The Annals of Statistics
2020, Vol. 48, No. 6, 3643–3671
https://doi.org/10.1214/20-AOS1946
© Institute of Mathematical Statistics, 2020

BEYOND GAUSSIAN APPROXIMATION: BOOTSTRAP FOR MAXIMA OF SUMS OF INDEPENDENT RANDOM VECTORS

BY HANG DENG[*] AND CUN-HUI ZHANG[†]

Department of Statistics, Rutger University, [*]*hdeng@stat.rutgers.edu;* [†]*czhang@stat.rutgers.edu*

The Bonferroni adjustment, or the union bound, is commonly used to study rate optimality properties of statistical methods in high-dimensional problems. However, in practice, the Bonferroni adjustment is overly conservative. The extreme value theory has been proven to provide more accurate multiplicity adjustments in a number of settings, but only on an ad hoc basis. Recently, Gaussian approximation has been used to justify bootstrap adjustments in large scale simultaneous inference in some general settings when $n \gg (\log p)^7$, where p is the multiplicity of the inference problem and n is the sample size. The thrust of this theory is the validity of the Gaussian approximation for maxima of sums of independent random vectors in high dimension. In this paper, we reduce the sample size requirement to $n \gg (\log p)^5$ for the consistency of the empirical bootstrap and the multiplier/wild bootstrap in the Kolmogorov–Smirnov distance, possibly in the regime where the Gaussian approximation is not available. New comparison and anticoncentration theorems, which are of considerable interest in and of themselves, are developed as existing ones interweaved with Gaussian approximation are no longer applicable or strong enough to produce desired results.

1. Introduction. Let $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^{n\times p}$ be a random matrix with independent rows $X_i = (X_{i,1}, \dots, X_{i,p})^T \in \mathbb{R}^p$, $i = 1, \dots n$, where $p \equiv p_n$ is allowed to depend on n. Let

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = (\overline{X}_{n,1}, \dots, \overline{X}_{n,p})^T.$$

We are interested in the consistency of the bootstrap for the maxima

$$T_n = \max_{1\le j\le p} \sqrt{n}(\overline{X}_{n,j} - \mathbb{E}\overline{X}_{n,j}) \tag{1}$$

in the case of large p, including exponential growth of p at certain rate as $n \to \infty$.

The consistency of the bootstrap for the maxima T_n can be directly used to construct simultaneous confidence intervals in the many means problem, but the spectrum of its application is much broader. Examples include sure screening (Fan and Lv (2008)), removing spurious correlation (Fan and Zhou (2016)), testing the equality of two matrices (Cai, Liu and Xia (2013), Chang et al. (2017)), detecting ridges and estimating level sets (Chen, Genovese and Wasserman (2015, 2017)) and many more. It can be also used in time series settings (Zhang and Wu (2017)) and high-dimensional regression (Zhang and Zhang (2014), Belloni, Chernozhukov and Kato (2014, 2015), Dezeure, Bühlmann and Zhang (2017), Zhang and Cheng (2017)). In such modern applications, $p = p_n$ is not fixed and can be much larger than n.

In closely related settings, Giné and Zinn (1990) proved the consistency of bootstrap for Donsker classes of functions, Nagaev (1976), Senatov (1980), Sazonov (1981), Götze (1991)

Received May 2017; revised November 2019.
MSC2020 subject classifications. 62F40, 62E17.
Key words and phrases. Empirical bootstrap, multiplier bootstrap, wild bootstrap, Lindeberg interpolation, Gaussian approximation, multiple testing, simultaneous confidence intervals, maxima of sums, comparison theorem, anticoncentration.

and Bentkus (1986, 2003) for convex sets when $n \geq p^{7/2}$, and Zhilova (2016) for Euclidean balls. The set $\{T_n \leq t\}$ is convex but we are interested in potentially much larger p.

More recently, in a groundbreaking paper, Chernozhukov, Chetverikov and Kato (2013) used Gaussian approximation to prove the consistency of the bootstrap with a convergence rate of $((\log p)^7/n)^{1/8}$ under certain moment and tail probability conditions on $\{X_{i,j}\}$. This convergence rate was improved upon in Chernozhukov, Chetverikov and Kato (2017) to $((\log p)^7/n)^{1/6}$, with extensions to the uniform consistency for $\mathbb{P}\{\sqrt{n}(\overline{X}_n - \mathbb{E}\overline{X}_n) \in A\}$ in certain classes of hyperrectangular and sparse convex sets $A \subseteq \mathbb{R}^p$.

In this paper, we improve the convergence rate to $((\log p)^5/n)^{1/6}$ for the multiplier/wild bootstrap with third moment match (Liu (1988), Mammen (1993)) and the empirical bootstrap (Efron (1979)) of T_n, so that the sample size requirement is reduced from $n \gg (\log p)^7$ to $n \gg (\log p)^5$. We establish this sharper rate by exploiting the fact that under suitable conditions, the average third moment tensor of X_i is well approximated by its bootstrapped version,

$$n^{-1}\sum_{i=1}^{n}\mathbb{E}^*\big(X_i^* - \mathbb{E}^*X_i^*\big)^{\otimes 3} \approx n^{-1}\sum_{i=1}^{n}\mathbb{E}(X_i - \mathbb{E}X_i)^{\otimes 3}, \tag{2}$$

