Bootstrap Standard Error Estimates and Inference*

Jinyong Hahn† Zhipeng Liao‡

UCLA UCLA

First Version: November 2019; This Version: December 2020

Abstract

Asymptotic justification of the bootstrap often takes the form of weak convergence of the bootstrap distribution to some limit distribution. Theoretical literature recognized that the weak convergence does not imply consistency of the bootstrap second moment or the bootstrap variance as an estimator of the asymptotic variance, but such concern is not always reflected in the applied practice. We bridge the gap between the theory and practice by showing that such common bootstrap based standard error in fact leads to a potentially conservative inference.

JEL Classification: C12, C14, C31, C32

Keywords: Asymptotic Size Control; Bootstrap; Standard Error Estimates

1 Introduction

A typical asymptotic justification of bootstrap takes the following form. First, it is shown that an estimator of interest (after subtracting the estimand and properly rescaling the difference) converges weakly to an asymptotic distribution, usually normal with mean equal to zero and variance equal to σ^2, say. Second, it is shown that the bootstrap version of the estimator (again, after proper centering and rescaling) also converges weakly to the same asymptotic distribution. See Bickel and Freedman (1981), Giné and Zinn (1990), Arcones and Giné (1992), Hahn (1995, 1996), Chen, Linton and Van Keilegom (2003), etc. These results can be used to justify the bootstrap inference based on percentile methods.

Given the weak convergence result, it may appear intuitive to estimate the asymptotic variance by the second moment of the bootstrap distribution, but such an intuition is not supported by theory.

*We thank the Editor and two anonymous referees for their suggestions which have greatly improved the paper.

†Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: hahn@econ.ucla.edu

‡Department of Economics, UCLA, Los Angeles, CA 90095-1477 USA. Email: zhipeng.liao@econ.ucla.edu

Weak convergence concerns expectations of all continuous and bounded functions of a random variable, but because moments concern unbounded functions of that variable, weak convergence does not imply convergence of moments. In particular, consistency of moments may not follow in settings in which large outliers may occur with a vanishing probability. In fact, the second moment of the bootstrap distribution was shown to be inconsistent in Ghosh, Parr, Singh, and Babu's (1984, p.1131) example,[1] where it was pointed out that under some conditions, the bootstrap second moment of the sample median diverges to infinity even though the asymptotic variance of the sample variance is finite,[2] and as a consequence, the implied confidence interval has an asymptotic coverage probability equal to one regardless of the nominal coverage probability.

Despite these theoretical concerns, the bootstrap second moment is often used as an estimator of σ^2.[3] We note that the moment-based bootstrap does not have any inconsistency problem when the mean is the parameter of interest, which is often used as an example in introductory discussions of the bootstrap. Historically, bootstrap was originally developed by Efron (1979), and he devoted a fair amount of space for discussion of bootstrap moments. This may be partly why applied researchers often choose to use the moment-based bootstrap, even in complicated models where the moment-based bootstrap is not justified. Regardless of the reason, the bootstrap second moment is often used in econometric inference, and there is an obvious gap between theory and practice.

The purpose of this paper is to fill the gap by showing that the bootstrap second moment and the bootstrap variance often lead to conservative inference. Loosely speaking, we show that the bootstrap second moment (or the bootstrap variance) cannot be smaller than σ^2, and as such, the resultant inference would be more conservative than is suggested by the nominal significance level.

The paper is organized as follows. Section 2 presents the main results of the paper under the assumption that the data are i.i.d. Section 3 discusses the extension of the main results under weaker conditions and their applicability to time series data. Section 4 reports simulation results. Section 5 concludes. The appendix contains proofs of the main results and additional technical details.

[1] See also Shao's (1992, p.96) example. Although Ghosh, Parr, Singh, and Babu (1984), and Babu (1985) discuss additional regularity conditions under which bootstrap second moments are consistent for σ^2, such consistency results for the bootstrap second moments are not commonly available for other econometric estimators, and often the only available results in the literature take the form of weak convergence of bootstrap distribution of various estimators.

[2] The asymptotic variance of the sample median is known to be $1/\left(4f^2\right)$, where f is the density at the population median.

[3] Shao (1992) and Gonçalves and White (2005) propose modification of the bootstrap to make sure that the resultant estimator of σ^2 is consistent, but their proposals do not always seem to be adopted in applied practice.

2 Main Results

Suppose that we are interested in estimating a scalar parameter θ_0 with a random sample $\{W_i\}_{i=1}^n$ from a random sequence $\{W_i\}_{i=1}^\infty$ which takes values in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let ω denote a generic point in Ω, which can be understood as one of the realizations of the infinite sequence $\{W_i\}_{i=1}^\infty$. Given the sample $\{W_i\}_{i=1}^n$, let $\mathbb{P}_n(\omega)$ denote the empirical distribution function, i.e., it is a multinomial distribution that puts equal weights on each $W_1, \ldots, W_n$, where the $\omega \in \Omega$ emphasizes the conditional nature of the empirical distribution. An estimator $\widehat{\theta}_n$ of the parameter θ_0 of interest is understood to be some function $g_n(W_1, \ldots, W_n)$ of the sample $\{W_i\}_{i=1}^n$. We assume that

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \Rightarrow Z, \tag{1}$$

where Z is a random variable with mean zero and variance $\sigma^2 > 0$, and the arrow "$\Rightarrow$" denotes weak convergence. In many econometric applications, it happens that $Z \sim N(0, \sigma^2)$.[4]

Letting $\{W_i^*\}_{i=1}^n$ denote the n i.i.d. random variables from the empirical distribution $\mathbb{P}_n(\omega)$, we have a bootstrap version of the estimator $\widehat{\theta}_n^* = g_n(W_1^*, \ldots, W_n^*)$. Letting $\mathbb{P}_n^*(\omega)$ denote the empirical distribution of $\{W_i^*\}_{i=1}^n$, we can understand both $\widehat{\theta}_n$ and $\widehat{\theta}_n^*$ to be a function of the empirical distributions $\mathbb{P}_n(\omega)$ and $\mathbb{P}_n^*(\omega)$ respectively, and write $\widehat{\theta}_n(\omega)$ and $\widehat{\theta}_n^*(\omega)$ when we need to emphasize their dependence on ω. A typical asymptotic justification of the bootstrap establishes that

$$n^{1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right) \Rightarrow Z \quad \omega\text{-almost surely} \tag{2}$$

which means that $n^{1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right)$ converges in distribution to Z for almost all ω under the measure $\mathbb{P}$.

Given the result in (2), it may be tempting to estimate $\sigma^2 = \mathbb{E}[Z^2]$ by the second moment of the bootstrap distribution where $\mathbb{E}[\cdot]$ denote the expectation taken under $\mathbb{P}$. We let

$$\widehat{s}_n^{*2} \equiv \mathbb{E}^*\left[n(\widehat{\theta}_n^* - \widehat{\theta}_n)^2\right]$$

denote the bootstrap second moment around $\widehat{\theta}_n$, where $\mathbb{E}^*[\cdot]$ denotes the expectation taken under the bootstrap distribution.[5] As discussed in the introductory section, $\widehat{s}_n^{*2}$ should not be viewed as a consistent estimator of σ^2 with the weak convergence result (2) alone. It is a consequence of the fact that weak convergence does not necessarily imply convergence of moments, although it does not seem to be appreciated in practice. In other words, there is a gap between theory and practice. We argue that this gap can be bridged by using the following lemma:

[4] Throughout this paper, $A \sim B$ means that random variable/vector A has the same distribution as B.

