalent of the consistent estimator of the population variance to studentize the bootstrap test statistic, the exact variance of the leading term of the Taylor series expansion of the bootstrap test statistic could not be made equal to one without introducing extra terms into the bootstrap Edgeworth expansion[4] that would not be present in the population expansion. If, on the other hand, we used the exact bootstrap variance to studentize the bootstrap test statistic, the exact variance of the leading term of the Taylor series expansion of the test statistic would be equal to one, but again the structure of the population and the bootstrap Edgeworth expansions would not be the same. Thus, 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

[4]An *Edgeworth expansion* is an approximation to 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 of a given random variable. See Hall (1992) for detailed discussion of Edgeworth expansions.

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 paper is organized as follows: section 2 looks at the regularity conditions and introduces the test statistics of interest, section 3 lays out the theoretical results. This is followed by the appendix containing the relevant mathematical derivations.

2. Regularity conditions and test statistics

Let us introduce the notation by explaining the Carlstein's blocking rule. Notation will largely follow that laid out in Hall, et al (1995) and Hall and Horowitz (1996). Denote the sample by $\mathcal{X} = (X_1, \ldots, X_n)$, where $X_i \in \mathcal{R}^d$ is a $d \times 1$ random variable. Let b, l denote integers such that $n = bl$. Carlstein's rule divides the sample $\mathcal{X}$ in b disjoint blocks, where the k^{th} block is $\mathcal{B}_k = (X_{(k-1)l+1}, \ldots, X_{kl})$ for $1 \leq k \leq b$. According to the Carlstein's rule bootstrap sample $\mathcal{X}^*$ 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.

2.1. Regularity conditions

In this paper we have established the optimal bootstrap block lengths by minimizing the error in the coverage probabilities of one- and two-sided block bootstrap confidence intervals of normalized and studentized smooth functions of sample averages. Many test statistics and estimators are smooth functions of sample averages or can be approximated by such functions with negligible error. Test statistics based on GMM estimators constitute an example of the latter case. To motivate the regularity conditions for the existence of the Edgeworth and Cornish-Fisher expansions that are employed later, let the test statistic of interest be equal to a GMM test statistic up to a negligibly small error. The following regularity conditions are a slightly modified version of those in Hall and Horowitz (1996).

Let the GMM estimation be based on the moment condition $Eg(X, \theta) = 0$, where g is a $L_g \times 1$ function, θ is a $L_\theta \times 1$ parameter vector whose true but unknown value is θ_0, and $L_g \geq L_\theta$. Assume that $\{X_i\}$ is a covariance stationary, ergodic stochastic process and that $Eg(X_i, \theta_0)g(X_j, \theta_0)' = 0$ if $|i - j| > k$ for some integer $k < \infty$. Also, assume that $Cov(X_i, X_j) = 0$ if $|i - j| > k$. The assumptions then are:

Assumption 1 *There is a sequence of iid vectors* $\{\varepsilon_i : i = -\infty, \ldots, \infty\}$ *of dimension* $L_\varepsilon \geq d$ *and a* $d \times 1$ *function* h *such that* $X_i = h(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$. *There is a constant*

$c > 0$ such that for all $n = 1, 2, \ldots$ and all $m > c^{-1}$

$$E\|h(\varepsilon_n, \varepsilon_{n-1}, \varepsilon_{n-2}, \ldots) - h(\varepsilon_n, \varepsilon_{n-1}, \varepsilon_{n-2}, \ldots, \varepsilon_{n-m}, 0, 0, \ldots)\| \leq c^{-1} exp(-cm).$$