in the supreme norm. Here and in the sequel, $\xi^{\otimes m} = (\xi_{i_1}\cdots\xi_{i_m})_{p\times\cdots\times p}$ denotes the m dimensional tensor/array generated by vector $\xi \in \mathbb{R}^p$. The benefit of the third and higher moment approximation in bootstrap is well understood in the case of fixed p (Hall (1988), Mammen (1993), Shao and Tu (1995), Singh (1981)). However, the classical higher order results on bootstrap were established based on the Edgeworth expansion associated with the central limit theorem, while we are interested in high-dimensional regimes in which the consistency of the Gaussian approximation is in question to begin with. Moreover, as existing approaches of studying the bootstrap in high dimension are very much interweaved with the approximation of the average second moment or the more restrictive approximation of the moments of individual vectors

$$\mathbb{E}^*\big(X_i^* - \mathbb{E}^*X_i^*\big)^{\otimes m} \approx \mathbb{E}(X_i - \mathbb{E}X_i)^{\otimes m}, \quad m = 2, 3, \forall i \leq n, \tag{3}$$

our analysis requires new comparison and anticoncentration theorems. These new comparison and anticoncentration theorems, also proved in this paper, are of considerable interest in their own right.

The difference between the existing and our analytical approaches can be briefly explained as follows. The first issue is the comparison between the expectation of smooth functions of the maxima and its bootstrapped version. The comparison theorems in Chernozhukov, Chetverikov and Kato (2013, 2017) were derived with a combination of the Slepian (1962) smart path interpolation and the Stein (1981) leave-one-out method. As this Slepian–Stein approach does not take advantage of the bootstrap approximation of the third moment, we opt for the Lindeberg approach (Chatterjee (2006), Lindeberg (1922)). In fact, the original Lindeberg method was briefly considered in Chernozhukov, Chetverikov and Kato (2013) without an expansion for the third or higher moment match. As a direct application of the original Lindeberg method requires the more restrictive condition (3), we develop a coherent Lindeberg interpolation to prove comparison theorems based on (2). This coherent Lindeberg approach and the resulting comparison theorems are new to the best of our knowledge. The second issue is the anticoncentration of the maxima, or an upper bound for the modulus of continuity for the distribution of the maxima, without a valid Gaussian approximation. We resolve this issue by applying the new comparison theorem to a mixed multiplier bootstrap with a Gaussian component and a perfect match in the first three moments, so that the anticoncentration of the Gaussian maxima can be utilized through the mixture. This solution to the anticoncentration problem is again new to the best of our knowledge. For the anticoncentration of the maximum of Gaussian vector $(\xi_1,\dots,\xi_p)^T$ with marginal distributions $\xi_j \sim N(\mu_j,\sigma_j^2)$, $1\le j\le p$, we sharpen the existing upper bound for the density of the maximum from $C(2+\sqrt{2\log p})/\sigma_{(1)}$ (based on Klivans, O'Donnell and Servedio (2008)) to the potentially much smaller $(2+\sqrt{2\log p})/\overline{\sigma}$, where

$$\overline{\sigma} = \min_{1\le j\le p} \frac{2+\sqrt{2\log p}}{1/\sigma_{(1)} + (1+\sqrt{2\log j})/\sigma_{(j)}} \tag{4}$$

and $\sigma_{(j)}^2$ is the jth smallest average variance among $\{\sigma_k^2 = \frac{1}{n}\sum_{i=1}^n \mathrm{Var}(X_{i,k}),$ $1\le k\le p\}$. Moreover, our anticoncentration bound is sharp up to explicit constants when ξ_j are correlated and/or noncentral. As more weights are given to the smaller $1/\sigma_{(j)}$ in the denominator in (4), $\sigma_{(1)}\le\overline{\sigma}\le\sigma_{(p)}$.

We organize the paper as follows. In Section 2, we state our bootstrap consistency theorems and discuss their implications and applications. In Section 3, we present new comparison theorems based on the coherent Lindeberg interpolation. In Section 4, we provide new anticoncentration theorems based mixtures with Gaussian components. In Section 5, we present some simulation results. The full proofs of all theorems, propositions and lemmas in this paper are relegated to the Supplementary Material (Deng and Zhang (2020)).

We use the following notation. We assume $n\to\infty$ and $p=p_n$ to allow $p\to\infty$ as $n\to\infty$. We assume $p>1$ for notational simplicity; our analysis remain true for $p=1$ if we replace $\log p$ with $1\vee(\log p)$. To shorten mathematical expressions, we write moments as tensors as in (2) and (3). We also write partial derivative operators as tensors $(\frac{\partial}{\partial x})^{\otimes m} = ((\frac{\partial}{\partial x_{i_1}})\cdots(\frac{\partial}{\partial x_{i_m}}))_{p\times\cdots\times p}$ for $x=(x_1,\dots,x_p)^T$, so that $f^{(m)} = (\partial/\partial x)^{\otimes m} f(x)$ is a tensor for functions $f(x)$ of input $x\in\mathbb{R}^p$, and for two mth order tensors f and g in $\mathbb{R}^{p\times\cdots\times p}$, the vectorized inner product is denoted by

$$\langle f,g\rangle = \sum_{j_1=1}^p \cdots \sum_{j_m=1}^p f_{j_1,\dots,j_m} g_{j_1,\dots,j_m}$$

and $|f|\le|g|$ means $|f_{j_1,\dots,j_m}|\le|g_{j_1,\dots,j_m}|$ for all indices $j_1,\dots,j_m$. We denote by $\|\cdot\|_q$ the ℓ_q norm for vectors, $\|\cdot\|_{L_q} = \|\cdot\|_{L_q(\mathbb{P})}$ the $L_q(\mathbb{P})$ norm for random variables under probability $\mathbb{P}$, and $\|\cdot\|_{\max}$ the ℓ_∞ norm for matrices and tensors after vectorization.