[5] Throughout this paper, we use $a \equiv b$ to denote that a is defined as b.

Lemma 1 *Suppose that F_n is a sequence of distributions on a metric space. Also suppose that $F_n \Rightarrow F$, where the support of F is separable. We then have* $\liminf_{n\to\infty} \int z^2 F_n(dz) \geq \int z^2 F(dz)$.

Proof. See appendix.[6] ■

Taking F_n and F to denote the distribution of $n^{1/2}\left(\widehat{\theta}^*(\omega) - \widehat{\theta}(\omega)\right)$ and Z respectively, we can conclude from (2) and Lemma 1 that $\liminf_{n\to\infty} \widehat{s}_n^{*2} \geq \sigma^2$ on each ω such that $n^{1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right) \Rightarrow Z$. Because the set of such ω has probability equal to 1, we conclude that

$$\liminf_{n\to\infty} \widehat{s}_n^{*2} \geq \sigma^2, \ \omega\text{-almost surely.} \tag{3}$$

The result (3) essentially contains every practical implication about the conservative nature of the bootstrap second moment; because the bootstrap second moment $\widehat{s}_n^{*2}$ tends to be at least as large as σ^2, we can intuitively expect that the inference based on $\widehat{s}_n^{*2}$ should be conservative.

Another tempting approach to use (2) is to use the bootstrap variance to estimate σ^2. Let

$$\widehat{\sigma}_n^{*2} \equiv \widehat{s}_n^{*2} - \left(\mathbb{E}^*\left[n^{1/2}(\widehat{\theta}_n^* - \widehat{\theta}_n)\right]\right)^2 \tag{4}$$

denote the bootstrap variance. Lemma 3 in the appendix, which was not straightforward to us, provides a counterpart of Lemma 1. Using the same argument that led to (3) along with Lemma 3, we can obtain

$$\liminf_{n\to\infty} \widehat{\sigma}_n^{*2} \geq \sigma^2, \ \omega\text{-almost surely.} \tag{5}$$

This result has practical significance. Because of (3), a practitioner may want to use an estimator of σ^2 which is less conservative than the bootstrap second moment $\widehat{s}_n^{*2}$. Because variances are bounded above by second moments, the bootstrap variance $\widehat{\sigma}_n^{*2}$ is a natural candidate. On the other hand, the practitioner may be concerned that the bootstrap variance may potentially under-estimate σ^2, and therefore, the inference based on bootstrap variance may not be valid even from the conservative point of view. However, (5) indicates that it is not the case.

From (5) we derive the main result formally:

Theorem 1 *Suppose that (1) and (2) hold, and that Z is continuously distributed. Then for any finite $z > 0$, we have:*

(i) $\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}(\widehat{\theta}_n - \theta_0)\big/\, \widehat{\sigma}_n^* > z\right) \leq \mathbb{P}\left(Z/\sigma > z\right)$;

(ii) $\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}(\widehat{\theta}_n - \theta_0)\big/\, \widehat{\sigma}_n^* < -z\right) \leq \mathbb{P}\left(Z/\sigma < -z\right)$;

(iii) $\limsup_{n\to\infty} \mathbb{P}\left(\left|n^{1/2}(\widehat{\theta}_n - \theta_0)\big/\, \widehat{\sigma}_n^*\right| > z\right) \leq \mathbb{P}\left(|Z/\sigma| > z\right)$.

[6] Lemma 1 is available in standard literature, e.g. Lehmann and Casella (1998, Lemma 1.14). We present our proof in the appendix for ease of reading.

Proof. In Appendix. ■

Theorem 1 implies that the inference based on $\widehat{\sigma}_n^*$ may be more conservative than is expected by the nominal significance level. Suppose that the asymptotic distribution of $n^{1/2}(\widehat{\theta}_n - \theta_0)$ is $N\left(0, \sigma^2\right)$. Let z_α denote the upper α-quantile of the standard normal random variable.[7] Then Theorem 1(i) implies that

$$\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}(\widehat{\theta}_n - \theta_0) \Big/ \widehat{\sigma}_n^* > z_\alpha \right) \leq \alpha,$$

which shows the size control of the one-sided test based on $n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n^*$. Similarly by Theorem 1(ii, iii), we can show that

$$\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}(\widehat{\theta}_n - \theta_0) \Big/ \widehat{\sigma}_n^* < -z_\alpha \right) \leq \alpha \text{ and } \limsup_{n\to\infty} \mathbb{P}\left(\left| n^{1/2}(\widehat{\theta}_n - \theta_0) \Big/ \widehat{\sigma}_n^* \right| > z_{\alpha/2} \right) \leq \alpha.$$

Because $\widehat{s}_n^{*2} \geq \widehat{\sigma}_n^{*2}$, we can see that the same is true for the inference based on $\widehat{s}_n^*$. In other words, Theorem 1(i, ii, iii) remain true even if $\widehat{\sigma}_n^*$ is replaced by $\widehat{s}_n^*$.[8]

In many empirical studies, the bootstrap estimator $\widehat{\sigma}_n^*$ (or $\widehat{s}_n^*$) is reported, and is often plugged in the test statistic $n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n^*$ (or $n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{s}_n^*$) for one-sided or two-sided test of the null $\theta_0 = 0$ using critical values from the standard normal distribution. Such convention is not theoretically justified if the purpose is to construct confidence intervals whose asymptotic coverage probability is the same as the nominal coverage probability, e.g., as was noted in the introductory section. On the other hand, Theorem 1 implies that the one-sided and two-sided tests based on $n^{1/2}(\widehat{\theta}_n - \theta_0)/\widehat{\sigma}_n^*$ lead to potentially conservative inference, thereby establishing the direction of the possible size distortion for these tests.

We now consider the multivariate generalization where $d \equiv \dim(\theta) > 1$ (yet finite). We assume that the random variable Z in (1) and (2) has a mean equal to 0 and variance-covariance matrix equal to Σ, which we assume is a finite positive definite matrix. Let

$$\hat{S}_n^* \equiv \mathbb{E}^* \left[n(\widehat{\theta}_n^* - \widehat{\theta}_n)(\widehat{\theta}_n^* - \widehat{\theta}_n)' \right] \quad \text{and} \quad \hat{\Sigma}_n^* \equiv \hat{S}_n^* - \mathbb{E}^* \left[n^{1/2}(\widehat{\theta}_n^* - \widehat{\theta}_n) \right] \mathbb{E}^* \left[n^{1/2}(\widehat{\theta}_n^* - \widehat{\theta}_n)' \right]$$

denote the bootstrap second moment and the bootstrap variance-covariance matrix, which are often used for joint inference of θ_0 based on the Wald test statistics

$$n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) \quad \text{or} \quad n(\widehat{\theta}_n - \theta_0)(\hat{\Sigma}_n^*)^{-1}(\widehat{\theta}_n - \theta_0).$$

We can derive the analogs of Theorem 1:

[7] We assume that $\alpha < 0.5$ to make sure that $z_\alpha > 0$.