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

Assumption 3 *(a) $E\|g(X, \theta)\| < \infty$ for all $\theta \in \Theta$, where $\|\cdot\|$ is the Euclidean norm. (b) $Eg(X_1, \theta_0)g(X_{1+i}, \theta_0) = 0$ if $i > k$ for some $k < \infty$. (c) $E\sum_{i=0}^{k}[g(X_1, \theta)g(X_{1+i}, \theta)' + g(X_{1+i}, \theta)g(X_1, \theta)']$ exists for all $\theta \in \Theta$. Its smallest eigenvalue is bounded away from zero uniformly over θ in an open sphere, N_0, centered on θ_0. (d) There is a function $C_g(x)$ such that $\|g(x, \theta_1) - g(x, \theta_2)\| \leq C_g(x)\|\theta_1 - \theta_2\|$ for any $\theta_1, \theta_2 \in \Theta$. (e) g is 7-times differentiable with respect to the components of θ everywhere in N_0. (f) Let $\bar{g}(x, \theta)$ be a vector containing the unique components of the derivatives of $g(x, \theta)$ through order 7 with respect to θ. There is a function $C^*(x)$ such that $\|\bar{g}(x, \theta_1) - \bar{g}(x, \theta_2)\| \leq C^*(x)\|\theta_1 - \theta_2\|$ for any $\theta_1, \theta_2 \in N_0$. (g) Let C denote C_g or C^*. Then $P[C(X) > z] = O(z^{-49})$ as $z \to \infty$.*

Let $w(\tilde{x}_1, \theta)$ be a vector containing the unique components of $g(X_1, \theta)$, $g(X_1, \theta)g(X_j, \theta)'$ $(1 \leq j \leq k + 1)$, and their derivatives through order 7 with respect to the components of θ.

Assumption 4 *$w(\tilde{X}_1, \theta_0)$ is a Lipschitz continuous function of $\tilde{x}$. As $z \to \infty$, $P[\|w(\tilde{X}_1, \theta_0)\| > z] = \mathcal{O}(z^{-49})$.*

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

Assumption 6 *There exist $r > 0$ and $\delta > 0$ such that for all integers m satisfying $\delta^{-1} < m + 1 < n$ and all $t \in \mathcal{R}^{dim(w)}$ with $\|t\| > \delta$*

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

2.2. Normalized statistic

The random variable of interest here is equal to (up to a negligible error[5]) the standardized/studentized rth component of the GMM estimator of vector θ. The GMM estimation can be carried out either with a fixed weight matrix or with an estimate of the

[5]Under the above assumptions one can extend Propositions 1 and 2 of Hall and Horowitz (1996) to a case, where the GMM estimator and its bootstrap equivalent are approximated by smooth functions of sample moments with an error of size $o(n^{-2})$ in an appropriate sense.

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

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 $\mathcal{X}$.

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

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

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[6] of U_N, and they are both of order $\mathcal{O}(1)$.

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

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 distribution function (conditional on the sample) induced by the bootstrap sampling. The expansions are asymptotic series, i.e., if the series is stopped after a given number of terms then the remainder is of smaller order than the last term that has been included (see, e.g., Hall (1992) for an extensive description of the Edgeworth expansions and the bootstrap).

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

$$\sup_{\varepsilon \leq \alpha \leq 1-\varepsilon} \left| u_\alpha - z_\alpha - n^{-1/2}p_{11}(z_\alpha) - n^{-1}p_{21}(z_\alpha) \right| = o(n^{-1}),$$

where $0 < \varepsilon < 1/2$.[7] Similarly, define $\hat{u}_\alpha$ as $P(U_N^* < \hat{u}_\alpha) = \alpha$.[8] Then for $\varepsilon > 0$,

$$\sup_{n^{-\varepsilon} \leq \alpha \leq 1-n^{-\varepsilon}} \left| \hat{u}_\alpha - z_\alpha - n^{-1/2}\hat{p}_{11}(z_\alpha) - n^{-1}\hat{p}_{21}(z_\alpha) \right| = o(n^{-1}),$$

[6] 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 $\chi(t) = \exp(k_1 it + \frac{1}{2}k_2(it)^2 + \ldots + \frac{1}{j!}k_j(it)^j + \ldots +)$.

[7] 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.