We define quantities M_n, $\mathfrak{M}_m$, $\mathfrak{M}_{m,1}$ and $\mathfrak{M}_{m,2}$ as follows for the average centered moments of X_{ij} under different ways of maximization: The maximum average centered moments and the average moments of the maximum are respectively

$$M_m^m = \max_{1\le j\le p} \frac{1}{n}\sum_{i=1}^n \mathbb{E}|X_{i,j}-\mathbb{E}X_{i,j}|^m, \qquad \mathfrak{M}_m^m = \frac{1}{n}\sum_{i=1}^n \mathbb{E}\max_{1\le j\le p}|X_{i,j}-\mathbb{E}X_{i,j}|^m, \tag{5}$$

and the average of the maximum moment and the expected maximum average power are respectively

$$\mathfrak{M}_{m,1}^m = \frac{1}{n}\sum_{i=1}^n \max_{1\le j\le p}\mathbb{E}|X_{i,j}-\mathbb{E}X_{i,j}|^m, \qquad \mathfrak{M}_{m,2}^m = \mathbb{E}\max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n |X_{i,j}-\mathbb{E}X_{i,j}|^m. \tag{6}$$

Clearly, $M_m\le\mathfrak{M}_{m,j}\le\mathfrak{M}_m$, $j=1,2$.

In what follows, we denote by C_0 a numerical constant and C_{index} a constant depending on the "index" only. For example, $C_{a,b,c}$ is a constant depending on (a,b,c) only. To avoid cumbersome calculation of explicit expressions of these constants, they will be allowed to take different values from one appearance to the next in the proofs. Finally, we denote by $\Phi(\cdot)$ the standard normal cumulative distribution function and $\Phi^{-1}(\cdot)$ the corresponding quantile function.

2. Consistency of bootstrap. Let T_n be the maximum of normalized sum of n independent random vectors $X_i \in \mathbb{R}^p$ as defined in (1). In this section, we present our main theorems on the consistency of bootstrap in approximating the distribution of T_n. We consider this consistency in two somewhat different perspectives. In simultaneous inference about the average mean $\mathbb{E}\sum_{i=1}^n X_{i,j}/n$, we are interested in the performance of the bootstrapped quantile

$$t_\alpha^* = \inf\bigl[t : \mathbb{P}^*\bigl\{T_n^* > t\bigr\} \le \alpha\bigr]$$

at a prespecified significance level α, where T_n^* is the bootstrapped version of T_n and $\mathbb{P}^*$ is the conditional expectation given the original data. As an approximation of the $1-\alpha$ quantile of T_n, the performance of such t_α^* is measured by

$$\bigl|\mathbb{P}\{T_n > t_\alpha^*\} - \alpha\bigr|.$$

On the other hand, if we are interested in recovering the entire distribution function of T_n, it is natural to consider the Kolmogorov–Smirnov distance

$$\eta_n^*(T_n, T_n^*) = \sup_t \bigl|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\bigl\{T_n^* \le t\bigr\}\bigr|.$$

We shall consider Efron's (1979) empirical bootstrap and the wild bootstrap in separate subsections.

It seems possible to extend our ideas and analysis to more general settings, for example the bootstrap schemes in Hall and Presnell (1999) and Præstgaard and Wellner (1993) and the consistency in rectangular sets (Chernozhukov, Chetverikov and Kato (2017)). However, we would not pursue these extensions here as they would make the paper more technical.

2.1. *Empirical bootstrap.* In the empirical bootstrap, we generate i.i.d. vectors $X_1^*, \dots, X_n^*$ from the empirical distribution of the centered data points $X_1 - \overline{X}, \dots, X_n - \overline{X}$ from the original sample: Under the conditional probability $\mathbb{P}^*$ given the original data $\mathbf{X} = (X_1, \dots, X_n)^T$,

$$\mathbb{P}^*\bigl\{X_i^* = X_k - \overline{X}\bigr\} = \frac{\#\{j : 1 \le j \le n : X_j = X_k\}}{n}, \quad 1 \le k \le n, 1 \le i \le n, \tag{7}$$

where $\overline{X} = \sum_{i=1}^n X_i/n$ is the sample mean. The bootstrapped version of T_n is defined as

$$T_n^* = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,j}^*. \tag{8}$$

We state our main theorem on the consistency of empirical bootstrap as follows.