[8] We are grateful for the editor's insight which simplified the argument here.

Theorem 2 *Suppose that (1) and (2) hold, and that Z is continuously distributed. Then for any finite $z > 0$, we have:*

(i) $\limsup_{n\to\infty} \mathbb{P}\left(n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > z\right) \leq \mathbb{P}\left(Z'\Sigma^{-1}Z > z\right);$

(ii) $\limsup_{n\to\infty} \mathbb{P}\left(n(\widehat{\theta}_n - \theta_0)'(\hat{\Sigma}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > z\right) \leq \mathbb{P}\left(Z'\Sigma^{-1}Z > z\right).$

Proof. In Appendix. ■

Suppose that the asymptotic distribution of $n^{1/2}(\widehat{\theta}_n - \theta_0)$ is $N(0, \Sigma)$. Let $\chi_\alpha(d)$ denote the upper α-quantile of the Chi-square random variable with degree of freedom d. Then Theorem 2 implies that

$$\limsup_{n\to\infty} \mathbb{P}\left(n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > \chi_\alpha(d)\right) \leq \alpha, \text{ and}$$
$$\limsup_{n\to\infty} \mathbb{P}\left(n(\widehat{\theta}_n - \theta_0)'(\hat{\Sigma}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > \chi_\alpha(d)\right) \leq \alpha$$

which shows that the Chi-square tests based on

$$n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) \text{ and } n(\widehat{\theta}_n - \theta_0)'(\hat{\Sigma}_n^*)^{-1}(\widehat{\theta}_n - \theta_0)$$

lead to potentially conservative inference.[9]

3 Discussion

Our result was predicated on the assumption that a typical bootstrap result takes the form (2). It can be easily extended to the case where a bootstrap result takes the form

$$\delta\left(n^{1/2}(\widehat{\theta}_n^* - \widehat{\theta}_n), Z\right) = o_p(1), \tag{6}$$

where δ denotes any metric that metrizes weak convergence such as Prokhorov metric. This is because of the following reasoning. For any subsequence $\{n_k\}$, there is a further subsequence $\{n_p\}$ such that $\delta\left(n_p^{1/2}(\widehat{\theta}_{n_p}^* - \widehat{\theta}_{n_p}), Z\right) \to 0$ almost surely. We can then see that our main results hold true along this subsequence $\{n_p\}$, which implies that our main results are valid along the original sequence $\{n\}$.

For simplicity, we assumed that our sample $\{W_i\}_{i=1}^n$ is independent and identically distributed, but this is not necessary either. Our result only requires that the bootstrap result takes the form (2) or (6), so it applies to various modified bootstrap approaches. The modified bootstrap methods include (but not limited to) the subsampling (see, e.g., Politis and Romano (1994), Hong and Scaillet (2004) and

[9]For any nonzero $\alpha \in \mathbb{R}^d$, by (1) and (2), $n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0) \Rightarrow \alpha' Z$ and $n^{1/2}(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)) \Rightarrow \alpha' Z$ ω-almost surely, based on which we can also obtain straightforward generalization of Theorem 1 for inference of the linear combination $\alpha'\theta_0$. See Appendix B.

Hong and Li (2019)), the residual bootstrap (see, e.g., Bose (1988, 1990) and Franke and Kreiss (1992)), the block bootstrap (see, e.g., Künsch (1989) and Liu and Singh (1992)), the wild bootstrap (see, e.g., Wu (1986) and Liu (1988)) for the dependent and/or non-identically distributed observations, as well as the alternative methods to bootstrap for ease of computation (see, e.g., Armstrong, Bertanha and Hong (2014) and Honoré and Hu (2017)). Our results apply to these modified or alternative bootstrap methods as well.[10]

4 Simulation Study

In this section, we study the properties of the bootstrap standard deviation and related inference by simulation. We consider the standard linear simultaneous model

$$y_i = x_i\beta + u_i,$$
$$x_i = z_i\pi + v_i,$$

where u_i, v_i and z_i are all standard normal and (u_i, v_i) is independent of z_i. We let $\rho \equiv \mathrm{Cov}\,(u_i, v_i)$ and $\beta = 0$ in our simulation. For the values of π and ρ, we follow Hansen, Hausman and Newey (2008, Section 4) for guidance. Surveying articles published in various venues, they note that the medians of the concentration parameter (i.e., $\mu^2 \equiv \sum_{i=1}^n (z_i\pi)^2/\mathbb{E}[v_i^2]$) and ρ are 23.6 and 0.279, respectively. Inspired by this, we calibrate π (with $\pi > 0$) so that $\mu^2 = 25$ for $n = 100$ and $n = 1,000$, which implies that

$$R^2 \equiv \frac{\mathrm{Var}\,(z_i\pi)}{\mathrm{Var}\,(z_i\pi) + \mathrm{Var}\,(v_i)} = 0.2 \text{ and } 0.025.$$

As for ρ, we choose $\rho = 0.25$ and 0.5 respectively.

We consider the case that β is exactly identified here for two reasons. First, we would like to work with the limited information maximum likelihood (LIML) estimator while ensuring that it is well defined for all bootstrap samples. If β is overidentified, the LIML estimator would involve the inverse of $\sum_{i=1}^n z_i z_i'$. When we consider the nonparametric bootstrap, the bootstrap counterpart of $\sum_{i=1}^n z_i z_i'$ is singular with positive probability, which leads to an issue of defining the bootstrap moments. With exact identification, the LIML estimator becomes $\widehat{\beta} \equiv \sum_{i=1}^n z_i y_i / \sum_{i=1}^n z_i x_i$ and the problem disappears. Second, exactly identified models are not unusual in empirical practice. Hansen, Hausman and Newey (2008, p.405) note

[10] Our focus is the first-order validity of the inference based on the boostrap variance, and the higher-order consideration is outside the scope of this paper. For higher order refinements, see, e.g., Hall and Horowitz (1996) and Andrews (2002, 2004).

"50% of the articles had at least one overidentifying restrictions", which we take as evidence that 50% of the articles were exactly identified.

We investigate the properties of two estimators of the asymptotic standard deviation of $\sqrt{n}(\widehat{\beta} - \beta)$ based on the nonparametric bootstrap. The first is the bootstrap estimator $\widehat{\sigma}_n^*$ defined in (4). The second is defined as $\sqrt{n}\left(\beta^*_{\alpha_2} - \beta^*_{\alpha_1}\right)/(z_{\alpha_2} - z_{\alpha_1})$, where $\beta^*_{\alpha_1}$ and $\beta^*_{\alpha_2}$ (or z_{α_1} and z_{α_2}) denote the α_1 and α_2 quantiles of the bootstrap distribution (or the standard normal distribution), respectively. This estimator is considered in Machado and Parente (2005). We work with $(\alpha_1, \alpha_2) = (1/4,\ 3/4)$ in the simulation and the corresponding estimator is called the inter-quartile range (IQR) estimator. The two estimators are normalized/divided by the asymptotic standard deviation of the LIML estimator $\widehat{\beta}$, which equals to π^{-1} $(\pi > 0)$ in our simulation design. In Table 1 below, we will call them normalized standard error estimators.