[8] To obtain an empirical estimate of $\hat{u}_\alpha$, one can carry out a Monte Carlo experiment that consists of resampling the original sample $\mathcal{X}$, calculating the bootstrap test statistic U_N^*, and forming the empirical distribution of U_N^*. The αth quantile of the empirical distribution of the bootstrap test statistic is the empirical estimate of $\hat{u}_\alpha$.

except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-1})$. Let k_i denote the ith cumulant of U_N. Then,

$$\begin{aligned}
n^{-1/2}p_{11}(x) &= -n^{-1/2}p_1(x) \\
n^{-1}p_{21}(x) &= n^{-1/2}p_1(x)n^{-1/2}p_1^{'}(x) - \frac{1}{2}xn^{-1}p_1(x)^2 - n^{-1}p_2(x) \qquad (1)\\
n^{-1/2}p_1(x) &= -k_1 - \frac{k_3}{6}(x^2-1) \\
n^{-1}p_2(x) &= -\frac{1}{2}k_1^2 x + \left(\frac{k_4}{24} + \frac{k_1 k_3}{6}\right)(3x - x^3) - \frac{k_3^2}{72}(x^5 - 10x^3 + 15x),
\end{aligned}$$

with obvious modifications for $\hat{p}_{11}(x)$ and $\hat{p}_{21}(x)$. 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 $E(U_N^2) = \mathcal{O}(1) + \mathcal{O}(n^{-1})$.

Let us introduce also 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:

$$\begin{aligned}
\sup_x \left| P(|U_N| < x) - 2\Phi(x) + 1 - 2n^{-1}p_2(x) \right| &= o(n^{-2}), \\
\sup_x \left| P^*(|U_N^*| < x) - 2\Phi(x) + 1 - 2n^{-1}\hat{p}_2(x) \right| &= o(n^{-2}),
\end{aligned}$$

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:

$$\sup_{\varepsilon \le \alpha \le 1-\varepsilon} \left| w_\alpha - z_\xi - n^{-1}p_{12}(z_\xi) \right| = o(n^{-2}),$$

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

$$\sup_{n^{-\varepsilon}\leq\alpha\leq 1-n^{-\varepsilon}} \left|\hat{w}_\alpha - z_\xi - n^{-1}\hat{p}_{12}(z_\xi)\right| = o(n^{-2}),$$

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

2.3. 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:[9]

$$s^2 \sim \sum_{i=1}^{d} C_i^2 \cdot V(\bar{X}_i) + 2 \sum_{1\leq j<i\leq d}\sum C_i \cdot C_j \cdot Cov(\bar{X}_i, \bar{X}_j),$$

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 $f(\cdot)$. Note that $Cov(\bar{X}_i, \bar{X}_j)$ is dominated by $V(\bar{X}_i)$. A consistent estimator of s^2 is given by:

$$\hat{s}^2 \sim \sum_{i=1}^{d} C_i^2 \cdot \hat{V}(\bar{X}_i) + 2 \sum_{1\leq j<i\leq d}\sum C_i \cdot C_j \cdot \hat{C}ov(\bar{X}_i, \bar{X}_j),$$

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

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

$$\hat{s}^{*2} \sim \sum_{i=1}^{d} C_i^2 \cdot \hat{V}(\bar{X}_i^*) + 2 \sum_{1\leq j<i\leq d}\sum C_i \cdot C_j \cdot \hat{C}ov(\bar{X}_i^*, \bar{X}_j^*).$$

Here $\hat{V}(\bar{X}_i^*) = n^{-2}\sum_{j=1}^{n}(X_{ij}^*-\bar{X}_i^*)^2 + (2/n)\sum_{l=1}^{k}(1-n^{-1}l)\sum_{j=1}^{n-l}(X_{ij}^*-\bar{X}_i^*)(X_{i,j+l}^*-\bar{X}_i^*)/n$, $\hat{C}ov(\bar{X}_i^*, \bar{X}_j^*)$ is dominated by $\hat{V}(\bar{X}_i^*)$, X_{ij}^* is the jth observation of the ith argument of

[9] 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.