THEOREM 1 (Empirical bootstrap). *Let $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^{n\times p}$ be a random matrix with independent rows $X_i \in \mathbb{R}^p$, X_i^* the empirical bootstrapped X_i as in (7), and T_n and T_n^* as in (1) and (8), respectively. Let M_4 and $\mathfrak{M}_4$ be as in (5), and $\overline{\sigma}$ be as in (4). Define*

$$\gamma_{\delta,M_0}^* = \left(\frac{(\log p)^2(\log(np/\delta))^3}{n} \frac{M_0^4}{\overline{\sigma}^4}\right)^{1/6}. \tag{9}$$

Then, with $M \ge M_4$ satisfying

$$\mathbb{P}\left\{\|\mathbf{X} - \mathbb{E}\mathbf{X}\|_{\max} > \frac{n^{1/3}\overline{\sigma}^{1/3}M^{2/3}}{(\log p)^{1/6}(\log(4np/\delta))^{1/2}}\right\} \le \frac{1}{2}\min\{\delta, \gamma_{\delta,M}^*\}, \tag{10}$$

there exists a numerical constant C_0 such that the Kolmogorov–Smirnov distance between the distributions of T_n and T_n^* is bounded by

$$\sup_{t\in\mathbb{R}} \bigl|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\bigl\{T_n^* \le t\bigr\}\bigr| \le C_0 \min\bigl\{\gamma_{\delta,M}^*, \gamma_{\delta,\mathfrak{M}_4}^*\bigl[1 \vee \bigl(\gamma_{\delta,\mathfrak{M}_4}^*/\delta\bigr)^{1/5}\bigr]\bigr\} \tag{11}$$

with at least probability $1-\delta$. *Moreover, with* $M \geq M_4$ *satisfying* (10) *for* $\delta = 1$,

$$\big|\mathbb{P}\{T_n \leq t_\alpha^*\} - (1-\alpha)\big| \leq C_0 \min\{\gamma_{1,M}^*, \gamma_{1,\mathfrak{M}_4}^*\}. \tag{12}$$

Note that the tail probability condition (10) is needed only when the first component on the right-hand side of (11) and (12) is smaller. Theorem 1 asserts that under the fourth moment and tail probability conditions, Efron's empirical bootstrap provides a consistent estimate of the distribution of T_n when

$$n \gg (\log p)^5.$$

This should be compared with the existing results on the Gaussian wild bootstrap and empirical bootstrap where

$$n \gg (\log p)^7$$

is required (Chernozhukov, Chetverikov and Kato (2013, 2017)). In practice, the significance of the difference between $(\log p)^5$ and $(\log p)^7$ would depend on applications even if we ignore the constant factors involved in different theorems. If the above conditions are viewed as sample size requirements, it would be fair to say that the difference could be quite significant, that is, a $(\log p)^2$ fold increase in n, when data are not dirt cheap. More important, our results prove theoretical advantages of bootstrap schemes with third moment match in high dimension, compared with methods based on Gaussian approximation, as supported by our simulation results in Section 5 for moderately large p. Moreover, as we show in Corollary 1 below, our theory either requires just the fourth moment M_4 or provides the rate $\gamma_n^* \asymp ((B_n/\overline{\sigma})^2(\log(np))^5/n)^{1/2}$ where B_n is the maximum Orlicz norm of X_{ij}.

2.2. *Wild bootstrap*. In wild bootstrap (Wu (1986)), we generate

$$X_i^* = W_i(X_i - \overline{X}), \tag{13}$$

where $\overline{X} = \sum_{i=1}^n X_i/n$ is the sample mean, $W_1, \dots, W_n$ are i.i.d. variables with

$$\mathbb{E}W_i = 0, \qquad \mathbb{E}W_i^2 = 1, \tag{14}$$

and the sequence $\{W_i\}$ is independent of the original data $\mathbf{X} = (X_1, \dots, X_n)^T$.

This general formulation of the wild bootstrap allows broad choices of the multiplier W_i among them the Gaussian $W_i \sim N(0,1)$ and Rademacher $\mathbb{P}\{W_i = \pm 1\} = 1/2$ are the most obvious. Liu (1988) suggested the use of multipliers satisfying

$$\mathbb{E}W_i = 0, \qquad \mathbb{E}W_i^2 = 1, \qquad \mathbb{E}W_i^3 = 1, \tag{15}$$

to allow the third moment match $\mathbb{E}(X_i^*)^{\otimes 3} \approx \mathbb{E}X_i^{\otimes 3}$, and explored the benefits of such schemes. Mammen (1993) proposed a specific choice of the multiplier W_i satisfying (15),

$$\mathbb{P}\left\{W_i = \frac{1 \pm \sqrt{5}}{2}\right\} = \frac{\sqrt{5} \mp 1}{2\sqrt{5}}, \tag{16}$$

and studied extensively the benefit of the third moment match in wild bootstrap. We note here that while (15) holds for many choices of W_i, the Gaussian and Rademacher multipliers do not possess this property. In the following theorem, we assume the sub-Gaussian condition

$$\mathbb{E}\exp(tW_1) \leq \exp(\tau_0^2 t^2/2), \quad \forall t \in \mathbb{R}, \tag{17}$$

in addition to the third moment condition (15).