Second, we study and compare the rejection probabilities of four confidence intervals at the 10% nominal significance level. The first (denoted as B-CI) is based on the bootstrap estimator $\widehat{\sigma}_n^*$, i.e., $\widehat{\beta} \pm 1.645\widehat{\sigma}_n^* n^{-1/2}$. For the second (denoted as IQR-CI), we replace the bootstrap estimator $\widehat{\sigma}_n^*$ in the first confidence interval by the IQR estimator. The third and fourth confidence intervals (denoted as BP-CI and BPT-CI, respectively) are based on the percentile method and the percentile-t method, respectively. All the simulation results are based on 999 bootstrap replications with $100,000$ simulation replications.

Table 1. Properties of the Normalized Standard Error Estimators

			B-SD				IQR-SD			
n	R^2	ρ	Mean	SD	Median	IQR	Mean	SD	Median	IQR
100	0.025	0.25	220.50	7567.03	21.63	55.96	1.46	1.10	1.12	1.07
$1,000$	0.025	0.25	4.51	469.09	1.10	0.43	1.09	0.33	1.02	0.32
100	0.200	0.25	4.92	152.71	1.10	0.52	1.08	0.38	1.01	0.38
$1,000$	0.200	0.25	1.01	0.09	1.01	0.12	1.01	0.09	1.00	0.12
100	0.025	0.50	239.53	21891.88	21.79	57.28	1.45	1.24	1.07	1.06
$1,000$	0.025	0.50	6.20	357.48	1.11	0.54	1.11	0.45	1.01	0.38
100	0.200	0.50	6.87	317.41	1.12	0.64	1.10	0.47	1.00	0.42
$1,000$	0.200	0.50	1.02	0.10	1.01	0.14	1.01	0.10	1.00	0.14

Table 1 presents the mean, the standard deviation (SD), the median and the IQR of the normalized bootstrap standard error estimator (denoted as B-SD), and the normalized IQR standard error estimator (denoted as IQR-SD). From Table 1, we see that the variability of the bootstrap standard error estimator is much larger than that of the IQR standard error estimator when the sample size is small and/or the instrumental variable (IV) is not strong, which reflects the moment issue of the LIML estimator under exact identification. When the sample size is large and the IV is strong, i.e., $n = 1,000$ and $R^2 = 0.2$, the properties of these two standard error estimators are almost the same.

Table 2. Empirical Rejection Probabilities (Nominal Size = 0.10)

n	R^2	ρ	B-CI	IQR-CI	BP-CI	BPT-CI
100	0.025	0.25	0.0056	0.0763	0.0414	0.4176
1,000	0.025	0.25	0.0518	0.0772	0.0449	0.1507
100	0.200	0.25	0.0582	0.0886	0.0541	0.1598
1,000	0.200	0.25	0.0959	0.0985	0.0937	0.1041
100	0.025	0.50	0.0149	0.1222	0.0683	0.4227
1,000	0.025	0.50	0.0613	0.0795	0.0771	0.1574
100	0.200	0.50	0.0654	0.0893	0.0823	0.1629
1,000	0.200	0.50	0.0969	0.0994	0.0966	0.1053

Table 2 provides the empirical rejection probabilities of the four confidence intervals mentioned above. From this table, we see that the B-CI under-rejects in all cases we considered, and the size distortion is sometimes quite severe. The inference based on the bootstrap percentile method also under-rejects in all cases we considered. It is slightly more conservative than the B-CI in some cases, but avoids the severe size distortion in B-CI. The inference based on the IQR-CI also under-rejects except when $n = 100$, $R^2 = 0.025$ and $\rho = 0.5$. Overall, the IQR-CI seems to have the smallest size distortion. Finally, the inference based on the bootstrap percentile-t method is over-rejecting in all cases we considered.

5 Summary

Theoretical literature on the bootstrap often establishes that the bootstrap distribution converges weakly to a desired limit distribution, but not much more. Given such state of the theory literature, theoretical econometricians and statisticians have long advocated the use of the percentile method, which would produce confidence intervals with correct asymptotic coverage probabilities. Theoretical econometricians and statisticians have also developed a few variants of bootstrap to get more accurate confidence intervals (i.e., refinements). We recommend these standard bootstrap procedures to be used in practice.

Despite the lack of theoretical justification, the bootstrap second moment or the bootstrap variance is often used as a basis of inference in practice. Therefore, it may be of interest to look for an understanding of the inference based on bootstrap moments. Our result clarifies the conservative nature of such inference.

Appendix

A Proof of the results in Section 2

Proof of Lemma 1. By Skorokhod's representation theorem, there exist random variables Z_n and Z defined on a common probability space such that $Z_n \sim F_n$, $Z \sim F$, and $Z_n \to Z$ almost surely. It follows that $Z_n^2 \to Z^2$ almost surely. So by Fatou's lemma, we have $\liminf_{n\to\infty} E\left[Z_n^2\right] \geq E\left[Z^2\right]$. The conclusion follows by observing that $E\left[Z_n^2\right] = \int z^2 F_n(dz)$ and $E\left[Z^2\right] = \int z^2 F(dz)$. ■

Lemma 2 *Suppose that* $Z_n \Rightarrow Z$ *and* $\lim_{n\to\infty} \mathrm{Var}(Z_n) < \mathrm{Var}(Z)$. *Then* $\limsup_{n\to\infty} |E[Z_n]| \leq |\mu| + 4\sigma$, *where* $\mu \equiv E[Z]$ *and* $\sigma^2 \equiv \mathrm{Var}(Z)$

Proof. By the weak convergence and Chebyshev's inequality,

$$\liminf_{n\to\infty} P(-2\sigma < Z_n - \mu < 2\sigma) \geq P(-2\sigma < Z - \mu < 2\sigma) \geq 3/4, \tag{7}$$

which implies that for all large n,

$$P(|Z_n - \mu| < 2\sigma) \geq 3(1-\varepsilon)/4 \tag{8}$$

for $\varepsilon > 0$ sufficiently small. By the triangle inequality $|E[Z_n] - \mu| \leq |Z_n - E[Z_n]| + |Z_n - \mu|$, so we have

$$P(|Z_n - \mu| < 2\sigma) \leq P(|E[Z_n] - \mu| < |Z_n - E[Z_n]| + 2\sigma), \tag{9}$$

which can be combined with (8) to yield

$$\liminf_{n\to\infty} P(|E[Z_n] - \mu| < |Z_n - E[Z_n]| + 2\sigma) \geq 3(1-\varepsilon)/4. \tag{10}$$

Since $|E[Z_n] - \mu| \cdot I\{|E[Z_n] - \mu| < |Z_n - E[Z_n]| + 2\sigma\} \leq |Z_n - E[Z_n]| + 2\sigma$, we obtain

$$|E[Z_n] - \mu| \cdot P(|E[Z_n] - \mu| < |Z_n - E[Z_n]| + 2\sigma) \leq E[|Z_n - E[Z_n]|] + 2\sigma. \tag{11}$$