$f(\cdot)$ in the block bootstrap sample $\mathcal{X}^*$, 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 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:[10]

$$\begin{aligned}
U_S &= \left[\sum_{i=1}^{d} \frac{(D_i f)(\mu)(\bar{X}_i - \mu_i)}{s} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{(D_i D_j f)(\mu)(\bar{X}_i - \mu_i)(\bar{X}_j - \mu_j)}{s} + o_p(n^{-1/2})\right] \\
&\times \left[1 - \frac{\hat{s}^2 - s^2}{2s^2} + \frac{3}{8}\left(\frac{\hat{s}^2 - s^2}{s^2}\right)^2 + o_p(n^{-1})\right] \qquad (2) \\
U_S^* &= \left[\sum_{i=1}^{d} \frac{(D_i f)(\bar{X})(\bar{X}_i^* - \bar{X})}{\tilde{s}} + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \frac{(D_i D_j f)(\bar{X})(\bar{X}_i^* - \bar{X}_i)(\bar{X}_j^* - \bar{X}_j)}{\tilde{s}} + o_p(b^{-1/2})\right] \\
&\times \left[1 - \frac{\hat{s}^{*2} - \hat{s}^2}{2\hat{s}^2} + \frac{3}{8}\left(\frac{\hat{s}^{*2} - \hat{s}^2}{\hat{s}^2}\right)^2 + o_p(b^{-1})\right],
\end{aligned}$$

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. First brackets of the above two expressions are exactly equal to U_N and U_N^*, respectively. Therefore, the exact variances of the first brackets 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:

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

[10]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.

where

$$
\begin{aligned}
n^{-1/2}q_1(x) &= -k_1^{'} - \frac{k_3^{'}}{6}(x^2-1) \\
n^{-1}q_2(x) &= \left(-\frac{k_{2,2}^{'}}{2} - \frac{k_1^{'2}}{2}\right)x + \left(\frac{k_4^{'}}{24} + \frac{k_1^{'}k_3^{'}}{6}\right)(3x - x^3) \\
&- \frac{k_3^{'2}}{72}(x^5 - 10x^3 + 15x) \\
n^{-1/2}q_{11}(x) &= -n^{-1/2}q_1(x) \\
n^{-1}q_{21}(x) &= n^{-1/2}q_1(x)n^{-1/2}q_1^{'}(x) - \frac{1}{2}xn^{-1}q_1(x)^2 - n^{-1}q_2(x),
\end{aligned}
\tag{3}
$$

$k_i^{'}$ is the ith population cumulant of U_S, and $k_2^{'} = 1 + k_{2,2}^{'}/n + o(n^{-1})$. The functional forms of $n^{-1/2}\hat{q}_{11}(\cdot)$ 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, define the following polynomial for a two-sided confidence interval case: $n^{-1}q_{12}(\cdot) = -n^{-1}q_2(\cdot)$.

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.

The regularity conditions for the existence of all the above Edgeworth and Cornish-Fisher expansions are given in section 2.1.

3. Main results

The goal of this paper is to find the block length l that minimizes the coverage error of one- and two-sided confidence intervals when bootstrap critical values are used in the dependent data setting. 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 distribution function

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

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

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

Now the objective is to develop first 4 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[11] of U_N plus some error terms. Lastly, evaluate the resulting expression at $x \equiv \sum_{j=1}^{2} n^{-j/2} p_{j1}(z_\alpha) + z_\alpha$.