THEOREM 2 (Wild bootstrap). *Let $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^{n\times p}$ be a random matrix with independent rows $X_i \in \mathbb{R}^p$, and X_i^* be generated by the wild bootstrap as in* (13) *with multipliers satisfying the moment condition* (15) *and the sub-Gaussian condition* (17) *with a certain $\tau_0 < \infty$. Let T_n and T_n^* be as in* (1) *and* (8), *respectively. Define*

$$\gamma^*_{\delta,M_0} = \left(\frac{(\log p)^2(\log(np))(\log(np/\delta)^2)}{n}\frac{M_0^4}{\overline{\sigma}^4}\right)^{1/6}. \tag{18}$$

Then, with $M \ge M_4$ satisfying

$$\mathbb{P}\left\{\|\mathbf{X}-\mathbb{E}\mathbf{X}\|_{\max} > \frac{n^{1/3}\overline{\sigma}^{1/3}M^{2/3}}{(\log p)^{1/6}(\log(np))^{1/3}(\log(4np/\delta))^{1/6}}\right\} \le \frac{1}{2}\min\{\delta, \gamma^*_{\delta,M}\}, \tag{19}$$

there exists a numerical constant C_{τ_0} such that the Kolmogorov–Smirnov distance between the distributions of T_n and T_n^* is bounded by

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\{T_n^* \le t\}\right| \le C_{\tau_0}\min\{\gamma^*_{\delta,M}, \gamma^*_{\delta,\mathfrak{M}_{4,2}}[1 \vee (\gamma^*_{\delta,\mathfrak{M}_{4,2}}/\delta)^{1/5}]\} \tag{20}$$

with at least probability $1-\delta$, where $\mathfrak{M}_{4,2} \le \mathfrak{M}_4$ by its definition in (6). *Moreover, with $M \ge M_4$ satisfying* (19) *for $\delta = 1$,*

$$\left|\mathbb{P}\{T_n \le t^*_\alpha\} - (1-\alpha)\right| \le C_{\tau_0}\min\{\gamma^*_{1,M}, \gamma^*_{1,\mathfrak{M}_{4,2}}\}. \tag{21}$$

REMARK 1. A user friendly bound of $\mathfrak{M}_{4,2}$

$$\mathfrak{M}_{4,2}^4 \le K\left(M_4^4 + \frac{\log p}{n}\mathbb{E}\max_{i,j}|X_{i,j} - \mathbb{E}X_{i,j}|^4\right) \tag{22}$$

for some universal constant K can be found in Lemma 9 of Chernozhukov, Chetverikov and Kato (2015) and Lemma E.3 of Chernozhukov, Chetverikov and Kato (2017).

Theorem 2 asserts that with the third moment condition (15) on the multiplier, the conclusions of Theorem 1 are all valid for the wild bootstrap under weaker moment condition. Thus, the discussion below Theorem 1 about its significance also applies to Theorem 2.

While the statements of Theorems 1 and 2 are almost identical, the smaller quantity $\mathfrak{M}_{4,2}$ is used in (20) and (21) in Theorem 2, compared with the larger $\mathfrak{M}_4$ in (11) and (12) in Theorem 1. Theorem 2 can be further sharpened if Theorems 7 and 8 in Section 3 are applied in full strength.

As briefly discussed below Theorem 1, a key point in our theory is the benefit of the third or higher moment match in both the empirical bootstrap and wild bootstrap. Efron's empirical bootstrap can always match moments but not exactly,

$$\mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i^{\otimes m} - \frac{1}{n}\sum_{i=1}^n \mathbb{E}^*(X_i^*)^{\otimes m}\right\} \approx 0, \quad m = 1, 2, \dots$$

An alternative wild bootstrap scheme, $X_i^* = W_i X_i$, which approximates (13) with negligible difference in our analysis under the assumption of $\mathbb{E}X_i = 0$, matches the moments of X_i perfectly,

$$\mathbb{E}\{\mathbb{E}X_i^{\otimes m} - \mathbb{E}^*(X_i^*)^{\otimes m}\} = 0, \tag{23}$$

but only up to a certain order; $m = 1, 2$ for the Gaussian and Rademacher wild bootstrap, and $m = 1, 2, 3$ for Mammen's and other wild bootstrap schemes satisfying (15). Thus, compared with the proof of Theorem 2 which directly applies the exact moment match in (23), the proof of Theorem 1 requires an additional analysis of the the difference in the moments, leading to the stronger condition involving $\mathfrak{M}_4$.

If $X_i \in \mathbb{R}^p$ have symmetric distributions, condition (23) holds for all m for the Rademacher wild bootstrap. In this case, the sample size condition $n \gg (\log p)^4$ is sufficient for the consistency of the bootstrap under sixth moment and tail probability conditions and an anticoncentration condition.

THEOREM 3 (Rademacher wild bootstrap). *Let* $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^{n\times p}$ *be a random matrix with independent rows* $X_i \in \mathbb{R}^p$. *Suppose* $\mathbb{E}(X_i - \mathbb{E}X_i)^{\otimes m} = 0$ *for* $m = 3$ *and* $m = 5$. *Let* X_i^* *be generated by the Rademacher wild bootstrap, with* $\mathbb{P}\{W_i = \pm 1\} = 1/2$ *for the multiplier in* (13). *Then, for any given constants* c_0, c_1 *and* $M \ge M_6$,

$$
\begin{aligned}
&\big|\mathbb{P}\{T_n \le t_\alpha^*\} - (1-\alpha)\big| + \Big(\mathbb{E}\sup_{t\in\mathbb{R}}\big|\mathbb{P}\{T_n < t\} - \mathbb{P}^*\{T_n^* < t\}\big|^2\Big)^{1/2} \\
(24)\quad &\le C_{c_0,c_1}\Big(\frac{\log p}{n^{1/4}}\Big)^{4/7} + \sup_{t\in\mathbb{R}}\mathbb{P}\Big\{t - c_0\Big(\frac{\log p}{n^{1/4}}\Big)^{4/7} \le \sqrt{\log p}\,\frac{T_n}{M} \le t\Big\} \\
&\quad + \Big[\mathbb{E}\min\Big\{4, C_{c_0,c_1}\Big(\frac{\log p}{n^{1/4}}\Big)^{32/7}\max_{1\le j\le p}\sum_{i=1}^n \frac{(X_{i,j} - \mathbb{E}X_{i,j})^6}{M^6 n} I_{\{|X_{i,j}-\mathbb{E}X_{i,j}|>a_n\}}\Big\}\Big]^{1/3},
\end{aligned}
$$

