Bootstrap in High Dimension with Low Computation

Henry Lam [1] **Zhenyuan Liu** [1]

Abstract

The bootstrap is a popular data-driven method to quantify statistical uncertainty, but for modern high-dimensional problems, it could suffer from huge computational costs due to the need to repeatedly generate resamples and refit models. We study the use of bootstraps in high-dimensional environments with a small number of resamples. In particular, we show that with a recent "cheap" bootstrap perspective, using a number of resamples as small as one could attain valid coverage even when the dimension grows closely with the sample size, thus strongly supporting the implementability of the bootstrap for large-scale problems. We validate our theoretical results and compare the performance of our approach with other benchmarks via a range of experiments.

1. Introduction

The bootstrap is a widely used method for statistical uncertainty quantification, notably confidence interval construction (Efron & Tibshirani, 1994; Davison & Hinkley, 1997; Shao & Tu, 2012; Hall & Martin, 1988). Its main idea is to resample data and use the distribution of resample estimates to approximate a sampling distribution. Typically, this approximation requires running many Monte Carlo replications to generate the resamples and refit models. This is affordable for classical problems, but for modern large-scale problems, this repeated fitting could impose tremendous computation concerns. This issue motivates an array of recent works to curb the computation effort, mostly through a "subsampling" perspective that fits models on smaller data sets in the bootstrap process, e.g., Kleiner et al. (2012); Lu et al. (2020); Giordano et al. (2019); Schulam & Saria (2019); Alaa & Van Der Schaar (2020).

In contrast to subsampling, we consider in this paper the reduction in bootstrap computation cost by using a fewer number of Monte Carlo replications or resamples. In particular, we target the following question: *Is it possible to run a valid bootstrap for high-dimensional problems with very little Monte Carlo computation?* While conventional bootstraps rely heavily on adequate resamples, recent work (Lam, 2022a;b) shows that it is possible to reduce the resampling effort dramatically, even down to one Monte Carlo replication. The rough idea of this "cheap" bootstrap is to exploit the approximate independence among the original and resample estimates, instead of their distributional closeness utilized in the conventional bootstraps. We will leverage this recent idea in this paper. However, since Lam (2022a;b) is based purely on asymptotic derivation, giving an affirmative answer to the above question requires the study of new finite-sample bounds to draw understanding on bootstrap behaviors jointly in terms of problem dimension p, sample size n and number of resamples B.

To this end, our main theoretical contribution in this paper is three-fold:

General Finite-Sample Bootstrap Bounds: We derive general finite-sample bounds on the coverage error of confidence intervals aggregated from B resample estimates, where B is small using the "cheap" bootstrap idea, and $B = \infty$ for traditional quantile-based bootstrap methods including the basic and percentile bootstraps (e.g., Davison & Hinkley (1997) Section 5.2-5.3). Our bounds reveal that, given the same primitives on the approximate normality of the original and each resample estimate, the cheap bootstrap with fixed small B achieves similar coverage error bounds as conventional bootstraps using infinite resamples. This also simultaneously recovers the main result in Lam (2022a), but stronger in terms of the finite-sample guarantee.

Bootstrap Bounds on Function-of-Mean Models Explicit in p, n and B: We specialize our general bounds above to the function-of-mean model that is customary in the high-dimensional Berry-Esseen and central limit theorem (CLT) literature (Pinelis & Molzon, 2016; Zhilova, 2020). In particular, our bounds explicit on p, n and B conclude vanishing coverage errors for the cheap bootstrap when $p = o(n)$, for any given $B \geq 1$. Note that the function-of-mean model does not capture all interesting problems, but it has been commonly used – and in fact, appears the only model used

[1]Department of Industrial Engineering and Operations Research, Columbia University, New York, NY, USA. Correspondence to: Henry Lam <khl2114@columbia.edu>.

Proceedings of the 40^{th} International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023.

in deriving finite-sample CLT errors for technicality reasons. Our bounds shed light that, at least for this wide class of models, using a small number of resamples can achieve a good coverage even in a dimension p growing closely with n.

Bootstrap Bounds on Linear Models Independent of p: We further specialize our bounds to linear functions with weaker tail conditions, which have orders independent of p under certain conditions on the L_p norm or Orlicz norm of the linearly scaled random variable.

In addition to theoretical bounds, we investigate the empirical performances of bootstraps using few resamples on large-scale problems, including high-dimensional linear regression, high-dimensional logistic regression, computational simulation modeling, and a real-world data set RCV1-v2 (Lewis et al., 2004). To give a sense of our comparisons that support using the cheap bootstrap in high dimension, here is a general conclusion observed in our experiments: Figure 1(a) shows the coverage probabilities of 95%-level confidence intervals for three regression coefficients with corresponding true values $0, 2, -1$ in a 9000-dimensional linear regression (in Section 4). The cheap bootstrap coverage probabilities are close to the nominal level 95% even with one resample, but the basic and percentile bootstraps only attain around 80% coverage with ten resamples. In this example, one Monte Carlo replication to obtain each resample estimate takes around 4 minutes in the virtual machine e2-highmem-2 in Google Cloud Platform. Therefore, the cheap bootstrap only requires 4 minutes to obtain a statistically valid interval, but the standard bootstrap methods are still far from the nominal coverage even after more than a 40-minute run. Figure 1(b) shows the average interval widths. This reveals the price of a wider interval for the cheap bootstrap when the Monte Carlo budget is very small, but considering the low coverages in the other two methods and the fast decay of the cheap bootstrap width for the first few number of resamples, such a price appears secondary.

Notation. For a random vector X, we write X^k as the tensor power $X^{\otimes k}$. The vector norm is taken as the usual Euclidean norm. The matrix and tensor norms are taken as the operator norm. For a square matrix M, $tr(M)$ denotes the trace of M. $I_{p\times p}$ denotes the identity matrix in $\mathbb{R}^{p\times p}$ and $\mathbf{1}_p$ denotes the vector in $\mathbb{R}^p$ whose components are all 1. Φ denotes the cumulative distribution function of the standard normal. $C^2(\mathbb{R}^p)$ denotes the set of twice continuously differentiable functions on $\mathbb{R}^p$. Throughout the whole paper, we use $C > 0$ (without subscripts) to denote a universal constant which may vary each time it appears. We use $C_1, C_2, \dots$ to denote constants that could depend on other parameters and we will clarify their dependence when using them.

2. Background on Bootstrap Methods

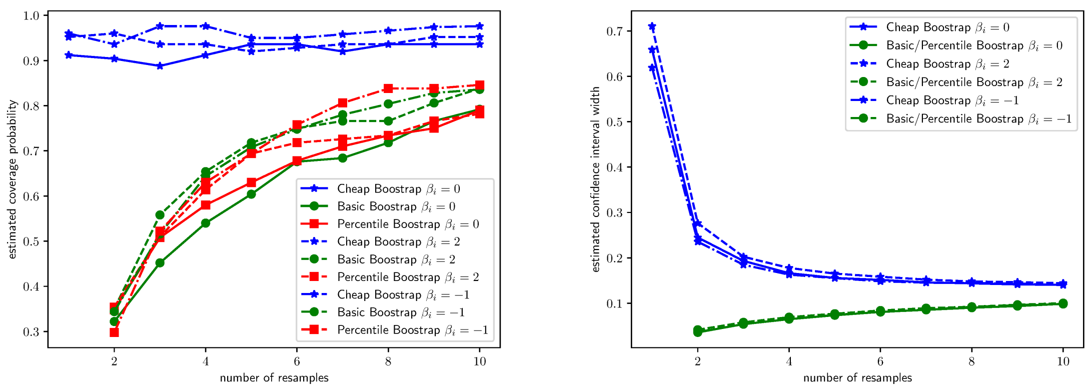
We briefly review standard bootstrap methods and from there the recent cheap bootstrap. Suppose we are interested in estimating a target statistical quantity $\psi := \psi(P_X)$ where $\psi(\cdot) : \mathcal{P} \mapsto \mathbb{R}$ is a functional defined on the probability measure space $\mathcal{P}$. Given i.i.d. data $X_1, \dots, X_n \in \mathbb{R}^p$ following the unknown distribution P_X, we denote the empirical distribution as $\hat{P}_{X,n}(\cdot) := (1/n)\sum_{i=1}^n I(X_i \in \cdot)$. A natural point estimator is $\hat{\psi}_n := \psi(\hat{P}_{X,n})$.

To construct a confidence interval from $\hat{\psi}_n$, a typical beginning point is the distribution of $\hat{\psi}_n - \psi$ from which we can pivotize. As this distribution is unknown in general, the bootstrap idea is to approximate it using the resample counterpart, as if the empirical distribution was the true distribution. More concretely, conditional on $X_1, \dots, X_n$, we repeatedly, say for B times, resample (i.e., sample with replacement) the data n times to obtain resamples $\{X_1^{*b}, \dots, X_n^{*b}\}, b = 1, \dots, B$. Denoting $\hat{P}_{X,n}^{*b}$ as the resample empirical distributions, we construct B resample estimates $\hat{\psi}_n^{*b} := \psi(\hat{P}_{X,n}^{*b})$. Then we use the $\alpha/2$ and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^{*b} - \hat{\psi}_n$, called $q_{\alpha/2}$ and $q_{1-\alpha/2}$, to construct $[\hat{\psi}_n - q_{1-\alpha/2}, \hat{\psi}_n - q_{\alpha/2}]$ as a $(1-\alpha)$-level confidence interval, which is known as the basic bootstrap (Davison & Hinkley (1997) Section 5.2). Alternatively, we could also use the $\alpha/2$ and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^{*b}$, say $q'_{\alpha/2}$ and $q'_{1-\alpha/2}$, to form $[q'_{\alpha/2}, q'_{1-\alpha/2}]$, which is known as the percentile bootstrap (Davison & Hinkley (1997) Section 5.3). There are numerous other variants in the literature, such as studentization (Hall, 1988), calibration or iterated bootstrap (Hall, 1986a; Beran, 1987), and bias correction and acceleration (Efron, 1987; DiCiccio et al., 1996; DiCiccio & Tibshirani, 1987), with the general goal of obtaining more accurate coverage.

All the above methods rely on the principle that $\hat{\psi}_n - \psi$ and $\hat{\psi}_n^* - \hat{\psi}_n$ (conditional on $X_1, \dots, X_n$) are close in distribution. Typically, this means that, with a $\sqrt{n}$-scaling, they both converge to the same normal distribution. In contrast, the cheap bootstrap proposed in Lam (2022a;b) constructs a $(1-\alpha)$-level confidence interval via

$$\left[\hat{\psi}_n - t_{B,1-\alpha/2} S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2} S_{n,B}\right], \qquad (1)$$

where $S_{n,B}^2 = (1/B)\sum_{b=1}^B (\hat{\psi}_n^{*b} - \hat{\psi}_n)^2$, and $t_{B,1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of t_B, the t-distribution with degree of freedom B. The quantity $S_{n,B}^2$ resembles the sample variance of the resample estimates $\hat{\psi}_n^{*b}$'s, in the sense that as $B \to \infty$, $S_{n,B}^2$ approaches the bootstrap variance $Var_*(\hat{\psi}_n^*)$ (where $Var_*(\cdot)$ denotes the variance of a resample conditional on the data). In this way, (1) reduces to the normality interval with a "plug-in" estimator of the standard error term when B and n are both large. However, intriguingly, B *does not need to be large*, and $S_{n,B}^2$ is not necessarily close to



(a) Coverage probability

(b) Confidence interval width

Figure 1. Empirical coverage probabilities and confidence interval widths for different numbers of resamples in a linear regression.

the bootstrap variance. Instead, the idea is to consider the joint distribution of $\hat{\psi}_n - \psi, \hat{\psi}_n^{*1} - \hat{\psi}_n, \ldots, \hat{\psi}_n^{*B} - \hat{\psi}_n$ that is argued to be asymptotically independent as $n \to \infty$ and B fixed, which subsequently allows the construction of a pivotal t-statistic and gives rise to (1) for fixed B. A more detailed explanation on the cheap bootstrap in the low-dimensional case (i.e., p is fixed) is given in Appendix A.

3. High-Dimensional Bootstrap Bounds

As explained in the introduction, we aim to study the coverage performance of bootstraps in high-dimensional problems, focusing on the cheap bootstrap approach that allows the use of small computation effort. We describe our results at three levels, first under general assumptions (Section 3.1), then more explicit bounds under the function-of-mean model and sub-Gaussianity of X (Section 3.2), finally bounds for a linear function under weaker tail assumptions on X (Section 3.3).

3.1. General Finite-Sample Bounds

We have the following finite-sample bound for the cheap bootstrap:

Theorem 3.1. *Suppose we have the finite-sample accuracy for the estimator* $\hat{\psi}_n$

$$\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) \le x) - \Phi(x/\sigma) \right| \le \mathcal{E}_1, \quad (2)$$

and with probability at least $1-\beta$ *we have the finite-sample accuracy for the bootstrap estimator* $\hat{\psi}_n^*$

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n) \le x) - \Phi(x/\sigma) \right| \le \mathcal{E}_2, \quad (3)$$

where $\sigma > 0$, $\mathcal{E}_1$ *and* $\mathcal{E}_2$ *are deterministic quantities, and* P^* *denotes the probability on a resample conditional on the data. Then the coverage error of* (1) *satisfies, for any* $B \ge 1$,

$$\begin{aligned} &\left| P(|\psi - \hat{\psi}_n| \le t_{B,1-\alpha/2} S_{n,B}) - (1-\alpha) \right| \\ &\le 2\mathcal{E}_1 + 2B\mathcal{E}_2 + \beta. \end{aligned} \quad (4)$$

Condition (2) is a Berry-Esseen bound (Bentkus, 2003; Pinelis & Molzon, 2016) that gauges the normal approximation for the original estimate $\hat{\psi}_n$. Condition (3) manifests a similar normal approximation for the resample estimate $\hat{\psi}_n^*$, and has been a focus in the high-dimensional CLT literature (Zhilova, 2020; Lopes, 2022; Chernozhukov et al., 2020). Both conditions (in their asymptotic form) are commonly used to establish the validity of standard bootstrap methods. Theorem 3.1 shows that, under these conditions, the coverage error of the cheap bootstrap interval (1) with any $B \ge 1$ can be controlled. Note that Theorem 3.1 is very general in the sense that there is no direct assumption applied on the form of $\psi(\cdot)$ – All we assume is approximate normality in the sense of (2) and (3). Due to technical delicacies, in the bootstrap literature, finite-sample or higher-order coverage errors are typically obtainable only with specific models (Hall, 2013; Zhilova, 2020; Lopes, 2022; Chernozhukov et al., 2020), the most popular being the function-of-mean model (see Section 3.2) or even simply the sample mean. In contrast, the bound in Theorem 3.1 that concludes the sufficiency in using a very small B is a general statement that does not depend on the delicacies of $\psi(\cdot)$. Moreover, by plugging in suitable bounds for $\mathcal{E}_1, \mathcal{E}_2, \beta$ under regularity conditions, Theorem 3.1 also recovers the main result (Theorem 1) in Lam (2022a). The detailed proof of Theorem 3.1 is in Appendix D (and so are the proofs of all other theorems). Below we give a sketch of the main argument.

Proof sketch of Theorem 3.1. *Step 1:* We write the coverage probability as the expected value (with respect to data) of a multiple integral with respect to the distributions of $\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n)$ (denoted by Q^*, conditional on data), i.e.,

$$\begin{aligned}
&P(|\psi - \hat{\psi}_n| \le t_{B,1-\alpha/2} S_{n,B}) \\
&= P\left(\left| \frac{\sqrt{n}(\hat{\psi}_n - \psi)}{\sqrt{\frac{1}{B}\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2}} \right| \le t_{B,1-\alpha/2} \right) \\
&= E\left[\int_{\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sqrt{\sum_{b=1}^{B} z_b^2/B}} \le t_{B,1-\alpha/2}} dQ^*(z_B) \cdots dQ^*(z_1) \right].
\end{aligned} \tag{5}$$

Step 2: Suppose (3) happens and denote this event by $\mathcal{A}$ which satisfies $P(\mathcal{A}^c) \le \beta$. For each $b = 1, \ldots, B$, given all other $z_{b'}, b' \ne b$, the integration region is of the form $z_b \in (-\infty, -q] \cup [q, \infty)$ for some $q \ge 0$. Then we can replace the distribution Q^* by the distribution of $N(0, \sigma^2)$ (denoted by P_0) with controlled error given in (3) and obtain

$$\begin{aligned}
&E\left[\int_{\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sqrt{\sum_{b=1}^{B} z_b^2/B}} \le t_{B,1-\alpha/2}} dQ^*(z_B) \cdots dQ^*(z_1) \right] \\
&= E\left[\int_{\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sqrt{\sum_{b=1}^{B} z_b^2/B}} \le t_{B,1-1-\alpha/2}} dP_0(z_B) \cdots dP_0(z_1) \right] \\
&+ R_1,
\end{aligned} \tag{6}$$

where $|R_1| \le 2B\mathcal{E}_2 + \beta$ accounts for the error from (3) and the small probability event $\mathcal{A}^c$.

Step 3: Following the same logic in Step 2 and noticing that the integration region for $\sqrt{n}(\hat{\psi}_n - \psi)$ is $[-q, q]$ for some $q \ge 0$, we can also replace the distribution of $\sqrt{n}(\hat{\psi}_n - \psi)$ by the distribution P_0 with controlled error $|R_2| \le 2\mathcal{E}_1$ according to (2):

$$\begin{aligned}
&E\left[\int_{\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sqrt{\sum_{b=1}^{B} z_b^2/B}} \le t_{B,1-\alpha/2}} dP_0(z_B) \cdots dP_0(z_1) \right] \\
&= \int_{\frac{|z_0|}{\sqrt{\sum_{b=1}^{B} z_b^2/B}} \le t_{B,1-\alpha/2}} dP_0(z_B) \cdots dP_0(z_1) dP_0(z_0) \\
&+ R_2 \\
&= 1 - \alpha + R_2.
\end{aligned} \tag{7}$$

Step 4: Plugging (6) and (7) back into (5), we can express the coverage probability as a sum of the nominal level and the remainder term:

$$P(|\psi - \hat{\psi}_n| \le t_{B,1-\alpha/2} S_{n,B}) = 1 - \alpha + R_1 + R_2$$

with error $|R_1 + R_2| \le 2\mathcal{E}_1 + 2B\mathcal{E}_2 + \beta$. This gives our conclusion. □

Theorem 3.1 is designed to work well for small B (our target scenario), but deteriorates when B grows. However, in the latter case, we can strengthen the bound to cover the large-B regime with additional conditions on the variance estimator (see Appendix B.1).

We compare with standard basic and percentile bootstraps using $B = \infty$. Below is a generalization of Zhilova (2020) which focuses only on the basic bootstrap under the function-of-mean model.