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:

$$\begin{aligned} P(U_N \leq \hat{u}_\alpha) &= P(S_N \leq x) + o(n^{-1}) \\ &= P(U_N \leq x) + n^{-1/2}E(\Delta_N)\phi(x) + n^{-1/2}E(U_N\Delta_N)x\phi(x) \end{aligned}$$

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

$$+ \quad \left(\frac{1}{2}n^{-1/2}E(U_N^2\Delta_N) - \frac{1}{2}n^{-1/2}E(U_N^2)E(\Delta_N)\right)(x^2-1)\phi(x)$$
$$+ \quad \left(\frac{1}{2}n^{-1/2}E(U_N^2)E(U_N\Delta_N) - \frac{1}{6}n^{-1/2}E(U_N^3\Delta_N)\right)(3x - x^3)\phi(x) + o(n^{-1}).$$

The next step is to evaluate the above equation at $x \equiv \sum_{j=1}^{2} n^{-j/2}p_{j1}(z_\alpha) + z_\alpha$:

$$\begin{aligned}
&P(U_N < \hat{u}_\alpha) \\
&= \quad P\left(S_N \leq \sum_{j=1}^{2} n^{-j/2}p_{j1}(z_\alpha) + z_\alpha\right) + o(n^{-1}) \\
&= \quad P\left(U_N \leq \sum_{j=1}^{2} n^{-j/2}p_{j1}(z_\alpha) + z_\alpha\right) + n^{-1/2}E(\Delta_N)\phi(z_\alpha) \\
&+ \quad n^{-1/2}E(U_N\Delta_N)z_\alpha\phi(z_\alpha) + \left(\frac{1}{2}n^{-1/2}E(U_N^2\Delta_N) - \frac{1}{2}n^{-1/2}E(U_N^2)E(\Delta_N)\right)(z_\alpha^2 - 1)\phi(z_\alpha) \\
&+ \quad \left(\frac{1}{2}n^{-1/2}E(U_N^2)E(U_N\Delta_N) - \frac{1}{6}n^{-1/2}E(U_N^3\Delta_N)\right)(3z_\alpha - z_\alpha^3)\phi(z_\alpha) + o(n^{-1}).
\end{aligned}$$

Further, note that $P\left(U_N \leq \sum_{j=1}^{2} n^{-j/2}p_{j1}(z_\alpha) + z_\alpha\right) = \alpha + \mathcal{O}(n^{-1})$, i.e., it does not depend on block length, l, and therefore can be dropped from the minimization function. Now the objective function takes on the following form:

$$\begin{aligned}
n^{-1/2}E(\Delta_N)C_1 + n^{-1/2}E(U_N\Delta_N)C_2 + n^{-1/2}E(U_N^2\Delta_N)C_3 + n^{-1/2}E(U_N^2)E(\Delta_N)C_4 \\
+ n^{-1/2}E(U_N^3\Delta_N)C_5 + n^{-1/2}E(U_N^2)E(U_N\Delta_N)C_6 + o(n^{-1}).
\end{aligned}$$

Note that $n^{-1/2}E(\Delta_N) \sim n^{-1/2}E(U_N^2)E(\Delta_N)$ and $n^{-1/2}E(U_N\Delta_N) \sim n^{-1/2}E(U_N^2)E(U_N\Delta_N)$. Also, note that terms $n^{-1/2}E(U_N^2\Delta_N)$ and $n^{-1/2}E(U_N^3\Delta_N)$ are dominated by $n^{-1/2}E(U_N\Delta_N)$. Thus, we are left with two terms: $n^{-1/2}E(\Delta_N)$ and $n^{-1/2}E(U_N\Delta_N)$. Appendix shows that these terms are of the following orders:

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

Therefore the error in the bootstrap coverage probability of a one-sided block bootstrap confidence interval is: $\mathcal{O}(n^{-1/2}l^{-1}) + \mathcal{O}(n^{-1}l)$. 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. Two-sided distribution function

The solution methods in one- and two-sided distribution function 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:

$$\begin{aligned}
P(|U_N| < \hat{w}_\alpha) &= P(|U_N| < w_\alpha + n^{-1}\Delta_N^A + r_N^A) \\
&= P(|U_N| < w_\alpha + n^{-1}\Delta_N^A) + o(n^{-2}) \\
&= P(U_N < w_\alpha + n^{-1}\Delta_N^A) - P(U_N < -w_\alpha - n^{-1}\Delta_N^A) + o(n^{-2}),
\end{aligned}$$