where $a_n = c_1 M\sqrt{\log p}(n^{1/4}/\log p)^{10/7}$ *and* C_{c_0,c_1} *is a constant depending on* $\{c_0, c_1\}$ *only.*

The discussion below Theorem 1 about its significance also applies here, although $(\log p)^5$ is further improved to $(\log p)^4$ and an anticoncentration condition is required in Theorem 3. In Section 4, we prove that the anti-concentration condition

$$
\sup_t \mathbb{P}\Big\{t - \epsilon_n \le \sqrt{\log p}\,\frac{T_n}{M} \le t\Big\} = o(1) \quad \forall \epsilon_n = o(1)
$$

holds when $\sum_{i=1}^n X_i/\sqrt{n}$ is conditionally a Gaussian vector given a certain sigma field $\mathcal{A}$, with $\mathrm{Var}(\sum_{i=1}^n X_{i,j}/\sqrt{n}|\mathcal{A}) = \sigma_j^2$ such that $\mathbb{P}\{\min_j \sigma_j^2 \ge \underline{\sigma}^2\} \to 1$ for a certain constant $\underline{\sigma} > 0$.

The condition $\mathbb{E}(X_i - \mathbb{E}X_i)^{\otimes m} = 0$ holds for the leading odd $m \in \{3, 5\}$ when X_i are symmetric about its mean, that is, $\mathbb{P}\{X_i - \mathbb{E}X_i \in A\} = \mathbb{P}\{\mathbb{E}X_i - X_i \in A\}$ for all Boreal sets $A \subset \mathbb{R}^p$. In practice, such conditions could be imposed by the application itself. If the validity of such conditions is uncertain, we may also test the moment condition when X_i are i.i.d. However, a theoretical analysis of such tests and the validity of (24) for the Rademacher wild bootstrap after such tests is beyond the scope of this paper.

2.3. *Examples.* In this subsection, we consider some specific examples in which the moment and tail probability conditions of our theorems hold. These examples cover many practical problems and applications as discussed in Chernozhukov, Chetverikov and Kato (2013, 2017), and many publications citing their work (Blanchet, Kang and Murthy (2019), Chen (2018), Dezeure, Bühlmann and Zhang (2017), Ning and Liu (2017), Zhang and Wu (2017), Horowitz (2019)). Throughout this subsection, we assume the following:

Cond-1: $0 < \overline{\sigma} \le (2 + \sqrt{2\log p})/\{1/\sigma_{(1)} + (1 + \sqrt{2\log j})/\sigma_{(j)}\}, \forall j = 1, \dots, p,$
Cond-2: $n^{-1}\sum_{i=1}^n \mathbb{E}|X_{i,j} - \mathbb{E}X_{i,j}|^4 \le M_4^4, \forall j = 1, \dots, p,$

where $\sigma_{(1)} \le \cdots \le \sigma_{(p)}$ are the ordered values of $\sigma_j = (n^{-1}\sum_{i=1}^n \mathbb{E}(X_{i,j} - \mathbb{E}X_{i,j})^2)^{1/2}$. Here, $\overline{\sigma}$ and M_4 are allowed to depend on n and to diverge to 0 or ∞, but they can also be treated as constants for simplicity. Under the above moment conditions, we consider three examples specified by certain measure B_n of the tail of $\{|X_{i,j}|\}$, possibly with unbounded B_n.

2.3.1. *Exponential tail.* Here, we impose one additional condition on the tail of $X_{i,j}$ in the form of a uniform bound on their Orlicz norm with respect to $\psi_1(x) = e^x - 1$: with $\inf \varnothing = \infty$,

(E.1) $\|X_{i,j}\|_{\psi_1} = \inf\{B : \mathbb{E}\psi_1(|X_{i,j} - \mathbb{E}X_{i,j}|/B) \le 1\} \le B_n, \forall i, j.$

COROLLARY 1. *Suppose X_i are independent. Let T_n and T_n^* be as in* (1) *and* (8), *respectively, and B_n be as in* (E.1).