Theorem 3.2. *Suppose the conditions in Theorem 3.1 hold. If $q_{\alpha/2}$, $q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^* - \hat{\psi}_n$ respectively given $X_1, \ldots, X_n$, then a finite-sample bound on the basic bootstrap coverage error is*

$$\begin{aligned}
&|P(\hat{\psi}_n - q_{1-\alpha/2} \le \psi \le \hat{\psi}_n - q_{\alpha/2}) - (1-\alpha)| \\
&\le 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta.
\end{aligned} \tag{8}$$

If $q'_{\alpha/2}$, $q'_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^*$ respectively given $X_1, \ldots, X_n$, then a finite-sample bound on the percentile bootstrap coverage error is

$$|P(q'_{\alpha/2} \le \psi \le q'_{1-\alpha/2}) - (1-\alpha)| \le 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta. \tag{9}$$

In view of Theorems 3.1 and 3.2, the cheap bootstrap with any fixed B can achieve the same order of coverage error bound as the basic and percentile bootstraps with $B = \infty$, in the sense that

$$(1/2)\ \mathrm{EB}_{\text{Quantile}} \le \mathrm{EB}_{\text{Cheap}} \le B\ \mathrm{EB}_{\text{Quantile}}, \tag{10}$$

where $\mathrm{EB}_{\text{Cheap}}$ is the RHS error bound of (4) and $\mathrm{EB}_{\text{Quantile}}$ is that of (8) or (9). This shows that, to attain a good coverage that is on par with standard basic/percentile bootstraps, it suffices to use the cheap bootstrap with a small B which could save computation dramatically.

Besides coverage, another important quality of confidence interval is its width. To this end, note that for any fixed B, (3) ensures that $\sqrt{n}S_{n,B} \Rightarrow \sigma\sqrt{\chi_B^2/B}$ (unconditionally as $n \to \infty$ with proper model assumptions) . Therefore, the half-width of (1) is approximately $t_{B,1-\alpha/2}\sigma\sqrt{\chi_B^2/(nB)}$ with expected value

$$E\left[t_{B,1-\alpha/2}\sigma\sqrt{\frac{\chi_B^2}{nB}} \right] = t_{B,1-\alpha/2}\sigma\sqrt{\frac{2}{Bn}}\frac{\Gamma((B+1)/2)}{\Gamma(B/2)}, \tag{11}$$

where $\Gamma(\cdot)$ is the gamma function. Since the dimensional impact is hidden in σ which is a common factor in the expected width as B varies, we can see p does not affect the relative width behavior as B changes. In particular, from (11) the inflation of the expected width relative to the case

$B=\infty$ is 417.3% for $B=1$, and dramatically reduces to 94.6%, 24.8% and 10.9% for $B=2,5,10$, thus giving an interval with both correct coverage and short width using a small computation budget B.

In the next sections, we will apply Theorem 3.1 to obtain explicit bounds for specific high-dimensional models. Here, in relation to (10), we briefly comment that the order of the coverage error bounds for these models is of order $1/\sqrt{n}$, both for the cheap bootstrap (which we will derive) and state-of-the-art high-dimensional bootstrap CLT. This is in contrast to the typical $1/n$ coverage error in two-sided bootstrap confidence intervals in low dimension (see Hall (2013) Section 3.5 for quantile-based bootstraps and Lam (2022a) Section 3.2 for the cheap bootstrap).

3.2. Function-of-Mean Models

We now specialize to the function-of-mean model $\psi=g(\mu)$ for a mean vector $\mu=E[X]\in\mathbb{R}^p$ and smooth function $g:\mathbb{R}^p\mapsto\mathbb{R}$, which allows us to construct more explicit bounds. The original estimate $\hat{\psi}_n$ and resample estimate $\hat{\psi}_n^{*b}$ are now given by $g(\bar{X}_n)$ and $g(\bar{X}_n^{*b})$ respectively, where $\bar{X}_n$ denotes the sample mean of data and $\bar{X}_n^{*b}$ denotes the resample mean of $X_1^{*b},\dots,X_n^{*b}$. We assume:

Assumption 3.3. The function $g(x)\in C^2(\mathbb{R}^p)$ has Hessian matrix $H_g(x)$ with uniformly bounded eigenvalues, that is, $\exists$ a constant $C_{H_g}>0$ s.t. $\sup_{x\in\mathbb{R}^p}|a^\top H_g(x)a|\le C_{H_g}||a||^2,\forall a\in\mathbb{R}^p$.

Assumption 3.4. X is sub-Gaussian, i.e., there is a constant $\tau^2>0$ s.t. $E[\exp(a^\top(X-\mu))]\le\exp(||a||^2\tau^2/2),\forall a\in\mathbb{R}^p$. Furthermore, X has a density bounded by a constant C_X and its covariance matrix Σ is positive definite with the smallest eigenvalue $\lambda_\Sigma>0$.

Based on Theorem 3.1, we derive the following explicit bound:

Theorem 3.5. *Suppose the function g satisfies Assumption 3.3 and random vector X satisfies Assumption 3.4. Moreover, assume $||\nabla g(\mu)||>C_{\nabla g}\sqrt{p}$ for some constant $C_{\nabla g}>0$. Then we have*

$$\begin{aligned}
&\left|P(|g(\mu)-g(\bar{X}_n)|\le t_{B,1-\alpha/2}S_{n,B})-(1-\alpha)\right|\\
&\le\frac{6}{n}+BC\left(\frac{m_{31}}{\sqrt{n}\sigma^3}+\frac{C_{H_g}m_{31}^{1/3}tr(\Sigma)}{\sqrt{n}\sigma^2}+\frac{C_{H_g}m_{32}^{2/3}}{n^{5/6}\sigma}\right.\\
&+\frac{C_{H_g}m_{31}^{1/3}m_{32}^{2/3}}{n\sigma^2}+\frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1+\frac{\log n}{p}\right)\sqrt{\frac{p}{n}}\\
&+\frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1+\frac{\log n}{p}\right)^{3/2}\frac{1}{\sqrt{n}}\\
&+\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{\log n}{p}\right)^{1/2}\\
&\left.+\frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2}n}\left(1+\frac{\log n}{p}\right)\right)\\
&+BC_1\left(\frac{\tau^4(\log n)^{3/2}}{\lambda_\Sigma^2\sqrt{n}}+\frac{\tau^2(\log n)^{3/2}}{\lambda_\Sigma\sqrt{n}}\right.\\
&\left.+\frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right),
\end{aligned}$$

where $m_{31}=E[|\nabla g(\mu)^\top(X-\mu)|^3]$, $m_{32}:=E[||X-\mu||^3]$, $\sigma^2=\nabla g(\mu)^\top\Sigma\nabla g(\mu)$, C is a universal constant and C_1 is a constant only depending on C_X.

Theorem 3.5 is obtained by tracing the implicit quantities in Theorem 3.1 for the function-of-mean model, via extracting the dependence on problem parameters in the Berry-Esseen theorems for the multivariate delta method (Pinelis & Molzon, 2016) and the standard bootstrap (Zhilova, 2020). In particular, the sub-Gaussian assumption is required to derive finite-sample concentration inequalities, in a similar spirit as the state-of-the-art high-dimensional CLTs (e.g., Chernozhukov et al. (2017); Lopes (2022)). On the other hand, the third moments such as $||E[(X-\mu)^3]||$ (operation norm of the third order tensor $E[(X-\mu)^3]$), m_{31} and m_{32} are due to the use of the Berry-Esseen theorem and a multivariate higher-order Berry-Esseen inequality in Zhilova (2020), which generally requires this order of moments. The bound in Theorem 3.5 can be simplified with reasonable assumptions on the involved quantities:

Corollary 3.6. *Suppose the conditions in Theorem 3.5 hold. Moreover, suppose that $\tau,C_{H_g},C_1=O(1)$, $C_{\nabla g},\lambda_\Sigma=\Theta(1)$, $||\nabla g(\mu)||^2=O(p)$, $\sigma^2=\Theta(p)$ and $||E[(X-\mu)^3]||=O(1)$. Then as $p,n\to\infty$,*

$$\begin{aligned}
&\left|P(|g(\mu)-g(\bar{X}_n)|\le t_{B,1-\alpha/2}S_{n,B})-(1-\alpha)\right|\\
&=B\times O\left(\left(1+\frac{\log n}{p}\right)\sqrt{\frac{p}{n}}\right.\\
&\left.+\frac{1}{\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right).
\end{aligned}$$

Consequently, for any fixed $B\ge1$, the cheap bootstrap confidence interval is asymptotically exact provided $p=o(n)$, i.e.,

$$\lim_{\substack{p,n\to\infty\\ p=o(n)}}P(|g(\mu)-g(\bar{X}_n)|\le t_{B,1-\alpha/2}S_{n,B})=1-\alpha.$$

In Corollary 3.6, the cheap bootstrap coverage error shrinks to 0 as $n\to\infty$ if $p=o(n)$, i.e., the problem dimension grows slower than n in any arbitrary fashion. Although there is no theoretical guarantee that the choice of $p=o(n)$ is tight, we offer numerical evidence in Section 4 where the cheap bootstrap works with a small B when $p/n=0.09$ but it fails (i.e., over-covers the target with a quite large interval width) when $p/n = 0.25$. Such a difference indicates that $p = o(n)$ can be tight in some cases. Recall that $||E[(X-\mu)^3]||$ denotes the operator norm of the third order tensor $E[(X-\mu)^3]$, and so the assumption $||E[(X-\mu)^3]|| = O(1)$ holds if the components of X are independent (or slightly weakly dependent). Other order assumptions in Corollary 3.6 are natural. An example of the function-of-mean model is $g(\mu) = ||\mu||^2$, used also in Zhilova (2020), whose confidence interval becomes a simultaneous confidence region for the mean vector μ.

3.3. Linear Functions

We consider a further specialization to linear g where, instead of sub-Gaussanity of X, we are now able to use weaker tail conditions. Assume $g(x) = g_1^\top x + g_2$, where $g_1 \in \mathbb{R}^p$ and $g_2 \in \mathbb{R}$ are known. Then $g(\bar{X}_n)$ and $g(\bar{X}_n^{*b})$ are essentially the sample mean and resample mean of i.i.d. random variables $g_1^\top X_i + g_2, i = 1, \ldots, n$.

First, we consider the case where $g_1^\top(X-\mu)$ is sub-exponential, i.e., $||g_1^\top(X-\mu)||_{\psi_1} := \inf\{\lambda > 0 : E[\psi_1(|g_1^\top(X-\mu)|/\lambda)] \le 1\} < \infty$, where $||\cdot||_{\psi_1}$ is the Orlicz norm induced by the function $\psi_1(x) = e^x - 1$. Sub-exponential property is a weaker tail condition than sub-Gaussianity; see e.g. Vershynin (2018) Sections 2.5 and 2.7. Under this condition, we have:

Theorem 3.7. *Suppose g is a linear function in the form $g(x) = g_1^\top x + g_2$. Assume that $\sigma^2 = g_1^\top \Sigma g_1 > 0$ and $||g_1^\top(X-\mu)||_{\psi_1} < \infty$. Then for any $n \ge 3$, we have the following finite-sample bound on the cheap bootstrap coverage error*

$$\begin{aligned}
&\left|P(|g(\mu) - g(\bar{X}_n)| \le t_{B,1-\alpha/2} S_{n,B}) - (1-\alpha)\right| \\
&\le \frac{C}{n} + BC\frac{E[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}} \\
&+ BC\frac{||g_1^\top(X-\mu)||_{\psi_1}^4 \log^{11}(n)}{\sigma^4\sqrt{n}},
\end{aligned}$$

where C is a universal constant.

Note that the bootstrap in Theorem 3.7 effectively applies on the univariate $g_1^\top(X-\mu)$. Nonetheless, proving Theorem 3.7 requires tools from high-dimensional CLT (Lopes, 2022; Chernozhukov et al., 2020), as this appears the only line of work that investigates finite-sample bootstrap errors (for mean estimation). The order of the bound in terms of p is controlled by $g_1^\top(X-\mu)$, and so if the latter is well-scaled by its standard deviation σ in the sense that $E[|g_1^\top(X-\mu)/\sigma|^3], ||g_1^\top(X-\mu)/\sigma||_{\psi_1} = O(1)$ (e.g., X follows a multivariate normal distribution), then the order is independent of p, which means the error goes to 0 for any p as long as $n \to \infty$. However, if the orders of $E[|g_1^\top(X-\mu)/\sigma|^3]$ and $||g_1^\top(X-\mu)/\sigma||_{\psi_1}$ depend on p, then the growing rate of p must be restricted by n to ensure the error goes to 0.

Next, we further weaken the tail condition on $g_1^\top(X-\mu)$. We only assume $E[|g_1^\top(X-\mu)|^q] < \infty$ for some $q \ge 4$. In this case, we have the following:

Theorem 3.8. *Suppose g is a linear function in the form of $g(x) = g_1^\top x + g_2$. Assume that $\sigma^2 = g_1^\top \Sigma g_1 > 0$ and $E[|g_1^\top(X-\mu)|^q] < \infty$ for some $q \ge 4$. Then for any $n \ge 3$, we have the following finite-sample bound on the cheap bootstrap coverage accuracy*

$$\begin{aligned}
&\left|P(|g(\mu) - g(\bar{X}_n)| \le t_{B,1-\alpha/2} S_{n,B}) - (1-\alpha)\right| \\
&\le \frac{BC_1\sqrt{\log n}}{n^{1/2-3/(2q)}} \max\Big\{E[|g_1^\top(X-\mu)/\sigma|^q]^{1/q}, \\
&\sqrt{E[|g_1^\top(X-\mu)/\sigma|^4]}\Big\} + C\frac{E[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}},
\end{aligned}$$

where C is a universal constant and C_1 is a constant depending only on q.

The implication of Theorem 3.8 on the choice of p is similar to Theorem 3.7. In parallel to the above, explicit finite-sample bounds for standard quantile-based bootstrap methods can also be obtained by means of Theorem 3.2 under the assumptions in Theorems 3.5, 3.7 or 3.8 (see Appendix B.2).

4. Numerical Experiments

We consider various high-dimensional examples:

Ellipsoidal estimation: The estimation target is $g(\mu) = ||\mu||^2$, where μ is the mean of $X \in \mathbb{R}^p$ with ground-truth distribution $N(0.02\mathbf{1}_p, 0.01I_{p\times p})$. Sample size $n = 10^5$ and dimension $p = 2.5 \times 10^4$.

Sinusoidal estimation: The estimation target is $g(\mu) = \sum_{i=1}^p \sin(\mu_i)$, where $\mu = (\mu_i)_{i=1}^p$ is the mean of $X \in \mathbb{R}^p$ with ground-truth distribution $N(\mathbf{0}, 0.01I_{p\times p})$. Sample size $n = 10^5$ and dimension $p = 2.5 \times 10^4$.

Linear regression with independent covariates: Consider the true model $Y = X^\top\beta + \varepsilon$, where $X \in \mathbb{R}^p$ follows $N(\mathbf{0}, 0.01I_{p\times p})$ and $\varepsilon \sim N(0,1)$ independent of X. The first, second and last $1/3$ components of $\beta = (\beta_i)_{i=1}^p$ are $0, 2, -1$ respectively. We estimate β given i.i.d. data $(X_i, Y_i)_{i=1}^n$ with $n = 10^5$ and $p = 9000$.

Logistic regression with independent covariates: Consider the true model $Y \in \{0,1\}$, $X \in \mathbb{R}^p$, $P(Y=1|X) = \exp(X^\top\beta)/(1+\exp(X^\top\beta))$, where $X \sim N(\mathbf{0}, 0.01I_{p\times p})$. The first 300 components of β_i's are 1, the second 300 components -1 and the rest 0. As suggested in Sur & Candès (2019), we choose such values of β_i's to make sure $Var(X^\top\beta) = 6$ does not increase with p so that $P(Y=1|X)$ is not trivially equal to 0 or 1 in most cases.

We estimate β given i.i.d. data $(X_i, Y_i)_{i=1}^n$ with $n = 10^5$ and $p = 9000$.

Stochastic simulation model: Consider a stochastic computer communication model used to calculate the steady-state average message delay (Cheng & Holland (1997); Lin et al. (2015); Lam & Qian (2022); see Appendix C.3 for details). This problem can be cast as computing $\psi(P_1, \dots, P_p)$ where ψ represents this expensive simulation model (due to the need to run very long time in order to reach steady state) and P_j's denote the input distributions, $p = 13$ in total. The data sizes for observing these 13 input models range from 3×10^4 to 6×10^4.

A real data example: We run logistic regression on the RCV1-v2 data in Lewis et al. (2004). This dataset contains $n = 804414$ manually categorized newswire stories with a total of $p = 47236$ features. "Economics" ("ECAT") is chosen as the $+1$ label. We target coefficient estimation.

Linear regression with dependent covariates: We consider the same linear regression setup as before but with two different distributions of X. One is $X \sim N(\mathbf{0}, 0.01\Sigma)$ where the components of Σ are $\Sigma_{ij} = 0.8^{|i-j|}$. For this distribution, the i-th component and j-th component of X are more dependent when i and j are closer to each other. The other distribution is $X \sim N(\mathbf{0}, 0.01AA^\top)$, where A is a random matrix whose components are i.i.d. from $U(0, 1)$. The two cases are referred to as "exponential decay" and "random covariance matrix" respectively.

Logistic regression with dependent covariates: We consider the same logistic regression setup as before but change the distribution of X into the exponential decay distribution mentioned above. Here we do not consider the random covariance matrix case because the significant noisiness of X makes $Var(X^\top\beta)$ so large that $P(Y = 1|X)$ is trivially equal to 0 or 1 in most cases, which is also avoided in other work (e.g., Sur & Candès (2019)).

Ridge regression with $p > n$: Consider the true linear model $Y = X^\top\beta + \varepsilon$, where $\varepsilon \sim N(0, 1)$ is independent of X. The first, second and last $1/3$ components of $\beta = (\beta_i)_{i=1}^p$ are $0, 2, -1$ respectively. We consider three distributions of X. The first one is $X \sim N(\mathbf{0}, 0.01I_{p\times p})$ (referred to as "independent"). The other two are exponential decay distribution and random covariance matrix mentioned above. Given i.i.d. data $(X_i, Y_i)_{i=1}^n$ with $n = 8000$ and $p = 9000$, we estimate β by means of ridge regression which minimizes $||Y - X\beta||^2 + \lambda||\beta||^2$.

(Regularized) logistic regression with varying dimensions p: We use the same setup of logistic regression with independent covariates as before (e.g. $n = 10^5$) but use $p \in \{12000, 15000, 18000, 21000, 25000\}$ to test the valid boundary of p (i.e., the maximum p that makes cheap bootstrap work) in this problem. Moreover, we also run a regularized version by adding the ℓ_2 regularization term $||\beta||^2/2$ to the log likelihood function of logistic regression to see the effect of regularization.