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 $P(U_N - n^{-1}\Delta_N^A < w_\alpha) - P(U_N + n^{-1}\Delta_N^A < -w_\alpha)$. The relevant error terms are: $n^{-1}E(\Delta_N^A)$, $n^{-1}E(U_N^2\Delta_N^A)$, and $n^{-1}E(U_N\Delta_N^A)E(U_N)$. Note that $n^{-1}E(U_N^2\Delta_N^A)$ is asymptotically equivalent to $n^{-1}E(U_N\Delta_N^A)E(U_N)$, which in turn is asymptotically equivalent to $n^{-3/2}E(U_N\Delta_N^A)$. Following the solution methods of one-sided distribution case (see Appendix) one can show that these terms are of the following orders:

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

Thus, the error in the coverage probability of a two-sided block bootstrap confidence interval is of order $\mathcal{O}(n^{-1}l^{-1}) + \mathcal{O}(n^{-3/2}l)$. The block length, l, that minimizes this error is proportional to $n^{1/4}$. The error is of size $\mathcal{O}(n^{-5/4})$.

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 two-sided distribution functions. The dominant error terms are: $n^{-1/2}E(\Delta_S)$ and $n^{-1/2}E(U_S\Delta_S)$ 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 $n^{-1}\Delta_S^A = n^{-1}(\hat{q}_{12}(z_\xi) - q_{12}(z_\xi))$.

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: $E[\hat{k}_1^{'} - k_1^{'}]$, $E[\hat{k}_3^{'} - k_3^{'}]$, $E[U_S \cdot (\hat{k}_1^{'} - k_1^{'})]$, and $E[U_S \cdot (\hat{k}_3^{'} - k_3^{'})]$, for two-sided case: $E[\hat{k}_2^{'} - k_2^{'}]$, $E[\hat{k}_1^{'2} - k_1^{'2}]$, $E[\hat{k}_4^{'} - k_4^{'}]$, $E[\hat{k}_1^{'}\hat{k}_3^{'} - k_1^{'}k_3^{'}]$, $E[\hat{k}_3^{'2} - k_3^{'2}]$, $E[U_S \cdot (\hat{k}_2^{'} - k_2^{'})]$, $E[U_S \cdot (\hat{k}_1^{'2} - k_1^{'2})]$, $E[U_S \cdot (\hat{k}_4^{'} - k_4^{'})]$, $E[U_S \cdot (\hat{k}_1^{'}\hat{k}_3^{'} - k_1^{'}k_3^{'})]$, and $E[U_S \cdot (\hat{k}_3^{'2} - k_3^{'2})]$.

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 term $E[\hat{k}_2^{'} - k_2^{'}]$. In the normalized statistic case 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 $E[\hat{k}_2^{'} - k_2^{'}]$ is dominated by the population and the bootstrap variances of the second brackets in equation 2. However, one can show (see Appendix) that $E(\hat{k}_2^{'} - k_2^{'})$ term is equal to $o(n^{-1}l^{-1})$. Thus, the error rates in the coverage probabilities of one- and two- sided block bootstrap confidence intervals of studentized statistics are $\mathcal{O}(n^{-1/2}l^{-1}) + \mathcal{O}(n^{-1}l)$ and $\mathcal{O}(n^{-1}l^{-1}) + \mathcal{O}(n^{-3/2}l)$, respectively. The optimal block lengths and the coverage error rates are the same for both, studentized and normalized cases.

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:

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

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. Noting that $n^{\lfloor (i+1)/2 \rfloor} E(\bar{X} - \mu)^i = \mathcal{O}(1)$, where $\lfloor \cdot \rfloor$ denotes the integer part function, we have

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

where $k_{i,j}$ are constants that do not depend on n.

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:

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

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.