The first element on the right-hand side of the inequality (11) can be bounded by $E[|Z_n - E[Z_n]|] \leq (E[|Z_n - E[Z_n]|^2])^{1/2} = (\mathrm{Var}(Z_n))^{1/2}$, which can be further bounded by σ for all large n using the condition $\lim_{n\to\infty} \mathrm{Var}(Z_n) < \sigma^2$. It follows that for all large n

$$|E[Z_n] - \mu| \leq \frac{3\sigma}{P(|E[Z_n] - \mu| < |Z_n - E[Z_n]| + 2\sigma)}$$

and therefore

$$\begin{aligned}\limsup_{n\to\infty}|E[Z_n]| &\le |\mu| + \limsup_{n\to\infty}|E[Z_n]-\mu| \\ &\le |\mu| + \frac{3\sigma}{\liminf_{n\to\infty} P\left(|E[Z_n]-\mu| < |Z_n - E[Z_n]| + 2\sigma\right)} \\ &\le |\mu| + \frac{4\sigma}{1-\varepsilon},\end{aligned} \tag{12}$$

where the last inequality is by (10). The claim of the lemma follows by letting $\varepsilon \to 0$ on the right-hand side of the last inequality in (12). ■

Lemma 3 *Suppose that* $Z_n \Rightarrow Z$*. Then* $\liminf_{n\to\infty} \mathrm{Var}(Z_n) \ge \mathrm{Var}(Z)$*.*

Proof. By Lemma 1, there exists a subsequence $\{n_k\}$ of $\{n\}$ such that

$$\lim_{k\to\infty} E\left[Z_{n_k}^2\right] = \liminf_{n\to\infty} E\left[Z_n^2\right] \ge E\left[Z^2\right]. \tag{13}$$

Suppose that $\liminf_{n\to\infty} \mathrm{Var}(Z_n) < \sigma^2$. Then there exists a further subsequence $\{n_p\}$ of $\{n_k\}$ such that

$$\lim_{p\to\infty} \mathrm{Var}\left(Z_{n_p}\right) = \liminf_{k\to\infty} \mathrm{Var}\left(Z_{n_k}\right) < \sigma^2 \tag{14}$$

and

$$\lim_{p\to\infty} \left|E[Z_{n_p}]\right| = \limsup_{k\to\infty} |E[Z_{n_k}]| \le |\mu| + 4\sigma, \tag{15}$$

where the inequality in (15) is by Lemma 2. Combining the results in (14) and (15), we get

$$\lim_{p\to\infty} E\left[Z_{n_p}^2\right] = \lim_{p\to\infty} \mathrm{Var}\left(Z_{n_p}\right) + \lim_{p\to\infty} (E[Z_{n_p}])^2 \le 33\sigma^2 + 2\mu^2. \tag{16}$$

Since $\left|Z_{n_p}\right| I\left\{\left|Z_{n_p}\right| \ge C\right\} \le C^{-1} Z_{n_p}^2$, by (16)

$$\limsup_{p\to\infty} E\left[\left|Z_{n_p}\right| I\left\{\left|Z_{n_p}\right| \ge C\right\}\right] \le C^{-1}(33\sigma^2 + 2\mu^2), \tag{17}$$

which implies that $\left|Z_{n_p}\right|$ is uniformly integrable. Therefore, $\lim_{p\to\infty} E[Z_{n_p}] = E[Z] = \mu$ which together with (14) implies that

$$\lim_{p\to\infty} E\left[Z_{n_p}^2\right] = \lim_{p\to\infty} \mathrm{Var}\left(Z_{n_p}\right) + \lim_{p\to\infty} (E[Z_{n_p}])^2 < \sigma^2 + \mu^2 = E[Z^2],$$

which contradicts (13). ■

Lemma 4 *Under (2), we have* $\limsup_{n\to\infty} \mathbb{P}\left(\widehat{\sigma}_n^{*2} < \sigma^2 - \epsilon\right) = 0$ *for any* $\epsilon > 0$*.*

Proof. Let $C_{\sigma,\epsilon} \equiv \left\{\omega : \liminf_{n\to\infty} \widehat{\sigma}_n^{*2}(\omega) < \sigma^2 - \epsilon\right\}$. By (2) and Lemma 3, $\liminf_{n\to\infty} \widehat{\sigma}_n^{*2}(\omega) \geq \sigma^2$, ω-almost surely, which implies that $\mathbb{P}(C_{\sigma,\epsilon}) = 0$. For any $\omega \notin C_{\sigma,\epsilon}$, there exists n_ω such that $\widehat{\sigma}_n^{*2}(\omega) \geq \sigma^2 - \epsilon$ for all $n \geq n_\omega$. This implies that $C_{\sigma,\epsilon}^c \subseteq \cup_{n\geq 1} \cap_{m\geq n} B_{m,\epsilon}^c$, where $B_{n,\epsilon} \equiv \{\omega : \widehat{\sigma}_n^{*2}(\omega) < \sigma^2 - \epsilon\}$, or $C_{\sigma,\epsilon} \supseteq \cap_{n\geq 1} \cup_{m\geq n} B_{m,\epsilon} = \limsup_{n\to\infty} B_{n,\epsilon}$. We therefore have $0 \leq \mathbb{P}(\limsup_{n\to\infty} B_{n,\epsilon}) \leq \mathbb{P}(C_{\sigma,\epsilon}) = 0$. Because $\limsup_{n\to\infty} \mathbb{P}(B_{n,\epsilon}) \leq \mathbb{P}(\limsup_{n\to\infty} B_{n,\epsilon})$, we obtain the desired conclusion. ■

Proof of Theorem 1. By the union bound of the probability, we have for any $\varepsilon > 0$

$$\begin{aligned}
\mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\widehat{\sigma}_n^*} > z\right) &= \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\sigma} > \frac{\widehat{\sigma}_n^*}{\sigma} z\right) \\
&= \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\sigma} > \frac{\widehat{\sigma}_n^*}{\sigma} z,\ \widehat{\sigma}_n^* \geq \sigma(1 - \sigma^{-1}\varepsilon)\right) \\
&\quad + \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\sigma} > \frac{\widehat{\sigma}_n^*}{\sigma} z,\ \widehat{\sigma}_n^* < \sigma(1 - \sigma^{-1}\varepsilon)\right) \\
&\leq \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\sigma} > z - z\sigma^{-1}\varepsilon\right) + \mathbb{P}\left(\widehat{\sigma}_n^* < \sigma - \varepsilon\right),
\end{aligned}$$

which together with Lemma 4 shows that

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\widehat{\sigma}_n^*} > z\right) \leq \limsup_{n\to\infty} \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\sigma} > z - z\sigma^{-1}\varepsilon\right)$$

for any $\varepsilon > 0$. Because $n^{1/2}(\widehat{\theta}_n - \theta_0) \Rightarrow Z$ and Z is a continuous random variable, we conclude that