Setups and comparison benchmarks. In each example above, our targets are 95%-level confidence intervals for the target parameters. We test four bootstrap confidence intervals: 1) cheap bootstrap (1); 2) basic bootstrap described in Section 2; 3) percentile bootstrap described in Section 2; 4) standard error bootstrap that uses standard normal quantile and standard deviation of $\hat{\psi}_n^{*b}$'s in lieu of $t_{B,1-\alpha/2}$ and $S_{n,B}$ respectively in (1). For each setup except the real-data example, we run 1000 experimental repetitions, each time generating a new data set from the ground truth distribution and construct the intervals. We report the empirical coverage and average interval width over these repetitions. For examples with more than one target estimation quantity, we further average the coverages and widths over all these targets. For the high-dimensional linear regression with independent covariates, we additionally show a box plot of the coverage probabilities and confidence interval widths of each individual β_i. We vary the number of resamples B from 1 to 10 in all examples and report the running time (i.e., model fitting time for one point estimate and B bootstrap estimates; the time for outputting the confidence intervals using these estimates is negligible compared to the model fitting time) in the virtual machine e2-highmem-2 in Google Cloud Platform. Some examples have larger scale and thus are run in the virtual machine e2-highmem-8 with larger memory and better CPU, whose running time will be starred (*).

Results and discussions. Tables 1-5 and Figure 2 describe our results (Tables 2-5 are delegated to Appendix C.1 due to the limitation of space), where we report $B = 1, 2, 5, 10$. "CB", "BB", "PB" and "SEB" in Figure 2 stand for the cheap bootstrap, basic bootstrap, percentile bootstrap and standard error bootstrap respectively.

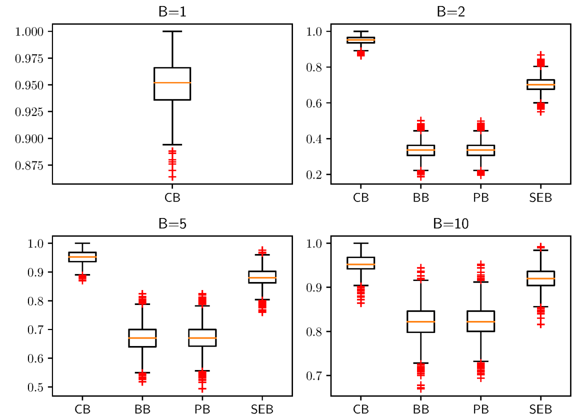
Coverage probability: According to Table 1, the cheap bootstrap performs the best in terms of the coverage probabilities in almost all cases (except the real-data example where we cannot validate and only report the interval widths). In all but three entries, the cheap bootstrap gives the closest coverages to the nominal 95% among all considered bootstrap methods, and in all but three entries the cheap bootstrap coverages are above 95%. In contrast, other approaches are substantially below the nominal level except for very few cases with $B = 10$. For example, in the ellipsoidal estimation, cheap bootstrap coverage probabilities are above 95% for all considered B's, while the highest coverage among other bootstrap methods is 82.1% even for $B = 10$. These observations corroborate with theory since unlike standard bootstrap methods, the cheap bootstrap gives small coverage errors even with very small B. Note that when $B = 1$, the entries of other bootstrap methods are all "N.A." since quantile-based approaches cannot even output two distinct finite numbers using one resample, and standard error bootstrap uses $B-1$ in the denominator of the sample variance. Similar results are also observed from Tables 2 and 3, which confirm that even for distributions with dependent components and ridge regression with $p > n$, the cheap bootstrap continues to work and outperform other bootstrap methods. In terms of the individual plot in Figure 2, the cheap bootstrap has coverages close to the nominal level 95% for almost all β_i's for any B. On the other hand, standard error bootstrap coverages are above 90% only when $B = 10$ while the quantile-based bootstrap coverages are still below 85% for most of the β_i's even for $B = 10$.

Interval width: Cheap bootstrap intervals are wider than other bootstrap intervals. However, these widths appear to decay very fast for the first few B's. In all examples, they decrease by around $2/3$ from $B = 1$ to $B = 2$ and by around $4/5$ from $B = 1$ to $B = 10$, hinting a quick enhancement of statistical efficiency as the computation budget increases. Note that while the other bootstrap intervals are shorter, they fall short in attaining the nominal coverage. It is thus reasonable to see the larger widths of the cheap bootstrap intervals which appear to push up the coverages by the right amounts.

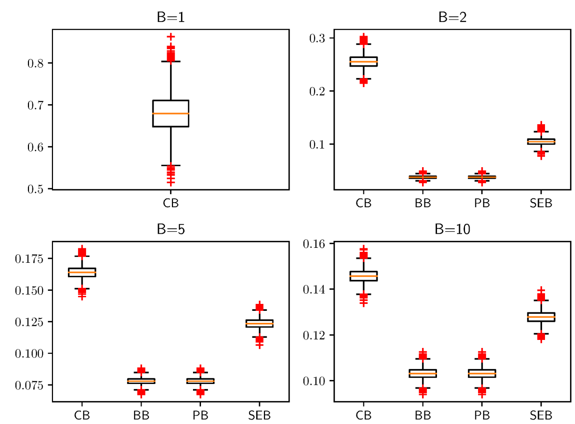
Valid boundary of p: Table 4 displays the results of logistic regression with varying p. We observe that the confidence widths have a dramatic increase when p increases to 21000 from 18000, which hints that the validity boundary of p under this setting should be at most 21000. Moreover, even for $p = 15000$ and $p = 18000$, the coverage probabilities are almost 100% which is overly large compared to the nominal level 95%. Combining with our existing favorable results for $p = 9000$, for this problem it appears the validity boundary could be indeed around 9000. As a comparison, Table 5 displays the results of regularized logistic regression. We can see although the interval widths seem to converge with the regularization, the coverage probabilities are still overly large for large p. For this regularized example, the validity boundary could be around 12000 or 15000, which is larger than the previous boundary. This contrast shows that regularization could be helpful to achieve a valid confidence interval when p is large. But no matter whether we add the regularization, it seems the validity boundary is always around the order $p = o(n)$ for logistic regression. The results here, combined with the favorable results for ridge regression with $p > n$ in Table 3, provide evidence that the precise validity threshold of p could be model-dependent.

5. Discussions and Other Connections

In this paper, we show how to run the bootstrap that achieves statistically valid coverage with very low cost, in terms of



(a) Coverage probability



(b) Confidence interval width

Figure 2. Box plots of empirical coverage probabilities and confidence interval widths of all β_i's for different numbers of resamples in a linear regression.

the number of resamples, even when the problem dimension grows closely with the data size. This is made possible by using sample-resample independence from a recent "cheap" bootstrap perspective. We provide general finite-sample coverage error bounds to support our validity, and specialize these bounds to explicit models including function-of-mean models with sub-Gaussian tails and linear models with weaker tail conditions. We discuss how our approach can operate with as low as one resample in attaining valid interval coverage in large sample. At the same time, the interval constructed with one resample tends to be wide, but fortunately shrinks quickly as the number of resamples increases from one even slightly. We run a wide set of numerical experiments to validate our theory and show the outperformance of our method compared to other benchmarks. As our numerical experiments hint that the tight growth rate of p in terms of n could be model-dependent, a future investigation is to establish more precise finite-sample bounds that

Table 1. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the first six numerical examples. The closest coverage probability to the nominal 95% level among all methods in each setting is **bold**.

Example	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
Ellipsoidal estimation	1	**96.0%**	0.069	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	9*
	2	**97.3%**	0.026	32.2%	0.002	5.5%	0.002	55.1%	0.006	15*
	5	**97.4%**	0.016	66.0%	0.005	13.6%	0.005	70.1%	0.007	36*
	10	**97.5%**	0.014	82.1%	0.006	20.8%	0.006	73.6%	0.008	70*
Sinusoidal estimation	1	**94.4%**	0.958	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	7*
	2	**95.2%**	0.384	29.6%	0.051	35.2%	0.051	63.2%	0.142	12*
	5	**93.6%**	0.248	71.2%	0.117	66.4%	0.117	86.4%	0.187	27*
	10	**94.4%**	0.222	84.0%	0.156	83.2%	0.156	89.6%	0.196	52*
Linear regression (independent covariates)	1	**95.1%**	0.68	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	443
	2	**95.1%**	0.256	33.5%	0.038	33.5%	0.038	70.2%	0.105	666
	5	**95.2%**	0.164	67.0%	0.078	67.0%	0.078	88.1%	0.123	1337
	10	**95.2%**	0.146	82.2%	0.103	82.2%	0.103	92.1%	0.128	2454
Logistic regression (independent covariates)	1	**96.1%**	2.866	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	50
	2	**96.9%**	1.074	39.7%	0.147	31.7%	0.147	73.4%	0.407	81
	5	**97.9%**	0.685	77.9%	0.302	63.3%	0.302	91.0%	0.479	175
	10	98.4%	0.609	91.9%	0.400	77.7%	0.400	**94.6%**	0.496	331
Stochastic simulation	1	**96.9%**	1.757×10^{-3}	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	1
	2	**98.8%**	6.417×10^{-4}	21.9%	6.962×10^{-5}	47.0%	6.962×10^{-5}	68.7%	1.930×10^{-4}	2
	5	99.7%	4.044×10^{-4}	43.2%	1.428×10^{-4}	**90.4%**	1.428×10^{-4}	87.1%	2.269×10^{-4}	3
	10	100%	3.591×10^{-4}	55.6%	1.915×10^{-4}	99.8%	1.915×10^{-4}	**92.6%**	2.375×10^{-4}	5
Real data	1	N.A.	3.594	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	156
	2	N.A.	1.361	N.A.	0.201	N.A.	0.201	N.A.	0.556	233
	5	N.A.	0.877	N.A.	0.414	N.A.	0.414	N.A.	0.658	464
	10	N.A.	0.779	N.A.	0.547	N.A.	0.547	N.A.	0.682	849

are tight in some appropriate sense.

We close this paper by positioning our results in the broader literature. First, our work is related to bootstrap coverage analysis. The commonest approach is to use the Edgeworth expansion that reveals the asymptotic higher-order terms in the coverage errors; see the comprehensive monograph (Hall, 2013). It is only until recently when finite-sample bounds on bootstrap appear, mostly in the high-dimensional CLT literature where the target is sample mean (Chernozhukov et al. (2017); Lopes (2022); Chernozhukov et al. (2020) and references therein). They aim to prove a uniform finite-sample bound of normal approximation of the (bootstrap) sample mean over all hyperrectangles. An alternative approach is to use Stein's method (Fang & Koike, 2021).

Second, within the bootstrap framework, various approaches have been proposed to reduce the Monte Carlo sampling effort by, e.g., variance reduction such as importance sampling (Booth & Do, 1993), or analytic approximation especially when applying iterated bootstraps (Booth & Hall, 1994; Lee & Young, 1995). These methods, however, require additional knowledge such as an explicit way to calculate variance, or focus on tail estimation issue. The closest work to the cheap bootstrap idea we utilize in this paper is Hall (1986b) who investigates the number of resamples for one-sided bootstrap intervals. Nonetheless, Hall (1986b) suggests a minimum of 19 for B in a 95%-level interval, obtained via an order-statistics calculation, which is still much larger than the minimal choice $B = 1$ in the cheap bootstrap.

Acknowledgement

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1834710 and IIS-1849280.

References

Alaa, A. and Van Der Schaar, M. Discriminative jackknife: Quantifying uncertainty in deep learning via higher-order influence functions. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 165–174. PMLR, 2020.

Balakrishnan, S. and Madigan, D. Algorithms for sparse linear classifiers in the massive data setting. *Journal of Machine Learning Research*, 9(10):313–337, 2008.

Bentkus, V. On the dependence of the Berry–Esseen bound on dimension. *Journal of Statistical Planning and Inference*, 113(2):385–402, 2003.

Beran, R. Prepivoting to reduce level error of confidence sets. *Biometrika*, 74(3):457–468, 1987.

Booth, J. G. and Do, K.-A. Simple and efficient methods for constructing bootstrap confidence intervals. *Computational Statistics*, 8:333–333, 1993.

Booth, J. G. and Hall, P. Monte Carlo approximation and the iterated bootstrap. *Biometrika*, 81(2):331–340, 1994.

Cheng, R. C. and Holland, W. Sensitivity of computer simulation experiments to errors in input data. *Journal of Statistical Computation and Simulation*, 57(1-4):219–241, 1997.

Chernozhukov, V., Chetverikov, D., and Kato, K. Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, 45(4):2309–2352, 2017.

Chernozhukov, V., Chetverikov, D., and Koike, Y. Nearly optimal central limit theorem and bootstrap approximations in high dimensions. *arXiv preprint arXiv:2012.09513*, 2020.

Davison, A. C. and Hinkley, D. V. *Bootstrap Methods and Their Application*. Cambridge University Press, 1997.

DiCiccio, T. and Tibshirani, R. Bootstrap confidence intervals and bootstrap approximations. *Journal of the American Statistical Association*, 82(397):163–170, 1987.

DiCiccio, T. J., Efron, B., et al. Bootstrap confidence intervals. *Statistical Science*, 11(3):189–228, 1996.

Efron, B. Better bootstrap confidence intervals. *Journal of the American statistical Association*, 82(397):171–185, 1987.

Efron, B. and Tibshirani, R. J. *An Introduction to the Bootstrap*. Chapman & Hall/CRC, 1994.

Fang, X. and Koike, Y. High-dimensional central limit theorems by Stein's method. *Annals of Applied Probability*, 31(4):1660–1686, 2021.

Giordano, R., Stephenson, W., Liu, R., Jordan, M., and Broderick, T. A Swiss army infinitesimal jackknife. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1139–1147. PMLR, 2019.

Hall, P. On the bootstrap and confidence intervals. *Annals of Statistics*, 14(4):1431–1452, 1986a.

Hall, P. On the number of bootstrap simulations required to construct a confidence interval. *Annals of Statistics*, pp. 1453–1462, 1986b.

Hall, P. Theoretical comparison of bootstrap confidence intervals. *The Annals of Statistics*, pp. 927–953, 1988.

Hall, P. *The Bootstrap and Edgeworth Expansion*. Springer Science & Business Media, 2013.

Hall, P. and Martin, M. A. On bootstrap resampling and iteration. *Biometrika*, 75(4):661–671, 1988.

Kleiner, A., Talwalkar, A., Sarkar, P., and Jordan, M. I. The big data bootstrap. In *Proceedings of the 29th International Conference on Machine Learning*, pp. 1787–1794, 2012.

Lam, H. A cheap bootstrap method for fast inference. *arXiv preprint arXiv:2202.00090*, 2022a.

Lam, H. Cheap bootstrap for input uncertainty quantification. In *2022 Winter Simulation Conference (WSC)*, pp. 2318–2329. IEEE, 2022b.

Lam, H. and Qian, H. Subsampling to enhance efficiency in input uncertainty quantification. *Operations Research*, 70(3):1891–1913, 2022.

Lee, S. M. and Young, G. A. Asymptotic iterated bootstrap confidence intervals. *Annals of Statistics*, 23(4):1301–1330, 1995.

Lewis, D. D., Yang, Y., Russell-Rose, T., and Li, F. Rcv1: A new benchmark collection for text categorization research. *Journal of Machine Learning Research*, 5(Apr):361–397, 2004.

Lin, Y., Song, E., and Nelson, B. Single-experiment input uncertainty. *Journal of Simulation*, 9(3):249–259, 2015.

Lopes, M. E. Central limit theorem and bootstrap approximation in high dimensions: Near $1/\sqrt{n}$ rates via implicit smoothing. *Annals of Statistics*, 50(5):2492–2513, 2022.

Lu, Z., Ie, E., and Sha, F. Uncertainty estimation with infinitesimal jackknife, its distribution and mean-field approximation. *arXiv preprint arXiv:2006.07584*, 2020.

Pinelis, I. and Molzon, R. Optimal-order bounds on the rate of convergence to normality in the multivariate delta method. *Electronic Journal of Statistics*, 10(1):1001–1063, 2016.

Rudelson, M. and Vershynin, R. Hanson-Wright inequality and sub-Gaussian concentration. *Electronic Communications in Probability*, 18:1–9, 2013.

Schulam, P. and Saria, S. Can you trust this prediction? Auditing pointwise reliability after learning. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 1022–1031. PMLR, 2019.

Shao, J. and Tu, D. *The Jackknife and Bootstrap*. Springer Science & Business Media, 2012.

Singh, S., Kubica, J., Larsen, S., and Sorokina, D. Parallel large scale feature selection for logistic regression. In *Proceedings of the 2009 SIAM International Conference on Data Mining*, pp. 1172–1183. SIAM, 2009.

Sur, P. and Candès, E. J. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525, 2019.

Van der Vaart, A. W. *Asymptotic Statistics*. Cambridge University Press, 2000.

Vershynin, R. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge University Press, 2018.

Zhilova, M. Nonclassical Berry–Esseen inequalities and accuracy of the bootstrap. *Annals of Statistics*, 48(4): 1922–1939, 2020.

A. A More Detailed Explanation on the Cheap Bootstrap in the Low-Dimensional Case

This section aims to give more details on the cheap bootstrap method in the low-dimensional case discussed in Section 2.

Suppose we are interested in estimating a target statistical quantity $\psi := \psi(P_X)$ where $\psi(\cdot) : \mathcal{P} \mapsto \mathbb{R}$ is a functional defined on the probability measure space $\mathcal{P}$. Given i.i.d. data $X_1, \ldots, X_n \in \mathbb{R}^p$ following the unknown distribution P_X, we denote the empirical distribution as $\hat{P}_{X,n}(\cdot) := (1/n)\sum_{i=1}^n I(X_i \in \cdot)$. A natural point estimator is $\hat{\psi}_n := \psi(\hat{P}_{X,n})$.

The cheap bootstrap confidence interval for ψ is constructed as follows. Conditional on $X_1, \ldots, X_n$, we independently resample, i.e., sample with replacement, the data for B times to obtain resamples $\{X_1^{*b}, \ldots, X_n^{*b}\}, b = 1, \ldots, B$. Denoting $\hat{P}^{*b}_{X,n}$ as the resample empirical distributions, we construct B resample estimates $\hat{\psi}_n^{*b} := \psi(\hat{P}^{*b}_{X,n})$. A $(1-\alpha)$-level confidence interval is then given by

$$\left[\hat{\psi}_n - t_{B,1-\alpha/2} S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2} S_{n,B}\right], \tag{12}$$

where $S^2_{n,B} = (1/B)\sum_{b=1}^B (\hat{\psi}_n^{*b} - \hat{\psi}_n)^2$, and $t_{B,1-\alpha/2}$ is the $(1-\alpha/2)$-th quantile of t_B, the t-distribution with degree of freedom B. Theorem 1 in Lam (2022a) shows that, under conditions on par with standard bootstrap methods, (12) is an asymptotically exact $(1-\alpha)$-level confidence interval *for any fixed* $B \geq 1$, i.e.,

$$P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2} S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2} S_{n,B}]) \to 1 - \alpha, \text{ as } n \to \infty \tag{13}$$

where P is the probability with respect to both the data and the randomness in the resampling process.