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

where $r_N = o(n^{-1})$, except, possibly, if $\mathcal{X}$ is contained in a set of probability $o(n^{-1})$. That is, $P(r_N > o(n^{-1})) = o(n^{-1})$. Therefore $\forall n \geq n_0$ and $\forall \varepsilon > 0$, $P(n \cdot |r_N| > \varepsilon) = o(n^{-1})$. Then,

$$\begin{aligned}
P(S_N \leq x + r_N) &= P(S_n \leq x + r_N, n \cdot |r_N| \leq \varepsilon) \\
&+ P(S_N \leq x + r_N, n \cdot |r_N| > \varepsilon) \\
&\leq P(S_N \leq x + r_N, n \cdot |r_N| \leq \varepsilon) + P(n \cdot |r_N| > \varepsilon) \\
&\leq P(S_N \leq x + \varepsilon \cdot n^{-1}) + o(n^{-1}) \\
&= F_{S_N}(x) + f_{S_N}(x) \cdot \varepsilon \cdot n^{-1} + o(n^{-1}) \\
&= P(S_N < x) + o(n^{-1}).
\end{aligned}$$

Also,

$$\begin{aligned}
P(S_N \geq x + r_N) &= P(S_N \geq x + r_N, n \cdot |r_N| \leq \varepsilon) \\
&+ P(S_N \geq x + r_N, n \cdot |r_N| > \varepsilon) \\
&\leq P(S_N \geq x + r_N, n \cdot |r_N| \leq \varepsilon) + P(n \cdot |r_N| > \varepsilon) \\
&\leq P(S_N \geq x + r_N, r_N \geq -\varepsilon \cdot n^{-1}) + o(n^{-1}) \\
&\leq P(S_N \geq x - \varepsilon \cdot n^{-1}) + o(n^{-1}).
\end{aligned}$$

Then the following inequalities hold:

$$\begin{aligned}
1 - P(S_N \leq x + r_N) &\leq 1 - P(S_N \leq x - \varepsilon \cdot n^{-1}) + o(n^{-1}) \\
P(S_N \leq x + r_N) &\geq P(S_N \leq x - \varepsilon \cdot n^{-1}) + o(n^{-1}) \\
&= F_{S_N}(x) - f_{S_N}(x) \cdot \varepsilon \cdot n^{-1} + o(n^{-1}) \\
&= P(S_N < x) + o(n^{-1}).
\end{aligned}$$

It follows then that $P(S_N < x + r_N) = P(S_N < x) + o(n^{-1})$.

Result 4 *Bounding of $n^{-1/2}E(\Delta_N)$ and $n^{-1/2}E(U_N\Delta_N)$.*

(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 $\hat{k}_1 \equiv \hat{\beta}/\tilde{s}$. Then

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

where $A_0 = C_1 n^{-2} l^{-2} + C_2 n^{-3} l$, $s^2 = \mathcal{O}(n^{-1})$, $\beta = \mathcal{O}(n^{-1})$, $\hat{\beta} - \beta = \mathcal{O}_p(A_0^{1/2})$, $\tilde{s}^2 - s^2 = \mathcal{O}_p(A_0^{1/2})$, $E(\hat{\beta} - \beta) \sim C_1 n^{-1} l^{-1} + C_2 n^{-2} l$, $E(\tilde{s}^2 - s^2) \sim C_1 n^{-1} l^{-1} + C_2 n^{-2} l$. The last four bounds are from Hall, et al (1995). Then $E(\hat{k}_1 - k_1) \sim C_1 n^{-1/2} l^{-1} + C_2 n^{-3/2} l$.

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

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

By Hall, et al (1995), p. 573:

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

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

(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 $o(A_3)$ follows 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 $\mathcal{O}(A_3)$. Some algebra:

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

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

$$\begin{aligned}
E\left[(f(\bar{X}) - f(\mu))(\hat{\beta} - \beta)\right] \;&\sim\; \frac{C}{b}E\left[(\bar{X} - \mu)(\bar{X}^{(2)} - \bar{X}^2) - (\bar{X} - \mu)(E(X^2) - \mu^2)\right] \\
&=\; \frac{C}{b}E\left[(\bar{X} - \mu)(\bar{X}^{(2)} - E(\bar{X}^{(2)})) - 2\mu(\bar{X} - \mu)^2 - (\bar{X} - \mu)^3\right] \\
&=\; \mathcal{O}(b^{-1}n^{-1}),
\end{aligned}$$

where the second equality follows from Taylor's theorem. 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 $E(U_N\hat{k}_1) - k_1^2 = \mathcal{O}(n^{-1}l)$.