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\widehat{\sigma}_n^*} > z\right) \leq \mathbb{P}\left(\frac{Z}{\sigma} > z - z\sigma^{-1}\varepsilon\right)$$

for any $\varepsilon > 0$. By letting $\varepsilon \to 0$ and using the continuity of Z, we conclude that

$$\limsup_{n\to\infty} \mathbb{P}\left(\frac{n^{1/2}(\widehat{\theta}_n - \theta_0)}{\widehat{\sigma}_n^*} > z\right) \leq \mathbb{P}\left(\frac{Z}{\sigma} > z\right),$$

which proves the first claim. The second and the third claims can be proved similarly and their proofs are omitted. ■

Theorem 3 *Let $\mathcal{S}^d \equiv \left\{\alpha \in \mathbb{R}^d : \alpha'\alpha = 1\right\}$. Then under (2), we have*

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\alpha\in\mathcal{S}^d} \alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha \geq 1 - \epsilon\right) = 1 \tag{18}$$

and

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\alpha\in\mathcal{S}^d} \alpha'(\Sigma^{-1/2}\hat{\Sigma}_n^*\Sigma^{-1/2})\alpha \geq 1 - \epsilon\right) = 1 \tag{19}$$

for any $\epsilon > 0$.

Proof. Let $D_0 \equiv \{\omega : n^{1/2}\Sigma^{-1/2}(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)) \Rightarrow \Sigma^{-1/2}Z\}$, and note that $\mathbb{P}(D_0) = 1$. Consider any given real vector $\alpha \in \mathcal{S}^d$. Letting F_n and F denote the distributions of $n^{1/2}\alpha'\Sigma^{-1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right)$ and $\alpha'\Sigma^{-1/2}Z$ respectively, we can conclude from (3) that $\liminf_{n\to\infty} \alpha'\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2}\alpha \geq \alpha'\alpha = 1$ on each $\omega \in D_0$. In particular, we have

$$\liminf_{n\to\infty} \alpha'\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2}\alpha \geq 1 \quad \text{for all } \alpha \in \mathcal{S}^d \tag{20}$$

for each $\omega \in D_0$. By the Slutzky theorem, we can also obtain a stronger result: As long as $\alpha_n \in \mathcal{S}^d$ converges to some $\alpha \in \mathcal{S}^d$, we would have $n^{1/2}\alpha_n'\Sigma^{-1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right) \Rightarrow \alpha'\Sigma^{-1/2}Z$ so we get

$$\liminf_{n\to\infty} \alpha_n'\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2}\alpha_n \geq 1 \tag{21}$$

for all $\alpha_n \to \alpha$ on each $\omega \in D_0$.

We now argue that

$$\liminf_{n\to\infty}\left(\inf_{\alpha\in\mathcal{S}^d} \alpha'\left(\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2}\right)\alpha\right) \geq 1 \quad \omega\text{-almost surely.} \tag{22}$$

Suppose that (22) is not satisfied. Then there exist an $\varepsilon > 0$ and a set D_1 of ωs with $P(D_1) > 0$ such that there exists a subsequence $\{n_k\}$ such that

$$\inf_{\alpha\in\mathcal{S}^d} \alpha'\Sigma^{-1/2}\hat{S}_{n_k}^*(\omega)\Sigma^{-1/2}\alpha < 1 - 2\varepsilon$$

for all $\omega \in D_1$. Because (21) is satisfied with probability 1, we may assume that (21) is satisfied at $\omega \in D_1$. For each fixed n_k, we can choose $\alpha_{n_k}(\omega) \in \mathcal{S}^d$ such that

$$\alpha_{n_k}(\omega)'\left(\Sigma^{-1/2}\hat{S}_{n_k}^*(\omega)\Sigma^{-1/2}\right)\alpha_{n_k}(\omega) < \inf_{\alpha\in\mathcal{S}^d} \alpha'\left(\Sigma^{-1/2}\hat{S}_{n_k}^*(\omega)\Sigma^{-1/2}\right)\alpha + \varepsilon.$$

This implies that for any $\omega \in D_1$

$$\alpha_{n_k}(\omega)'\left(\Sigma^{-1/2}\hat{S}_{n_k}^*(\omega)\Sigma^{-1/2}\right)\alpha_{n_k}(\omega) < 1 - \varepsilon.$$

Because $\mathcal{S}^d$ is compact, and there is a further subsequence $\{n_p\}$ of $\{n_k\}$ such that $\alpha_{n_p}(\omega) \to \alpha(\omega)$ for some $\alpha(\omega)$. This implies that

$$\alpha_{n_p}(\omega)'\left(\Sigma^{-1/2}\hat{S}_{n_p}^*(\omega)\Sigma^{-1/2}\right)\alpha_{n_p}(\omega) < 1 - \varepsilon$$

for any $\omega \in D_1$, which contradicts (21).

Now, let $C_\epsilon \equiv \left\{\omega : \liminf_{n\to\infty}(\inf_{\alpha\in\mathcal{S}^d} \alpha'(\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2})\alpha) \geq 1 - \epsilon\right\}$. Then by (22), $\mathbb{P}(C_\epsilon) = 1$. Let

$$A_{n,\epsilon} \equiv \left\{\omega : \inf_{\alpha\in\mathcal{S}^d} \alpha'\left(\Sigma^{-1/2}\hat{S}_n^*(\omega)\Sigma^{-1/2}\right)\alpha \geq 1 - \epsilon\right\}.$$

Then $C_\epsilon \subseteq \cup_{n\geq 1} \cap_{m\geq n} A_{m,\epsilon} = \liminf A_{n,\epsilon}$. We therefore have

$$1 \geq \liminf \mathbb{P}(A_{n,\epsilon}) \geq \mathbb{P}(\liminf A_{n,\epsilon}) \geq \mathbb{P}(C_\epsilon) = 1$$

which proves (18).

As for (19), we note that because $n^{1/2}\alpha_n'\Sigma^{-1/2}\left(\widehat{\theta}_n^*(\omega) - \widehat{\theta}_n(\omega)\right) \Rightarrow \alpha'\Sigma^{-1/2}Z$ for all $\alpha_n \to \alpha$ on each $\omega \in D_0$, by Lemma 3

$$\liminf_{n\to\infty} \alpha_n'\Sigma^{-1/2}\hat{\Sigma}_n^*\Sigma^{-1/2}\alpha_n \geq 1 \tag{23}$$

for all $\alpha_n \to \alpha$ on each $\omega \in D_0$. Using the same arguments of showing (22) but replacing (21) with (23) we can show that

$$\liminf_{n\to\infty}\left(\inf_{\alpha\in\mathcal{S}^d}\alpha'\left(\Sigma^{-1/2}\hat{\Sigma}_n^*(\omega)\Sigma^{-1/2}\right)\alpha\right) \geq 1 \quad \omega\text{-almost surely.} \tag{24}$$

The rest of the proof of (19) is the same as (18), and therefore is omitted. ■

Theorem 4 *Suppose that (2) holds. Then for any* $\epsilon > 0$, $\liminf_{n\to\infty}\mathbb{P}\left(\lambda_{\min}(\hat{S}_n^*) \geq (1-\epsilon)\lambda_{\min}(\Sigma)\right) = 1$ *and* $\liminf_{n\to\infty}\mathbb{P}\left(\lambda_{\min}(\hat{\Sigma}_n^*) \geq (1-\epsilon)\lambda_{\min}(\Sigma)\right) = 1$, *where* $\lambda_{\min}(A)$ *denotes the minimum eigenvalue of a real symmetric matrix* A.