(i) *Let X_i^* be generated by the empirical bootstrap as in* (7). *Then* (11) *and* (12) *hold with*

$$\gamma^*_{\delta,M} = \max\left\{\left(\frac{(\log p)^2(\log(np/\delta)^3}{n}\frac{M_4^4}{\overline{\sigma}^4}\right)^{1/6}, \left(\frac{(\log p)(\log(np/\delta))^4}{n}\right)^{1/2}\frac{B_n}{\overline{\sigma}}\right\}.$$

(ii) *Let X_i^* be generated by the wild bootstrap as in* (13). *Suppose the multipliers W_i satisfy the moment condition* (15) *and the sub-Gaussian condition* (17) *with a $\tau_0 < \infty$. Then* (20) *and* (21) *hold with*

$$\gamma^*_{\delta,M} = \max\left\{\left(\frac{(\log p)^2(\log(np)(\log(np/\delta)^2}{n}\frac{M_4^4}{\overline{\sigma}^4}\right)^{1/6}, \left(\frac{(\log p)(\log(np))(\log(np/\delta))^3}{n}\right)^{1/2}\frac{B_n}{\overline{\sigma}}\right\}.$$

REMARK 2. As $x^4 \le 5\psi_1(x)$ for $x \ge 0$, we have $M_4^4 \le 5B_n^4$, but B_n/M_4 could be unbounded. We may compare the above result under (E.1) with Chernozhukov, Chetverikov and Kato (2017) for the maxima. For the empirical bootstrap, Propositions 2.1 and 4.3 of Chernozhukov, Chetverikov and Kato (2017) yields the following Kolmogorov–Smirnov distance bound:

$$\sup_t\left|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\{T_n^* \le t\}\right| \le C_K\{\overline{B}_n^2(\log(np))^7/n\}^{1/6}$$

with probability at least $1-\delta$ when $\log(1/\delta) \le K\log(np)$, where $\overline{B}_n = \max\{M_4^2/\sigma_{(1)}^2, B_n/\sigma_{(1)}\}$ is a scale-free version of their constant factor with the B_n in (E.1) and $\sigma_{(1)} = \min_{j\le p}\sigma_j$. Corollary 1 (i) improves the rate of their upper bound by at least a factor of $\log^{1/3}(np)(\overline{\sigma}/\sigma_{(1)})^{2/3}$. When $M_4/\sigma_{(1)} = O(1)$ and $B_n/\sigma_{(1)} \asymp n^{\kappa_0}$ with nontrivial $\kappa_0 \in (0, 1/2)$, the rate improvement is by at least the following factor of polynomial order:

$$\min\{n^{\kappa_0/3}\log^{1/3}(np), n^{(1-2\kappa_0)/3}(\log(np))^{7/6-5/2}\}.$$

Similarly, for the Gaussian wild bootstrap, the combination of Proposition 2.1 and Corollary 4.2 of Chernozhukov, Chetverikov and Kato (2017) yields the following Kolmogorov–Smirnov distance bound:

$$\sup_t\left|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\{T_n^* \le t\}\right| \le C_0\{B_n^2\log^5(np)\log^2(np/\delta)/n\}^{1/6}$$

with probability at least $1-\delta$. With the third moment match in wild bootstrap, Corollary 1(ii) improves upon their rate by at least a factor of $\log^{1/3}(np)(\overline{\sigma}/\sigma_{(1)})^{2/3}$ in general, and by at least $\min\{n^{\kappa_0/3}, n^{(1-2\kappa_0)/3}\}\mathrm{polylog}(np/\delta)$ when $M_4/\sigma_{(1)} = O(1)$ and $B_n/\sigma_{(1)} \asymp n^{\kappa_0}$ with $\kappa_0 \in (0, 1/2)$.

We note that the product of sub-Gaussian variables satisfies the sub-exponential condition (E.1) imposed in Corollary 1. For example, for testing the equality of the population covariance matrices of two samples $\{Y_i\}$ and $\{Z_i\}$ in $\mathbb{R}^d$, we just need to set

$$X_i = \mathrm{vec}(Y_iY_i^T - Z_iZ_i^T) \quad \text{with } p = d(d+1)/2,$$

as in Cai, Liu and Xia (2013) and Chang et al. (2017).

2.3.2. *Conditionally Gaussian vectors with Gaussian tail.* Suppose $\sum_{i=1}^n X_i/\sqrt{n}$ is conditionally a Gaussian vector given a certain sigma

(E.2) : field $\mathcal{A}$, $(\sum_{i=1}^n X_{i,j}/\sqrt{n})|\mathcal{A} \sim N(\mu_j, \sigma_j^2)$, and with $\psi_2 = e^{x^2} - 1$,

$$\|X_{i,j}\|_{\psi_2} = \inf\{B : \mathbb{E}\psi_2(|X_{i,j} - \mathbb{E}X_{i,j}|/B) \le 1\} \le B_n.$$

Under (E.2), Theorem 3 is applicable, and a corollary of it is stated as follows.

COROLLARY 2. *Let $\mathbf{X} = (X_1, \dots, X_n)^T \in \mathbb{R}^{n\times p}$ be a random matrix with independent rows $X_i \in \mathbb{R}^p$. Suppose $\mathbb{E}(X_i - \mathbb{E}X_i)^{\otimes m} = 0$ for $m = 3$ and $m = 5$. Let X_i^* be generated by the Rademacher wild bootstrap, with $\mathbb{P}\{W_i = \pm 1\} = 1/2$ for the multiplier in* (13). *Then, under* (*E*.2), *we have*

$$\max\Big\{\big|\mathbb{P}\{T_n \le t_\alpha^*\} - (1-\alpha)\big|, \Big(\mathbb{E}\sup_{t\in\mathbb{R}}\big|\mathbb{P}\{T_n < t\} - \mathbb{P}^*\{T_n^* < t\}\big|^2\Big)^{1/2}\Big\}$$
$$\le C_0\bigg[\bigg(\frac{\log p}{n^{1/4}}\bigg)^{4/7}\frac{M_6}{\overline{\sigma}} + \bigg(\frac{\log p}{n^{1/4}}\bigg)^2\frac{B_n\sqrt{\log(np)}}{\overline{\sigma}\sqrt{\log p}}\bigg],$$

where $\overline{\sigma}$ is a constant upper bound for the soft minimum of $\{\sigma_1, \dots, \sigma_p\}$ in (*E*.2) *as in* (4).