Here we explain the asymptotic argument that gives rise to (13). Under suitable conditions, the sampling distribution of an estimate $\hat{\psi}_n$ and the distribution of a resample estimate $\hat{\psi}_n^*$ are approximately equal. More formally, they are equal in the asymptotic sense of two CLTs $\sqrt{n}(\hat{\psi}_n - \psi) \xrightarrow{d} N(0, \sigma^2)$ for some $\sigma^2 > 0$, and $\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n) \xrightarrow{d} N(0, \sigma^2)$ (conditional on $X_1, \ldots, X_n$ in probability) for the same σ^2. By means of a conditional argument, we can combine the two aforementioned CLTs to obtain the following joint convergence

$$\sqrt{n}(\hat{\psi}_n - \psi, \hat{\psi}_n^{*1} - \hat{\psi}_n, \ldots, \hat{\psi}_n^{*B} - \hat{\psi}_n) \xrightarrow{d} (\sigma Z_0, \sigma Z_1, \ldots, \sigma Z_B), \text{ as } n \to \infty \tag{14}$$

where $Z_b, b = 0, \ldots, B$ are i.i.d. standard normal. From (14), we can establish the convergence of a pivotal t-statistic $(\hat{\psi}_n - \psi)/S_{n,B} \xrightarrow{d} t_B$ which gives (13). The above shows that, with B fixed as small as 1, (12) already offers a coverage close to the nominal level as $n \to \infty$. In this argument, the approximation accuracy of $(\hat{\psi}_n - \psi)/S_{n,B}$ by the t_B random variable is crucial. However, in the high-dimensional case when $p \to \infty$ as $n \to \infty$, the joint CLT (14) may not hold and thus the techniques in this paper are needed to establish the validity of the cheap bootstrap method.

B. Additional Theoretical Results

This section provides additional theoretical results. Appendix B.1 establishes an alternative finite-sample bound for the cheap bootstrap that generalizes Theorem 3.1 to cover the large-B regime. Section B.2 provides finite-sample bounds for standard quantile-based bootstrap methods under the conditions in Sections 3.2 and 3.3.

B.1. Further Finite-Sample Bound for the Cheap Bootstrap

The following result generalizes Theorem 3.1 to include both the small and large-B regimes:

Theorem B.1. *Suppose we have the finite sample accuracy for the estimator* $\hat{\psi}_n$

$$\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) \leq x) - \Phi\left(\frac{x}{\sigma}\right)\right| \leq \mathcal{E}_1, \tag{15}$$

and with probability at least $1-\beta$ *we have the finite sample accuracy for the bootstrap estimator* $\hat{\psi}_n^*$

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right)\right| \leq \mathcal{E}_2,$$

where $\sigma > 0$ and P^* denotes the probability on a resample conditional on the data. Further, suppose that the following concentration inequality holds

$$P\left(\left|\sqrt{\frac{1}{B}\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2} - \sigma\right| \geq \mathcal{E}_3\right) \leq \mathcal{E}_4, \tag{16}$$

where $\mathcal{E}_3$ is deterministic and $\sigma - \mathcal{E}_3 > 0$. Then we have the following finite sample bound on the cheap bootstrap coverage error

$$\begin{aligned}
&|P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}]) - (1-\alpha)| \\
&\leq \min\left\{2\mathcal{E}_1 + 2B\mathcal{E}_2 + \beta, 2\mathcal{E}_1 + 2\mathcal{E}_4 + \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2}\right\},
\end{aligned}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal.

The finite sample accuracy in Theorem B.1 consists of two parts. The first one $2\mathcal{E}_1 + 2B\mathcal{E}_2 + \beta$ works well when B is small as shown in Sections 3.2 and 3.3 but it deteriorates when B grows. In contrast, the second part

$$2\mathcal{E}_1 + 2\mathcal{E}_4 + \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2} \tag{17}$$

vanishes as $B, n \to \infty$ but does not if B is bounded even if $n \to \infty$. Its behavior for bounded B is easy to see: The third term $\sqrt{2/\pi}|t_{B,1-\alpha/2} - z_{1-\alpha/2}|$ in (17) is bounded away from zero if B is bounded and thus (17) never converges to zero even if $n \to \infty$. To explain why (17) vanishes as $B, n \to \infty$, first note that the first term $2\mathcal{E}_1$ is independent of B and satisfies $\mathcal{E}_1 \to 0$ as $n \to \infty$ by the Berry-Esseen Theorem for a reasonable model $\psi(\cdot)$ such as the function-of-mean model in Section 3.2. Second, notice that $\sqrt{(1/B)\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2}$ is the bootstrap estimator of the asymptotic standard deviation σ. Therefore, (16) is the concentration inequality for the bootstrap principle applied to the estimation of σ and would hold with a choice of $\mathcal{E}_3$ and $\mathcal{E}_4$ satisfying $\mathcal{E}_3 \to 0, \mathcal{E}_4 \to 0$ as $B, n \to \infty$. Lastly, since $t_B \xrightarrow{d} N(0,1)$ as $B \to \infty$, by Lemma 21.2 in Van der Vaart (2000), we have $t_{B,1-\alpha/2} \to z_{1-\alpha/2}$ as $B \to \infty$. Therefore, we can see the second part (17) converges to zero as $B, n \to \infty$ at any rate.

Under concrete assumptions on X as in Sections 3.2 and 3.3, explicit forms of $\mathcal{E}_3$ and $\mathcal{E}_4$ depending on B, n and the distribution of X can be derived, based on similar arguments as the explicit bounds in Theorems 3.5-3.8. Then by studying the order of these explicit bounds with respect to B, p and n, we can deduce a proper growth rate of dimension $p = p(B, n)$ which ensures a vanishing error as $B, n \to \infty$. The concentration inequality (16) seems unexplored in the literature and we leave it as future work.

B.2. Explicit Finite-Sample Bounds for Quantile-Based Bootstrap Methods

In this section, in parallel to Theorems 3.5-3.8 for the cheap bootstrap, we provide a few explicit bounds for standard quantile-based bootstrap methods under the same conditions.

The first result is in parallel to Theorem 3.5 under the function-of-mean model:

Theorem B.2. *Suppose the conditions in Theorem 3.5 hold. If $q_{\alpha/2}$, $q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $g(\bar{X}_n^*) - g(\bar{X}_n)$ respectively given $X_1, \ldots, X_n$, then the finite-sample bound on the basic bootstrap coverage error is given by*

$$\begin{aligned}
&|P(g(\bar{X}_n) - q_{1-\alpha/2} \leq g(\mu) \leq g(\bar{X}_n) - q_{\alpha/2}) - (1-\alpha)| \\
&\leq \frac{12}{n} + C\left(\frac{m_{31}}{\sqrt{n}\sigma^3} + \frac{C_{H_g}m_{31}^{1/3}tr(\Sigma)}{\sqrt{n}\sigma^2} + \frac{C_{H_g}m_{32}^{2/3}}{n^{5/6}\sigma} + \frac{C_{H_g}m_{31}^{1/3}m_{32}^{2/3}}{n\sigma^2}\right. \\
&+ \frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1 + \frac{\log n}{p}\right)\sqrt{\frac{p}{n}} + \frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}} + \frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1 + \frac{\log n}{p}\right)^{3/2}\frac{1}{\sqrt{n}}
\end{aligned}$$

$$+\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{\log n}{p}\right)\Bigg)$$
$$+C_1\left(\frac{\tau^4(\log n)^{3/2}}{\lambda_\Sigma^2\sqrt{n}}+\frac{\tau^2(\log n)^{3/2}}{\lambda_\Sigma\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right),$$

where C is a universal constant and C_1 is a constant only depending on C_X. If $q'_{\alpha/2}$, $q'_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $g(\bar{X}_n^*)$ respectively given $X_1,\ldots,X_n$, then the finite-sample bound on the percentile bootstrap coverage error is given by

$$|P(q'_{\alpha/2}\le g(\mu)\le q'_{1-\alpha/2})-(1-\alpha)|$$
$$\le\frac{12}{n}+C\Bigg(\frac{m_{31}}{\sqrt{n}\sigma^3}+\frac{C_{H_g}m_{31}^{1/3}tr(\Sigma)}{\sqrt{n}\sigma^2}+\frac{C_{H_g}m_{32}^{2/3}}{n^{5/6}\sigma}+\frac{C_{H_g}m_{31}^{1/3}m_{32}^{2/3}}{n\sigma^2}$$
$$+\frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1+\frac{\log n}{p}\right)\sqrt{\frac{p}{n}}+\frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1+\frac{\log n}{p}\right)^{3/2}\frac{1}{\sqrt{n}}$$
$$+\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{\log n}{p}\right)\Bigg)$$
$$+C_1\left(\frac{\tau^4(\log n)^{3/2}}{\lambda_\Sigma^2\sqrt{n}}+\frac{\tau^2(\log n)^{3/2}}{\lambda_\Sigma\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right),$$

where C is a universal constant and C_1 is a constant only depending on C_X.

Our discussion below Theorem 3.2 shows that the cheap bootstrap error bound with any given B and quantile-based bootstrap error bounds with $B=\infty$ only differ up to a constant. Therefore, the order analysis for the cheap bootstrap in Corollary 3.6 also applies here, that is, under the conditions in Corollary 3.6, the quantile-based bootstrap coverage errors shrink to 0 as $n\to\infty$ if $p=o(n)$.

The second result is in parallel to Theorem 3.7 under the sub-exponential assumption and linearity of g:

Theorem B.3. *Suppose the conditions in Theorem 3.7 hold. If $q_{\alpha/2}$, $q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $g(\bar{X}_n^*)-g(\bar{X}_n)$ respectively given $X_1,\ldots,X_n$, then the finite-sample bound on the basic bootstrap coverage error is given by*

$$|P(g(\bar{X}_n)-q_{1-\alpha/2}\le g(\mu)\le g(\bar{X}_n)-q_{\alpha/2})-(1-\alpha)|$$
$$\le C\left(\frac{1}{n}+\frac{E[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}}+\frac{||g_1^\top(X-\mu)||_{\psi_1}^4\log^{11}(n)}{\sigma^4\sqrt{n}}\right)+\frac{CE[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}},$$

where C is a universal constant. If $q'_{\alpha/2}$, $q'_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $g(\bar{X}_n^*)$ respectively given $X_1,\ldots,X_n$, then the finite-sample bound on the percentile bootstrap coverage error is given by

$$|P(q'_{\alpha/2}\le g(\mu)\le q'_{1-\alpha/2})-(1-\alpha)|$$
$$\le C\left(\frac{1}{n}+\frac{E[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}}+\frac{||g_1^\top(X-\mu)||_{\psi_1}^4\log^{11}(n)}{\sigma^4\sqrt{n}}\right)+\frac{CE[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}},$$

where C is a universal constant.

The last result is in parallel to Theorem 3.8 under moment conditions and linearity of g:

Theorem B.4. *Suppose the conditions in Theorem 3.8 hold. If $q_{\alpha/2}$, $q_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha/2)$-quantiles of $g(\bar{X}_n^*)-g(\bar{X}_n)$ respectively given $X_1,\ldots,X_n$, then the finite-sample bound on the basic bootstrap coverage error is given by*

$$|P(g(\bar{X}_n)-q_{1-\alpha/2}\le g(\mu)\le g(\bar{X}_n)-q_{\alpha/2})-(1-\alpha)|$$

$$\le \frac{2}{\sqrt{n}} + C_1 \max\left\{E[|g_1^\top(X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top(X-\mu)/\sigma|^4]}\right\}\left(\frac{(\log n)^{3/2}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}}\right)$$
$$+ \frac{CE[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}},$$

where C is a universal constant and C_1 is a constant depending only on q. If $q'_{\alpha/2}$, $q'_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha/2)$-quantiles of $g(\bar{X}_n^*)$ respectively given $X_1, \ldots, X_n$, then the finite-sample bound on the percentile bootstrap coverage error is given by

$$|P(q'_{\alpha/2} \le g(\mu) \le q'_{1-\alpha/2}) - (1-\alpha)|$$
$$\le \frac{2}{\sqrt{n}} + C_1 \max\left\{E[|g_1^\top(X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top(X-\mu)/\sigma|^4]}\right\}\left(\frac{(\log n)^{3/2}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}}\right)$$
$$+ \frac{CE[|g_1^\top(X-\mu)|^3]}{\sigma^3\sqrt{n}},$$

where C is a universal constant and C_1 is a constant depending only on q.

The order analysis for the cheap bootstrap also applies to Theorems B.3 and B.4. If $g_1^\top(X-\mu)$ is well-scaled by its standard deviation σ in the sense that the L_p norm and Orlicz norm $||\cdot||_{\psi_1}$ is independent of p, then the errors shrink to 0 for any p as $n \to \infty$. Otherwise, the growth rate of p should depend on n to obtain a vanishing error.

C. Details of Numerical Experiments and Additional Numerical Results

In this section, we present additional results and details of the experiments in Section 4. We also report some additional experiments. Section C.1 presents tables for experimental results in Section 4 that have not been shown in previous sections. The following subsections, C.2, C.3 and C.4 provide additional details for logistic regression with independent covariates, the computer network and the real world problem presented in Section 4. Section C.5 further validates our performances by a simulation study with a lower nominal level 70%. Finally, Section C.6 studies the coverage error behavior as B and n vary for the sinusoidal estimation.

C.1. Additional Tables

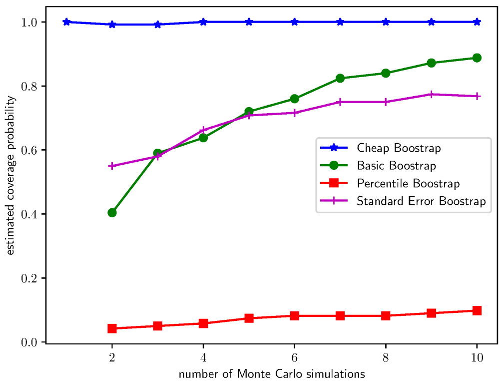
Tables 2-5 reports the experimental results of regression problems with dependent covariates, ridge regression and (regularized) logistic regression with different p presented in Section 4 respecitvely.

C.2. Logistic Regression with Independent Covariates

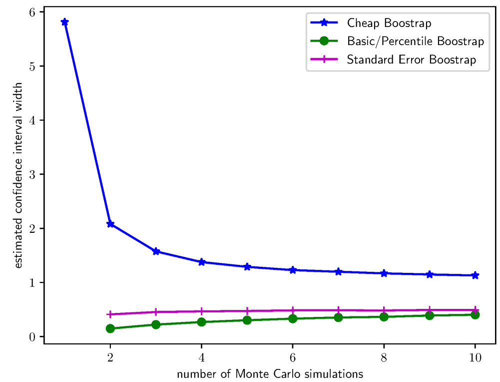
Figure 3 presents the coverage probabilities and confidence interval widths of 95%-level confidence intervals for three typical choices of parameters: $\beta_1 = 1$, $\beta_{301} = -1$ and $\beta_{601} = 0$. We observe that all cheap bootstrap coverage probabilities are close to or larger than the nominal level 95% while other bootstrap method coverages are below 90% except for the standard error bootstrap for $\beta_{601} = 0$ and $B \ge 5$. Besides, cheap bootstrap interval widths are larger than others but decay very fast for the first few B's, in line with our observation in the previous linear regression example. In fact, it is already quite close to other bootstrap widths for $\beta_{601} = 0$ and $B = 10$. Figure 4 reports the box plot of the coverage probabilities and confidence interval widths of all β_i's with $B = 1, 2, 5, 10$. We distinguish between $\beta_i \neq 0$ and $\beta_i = 0$ since the former has wider widths than the latter. For $\beta_i \neq 0$, the cheap bootstrap widths shrink more slowly so that almost all cheap bootstrap coverage probabilities are 100% but other bootstrap method coverages are still below 90% in almost all cases. For $\beta_i = 0$, the cheap bootstrap with any B, standard error bootstrap with $B = 5, 10$ and basic bootstrap with $B = 10$ have coverage probabilities close to the nominal level 95%. In other cases, most of the coverage probabilities are below 85%. A similar decay rate for the cheap bootstrap interval width is also observed here: it decreases by around $2/3$ from $B = 1$ to $B = 2$ and by around $4/5$ from $B = 1$ to $B = 10$.

C.3. Computer Network

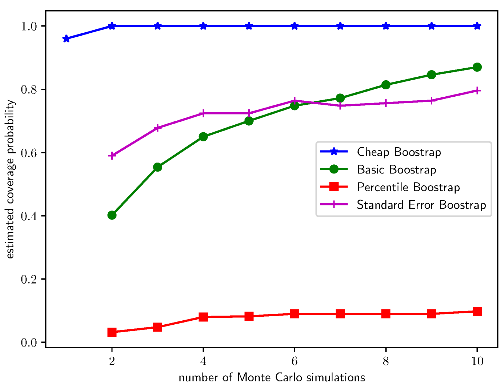
We detail the specifications of the computer communication network simulation model; similar models have been used in Cheng & Holland (1997); Lin et al. (2015); Lam & Qian (2022). This network can be represented by an undirected graph



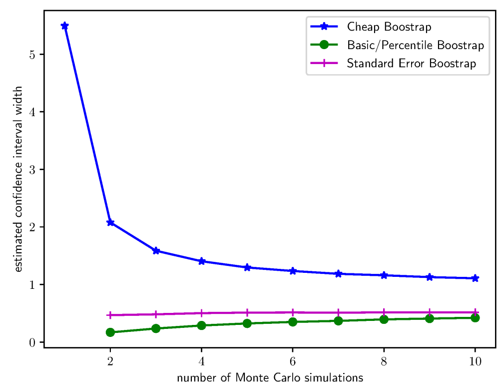
(a) Coverage probability for $\beta_1 = 1$



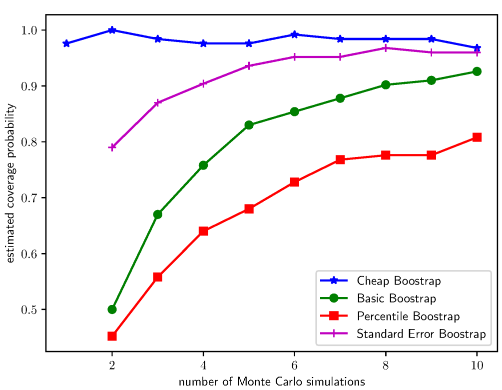
(b) Confidence interval width for $\beta_1 = 1$



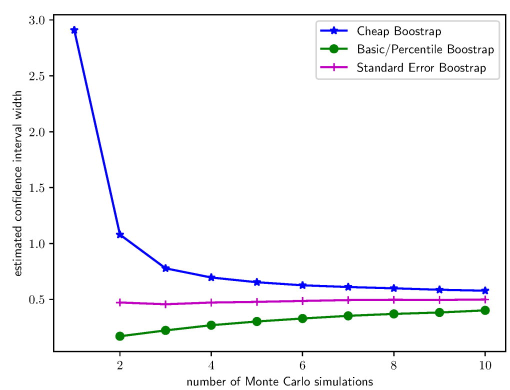
(c) Coverage probability for $\beta_{301} = -1$



(d) Confidence interval width for $\beta_{301} = -1$



(e) Coverage probability for $\beta_{601} = 0$



(f) Confidence interval width for $\beta_{601} = 0$

Figure 3. Empirical coverage probabilities and confidence interval widths for different numbers of resamples in a logistic regression.

Table 2. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the regression problems with dependent covariates. The closest coverage probability to the nominal 95% level among all methods in each setting is **bold**.