Let us examine $E(U_N\hat{k}_3) - k_1k_3$. 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

$$\begin{aligned}
E(U_N\hat{k}_3) - k_1k_3 \;&\sim\; k_1\left(\frac{l^{1/2}}{n^{1/2}}k_3^l - k_3\right) \\
&=\; \frac{k_1}{n^{1/2}} \cdot \mathcal{O}(l^{-1}) \\
&=\; \mathcal{O}(n^{-1}l^{-1}),
\end{aligned}$$

where the second to last line follows from Hall, et al (1995).

Thus, $n^{-1/2}E(U_N\Delta_N) = \mathcal{O}(n^{-1}l) + \mathcal{O}(n^{-1}l^{-1})$. Therefore $n^{-1/2}E(\Delta_N) + n^{-1/2}E(U_N\Delta_N) = \mathcal{O}(n^{-1/2}l^{-1}) + \mathcal{O}(n^{-1}l)$.

Result 5 *Bounding of* $E\left[V'(U_S^*) - V(U_S)\right]$.

$$\begin{aligned} E\left[V'(U_S^*) - V(U_S)\right] &\leq E\left(E'\left(\frac{C\cdot(\bar{X}^* - \bar{X})^3}{s^2}\right) - E\left(\frac{C\cdot(\bar{X}-\mu)^3}{s^2}\right)\right) \\ &\sim C\cdot\frac{l^2}{n}\sum_{i=1}^{b}\frac{E(\bar{X}_i-\mu)^3}{b} - C\cdot nE(\bar{X}-\mu)^3 \\ &= C\cdot\frac{l^2}{n}E(\bar{X}-\mu)^3 - C\cdot nE(\bar{X}-\mu)^3, \end{aligned}$$

where $E'\left(\bar{X}^* - \bar{X}\right)^3 = (1/b^2)\sum_{i=1}^{b}(\bar{X}_i - \bar{X})^3/b$ and $s^2 = \mathcal{O}(n^{-1})$. Note:

$$\begin{aligned} n^2E(\bar{X}-\mu)^3 &= E(X_1^3) + 3\sum_{j=1}^{k}(1-n^{-1}j)E(X_0X_j^2 + X_0^2X_j) \\ &+ 6\sum_{i,j\geq 1}\sum_{i+j\leq k}(1-n^{-1}(i+j))E(X_0X_iX_{i+j}). \end{aligned}$$

Then,

$$\begin{aligned} & C\cdot\frac{l^2}{n}E(\bar{X}_i-\bar{X})^3 - C\cdot nE(\bar{X}-\mu)^3 \\ \sim\ & \frac{1}{n}\left\{E(X_1^3) + 3\sum_{j=1}^{k}(1-l^{-1}j)E(X_0X_j^2+X_0^2X_j)\right. \\ +\ & \left. 6\sum_{i,j\geq 1}\sum_{i+j\leq k}(1-l^{-1}(i+j))E(X_0X_iX_{i+j})\right\} \\ -\ & \frac{1}{n}\left\{E(X_1^3) + 3\sum_{j=1}^{k}(1-n^{-1}j)E(X_0X_j^2+X_0^2X_j)\right. \\ +\ & \left. 6\sum_{i,j\geq 1}\sum_{i+j\leq k}(1-n^{-1}(i+j))E(X_0X_iX_{i+j})\right\} \\ =\ & o(n^{-1}l^{-1}) \end{aligned}$$

References

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

[2] 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.

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

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

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

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

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

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

[9] J.L. Horowitz. The bootstrap. In J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, vol. 5*. Elsevier, Forthcoming.

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

[11] 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.

[12] 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.