Proof. For any $\alpha \in \mathcal{S}^d$,

$$\begin{aligned}\alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha &= \frac{\alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha}{\alpha'\Sigma^{-1}\alpha}\alpha'\Sigma^{-1}\alpha \\ &\leq \frac{\alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha}{\alpha'\Sigma^{-1}\alpha}(\lambda_{\min}(\Sigma))^{-1}.\end{aligned}$$

Since $\inf_{\alpha\in\mathcal{S}^d}\frac{\alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha}{\alpha'\Sigma^{-1}\alpha} = \lambda_{\min}(\hat{S}_n^*)$, from the above inequality we obtain

$$\inf_{\alpha\in\mathcal{S}^d}\alpha'(\Sigma^{-1/2}\hat{S}_n^*\Sigma^{-1/2})\alpha \leq \lambda_{\min}(\hat{S}_n^*)(\lambda_{\min}(\Sigma))^{-1}$$

which together with Theorem 3 proves the first result. The second result can be proved similarly. ■

Remark 1 *The Wald test statistics based on* $\hat{S}_n^*$ *and* $\hat{\Sigma}_n^*$ *require these bootstrap matrices invertible to be well-defined. The invertibility of* $\hat{S}_n^*$ *and* $\hat{\Sigma}_n^*$ *are established by Theorem 4 above.*

Proof of Theorem 2. (i) For the ease of notations, we let $\widehat{c}_n \equiv \Sigma^{-1/2} n^{1/2}(\widehat{\theta}_n - \theta_0)$ and $\widehat{a}_n \equiv \widehat{c}_n \, (\widehat{c}_n' \widehat{c}_n)^{-1/2}$. Then for any ε in $(0,1)$,

$$
\begin{aligned}
&\left\{ n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > z \right\} \\
&= \left\{ (\widehat{c}_n' \widehat{c}_n) \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > z \right\} \\
&= \left\{ (\widehat{c}_n' \widehat{c}_n) \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > z, \; \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > 1 + \varepsilon \right\} \\
&\quad \cup \left\{ (\widehat{c}_n' \widehat{c}_n) \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > z, \; \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n \leq 1 + \varepsilon \right\} \\
&\subset \left\{ \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > 1 + \varepsilon \right\} \cup \left\{ \widehat{c}_n' \widehat{c}_n > \frac{z}{1+\varepsilon} \right\}.
\end{aligned}
$$

Note that

$$
\begin{aligned}
\left\{ \widehat{a}_n' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \widehat{a}_n > 1 + \varepsilon \right\} &\subset \left\{ \lambda_{\max} \left((\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2})^{-1} \right) > 1 + \varepsilon \right\} \\
&= \left\{ \lambda_{\min} \left(\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2} \right) < \frac{1}{1+\varepsilon} \right\} \\
&\subset \left\{ \inf_{\alpha \in \mathcal{S}^d} \alpha' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2}) \alpha < 1 - \frac{\varepsilon}{2} \right\}
\end{aligned}
$$

where $\lambda_{\max}(A)$ denotes the maximum eigenvalue of a real symmetric matrix A. Therefore, we have

$$
\left\{ n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > z \right\} \subset \left\{ \inf_{\alpha \in \mathcal{S}^d} \alpha' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2}) \alpha < 1 - \frac{\varepsilon}{2} \right\} \cup \left\{ \widehat{c}_n' \widehat{c}_n > \frac{z}{1+\varepsilon} \right\},
$$

which combined with the union bound of probability, the continuity of Z, (1) and (18) in Theorem 3 implies that

$$
\begin{aligned}
&\limsup_{n\to\infty} \mathbb{P} \left(n(\widehat{\theta}_n - \theta_0)'(\hat{S}_n^*)^{-1}(\widehat{\theta}_n - \theta_0) > z \right) \\
&\leq \limsup_{n\to\infty} \mathbb{P} \left(\widehat{c}_n' \widehat{c}_n > \frac{z}{1+\varepsilon} \right) + \limsup_{n\to\infty} \mathbb{P} \left(\inf_{\alpha \in \mathcal{S}^d} \alpha' (\Sigma^{-1/2} \hat{S}_n^* \Sigma^{-1/2}) \alpha < 1 - \frac{\varepsilon}{2} \right) \\
&= \mathbb{P} \left(Z' \Sigma^{-1} Z > \frac{z}{1+\varepsilon} \right).
\end{aligned} \tag{25}
$$

The claim of the lemma follows by the continuity of Z and letting $\varepsilon \to 0$.

(ii) This claim can be proved similarly as part (i) and its proof is hence omitted. ■

B Inference on linear combination of θ_0

We obtain the following generalization of Theorem 1, which shows that the inference of the linear combination of θ_0 based on the bootstrap variance-covariance matrix $\hat{\Sigma}_n^*$ may also be potentially conservative.

Theorem 5 *Suppose that (1) and (2) hold, and that Z is continuously distributed. Then for any $\alpha \in \mathcal{S}^d$ and any finite $z > 0$, we have:*

(i) $\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)\Big/ (\alpha'\hat{\Sigma}_n^*\alpha)^{1/2} > z\right) \le \mathbb{P}\left(\alpha' Z/ (\alpha'\Sigma\alpha)^{1/2} > z\right)$;

(ii) $\limsup_{n\to\infty} \mathbb{P}\left(n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)\Big/ (\alpha'\hat{\Sigma}_n^*\alpha)^{1/2} < -z\right) \le \mathbb{P}\left(\alpha' Z/ (\alpha'\Sigma\alpha)^{1/2} < -z\right)$;

(iii) $\limsup_{n\to\infty} \mathbb{P}\left(\left|n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)\Big/ (\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}\right| > z\right) \le \mathbb{P}\left(\left|\alpha' Z/ (\alpha'\Sigma\alpha)^{1/2}\right| > z\right)$.