Example	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
Linear	1	**95.1%**	1.451	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	167*
regression	2	**95.1%**	0.546	33.6%	0.081	33.5%	0.081	70.3%	0.224	258*
(exponential	5	**95.2%**	0.350	67.1%	0.166	67.1%	0.166	88.2%	0.264	531*
decay)	10	**95.2%**	0.311	82.2%	0.220	82.2%	0.220	92.1%	0.273	987*
Linear	1	**94.7%**	129.255	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	511
regression	2	**94.9%**	52.791	33.9%	7.575	34.5%	7.575	69.9%	20.995	798
(random	5	**94.8%**	30.733	67.8%	14.875	66.8%	14.875	87.5%	23.555	1659
cov matrix)	10	**95.0%**	27.048	82.1%	19.398	82.1%	19.398	91.9%	23.787	3095
Logistic	1	**95.3%**	7.411	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	114*
regression	2	**95.5%**	2.787	34.9%	0.404	32.3%	0.404	70.7%	1.119	175*
(exponential	5	**95.8%**	1.787	69.5%	0.832	64.8%	0.832	88.6%	1.318	360*
decay)	10	**96.0%**	1.588	84.6%	1.101	79.8%	1.101	92.6%	1.364	667*

in Figure 5. The four nodes denote message processing units and the four edges are transport channels. For every pair of nodes i, j ($i \neq j$), there are external messages which enter into node i from the external and are to be transmitted to node j through a prescribed path. Their arrival time follows a Poisson process with parameter $\lambda_{i,j}$ showed in Table 6. All the message lengths (unit: bits) are i.i.d. following a common exponential distribution with mean 300 bits. Suppose each unit spends 0.001 second to process a message passing it. We assume the node storage is unlimited but the channel storage is restricted to 275000 bits. Message speed in transport channels is 150000 miles per second and channel i has length $100i$ miles. Therefore, it takes $l/275000 + 100i/150000$ seconds for a message with length l bits to pass channel i. Suppose the network is empty at the beginning. The performance measure of interest is the steady-state average delay for the messages where delay means the time from the entering node to the destination node. It has approximate true value 7.05×10^{-3}. This example has 13 unknown input distributions, i.e., 12 inter-arrival time distributions $\mathrm{Exp}(\lambda_{i,j})$ and one message length distribution $\mathrm{Exp}(1/300)$, for which we have data sizes from 3×10^4 to 6×10^4. Given input distributions $P_1, \dots, P_{13}$, the performance measure of this system can be computed accurately by

$$\psi(P_1, \dots, P_{13}) = E_{P_1,\dots,P_{13}} \left[\frac{1}{9500} \sum_{k=501}^{10000} D_k \right],$$

where D_k is the delay for the k-th message. The point estimator of $\psi(P_1, \dots, P_{13})$ is taken as $\hat{\psi} = \psi(\hat{P}_{1,n_1}, \dots, \hat{P}_{13,n_{13}})$ where each $\hat{P}_{i,n_i}$ is the empirical distribution of n_i i.i.d. samples $\{X_{i,j}, j = 1, \dots, n_i\}$ from the i-th input distribution P_i.

Next we construct the bootstrap estimator $\hat{\psi}^{*b}$. For each $b = 1, \dots, B$ and $i = 1, \dots, 13$, we sample with replacement the data $\{X_{i,j}, j = 1, \dots, n_i\}$ to obtain the bootstrap resamples $\{X^{*b}_{i,j}, j = 1, \dots, n_i\}$ and denote the resample empirical distribution by $\hat{P}^{*b}_{i,n_i}$. The sampling procedure is conducted independently for different b and i. The bootstrap estimator is taken as $\hat{\psi}^{*b} = \psi(\hat{P}^{*b}_{1,n_1}, \dots, \hat{P}^{*b}_{13,n_{13}})$. The cheap bootstrap confidence interval is still constructed as in (1).

The results for the above configuration can be found in the row "Stochastic simulation" of Table 1 and the corresponding discussions can be found in Section 4.

To investigate the robustness of the cheap bootstrap or other methods, we consider an alternative configuration where computer network is the same but the input models are different. More concretely, all 13 input models (12 inter-arrival time distributions and one message length distribution) are changed to Gamma distributions $\mathrm{Gamma}(\alpha, \beta)$ which have densities of the form

$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0.$$

Table 3. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the ridge regression with $p = 9000$ and $n = 8000$. The closest coverage probabilities to the nominal 95% level among all methods are **bold**.

Example	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
Ridge	1	**96.7%**	15.572	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	48
regression	2	**97.2%**	5.719	23.1%	0.553	25.5%	0.553	68.8%	1.534	72
(independent;	5	**97.5%**	3.595	47.0%	1.141	51.8%	1.141	86.6%	1.807	144
$\lambda = 0.1$)	10	**97.6%**	3.171	60.2%	1.510	66.0%	1.510	90.7%	1.870	264
Ridge	1	**96.3%**	13.734	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	48
regression	2	**96.7%**	5.082	27.1%	0.539	24.8%	0.539	68.7%	1.495	72
(independent;	5	**97.0%**	3.212	54.8%	1.111	50.3%	1.111	86.5%	1.761	144
$\lambda = 1$)	10	**97.1%**	2.839	69.2%	1.471	64.3%	1.471	90.6%	1.822	264
Ridge	1	**96.9%**	9.588	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	47
regression	2	**97.7%**	3.453	11.6%	0.263	37.3%	0.263	48.0%	0.728	71
(exponential	5	**98.4%**	2.143	23.8%	0.541	73.9%	0.541	59.9%	0.857	142
decay; $\lambda = 0.1$)	10	**98.7%**	1.882	31.2%	0.716	88.5%	0.716	62.8%	0.887	260
Ridge	1	**95.7%**	5.776	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	48
regression	2	**96.3%**	2.145	19.5%	0.240	36.2%	0.240	61.6%	0.665	71
(exponential	5	**96.9%**	1.360	39.8%	0.495	71.8%	0.495	78.0%	0.784	142
decay; $\lambda = 1$)	10	**97.2%**	1.203	51.5%	0.654	86.7%	0.654	81.9%	0.811	261
Ridge	1	**96.7%**	15.232	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	49
regression	2	**96.8%**	5.527	19.4%	0.462	21.2%	0.462	64.4%	1.281	74
(random cov	5	**96.5%**	3.448	39.6%	0.953	43.4%	0.953	81.5%	1.510	151
matrix; $\lambda = 0.1$)	10	**96.3%**	3.032	51.2%	1.261	56.4%	1.261	85.6%	1.563	279
Ridge	1	**96.3%**	13.302	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	50
regression	2	**96.4%**	4.873	23.5%	0.457	20.9%	0.457	65.2%	1.267	76
(random cov	5	**96.2%**	3.060	47.9%	0.943	42.8%	0.943	82.5%	1.493	155
matrix; $\lambda = 1$)	10	**95.9%**	2.696	61.2%	1.247	55.7%	1.247	86.6%	1.545	286

The message length distribution follows $\mathrm{Gamma}(2.5, 1/200)$ and the parameters for the inter-arrival time distributions $\mathrm{Gamma}(\alpha_{i,j}, \beta_{i,j})$ are given in Table 7. Under the new input distributions, the true steady-state mean delay is approximately 0.0109. Figure 6 reports the results. The cheap bootstrap coverage probabilities are close to the nominal level 95% for any B while those of the basic bootstrap and standard error bootstrap are below 60% and 90% respectively even for $B = 10$. Percentile bootstrap also performs well when $B \geq 7$ perhaps because of the skewness of the estimates. But this is not always the case in view of the previous numerical results.

C.4. Real World Problem

The data we use is the RCV1-v2 data in Lewis et al. (2004). This dataset contains $n = 804414$ manually categorized newswire stories with a total of $p = 47236$ features. We compare the confidence interval widths of the logistic regression parameters for the four bootstrap methods, by using all the observations to run the logistic regression and estimate the parameters. That is, the observation matrix is of the size $804414 \times 47236 \approx 4 \times 10^{10}$. There are up to 103 different categories for all these newswire stories. As in Singh et al. (2009) and Balakrishnan & Madigan (2008), we only use the "economics" ("ECAT") as the +1 label, i.e., the label Y is 1 if the newswire story is in "economics" and 0 if not, which leads to 119920 positive labels. Besides, we add l_2 regularization to this logistic regression as in Singh et al. (2009) and Balakrishnan & Madigan (2008).

Table 4. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the logistic regression with $n = 10^5$ and different p. The closest coverage probability to the nominal 95% level among all methods in each setting is **bold**.

p	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
12000	1	**96.7%**	3.697	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	46*
	2	**97.7%**	1.378	42.8%	0.182	32.3%	0.182	75.9%	0.504	76*
	5	99.0%	0.878	83.1%	0.375	64.6%	0.375	**93.1%**	0.594	165*
	10	99.5%	0.778	**95.3%**	0.496	79.0%	0.496	96.2%	0.615	314*
15000	1	**97.5%**	5.368	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	84*
	2	**98.7%**	1.993	46.3%	0.251	33.3%	0.251	80.2%	0.697	142*
	5	99.8%	1.268	88.0%	0.518	66.4%	0.518	**95.9%**	0.821	316*
	10	100%	1.123	**96.2%**	0.686	80.9%	0.686	98.1%	0.849	606*
18000	1	**98.7%**	11.473	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	160*
	2	**99.7%**	4.249	47.2%	0.495	35.0%	0.495	88.3%	1.372	285*
	5	100%	2.700	89.1%	1.022	69.6%	1.022	**99.2%**	1.619	659*
	10	100%	2.389	**96.2%**	1.354	84.0%	1.354	99.9%	1.676	1283*
21000	1	**100%**	1272.497	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	143*
	2	100%	468.656	36.6%	48.271	36.5%	48.271	**99.9%**	133.799	242*
	5	**100%**	296.354	72.5%	99.900	72.2%	99.900	**100%**	158.067	537*
	10	**100%**	261.811	86.6%	132.870	86.3%	132.870	**100%**	163.747	1029*
25000	1	**99.9%**	201.611	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	142*
	2	100%	74.595	36.6%	7.597	35.3%	7.597	**98.8%**	21.057	215*
	5	**100%**	47.205	72.5%	15.737	70.1%	15.737	**100%**	24.922	433*
	10	**100%**	41.583	86.8%	20.819	84.7%	20.819	**100%**	25.750	797*

To run this logistic regression, we use sklearn.linear_model.LogisticRegression (a machine learning package in Python) in the virtual machine c2-standard-8 in Google Cloud Platform, which takes about 30-40 minutes to run one bootstrap resample. Therefore, the common bootstrap methods which require $B = 50$ or 100 would be computationlly expensive.

In Section 4, we report and discuss the average interval widths over all β_i's for the four bootstrap methods. Here we display the results for three individual parameters, namely the first three β_i's, in Figure 7. Since we are only able to run one experimental repetition in this real world example, the confidence interval widths contain some noises and thus we cannot observe the monotonicity of the widths when B increases. But we still see the cheap bootstrap confidence interval widths are wider than others, with general trends that resemble the average interval widths in our synthetic examples. This suggests that the cheap bootstrap confidence intervals would have higher and closer-to-nominal coverages than the other methods.

C.5. Numerical Experiments with a Lower Nominal Level

In this section, we conduct a simulation study with the nominal level 70% to further support the validity of the cheap bootstrap. We choose the ellipsoidal estimation and sinusoidal estimation presented in Section 4 as our model. All the settings are the same except that we use a different sample size $n = 4 \times 10^4$ and a different dimension $p = 9000$. Table 8 presents the empirical coverage and average interval width over 1000 experimental repetitions. We can observe that the cheap bootstrap coverage probabilities are still close to the nominal level, and are the closest among all methods in all cases except two (sinusoidal $B = 5$ and $B = 10$) where percentile and standard error bootstraps each outperforms slightly. Regarding these exceptional cases, we note that the outperformance of the percentile bootstrap is likely a coincidence, because as a quantile-based method it cannot construct a symmetric 70% level confidence interval from only 5 resamples. We have used the minimum and maximum of the 5 resamples to construct this percentile bootstrap confidence interval,

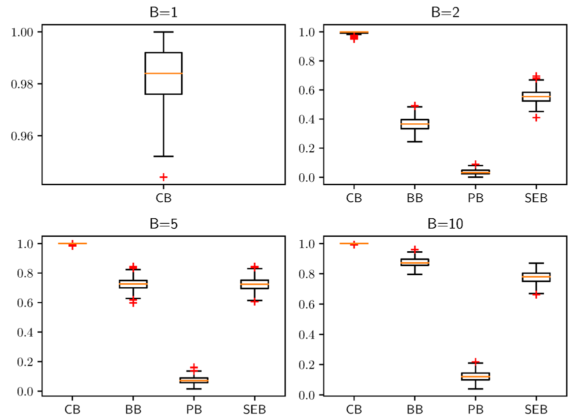
Table 5. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the ℓ_2-regularized logistic regression with $n = 10^5$ and different p. The closest coverage probability to the nominal 95% level among all methods in each setting is **bold**.

p	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
12000	1	**96.3%**	3.162	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	42*
	2	**97.3%**	1.182	41.0%	0.161	32.2%	0.161	74.7%	0.446	69*
	5	98.4%	0.755	80.3%	0.331	64.4%	0.331	**92.2%**	0.525	150*
	10	98.9%	0.669	93.5%	0.438	78.9%	0.438	**95.6%**	0.543	284*
15000	1	**96.8%**	3.943	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	83*
	2	**97.9%**	1.471	43.5%	0.196	32.9%	0.196	76.9%	0.543	141*
	5	99.1%	0.938	84.3%	0.404	65.6%	0.404	**93.9%**	0.640	315*
	10	99.6%	0.832	**95.9%**	0.534	80.1%	0.534	96.8%	0.662	606*
18000	1	**97.3%**	5.056	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	156*
	2	**98.5%**	1.884	45.7%	0.246	33.6%	0.246	79.3%	0.683	241*
	5	99.6%	1.202	87.6%	0.508	67.1%	0.508	**95.5%**	0.805	498*
	10	99.9%	1.065	**97.1%**	0.673	81.6%	0.673	97.8%	0.833	925*
21000	1	**97.6%**	6.400	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	191*
	2	**98.8%**	2.385	47.1%	0.310	34.5%	0.310	81.4%	0.859	319*
	5	99.8%	1.520	89.4%	0.639	68.6%	0.639	**96.5%**	1.012	703*
	10	100%	1.348	**97.5%**	0.845	83.1%	0.845	98.3%	1.047	1343*
25000	1	**97.4%**	7.259	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	208*
	2	**98.7%**	2.712	46.5%	0.362	35.2%	0.362	80.8%	1.004	350*
	5	99.7%	1.731	88.8%	0.746	69.9%	0.746	**96.0%**	1.183	777*
	10	99.9%	1.536	98.3%	0.988	84.5%	0.988	**97.9%**	1.224	1487*

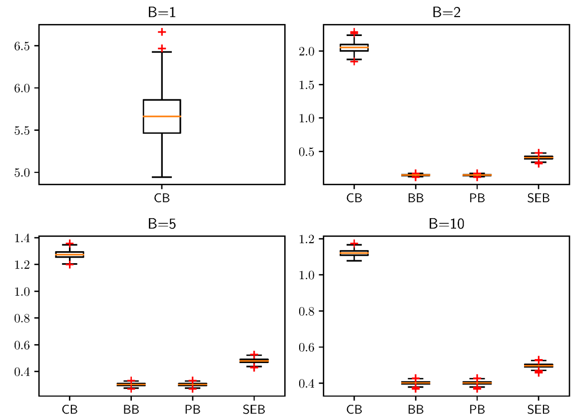
whose actual nominal level should be close to 100%. So it is likely by accident that percentile bootstrap coverage is closest to 70% and in fact this coverage is far away from its actual nominal level around 100%.

C.6. Coverage Error Behavior with Respect to B and n

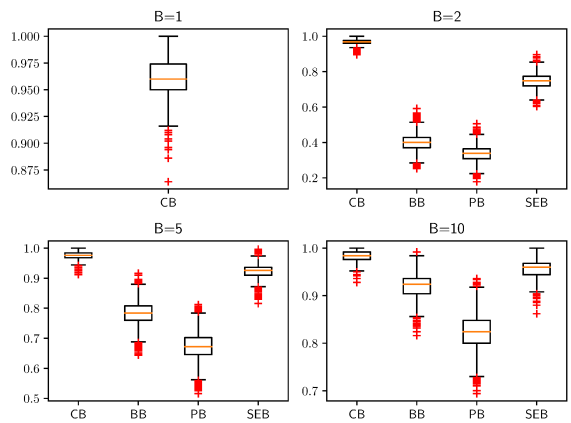
In this section, we numerically study the cheap bootstrap coverage error behavior with respect to B and n and illustrate how it aligns with our theoretical bounds. We choose the model as the sinusoidal estimation in Section 4. We fix the dimension $p = 9000$ and vary B and n. Figure 8 displays the colormaps of the absolute values of empirical coverage errors (nominal level 95%), where the x-axis represents B and y-axis represents n. We cut the results of the basic and percentile bootstraps for the first few B's because their errors are too large. From Figure 8 (a), it appears that the cheap bootstrap coverage error does not change much in this regime of n. This matches to some extent Theorem 3.5 and Corollary 3.6 that guarantee the coverage error would decrease as n increases with a slow rate $1/\sqrt{n}$. Further, when we fix n, the cheap bootstrap coverage error seems to lack clear trend and otherwise be quite stable as B changes. On the other hand, the basic and percentile bootstrap coverage errors show a clear decreasing trend as B increases. Their different behaviors are attributed to the different ideas behind them. The basic bootstrap and percentile bootstrap are quantile-based methods. As B increases, the bootstrap quantile estimate is closer and closer to the true quantile, which leads to the improvement on the coverage error. However, the cheap bootstrap method relies on a totally different idea, i.e., it relies on approximate independence of the resamples from the original estimator and thus a t-distribution-based (with degree of freedom B) confidence interval can be constructed. Different B just means a different pivotal t-distribution. There is no evident reason that the pivotal t-distribution with a larger degree of freedom B will lead to a smaller coverage error.



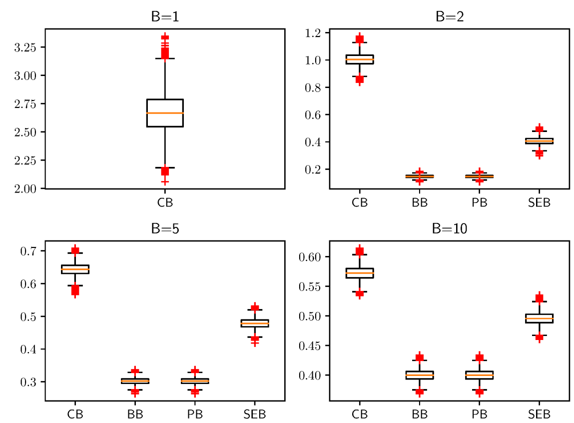
(a) Coverage probability $\beta_i \neq 0$



(b) Confidence interval width $\beta_i \neq 0$



(c) Coverage probability $\beta_i = 0$



(d) Confidence interval width $\beta_i = 0$

Figure 4. Box plots of empirical coverage probabilities and confidence interval widths of all β_i's for different numbers of resamples in a logistic regression.

D. Proofs

Proof of Theorem 3.1: We define Q^* as the distribution of $\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n)$ conditional on $X_1, \dots, X_n$. With the repeated bootstrap resampling, we have that $\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n), b = 1, \dots, B$ are independent conditional on $X_1, \dots, X_n$. Then we can write the coverage probability as

$$\begin{aligned}
&P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2} S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2} S_{n,B}]) \\
&= P\left(\left| \frac{\sqrt{n}(\hat{\psi}_n - \psi)}{\sqrt{\frac{1}{B}\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2}} \right| \leq t_{B,1-\alpha/2} \right) \\
&= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \leq t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dQ^*(z_B) \cdots dQ^*(z_1) \right]
\end{aligned}$$

where the expectation E is taken with respect to $X_1, \dots, X_n$. If we write $\mathcal{A}$ as the event that

$$\sup_{x \in \mathbb{R}} \left| Q^*((-\infty, x]) - \Phi\left(\frac{x}{\sigma}\right) \right| \equiv \sup_{x \in \mathbb{R}} \left| P^*(\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_2, \tag{18}$$

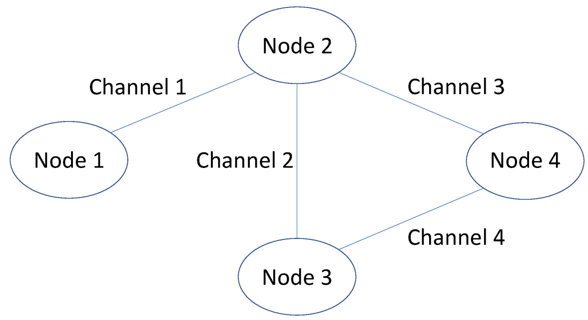
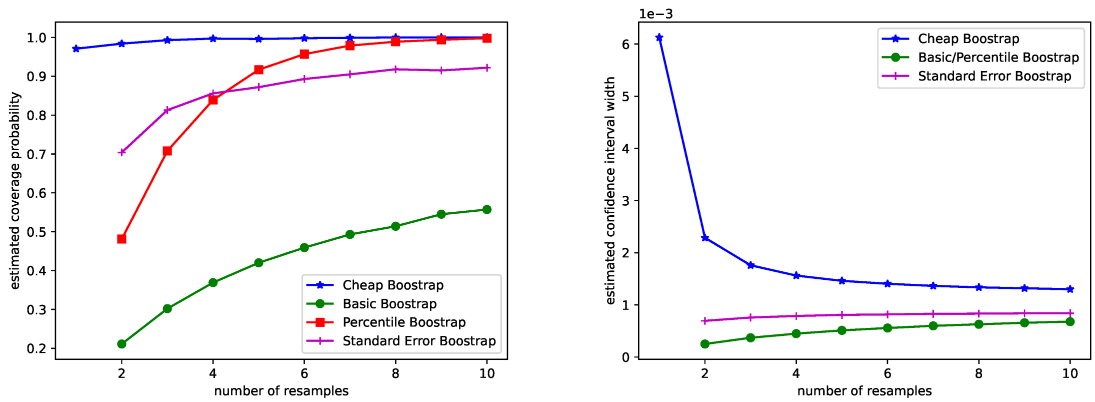


Figure 5. A computer network with four nodes and four channels.



(a) Coverage probability

(b) Confidence interval width

Figure 6. Empirical coverage probabilities and confidence interval widths for different numbers of resamples in a computer communication network.

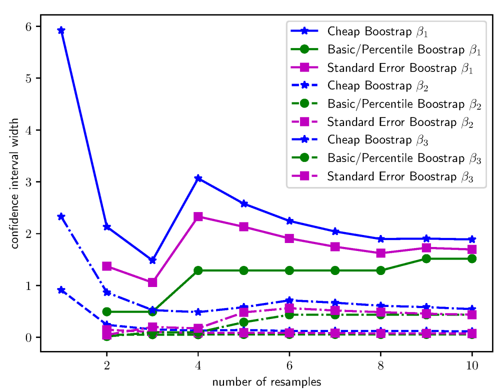


Figure 7. Confidence interval widths of the first three β's for different numbers of resamples in a real world logistic regression.

Table 6. Arrival rates $\lambda_{i,j}$ of messages to be transmitted from node i to node j.

	Node j			
Node i	1	2	3	4
1	N.A.	40	30	35
2	50	N.A.	45	15
3	60	15	N.A.	20
4	25	30	40	N.A.

Table 7. Parameters $(\alpha_{i,j}, \beta_{i,j})$ for the inter-arrival time distribution of messages to be transmitted from node i to node j.

	Node j			
Node i	1	2	3	4
1	N.A.	$(1.5, 60)$	$(0.7, 40)$	$(1.3, 50)$
2	$(2, 80)$	N.A.	$(1.5, 65)$	$(0.6, 20)$
3	$(3, 100)$	$(0.5, 25)$	N.A.	$(1.2, 30)$
4	$(0.8, 40)$	$(1.1, 50)$	$(0.9, 35)$	N.A.

then we know that $P(\mathcal{A}^c) \le \beta$. We consider the coverage probability intersected with $\mathcal{A}$, i.e.,

$$E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B) \cdots dQ^*(z_1); \mathcal{A}\right]. \tag{19}$$

Note that conditional on $X_1, \ldots, X_n$ and given $z_1, \ldots, z_{B-1}$, the integral region for z_B can be written as

$$\left\{z_B : |\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}\right\} = (-\infty, -q] \cup [q, \infty)$$

for some $q \ge 0$. Therefore, applying (18), we have

$$\left|\int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B) - \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\right| \le 2\mathcal{E}_2,$$

where P_0 is the distribution of $N(0, \sigma^2)$. Plugging it into (19), we have

$$\begin{aligned}
&E\left[\int \cdots \int_{|\sqrt{n}(g(\bar{X}_n) - g(\mu))| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B) \cdots dQ^*(z_1); \mathcal{A}\right] \\
&= E\left[\int \cdots \int \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B) dQ^*(z_{B-1}) \cdots dQ^*(z_1); \mathcal{A}\right] + R_B,
\end{aligned}$$

where the error R_B satisfies

$$|R_B| \le E\left[\int \cdots \int 2\mathcal{E}_2 dQ^*(z_{B-1}) \cdots dQ^*(z_1); \mathcal{A}\right] \le 2\mathcal{E}_2$$

By the same argument, we can further replace the remaining $Q^*(z_i)$'s by $P_0(z_i)$'s and obtain

$$E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B) \cdots dQ^*(z_1); \mathcal{A}\right]$$

Table 8. Coverage probabilities (Pro.), confidence interval widths (Wid.) and running time (unit: second) of the numerical examples. The closest coverage probability to the nominal 70% level among all methods in each setting is **bold**.

Example	B	Cheap Bootstrap		Basic Bootstrap		Percentile Bootstrap		Standard Error Bootstrap		Running Time
		Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	Pro.	Wid.	
Ellipsoidal estimation	1	**73.0%**	9.828×10^{-3}	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	2
	2	**70.0%**	7.568×10^{-3}	30.9%	2.197×10^{-3}	7.7%	2.197×10^{-3}	34.3%	3.220×10^{-3}	4
	5	**68.5%**	6.639×10^{-3}	64.7%	4.429×10^{-3}	16.5%	4.429×10^{-3}	42.3%	3.717×10^{-3}	10
	10	**66.9%**	6.323×10^{-3}	60.1%	3.756×10^{-3}	10.8%	3.756×10^{-3}	41.8%	3.804×10^{-3}	20
Sinusoidal estimation	1	**70.3%**	0.148	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	2
	2	**72.0%**	0.116	34.9%	0.052	33.3%	0.052	52.4%	0.077	4
	5	72.2%	0.104	67.6%	0.108	**68.6%**	0.108	65.6%	0.091	10
	10	72.3%	0.100	65.8%	0.093	63.7%	0.093	**68.8%**	0.094	19

$$= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\cdots dP_0(z_1); \mathcal{A}\right] + \sum_{b=1}^B R_b,$$

where each error R_b satisfies

$$|R_b| \le 2\mathcal{E}_2.$$

Therefore, the coverage probability satisfies

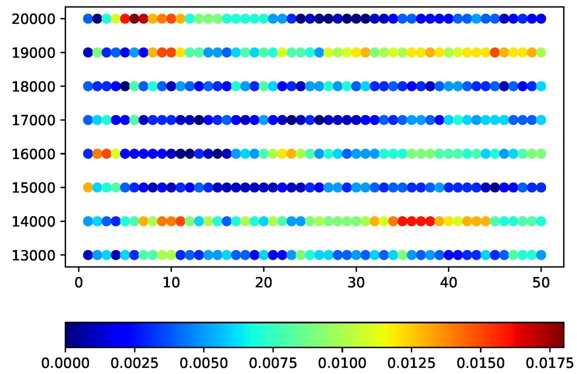
$$\begin{aligned}
&E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B)\cdots dQ^*(z_1)\right] \\
&= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B)\cdots dQ^*(z_1); \mathcal{A}\right] \\
&+ E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B)\cdots dQ^*(z_1); \mathcal{A}^c\right] \\
&= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\cdots dP_0(z_1); \mathcal{A}\right] + \sum_{b=1}^B R_b \\
&+ E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B)\cdots dQ^*(z_1); \mathcal{A}^c\right] \\
&= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\cdots dP_0(z_1)\right] + R_{\mathcal{A}^c} + \sum_{b=1}^B R_b, \qquad (20)
\end{aligned}$$

where the additional error $R_{\mathcal{A}^c}$ is given by

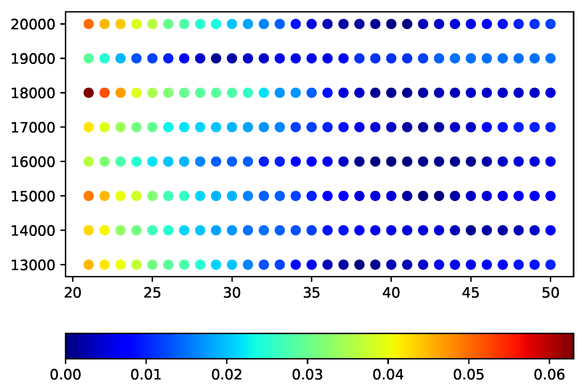
$$\begin{aligned}
R_{\mathcal{A}^c} &= E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dQ^*(z_B)\cdots dQ^*(z_1); \mathcal{A}^c\right] \\
&- E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\cdots dP_0(z_1); \mathcal{A}^c\right]
\end{aligned}$$

and it satisfies $|R_{\mathcal{A}^c}| \le P(\mathcal{A}^c) \le \beta$. Now we will handle the distribution of $\sqrt{n}(\hat{\psi}_n - \psi)$ which is denoted by Q_0. Note that by Fubini's theorem we have

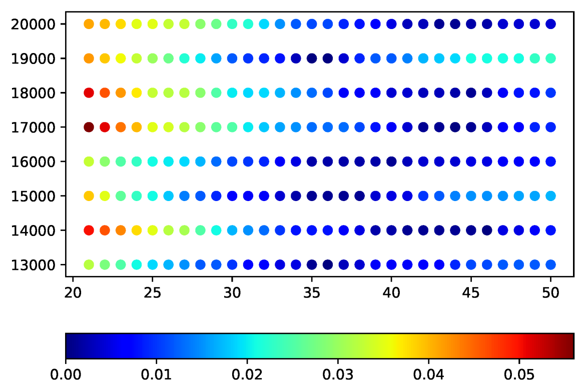
$$E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n-\psi)|\le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^B z_b^2}} dP_0(z_B)\cdots dP_0(z_1)\right]$$



(a) Cheap bootstrap



(b) Basic bootstrap



(c) Percentile bootstrap

Figure 8. Colormaps of the absolute values of empirical coverage errors (nominal level 95%) for the sinusoidal estimation.

$$= \int \cdots \int_{|z_0| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dQ_0(z_0) dP_0(z_B) \cdots dP_0(z_1).$$

Given $z_1, \ldots, z_B$, consider the innermost integral with respect to Q_0. By the finite-sample accuracy for Q_0, i.e.,

$$\sup_{x \in \mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right| \equiv \sup_{x \in \mathbb{R}} |Q_0((-\infty, x]) - P_0((-\infty, x])| \le \mathcal{E}_1,$$

we have

$$\left| \int_{|z_0| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dQ_0(z_0) - \int_{|z_0| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dP_0(z_0) \right| \le 2\mathcal{E}_1.$$

Therefore,

$$\begin{aligned} & E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dP_0(z_B) \cdots dP_0(z_1) \right] \\ & = \int \cdots \int_{|z_0| \le t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dP_0(z_0) dP_0(z_B) \cdots dP_0(z_1) + R_0 \\ & = 1 - \alpha + R_0, \end{aligned} \tag{21}$$

where the second equality follows from

$$\frac{Z_0}{\sqrt{\frac{1}{B}\sum_{b=1}^{B} Z_b^2}} \overset{d}{=} t_B$$

for i.i.d. $Z_i \sim N(0, \sigma^2), i = 0, \ldots, B$ and the error R_0 satisfies

$$|R_0| \leq 2\mathcal{E}_1.$$

Plugging (21) into (20), we have

$$\begin{aligned}
&E\left[\int \cdots \int_{|\sqrt{n}(\hat{\psi}_n - \psi)| \leq t_{B,1-\alpha/2}\sqrt{\frac{1}{B}\sum_{b=1}^{B} z_b^2}} dQ^*(z_B) \cdots dQ^*(z_1)\right] \\
&= 1 - \alpha + R_{\mathcal{A}^c} + \sum_{b=0}^{B} R_b := 1 - \alpha + R,
\end{aligned}$$

where the overall error satisfies

$$|R| \leq 2\mathcal{E}_1 + 2B\mathcal{E}_2 + \beta.$$

□

Proof of Theorem 3.2: Recall that for a cumulative distribution function F of a random variable, the q-th quantile is defined as $F^{-1}(q) = \inf\{x : F(x) \geq q\}$. We first prove a useful result: if the cumulative distribution functions of two random variables X and Y satisfy

$$\sup_{t \in \mathbb{R}} |F_X(t) - F_Y(t)| \leq \varepsilon, \tag{22}$$

then for any $\alpha \in [0, 1]$,

$$F_Y^{-1}(\alpha - \varepsilon) \leq F_X^{-1}(\alpha) \leq F_Y^{-1}(\alpha + \varepsilon). \tag{23}$$

To prove it, we note that if $\alpha - \varepsilon < 0$ then $-\infty = F_Y^{-1}(\alpha - \varepsilon) \leq F_X^{-1}(\alpha)$ trivially holds and if $\alpha + \varepsilon > 1$ then $F_X^{-1}(\alpha) \leq F_Y^{-1}(\alpha + \varepsilon) = \infty$ trivially holds. So we assume $0 \leq \alpha - \varepsilon \leq \alpha + \varepsilon \leq 1$. Now let's prove the first inequality $F_Y^{-1}(\alpha - \varepsilon) \leq F_X^{-1}(\alpha)$. By the definition of F_X^{-1} and right-continuity of F_X, we know that $F_X(F_X^{-1}(\alpha)) \geq \alpha$. Therefore, by (22), we know that $F_Y(F_X^{-1}(\alpha)) \geq \alpha - \varepsilon$, which implies $F_Y^{-1}(\alpha - \varepsilon) \leq F_X^{-1}(\alpha)$ by the definition of $F_Y^{-1}(\alpha - \varepsilon)$. This proves the first inequality in (23). Interchanging the role of X and Y, we have $F_X^{-1}(\alpha - \varepsilon) \leq F_Y^{-1}(\alpha)$. Replacing α by $\alpha + \varepsilon$, we obtain the second inequality $F_X^{-1}(\alpha) \leq F_Y^{-1}(\alpha + \varepsilon)$.

Now we consider the basic bootstrap. We write $\mathcal{A}$ as the event

$$\sup_{x \in \mathbb{R}} \left| P^*(\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_2.$$

By our assumption, we have $P(\mathcal{A}^c) \leq \beta$. Note that if $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^* - \hat{\psi}_n$ given $X_1, \ldots, X_n$, then $\sqrt{n}q_{\alpha/2}$ and $\sqrt{n}q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n)$ respectively given $X_1, \ldots, X_n$. By the inequality (23), when $\mathcal{A}$ happens, we have

$$\sigma z_{\alpha/2 - \mathcal{E}_2} \leq \sqrt{n} q_{\alpha/2} \leq \sigma z_{\alpha/2 + \mathcal{E}_2},$$

where z_q is the q-th quantile of the standard normal. This inequality implies

$$P(\sqrt{n}(\hat{\psi}_n - \psi) < \sigma z_{\alpha/2 - \mathcal{E}_2}; \mathcal{A}) \leq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sqrt{n} q_{\alpha/2}; \mathcal{A}) \tag{24}$$

and

$$P(\sqrt{n}(\hat{\psi}_n - \psi) < \sqrt{n} q_{\alpha/2}; \mathcal{A}) \leq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sigma z_{\alpha/2 + \mathcal{E}_2}; \mathcal{A}). \tag{25}$$

Next, we notice that

$$\begin{aligned}
&\sup_{x \in \mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_1 \\
\Leftrightarrow &\sup_{x \in \mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) < x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_1.
\end{aligned}$$

Thus, (24) implies that

$$\begin{aligned}
&P(\sqrt{n}(\hat{\psi}_n - \psi) < \sqrt{n} q_{\alpha/2}; \mathcal{A}) \\
&\geq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sigma z_{\alpha/2 - \mathcal{E}_2}; \mathcal{A}) \\
&\geq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sigma z_{\alpha/2 - \mathcal{E}_2}) - P(\mathcal{A}^c) \\
&\geq \Phi\left(\frac{\sigma z_{\alpha/2-\mathcal{E}_2}}{\sigma}\right) - \mathcal{E}_1 - \beta \\
&= \frac{\alpha}{2} - \mathcal{E}_1 - \mathcal{E}_2 - \beta.
\end{aligned}$$

Similarly, (25) implies that

$$P(\sqrt{n}(\hat{\psi}_n - \psi) < \sqrt{n} q_{\alpha/2}; \mathcal{A}) \leq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sigma z_{\alpha/2+\mathcal{E}_2}) \leq \frac{\alpha}{2} + \mathcal{E}_1 + \mathcal{E}_2.$$

Therefore, we have the following two-sided bound

$$\frac{\alpha}{2} - \mathcal{E}_1 - \mathcal{E}_2 - \beta \leq P(\sqrt{n}(\hat{\psi}_n - \psi) < \sqrt{n} q_{\alpha/2}; \mathcal{A}) \leq \frac{\alpha}{2} + \mathcal{E}_1 + \mathcal{E}_2.$$

For the $(1-\alpha/2)$-th quantile, we can also derive a similar bound

$$1 - \frac{\alpha}{2} - \mathcal{E}_1 - \mathcal{E}_2 - \beta \leq P(\sqrt{n}(\hat{\psi}_n - \psi) \leq \sqrt{n} q_{1-\alpha/2}; \mathcal{A}) \leq 1 - \frac{\alpha}{2} + \mathcal{E}_1 + \mathcal{E}_2.$$