Proof of Theorem 5. For any $\alpha \in \mathcal{S}^d$,

$$\alpha'(\hat{\Sigma}_n^* - \Sigma)\alpha = \alpha'\Sigma\alpha\left(\frac{\alpha'\hat{\Sigma}_n^*\alpha}{\alpha'\Sigma\alpha} - 1\right) \ge \alpha'\Sigma\alpha\left(\inf_{\alpha\in\mathcal{S}^d}\alpha'(\Sigma^{-1/2}\hat{\Sigma}_n^*\Sigma^{-1/2})\alpha - 1\right).$$

Therefore, $\alpha'(\hat{\Sigma}_n^* - \Sigma)\alpha \ge -\epsilon$ for any $\alpha \in \mathcal{S}^d$ whenever $\inf_{\alpha\in\mathcal{S}^d}\alpha'(\Sigma^{-1/2}\hat{\Sigma}_n^*\Sigma^{-1/2})\alpha - 1 \ge -\epsilon(\lambda_{\max}(\Sigma))^{-1}$, which together with (19) in Theorem 3 implies that

$$\liminf_{n\to\infty}\mathbb{P}\left(\inf_{\alpha\in\mathcal{S}^d}\alpha'(\hat{\Sigma}_n^* - \Sigma)\alpha \ge -\epsilon\right) = 1 \text{ for any } \epsilon > 0. \tag{26}$$

Consider any ϵ in $(0,1)$. By the union bound of probability, we have

$$\begin{aligned}
&\mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}} > z\right)\\
&= \mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\Sigma\alpha)^{1/2}} > \frac{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}}{(\alpha'\Sigma\alpha)^{1/2}}z\right)\\
&= \mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\Sigma\alpha)^{1/2}} > \frac{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}}{(\alpha'\Sigma\alpha)^{1/2}}z,\ \frac{\alpha'\hat{\Sigma}_n^*\alpha}{\alpha'\Sigma\alpha} \ge 1-\epsilon\right)\\
&+ \mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\Sigma\alpha)^{1/2}} > \frac{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}}{(\alpha'\Sigma\alpha)^{1/2}}z,\ \frac{\alpha'\hat{\Sigma}_n^*\alpha}{\alpha'\Sigma\alpha} < 1-\epsilon\right)\\
&\le \mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\Sigma\alpha)^{1/2}} > z(1-\epsilon)\right) + \mathbb{P}\left(\inf_{\alpha\in\mathcal{S}^d}\alpha'(\hat{\Sigma}_n^* - \Sigma)\alpha < -\lambda_{\min}(\Sigma)\epsilon\right),
\end{aligned}$$

which together with (26) shows that

$$\limsup_{n\to\infty}\mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}} > z\right) \le \limsup_{n\to\infty}\mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\Sigma\alpha)^{1/2}} > z(1-\epsilon)\right).$$

Because $n^{1/2}(\widehat{\theta}_n - \theta_0) \Rightarrow Z$ and Z is a continuous random variable, we conclude that

$$\limsup_{n\to\infty}\mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}} > z\right) \le \mathbb{P}\left(\frac{\alpha' Z}{(\alpha'\Sigma\alpha)^{1/2}} > z(1-\epsilon)\right).$$

By letting $\epsilon \to 0$ and using the continuity of Z, we conclude that

$$\limsup_{n\to\infty}\mathbb{P}\left(\frac{n^{1/2}\alpha'(\widehat{\theta}_n - \theta_0)}{(\alpha'\hat{\Sigma}_n^*\alpha)^{1/2}} > z\right) \le \mathbb{P}\left(\frac{\alpha' Z}{(\alpha'\Sigma\alpha)^{1/2}} > z\right),$$

which proves the first claim. The rest claims can be proved similarly and their proofs are omitted. ■

References

[1] Andrews, D.W.K (2002): "Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119-162.

[2] Andrews, D.W.K (2004): "The Block-Block Bootstrap: Improved Asymptotic Refinements," *Econometrica*, 72, 673-700.

[3] Arcones, M. and Giné, E. (1992): "On the Bootstrap of M-estimators and Other Statistical Functionals," in R. LePage and L. Billard (eds.), Exploring the Limits of Bootstrap. New York: Wiley.

[4] Armstrong, T., Bertanha, M. and Hong, H. (2014): "A Fast Resample Method for Parametric and Semiparametric Models," *Journal of Econometrics*, 179, 128-133.

[5] Babu, G. J. (1986): "A Note on Bootstrapping the Variance of Sample Quantile," *Annals of the Institute of Statistical Mathematics*, 38, Part A, 439-443.

[6] Bickel, P.J. and Freedman, D.A. (1981): "Some Asymptotic Theory for the Bootstrap," *Annals of Statistics*, 9, 1196-1217.

[7] Bose, A. (1988): "Edgeworth Correction by Bootstrap in Autoregressions," *Annals of Statistics*, 16, 1709-1722.

[8] Bose, A. (1990): "Bootstrap in Moving Average Models," *Annals of the Institute of Statistical Mathematics*, 42, 753-768.

[9] Chen, X., Linton, O., and Van Keilegom, I. (2003): "Estimation of Semiparametric Models When the Criterion Function Is Not Smooth," *Econometrica*, 71, 1591-1608.

[10] Efron, B. (1979): "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1–26.

[11] Efron, B. (1987): "Better Bootstrap Confidence Intervals,“ *Journal of the American Statistical Association*, 82, 171-200.

[12] Franke, J. and Kreiss, J.-P. (1992): "Bootstrapping Stationary Autoregressive Moving-average Models," *Journal of Time Series Analysis*, 13, 297-317.

[13] Giné, E. and Zinn, J. (1990): "Bootstrapping General Empirical Measures," *Annals of Probability*, 18, 851-869.

[14] Ghosh, M., Parr, W. C., Singh, K. and Babu, G. J. (1984): "A Note on Bootstrapping the Sample Median," *Annals of Statistics*,12, 1130-1135.

[15] Gonçalves, S., and White, H. (2005): "Bootstrap Standard Error Estimates for Linear Regression," *Journal of the American Statistical Association* 100, 970-979.

[16] Hahn, J. (1995): "Bootstrapping Quantile Regression Estimators," *Econometric Theory*, 11, 105-121.

[17] Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12, 187-197.

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

[19] Hansen, C., J. Hausman and W. Newey (2008): "Estimation With Many Instrumental Variables", *Journal of Business and Economic Statistics*, 26(4), 398-422.

[20] Hong H., and Scaillet, O. (2004): "A Fast Subsampling Method for Nonlinear Dynamic Models," *Journal of Econometrics*, 133, 557-578.

[21] Hong H., and Li J. (2019): "The Numerical Bootstrap," *Annals of Statistics*, Forthcoming.

[22] Honoré, B. and Hu, L. (2017): "Poor (Wo)man's Bootstrap," *Econometrica*, 84, 1277-1301.

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

[24] Lehmann, E.L., and Casella, G. (1998): *Theory of Point Estimation.* New York. Springer-Verlag.

[25] Liu, R.Y. (1988): "Bootstrap Procedures under Some Non-iid Models," *Annals of Statistics*, 16, 1696-1708.

[26] Liu, R.Y. and Singh, K. (1992): "Moving Blocks Jackknife and Bootstrap Capture Weak Dependence" in R. LePage and L. Billard (eds.), Exploring the Limits of Bootstrap. New York: Wiley.

[27] Machado, J.A.F. and Parente, P. (2005): "Bootstrap Estimation of Covariance Matrices via the Percentile Method,“ *Econometrics Journal*, 8, 70-78.

[28] Politis, D. and Romano, J. (1994): "Large Sample Confidence Regions Based on Subsamples under Minimal Assumptions," *Annals of Statistics*, 22, 2031-2050.

[29] Shao, J. (1992), "Bootstrap Variance Estimators With Truncation," *Statistics and Probability Letters*, 15, 95-101.

[30] Wu, C.F.G., (1986): "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis," *Annals of Statistics*, 14, 1261-1295.