So we have

$$|P(\sqrt{n} q_{\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq \sqrt{n} q_{1-\alpha/2}; \mathcal{A}) - (1-\alpha)| \leq 2\mathcal{E}_1 + 2\mathcal{E}_2 + \beta,$$

which gives rise to

$$\begin{aligned}
&|P(\sqrt{n} q_{\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq \sqrt{n} q_{1-\alpha/2}) - (1-\alpha)| \\
&\leq |P(\sqrt{n} q_{\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq \sqrt{n} q_{1-\alpha/2}; \mathcal{A}) - (1-\alpha)| \\
&+ P(\sqrt{n} q_{\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq \sqrt{n} q_{1-\alpha/2}; \mathcal{A}^c) \\
&\leq 2\mathcal{E}_1 + 2\mathcal{E}_2 + \beta + P(\mathcal{A}^c) \\
&\leq 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta,
\end{aligned}$$

or equivalently

$$|P(\hat{\psi}_n - q_{1-\alpha/2} \leq \psi \leq \hat{\psi}_n - q_{\alpha/2}) - (1-\alpha)| \leq 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta.$$

The result for the percentile bootstrap follows similarly but we need to use the symmetry of $N(0,\sigma^2)$. Note that

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_1 \\
\Leftrightarrow &\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) < x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \mathcal{E}_1
\end{aligned}$$

and the latter can be rewritten as

$$\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) < x) - \Phi\left(\frac{x}{\sigma}\right) \right| \\
=&\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\hat{\psi}_n - \psi) < -x) - \Phi\left(\frac{-x}{\sigma}\right) \right| \\
=&\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\psi - \hat{\psi}_n) > x) - \Phi\left(\frac{-x}{\sigma}\right) \right| \\
=&\sup_{x\in\mathbb{R}} \left| (1 - P(\sqrt{n}(\psi - \hat{\psi}_n) > x)) - \left(1 - \Phi\left(\frac{-x}{\sigma}\right)\right) \right|
\end{aligned}$$

$$= \sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(\psi - \hat{\psi}_n) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right| \le \mathcal{E}_1,$$

where the last equality uses the symmetry of $N(0,\sigma^2)$, i.e., $1-\Phi(-x/\sigma) = \Phi(x/\sigma)$. Moreover, if $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\hat{\psi}_n^*$ given $X_1,\dots,X_n$, then $\sqrt{n}(q_{\alpha/2} - \hat{\psi}_n)$ and $\sqrt{n}(q_{1-\alpha/2} - \hat{\psi}_n)$ are the $\alpha/2$-th and $(1-\alpha/2)$-th quantiles of $\sqrt{n}(\hat{\psi}_n^* - \hat{\psi}_n)$ given $X_1,\dots,X_n$. Therefore, the proof for the basic bootstrap also applies if we replace $\sqrt{n}q_{\alpha/2}$, $\sqrt{n}q_{1-\alpha/2}$ and $\sqrt{n}(\hat{\psi}_n - \psi)$ in that proof by $\sqrt{n}(q_{\alpha/2} - \hat{\psi}_n)$, $\sqrt{n}(q_{1-\alpha/2} - \hat{\psi}_n)$ and $\sqrt{n}(\psi - \hat{\psi}_n)$ respectively. In particular, the final result now reads as follows

$$|P(\sqrt{n}(q_{\alpha/2} - \hat{\psi}_n) \le \sqrt{n}(\psi - \hat{\psi}_n) \le \sqrt{n}(q_{1-\alpha/2} - \hat{\psi}_n)) - (1-\alpha)| \le 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta$$

or equivalently

$$|P(q_{\alpha/2} \le \psi \le q_{1-\alpha/2}) - (1-\alpha)| \le 2\mathcal{E}_1 + 2\mathcal{E}_2 + 2\beta.$$

This completes our proof. □

To prove Theorem 3.5, we first need to prove two lemmas regarding (2) and (3) respectively.

The following lemma is from Theorem 2.11 in Pinelis & Molzon (2016) which establishes the Berry-Esseen theorem in the multivariate delta method in the form of (2).

Lemma D.1. *Suppose that $X_1,\dots,X_n$ are i.i.d. random vectors in $\mathbb{R}^p$ satisfying $E[X] = \mu$, $Var(X) = \Sigma$, $m_{31} := E[|\nabla g(\mu)^\top (X-\mu)|^3] < \infty$ and $m_{32} := E[||X-\mu||^3] < \infty$. Suppose $g(x)$ satisfies Assumption 3.3 and $\sigma^2 := \nabla g(\mu)^\top \Sigma \nabla g(\mu) > 0$. Then there is a universal constant $C > 0$ s.t.*

$$\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(g(\bar{X}_n) - g(\mu)) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right|$$
$$\le C\left(\frac{m_{31}}{\sqrt{n}\sigma^3} + \frac{C_{H_g} m_{31}^{1/3} tr(\Sigma)}{\sqrt{n}\sigma^2} + \frac{C_{H_g} m_{32}^{2/3}}{n^{5/6}\sigma} + \frac{C_{H_g} m_{31}^{1/3} m_{32}^{2/3}}{n\sigma^2} \right).$$

Proof of Lemma D.1: Define $f(x) = g(x+\mu) - g(\mu)$ and its linearization $L(x) = \nabla g(\mu)^\top x$. Then by the second order Taylor expansion of $f(x)$ and boundedness property of H_g in Assumption 3.3, we can see (2.1) in Pinelis & Molzon (2016) holds for $M_\epsilon = C_{H_g}$ and any $\epsilon > 0$. By Theorem 2.11 in Pinelis & Molzon (2016) with $V = X - \mu$, $c_* = 1/2$ and $\epsilon \to \infty$, we have

$$\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(g(\bar{X}_n) - g(\mu)) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right|$$
$$\le \frac{\mathfrak{K}_0 + \mathfrak{K}_1 m_{31}/\sigma^3}{\sqrt{n}} + \frac{\mathfrak{K}_{20} + \mathfrak{K}_{21} m_{31}^{1/3}/\sigma}{\sqrt{n}} tr(\Sigma) + \frac{\mathfrak{K}_{30} + \mathfrak{K}_{31} m_{31}^{1/3}/\sigma}{\sqrt{n}} m_{32}^{2/3}, \tag{26}$$

where the additional term $\mathfrak{K}_\epsilon$ in Theorem 2.11 vanishes as $\epsilon \to \infty$. By the definition of these $\mathfrak{K}$'s with $c_* = 1/2$ in (2.30) in Pinelis & Molzon (2016), we can see there is a universal constant $C > 0$ s.t.

$$\mathfrak{K}_0 \le C, \mathfrak{K}_1 \le C, \mathfrak{K}_{20} \le C\frac{C_{H_g}}{\sigma}, \mathfrak{K}_{21} \le C\frac{C_{H_g}}{\sigma}, \mathfrak{K}_{30} \le C\frac{C_{H_g}}{\sigma n^{1/3}}, \mathfrak{K}_{31} \le C\frac{C_{H_g}}{\sigma n^{1/2}}.$$

Moreover, by Holder's inequality, $m_{31} = E[|\nabla g(\mu)^\top (X-\mu)|^3] \ge E[|\nabla g(\mu)^\top (X-\mu)|^2]^{3/2} = \sigma^3$, which implies that $\mathfrak{K}_0, \mathfrak{K}_{20}$ can be absorbed into $\mathfrak{K}_1 m_{31}/\sigma^3, \mathfrak{K}_{21} m_{31}^{1/3}/\sigma$ respectively by choosing a larger C. Therefore, (26) can be written as

$$\sup_{x\in\mathbb{R}} \left| P(\sqrt{n}(g(\bar{X}_n) - g(\mu)) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right|$$
$$\le C\left(\frac{m_{31}}{\sqrt{n}\sigma^3} + \frac{C_{H_g} m_{31}^{1/3} tr(\Sigma)}{\sqrt{n}\sigma^2} + \frac{C_{H_g} m_{32}^{2/3}}{n^{5/6}\sigma} + \frac{C_{H_g} m_{31}^{1/3} m_{32}^{2/3}}{n\sigma^2} \right).$$

This concludes our proof. □

Next we prove the finite-sample accuracy (3) for the bootstrap estimator by extracting the dependence on problem parameters in Theorem 4.2 in Zhilova (2020) and combining it with Lemma D.1.

Lemma D.2. *Suppose the conditions in Theorem 3.5 hold. Then with probability at least $1 - 6/n$ we have*

$$\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(g(\bar{X}_n^*) - g(\bar{X}_n)) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right|$$
$$\le C \left(\frac{m_{31}}{\sqrt{n}\sigma^3} + \frac{C_{H_g} m_{31}^{1/3} tr(\Sigma)}{\sqrt{n}\sigma^2} + \frac{C_{H_g} m_{32}^{2/3}}{n^{5/6}\sigma} + \frac{C_{H_g} m_{31}^{1/3} m_{32}^{2/3}}{n\sigma^2} \right.$$
$$+ \frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}} \left(1 + \frac{\log n}{p}\right)\sqrt{\frac{p}{n}} + \frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}} \frac{1}{\sqrt{n}} + \frac{\tau^3}{\lambda_\Sigma^{3/2}} \left(1 + \frac{\log n}{p}\right)^{3/2} \frac{1}{\sqrt{n}}$$
$$\left. + \frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n} \left(1 + \frac{\log n}{p}\right)^{1/2} + \frac{\tau^2\sqrt{p}}{\lambda_\Sigma n} \left(1 + \frac{\log n}{p}\right)^{1/2} + \frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n} \left(1 + \frac{\log n}{p}\right) \right)$$
$$+ C_1 \left(\frac{\tau^4 (\log n)^{3/2}}{\lambda_\Sigma^2 \sqrt{n}} + \frac{\tau^2 (\log n)^{3/2}}{\lambda_\Sigma \sqrt{n}} + \frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}} \left(1 + \frac{\log n}{p}\right)^{1/2} (\log n + \log p)\sqrt{\log n} \right),$$

where m_{31}, m_{32} and σ^2 are defined in Lemma D.1, C is a universal constant and C_1 only depends on C_X.

Proof of Lemma D.2: We use Theorem 4.2 in Zhilova (2020) to prove this lemma. First, we verify the conditions of Theorem 4.2 with $K = 3$. For $K = 3$, by Remark 2.1, we can take $U_i \equiv 0, Y_i - \mu = Z_i - \mu \sim N(0, \Sigma_z)$ independent of $X_1, \dots, X_n$ with $\Sigma_z = \Sigma$ which satisfies (2.1) and (2.2) for approximating $X_i - \mu$ (since X_i is not centered in Theorem 4.2, Y_i should also be non-centered). In this case, we can see $C_z := ||\Sigma_z^{-1/2}|| = ||\Sigma^{-1/2}|| = 1/\sqrt{\lambda_\Sigma}$. Similarly, $C_X := ||\Sigma^{-1/2}|| = 1/\sqrt{\lambda_\Sigma}$. Other conditions about f in Theorem 4.2 have already been assumed in the statement of this lemma. Thus, by Theorem 4.2, it holds with probability at least $1 - 6e^{-x}$ for $x > 0$:

$$\sup_{t\in\mathbb{R}} |P(\sqrt{n}(g(\bar{X}_n) - g(\mu)) \le t) - P^*(\sqrt{n}(g(\bar{X}_n^*) - g(\bar{X}_n)) \le t)|$$
$$\le 2C_{H_g}\sqrt{\frac{1}{\lambda_\Sigma}}\tau^2 \left(1 + 2\sqrt{\frac{x}{p}} + \frac{2x}{p}\right) \frac{1}{C_{\nabla g}}\sqrt{\frac{p}{n}} + C_{\mathcal{B},\text{i.i.d.}} \frac{1}{\lambda_\Sigma^{3/2}} C_{M,3} \frac{1}{\sqrt{n}}$$
$$+ 2C_{\mathcal{B},\text{i.i.d.}} \left(1 + 2\sqrt{\frac{x}{p}} + \frac{2x}{p}\right)^{3/2} \frac{1}{\lambda_\Sigma^{3/2}} \tau^3 \frac{1}{\sqrt{n}} + R_{1,n,3} \tag{27}$$

where $C_{\mathcal{B},\text{i.i.d.}} > 0$ is a constant only depending on K and thus is a universal constant for $K = 3$, $C_{M,3}$ is defined as

$$C_{M,3} = ||E[(X-\mu)^3]|| + ||E[(Y_1-\mu)^3]||$$

and $R_{1,n,3}$ is defined in (B.14) as

$$R_{1,n,3} = \frac{\tau^4 C_{x,2}}{\lambda_\Sigma^2\sqrt{2n}} + \frac{4\tau^2 C_{x,2}}{\lambda_\Sigma\sqrt{2n}} + \frac{\tilde{C}_{\phi,2}\tau^3 C_{x,3}}{2\lambda_\Sigma^{3/2}\sqrt{n}}. \tag{28}$$

Since $Y_1 - \mu \sim N(0, \Sigma_z)$, the tensor power $(Y_1 - \mu)^3$ has expectation zero and thus $C_{M,3}$ can be simplified as $C_{M,3} = ||E[(X-\mu)^3]||$. $\tilde{C}_{\phi,2}$ in (28) is defined in (A.13) and according to Remark A.1, it depends on the choice of $\phi(t)$ which is a $K = 3$ times continuously differentiable smooth approximation of the indicator function $1\{t \le 0\}$. Once we fix such $\phi(t)$, $\tilde{C}_{\phi,2}$ is a universal constant. $C_{x,2}$ and $C_{x,3}$ in (28) are defined in (B.27) of Theorem B.1 as

$$C_{x,2} = C_1((x + \log n) \vee \sqrt{x + \log n})\sqrt{2x} + 2\sqrt{\frac{p}{n}} \left(1 + 2\sqrt{\frac{x + \log n}{p}} + \frac{2(x + \log n)}{p}\right)^{1/2},$$

$$C_{x,3} = C_1 \left(1 + 2\sqrt{\frac{x + \log n + \log p}{p}} + \frac{2(x + \log n + \log p)}{p}\right)^{1/2}$$

$$
\begin{aligned}
&\times ((x+\log n+\log p) \vee \sqrt{x+\log n+\log p})\sqrt{2x} \\
&+ 3\sqrt{\frac{p}{n}}\left(1+2\sqrt{\frac{x+\log n}{p}}+\frac{2(x+\log n)}{p}\right),
\end{aligned}
$$

where C_1 is a constant only depending on the density bound C_X based on the proof of Theorem B.1. Besides, since we assume $n \geq 3$, we know that $x+\log n \geq \sqrt{x+\log n}$ and $x+\log n+\log p \geq \sqrt{x+\log n+\log p}$ holds for any $x>0$. Therefore, $C_{x,2}$ and $C_{x,3}$ can be simplified as

$$
C_{x,2} = C_1(x+\log n)\sqrt{2x} + 2\sqrt{\frac{p}{n}}\left(1+2\sqrt{\frac{x+\log n}{p}}+\frac{2(x+\log n)}{p}\right)^{1/2},
$$

$$
\begin{aligned}
C_{x,3} = &\, C_1\left(1+2\sqrt{\frac{x+\log n+\log p}{p}}+\frac{2(x+\log n+\log p)}{p}\right)^{1/2}(x+\log n+\log p)\sqrt{2x} \\
&+ 3\sqrt{\frac{p}{n}}\left(1+2\sqrt{\frac{x+\log n}{p}}+\frac{2(x+\log n)}{p}\right).
\end{aligned}
$$

Moreover, for any $y>0$, we always have $1+y \leq 1+2\sqrt{y}+2y \leq 4(1+y)$. Therefore, the remainder term (28) can be bounded in a more compact way up to some constants as follows:

$$
\begin{aligned}
R_{1,n,3} \leq &\, C_1\left(\frac{\tau^4}{\lambda_\Sigma^2\sqrt{n}}(x+\log n)\sqrt{x} + \frac{\tau^2}{\lambda_\Sigma\sqrt{n}}(x+\log n)\sqrt{x}\right. \\
&\left. + \frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{x+\log n+\log p}{p}\right)^{1/2}(x+\log n+\log p)\sqrt{x}\right) \\
&+ C\left(\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{x+\log n}{p}\right)^{1/2} + \frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{x+\log n}{p}\right)^{1/2} + \frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{x+\log n}{p}\right)\right),
\end{aligned}
$$

where C is a universal constant and C_1 only depends on C_X. Plugging it back to (27), we can similarly write (27) in a more compact way:

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}} |P(\sqrt{n}(g(\bar{X}_n)-g(\mu)) \leq t) - P^*(\sqrt{n}(g(\bar{X}_n^*)-g(\bar{X}_n)) \leq t)| \\
&\leq C\left(\frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1+\frac{x}{p}\right)\sqrt{\frac{p}{n}} + \frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}} + \frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1+\frac{x}{p}\right)^{3/2}\frac{1}{\sqrt{n}}\right. \\
&\left. + \frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{x+\log n}{p}\right)^{1/2} + \frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{x+\log n}{p}\right)^{1/2} + \frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{x+\log n}{p}\right)\right) \\
&+ C_1\left(\frac{\tau^4}{\lambda_\Sigma^2\sqrt{n}}(x+\log n)\sqrt{x} + \frac{\tau^2}{\lambda_\Sigma\sqrt{n}}(x+\log n)\sqrt{x}\right. \\
&\left. + \frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{x+\log n+\log p}{p}\right)^{1/2}(x+\log n+\log p)\sqrt{x}\right),
\end{aligned}
$$

where C is a universal constant and C_1 only depends on C_X. Now we choose $x=\log n$. Then with probability at least $1-6/n$, we have

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}} |P(\sqrt{n}(g(\bar{X}_n)-g(\mu)) \leq t) - P^*(\sqrt{n}(g(\bar{X}_n^*)-g(\bar{X}_n)) \leq t)| \\
&\leq C\left(\frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1+\frac{\log n}{p}\right)\sqrt{\frac{p}{n}} + \frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}} + \frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1+\frac{\log n}{p}\right)^{3/2}\frac{1}{\sqrt{n}}\right.
\end{aligned}
$$

$$+\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{\log n}{p}\right)\Bigg)$$

$$+C_1\left(\frac{\tau^4(\log n)^{3/2}}{\lambda_\Sigma^2\sqrt{n}}+\frac{\tau^2(\log n)^{3/2}}{\lambda_\Sigma\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right),$$

where C is a universal constant, C_1 only depends on C_X and we have absorbed $\log p/p$ into the constant term 1 due to $\log p/p \leq 1$.

Now we combine the above bound with Lemma D.1 in this paper. Since X is sub-Gaussian, the moment conditions in Lemma D.1 hold. Moreover, since Σ is positive definite and $||\nabla g(\mu)|| > 0$, $\sigma^2 = \nabla g(\mu)^\top \Sigma \nabla g(\mu) > 0$ is also satisfied. Therefore, by Lemma D.1 and triangular inequality, we obtain the desired bound with probability at least $1 - 6/n$

$$\sup_{x\in\mathbb{R}}\left|P^*(\sqrt{n}(g(\bar{X}_n^*)-g(\bar{X}_n))\leq x)-\Phi\left(\frac{x}{\sigma}\right)\right|$$

$$\leq C\left(\frac{m_{31}}{\sqrt{n}\sigma^3}+\frac{C_{H_g}m_{31}^{1/3}tr(\Sigma)}{\sqrt{n}\sigma^2}+\frac{C_{H_g}m_{32}^{2/3}}{n^{5/6}\sigma}+\frac{C_{H_g}m_{31}^{1/3}m_{32}^{2/3}}{n\sigma^2}\right.$$

$$+\frac{C_{H_g}\tau^2}{C_{\nabla g}\sqrt{\lambda_\Sigma}}\left(1+\frac{\log n}{p}\right)\sqrt{\frac{p}{n}}+\frac{||E[(X-\mu)^3]||}{\lambda_\Sigma^{3/2}}\frac{1}{\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}}\left(1+\frac{\log n}{p}\right)^{3/2}\frac{1}{\sqrt{n}}$$

$$+\frac{\tau^4\sqrt{p}}{\lambda_\Sigma^2 n}\left(1+\frac{\log n}{p}\right)^{1/2}+\frac{\tau^2\sqrt{p}}{\lambda_\Sigma n}\left(1+\frac{\log n}{p}\right)^{1/2}+\left.\frac{\tau^3\sqrt{p}}{\lambda_\Sigma^{3/2} n}\left(1+\frac{\log n}{p}\right)\right)$$

$$+C_1\left(\frac{\tau^4(\log n)^{3/2}}{\lambda_\Sigma^2\sqrt{n}}+\frac{\tau^2(\log n)^{3/2}}{\lambda_\Sigma\sqrt{n}}+\frac{\tau^3}{\lambda_\Sigma^{3/2}\sqrt{n}}\left(1+\frac{\log n}{p}\right)^{1/2}(\log n+\log p)\sqrt{\log n}\right),$$

where C is a universal constant and C_1 only depends on C_X. □

Proof of Theorem 3.5: Plugging the bounds in Lemma D.1 and Lemma D.2 into Theorem 3.1, we obtain the desired finite sample bound for the cheap bootstrap coverage accuracy. Besides, the error $\mathcal{E}_1$ can be absorbed in $\mathcal{E}_2$. □

We make a remark about the proof of Theorem 3.5. In Zhilova (2020), it appears that a generalized version of the Hanson-Wright inequality (equation (B.32)) that allows dependent components is derived as a middle step of the proof of Theorem B.1. If the classical Hanson-Wright inequality (Theorem 1.1 in Rudelson & Vershynin (2013)) is used in that proof instead, then our Theorem 3.5 would be changed accordingly to have an additional assumption that X has independent components but no longer needs to be continuously distributed with a bounded density. In this case, C_1 can be taken as a universal constant.

Proof of Corollary 3.6: It suffices to show that if $\tau = O(1)$ and $||\nabla g(\mu)||^2 = O(p)$, then $tr(\Sigma) = O(p)$, $m_{31} = O(p^{3/2})$ and $m_{32} = O(p^{3/2})$. In fact, if these orders hold, with other orders assumed in Corollary 3.6, we can easily get the desired order after absorbing the small order terms into large order terms.

Now let us prove $tr(\Sigma) = O(p)$, $m_{31} = O(p^{3/2})$ and $m_{32} = O(p^{3/2})$ provided $\tau = O(1)$ and $||\nabla g(\mu)||^2 = O(p)$. Recall that X is assumed to be sub-Gaussian, i.e.,

$$E[\exp(a^\top(X-\mu))] \leq \exp(||a||^2\tau^2/2), \forall a \in \mathbb{R}^p. \tag{29}$$

for some $\tau^2 > 0$. Therefore, $X_i - \mu_i, i = 1, \dots, p$ are sub-Gaussian random variables with sub-Gaussian norm τ up to a universal constant (see Vershynin (2018) Section 2.5). For simplicity, we write $a \lesssim b$ if $a \leq Cb$ for a universal constant $C > 0$. By Proposition 2.5.2 (ii) in Vershynin (2018), $E[|X_i - \mu_i|^2] = \Sigma_{ii} \lesssim \tau^2$ and $E[|X_i - \mu_i|^4] \lesssim \tau^4$. Therefore, we can see $tr(\Sigma) \lesssim \tau^2 p = O(p)$. By Hölder's inequality,

$$m_{32} = E[||X-\mu||^3] \leq E[||X-\mu||^4]^{3/4} = \left(\sum_{i,j=1}^{p} E\left[(X_i-\mu_i)^2(X_j-\mu_j)^2\right]\right)^{3/4}$$

$$\leq \left(\sum_{i,j=1}^{p} \sqrt{E\left[(X_i - \mu_i)^4\right] E[(X_j - \mu_j)^4]} \right)^{3/4} \lesssim \left(\sum_{i,j=1}^{p} \tau^4 \right)^{3/4} = \tau^3 p^{3/2} = O(p^{3/2}).$$

Moreover, (29) also implies that $\nabla g(\mu)^\top (X - \mu)$ is sub-Gaussian with sub-Gaussian norm $||\nabla g(\mu)||\tau$ up to a universal constant. By Proposition 2.5.2 (ii) in Vershynin (2018), we have $m_{31} = E[|\nabla g(\mu)^\top (X-\mu)|^3] \lesssim ||\nabla g(\mu)||^3 \tau^3 = O(p^{3/2})$. This concludes our proof. □

Proof of Theorem 3.7: We will apply Theorem 3.1 to prove this theorem. The finite-sample bound (2) can be obtained by the Berry-Esseen theorem:

$$\sup_{x \in \mathbb{R}} \left| P(\sqrt{n}(g_1^\top \bar{X}_n - g_1^\top \mu) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \frac{CE[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}}. \tag{30}$$

Next, we need to find a bound for

$$\sup_{x \in \mathbb{R}} \left| P^*(\sqrt{n}(g_1^\top \bar{X}_n^* - g_1^\top \bar{X}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right|.$$

We consider the i.i.d. centered random variables $g_1^\top (X_i - \mu), i = 1, \ldots, n$ which have non-degenerate variance $\sigma^2 = g_1^\top \Sigma g_1 > 0$. We apply Theorem 2.5 in Lopes (2022) to $g_1^\top (X_i - \mu)$'s by choosing $Y = N(0, g_1^\top \Sigma g_1)$ such that $\varrho = 1$, $\Delta = 0$, $\omega_1 = ||g_1^\top (X-\mu)/\sigma||_{\psi_1} = ||g_1^\top (X-\mu)||_{\psi_1}/\sigma$ and obtain that with probability at least $1 - C/n$,

$$\sup_{x \in \mathbb{R}} |P^*(\sqrt{n}(g_1^\top \bar{X}_n^* - g_1^\top \bar{X}_n) \leq x) - P(\sqrt{n} g_1^\top (\bar{X}_n - \mu) \leq x)| \leq \frac{C||g_1^\top (X-\mu)||_{\psi_1}^4 \log^{11}(n)}{\sigma^4 \sqrt{n}},$$

where $C > 0$ is a universal constant. By the triangle inequality and Berry-Esseen bound (30), the following holds with probability at least $1 - C/n$

$$\sup_{x \in \mathbb{R}} \left| P^*(\sqrt{n}(g_1^\top \bar{X}_n^* - g_1^\top \bar{X}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right| \leq \frac{CE[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}} + \frac{C||g_1^\top (X-\mu)||_{\psi_1}^4 \log^{11}(n)}{\sigma^4 \sqrt{n}}, \tag{31}$$

where $C > 0$ is a universal constant.

By Theorem 3.1 and the bounds (30) and (31), we finally get

$$\begin{aligned} &\left| P(g(\mu) \in [g(\bar{X}_n) - t_{B,1-\alpha/2} S_{n,B}, g(\bar{X}_n) + t_{B,1-\alpha/2} S_{n,B}]) - (1-\alpha) \right| \\ &\leq \frac{C}{n} + BC \left(\frac{E[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}} + \frac{||g_1^\top (X-\mu)||_{\psi_1}^4 \log^{11}(n)}{\sigma^4 \sqrt{n}} \right) + \frac{CE[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}}, \end{aligned}$$

where C is a universal constant. Furthermore, the last term can be absorbed into the second term by using a larger constant $2C$, which completes our proof. □

Proof of Theorem 3.8: We use Theorem 3.1 to prove this theorem. As in the proof of Theorem 3.7, (2) is given by the Berry-Esseen theorem in (30) and we only need to bound

$$\sup_{x \in \mathbb{R}} \left| P^*(\sqrt{n}(g_1^\top \bar{X}_n^* - g_1^\top \bar{X}_n) \leq x) - \Phi\left(\frac{x}{\sigma}\right) \right|.$$

In this regard, we will use the results in Chernozhukov et al. (2020). Note that their results only apply for at least three-dimensional random vectors so we consider the following setting. Suppose we have $3n$ i.i.d. random variables $g_1^\top (X_{ij} - \mu), i = 1, \ldots, n, j = 1, 2, 3$ where each X_{ij} has the same distribution as X_1. Then we can construct n three-dimensional i.i.d. random vectors as $\tilde{X}_i := (g_1^\top (X_{i1} - \mu), g_1^\top (X_{i2} - \mu), g_1^\top (X_{i3} - \mu))^\top, i = 1, \ldots, n$ whose components have common variance $\sigma^2 = g_1^\top \Sigma g_1$. Then we can see that conditions (E.3) and (M) are satisfied for

$$B_n = \max\left\{ 3E[|g_1^\top (X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top (X-\mu)/\sigma|^4]} \right\}.$$

Therefore, by Corollary 3.2 in Chernozhukov et al. (2020) ($\sigma_{*,W} = 1$ since $\tilde{X}_i$ has independent components), we have with probability at least $1 - 1/\sqrt{n}$ that

$$
\begin{aligned}
&\sup_{A\in\mathcal{R}} |P^*(\sqrt{n}(\bar{\tilde{X}}_n^* - \bar{\tilde{X}}_n) \in A) - P(N(0, \sigma^2 I_{3\times 3}) \in A)| \\
&\le C_1 B_n \left(\frac{\log 3 \log n \sqrt{\log(3\sqrt{n})}}{\sqrt{n}} + \frac{\log 3 \sqrt{\log(3n)}}{n^{1/2-3/(2q)}} \right) \\
&\le C_1 \max\left\{ 3E[|g_1^\top (X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top (X-\mu)/\sigma|^4]} \right\} \left(\frac{(\log n)^{3/2}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}} \right),
\end{aligned}
$$

where $C_1 > 0$ denotes a constant depending only on q which are different for its two appearances and $\mathcal{R}$ contains all the hyperrectangles in $\mathbb{R}^3$. In particular, if we only focus on the first component of $\tilde{X}_i$, that is, we choose $A = (-\infty, x] \times \mathbb{R} \times \mathbb{R}$, we have with probability at least $1 - 1/\sqrt{n}$ that

$$
\begin{aligned}
&\sup_{x\in\mathbb{R}} \left| P^*(\sqrt{n}(g_1^\top \bar{X}_n^* - g_1^\top \bar{X}_n) \le x) - \Phi\left(\frac{x}{\sigma}\right) \right| \\
&\le C_1 \max\left\{ 3E[|g_1^\top (X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top (X-\mu)/\sigma|^4]} \right\} \left(\frac{(\log n)^{3/2}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}} \right).
\end{aligned} \tag{32}
$$

By Theorem 3.1 and the bounds (30) and (32), we then obtain

$$
\begin{aligned}
&\left| P(g(\mu) \in [g(\bar{X}_n) - t_{B,1-\alpha/2} S_{n,B}, g(\bar{X}_n) + t_{B,1-\alpha/2} S_{n,B}]) - (1-\alpha) \right| \\
&\le \frac{2}{\sqrt{n}} + BC_1 \max\left\{ E[|g_1^\top (X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top (X-\mu)/\sigma|^4]} \right\} \\
&\times \left(\frac{(\log n)^{3/2}}{\sqrt{n}} + \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}} \right) + \frac{CE[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}},
\end{aligned}
$$

where C is a universal constant and C_1 is a constant depending only on q. Finally notice that

$$
\frac{(\log n)^{3/2}}{\sqrt{n}} = o\left(\frac{\sqrt{\log n}}{n^{1/2-3/(2q)}} \right),
$$

and $E[|g_1^\top (X-\mu)|^3]/\sigma^3 \ge 1$. We can absorb $(\log n)^{3/2}/\sqrt{n}$ into $\sqrt{\log n}/n^{1/2-3/(2q)}$ and absorb $2/\sqrt{n}$ into $CE[|g_1^\top (X-\mu)|^3]/\sigma^3\sqrt{n}$ (with larger constants C_1 and C), which leads to

$$
\begin{aligned}
&\left| P(g(\mu) \in [g(\bar{X}_n) - t_{B,1-\alpha/2} S_{n,B}, g(\bar{X}_n) + t_{B,1-\alpha/2} S_{n,B}]) - (1-\alpha) \right| \\
&\le BC_1 \max\left\{ E[|g_1^\top (X-\mu)/\sigma|^q]^{1/q}, \sqrt{E[|g_1^\top (X-\mu)/\sigma|^4]} \right\} \frac{\sqrt{\log n}}{n^{1/2-3/(2q)}} + \frac{CE[|g_1^\top (X-\mu)|^3]}{\sigma^3 \sqrt{n}}.
\end{aligned}
$$

□

Proof of Theorem B.1: In view of Theorem 3.1, it suffices to show

$$
\begin{aligned}
&|P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2} S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2} S_{n,B}]) - (1-\alpha)| \\
&\le 2\mathcal{E}_1 + 2\mathcal{E}_4 + \sqrt{\frac{2}{\pi}} |t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}} \frac{\mathcal{E}_3}{\sigma} t_{B,1-\alpha/2}.
\end{aligned}
$$

Then taking the minimum of the two bounds, we can get the desired result. We write $\mathcal{A}$ as the event that

$$
\left| \sqrt{\frac{1}{B} \sum_{b=1}^{B} (\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2} - \sigma \right| \le \mathcal{E}_3
$$

$$\Leftrightarrow \sigma - \mathcal{E}_3 \leq \sqrt{\frac{1}{B}\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2} \leq \sigma + \mathcal{E}_3.$$

Then we have $P(\mathcal{A}^c) \leq \mathcal{E}_4$. Note that the confidence interval can be written as

$$\begin{aligned}
&\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}] \\
\Leftrightarrow &\left\{\psi : \left|\frac{\sqrt{n}(\hat{\psi}_n - \psi)}{\sqrt{\frac{1}{B}\sum_{b=1}^{B}(\sqrt{n}(\hat{\psi}_n^{*b} - \hat{\psi}_n))^2}}\right| \leq t_{B,1-\alpha/2}\right\}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma - \mathcal{E}_3} \leq t_{B,1-\alpha/2}; \mathcal{A}\right) \\
&\leq P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}]; \mathcal{A}) \\
&\leq P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma + \mathcal{E}_3} \leq t_{B,1-\alpha/2}; \mathcal{A}\right).
\end{aligned}$$

For the upper bound, we can further bound it as follows

$$\begin{aligned}
&P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma + \mathcal{E}_3} \leq t_{B,1-\alpha/2}; \mathcal{A}\right) \\
&\leq P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma + \mathcal{E}_3} \leq t_{B,1-\alpha/2}\right) \\
&= P(-(\sigma + \mathcal{E}_3)t_{B,1-\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq (\sigma + \mathcal{E}_3)t_{B,1-\alpha/2}).
\end{aligned}$$

By means of the finite-sample accuracy in (15), we have

$$\begin{aligned}
&P(-(\sigma + \mathcal{E}_3)t_{B,1-\alpha/2} \leq \sqrt{n}(\hat{\psi}_n - \psi) \leq (\sigma + \mathcal{E}_3)t_{B,1-\alpha/2}) \\
&\leq \Phi\left(\frac{\sigma + \mathcal{E}_3}{\sigma}t_{B,1-\alpha/2}\right) - \Phi\left(-\frac{\sigma + \mathcal{E}_3}{\sigma}t_{B,1-\alpha/2}\right) + 2\mathcal{E}_1 \\
&\leq \Phi(z_{1-\alpha/2}) - \Phi(-z_{1-\alpha/2}) + 2\mathcal{E}_1 + \sqrt{\frac{2}{\pi}}\left|\frac{\sigma + \mathcal{E}_3}{\sigma}t_{B,1-\alpha/2} - z_{1-\alpha/2}\right| \\
&\leq 1 - \alpha + 2\mathcal{E}_1 + \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2},
\end{aligned}$$

where $z_{1-\alpha/2}$ is the $(1 - \alpha/2)$-th quantile of the standard normal and the second inequality is due to the $1/\sqrt{2\pi}$-Lipschitz property of $\Phi(\cdot)$. For the lower bound, by a similar argument we can obtain

$$\begin{aligned}
&P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma - \mathcal{E}_3} \leq t_{B,1-\alpha/2}; \mathcal{A}\right) \\
&\geq P\left(\frac{|\sqrt{n}(\hat{\psi}_n - \psi)|}{\sigma - \mathcal{E}_3} \leq t_{B,1-\alpha/2}\right) - P(\mathcal{A}^c) \\
&\geq 1 - \alpha - 2\mathcal{E}_1 - \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| - \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2} - P(\mathcal{A}^c) \\
&\geq 1 - \alpha - 2\mathcal{E}_1 - \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| - \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2} - \mathcal{E}_4.
\end{aligned}$$

Thus, by combining the upper and lower bounds, we have the following bound for the coverage error when $\mathcal{A}$ happens

$$\begin{aligned}&|P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}]; \mathcal{A}) - (1-\alpha)|\\ &\leq 2\mathcal{E}_1 + \mathcal{E}_4 + \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2}.\end{aligned}$$

Finally, the overall coverage error can be bounded by

$$\begin{aligned}&|P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}]) - (1-\alpha)|\\ &\leq |P(\psi \in [\hat{\psi}_n - t_{B,1-\alpha/2}S_{n,B}, \hat{\psi}_n + t_{B,1-\alpha/2}S_{n,B}]; \mathcal{A}) - (1-\alpha)| + P(\mathcal{A}^c)\\ &\leq 2\mathcal{E}_1 + 2\mathcal{E}_4 + \sqrt{\frac{2}{\pi}}|t_{B,1-\alpha/2} - z_{1-\alpha/2}| + \sqrt{\frac{2}{\pi}}\frac{\mathcal{E}_3}{\sigma}t_{B,1-\alpha/2},\end{aligned}$$

which, combined with Theorem 3.1, gives us the desired bound. □

Proof of Theorem B.2: By means of Lemmas D.1 and D.2, this directly follows from Theorem 3.2. Besides, the error $\mathcal{E}_1$ can be absorbed into $\mathcal{E}_2$. □

Proof of Theorem B.3: Plugging the bounds (30) and (31) into Theorem 3.2, we get the desired result. □

Proof of Theorem B.4: Plugging the bounds (30) and (32) into Theorem 3.2, we get the desired result. □