2.3.3. *Moment conditions.* Consider the following conditions on moments of the maxima:

(E.3): $\mathfrak{M}_q^q = n^{-1}\sum_{i=1}^n \mathbb{E}\max_{1\le j\le p}|X_{i,j} - \mathbb{E}X_{i,j}|^q \le B_n^q$,
(E.4): $\mathfrak{M}_{4,2}^4 = \mathbb{E}\max_{1\le j\le p}\sum_{i=1}^n |X_{i,j} - \mathbb{E}X_{i,j}|^4/n \le B_n^4$,
(E.5): $\mathfrak{M}_{6,2}^6 = \mathbb{E}\max_{1\le j\le p}\sum_{i=1}^n |X_{i,j} - \mathbb{E}X_{i,j}|^6/n \le B_n^6$.

Theorems 1, 2 and 3, respectively, imply the following corollary.

COROLLARY 3. *Suppose X_i are independent. Let T_n and T_n^* be as in* (1) *and* (8), *respectively.*

(i) *Let X_i^* be generated by the empirical bootstrap as in* (7). *Then, under* (*E*.3), (11) *and* (12) *hold with constant C_q and*

$$\gamma_{\delta,M}^* = \max\bigg\{\gamma_{\delta,M_4}^*, \gamma_1^*(B_n)\bigg[1 \vee \bigg(\frac{\gamma_1^*(B_n)}{\delta}\bigg)^{1/q}\bigg]\bigg\},$$

where

$$\gamma_1^*(B_n) = \bigg(\frac{(\log p)^{1/2}(\log(np/\delta))}{n^{1/2-1/q}}\frac{B_n}{\overline{\sigma}}\bigg)^{q/(q+1)}.$$

Moreover, if (*E*.3) *holds with $q = 4$, then $\gamma_{\delta,\mathfrak{M}_4}^* = \gamma_{\delta,B_n}^*$ in* (11) *and* (12).

(ii) *Let X_i^* be generated by the wild bootstrap as in* (13) *with multipliers satisfying the moment condition* (15) *and the sub-Gaussian condition* (17) *with a certain $\tau_0 < \infty$. Then, under* (*E*.3), (20) *and* (21) *hold with constant $C_{\tau_0,q}$ and*

$$\gamma_{\delta,M}^* = \max\bigg\{\gamma_{\delta,M_4}^*, \gamma_2^*(B_n)\bigg[1 \vee \bigg(\frac{\gamma_2^*(B_n)}{\delta}\bigg)^{1/q}\bigg]\bigg\},$$

where

$$\gamma_2^*(B_n) = \bigg(\frac{(\log p)^{1/2}(\log(np))^{1/2}(\log(np/\delta))^{1/2}}{n^{1/2-1/q}}\frac{B_n}{\overline{\sigma}}\bigg)^{q/(q+1)}.$$

However, under (*E*.4), (20) *and* (21) *hold with $\gamma_{\delta,\mathfrak{M}_{4,2}}^* = \gamma_{\delta,B_n}^*$.*

(iii) *Suppose that* $\mathbf{X}$ *satisfies the conditions of Theorem* 3 *and* (E.5), *and that* $\sum_{i=1}^n X_i/\sqrt{n}$ *satisfies the conditional Gaussian condition in* (E.2) *with a constant lower bound* $\overline{\sigma}$ *for the soft minimum of the conditional standard deviation as in* (4). *Let* X_i^* *be generated by the Rademacher wild bootstrap as in Theorem* 3. *Then*

$$\begin{aligned}&\left|\mathbb{P}\{T_n \le t_\alpha^*\} - (1-\alpha)\right| + \Bigl(\mathbb{E}\sup_{t\in\mathbb{R}}\left|\mathbb{P}\{T_n < t\} - \mathbb{P}^*\{T_n^* < t\}\right|^2\Bigr)^{1/2}\\ &\quad\le C_0\biggl[\Bigl(\frac{\log p}{n^{1/4}}\Bigr)^{4/7}\frac{B_n\sqrt{\log(np)}}{\overline{\sigma}\sqrt{\log p}} + \Bigl(\frac{\log p}{n^{1/4}}\Bigr)^{32/21}\biggr].\end{aligned}$$

REMARK 3. We compare the above result under (E.3) with Chernozhukov, Chetverikov and Kato (2017) for the maxima. For the empirical bootstrap, Corollary 3(i) implies with at least probability $1-\delta$, the Kolmogorov–Smirnov distance in (11) is bounded by

$$C_q \max\biggl\{\Bigl(\frac{(\log p)^2(\log(np))^3}{n}\frac{M_4^4}{\overline{\sigma}^4}\Bigr)^{1/6}, \Bigl(\frac{(\log p)(\log(np))^2}{n^{1-2/q}}\frac{B_n^2}{\overline{\sigma}^2}\Bigr)^{\frac{q}{2(q+1)}}\biggr\} \tag{25}$$

when δ is greater than the second component, and

$$C_q \max\biggl\{\Bigl(\frac{(\log p)^2(\log(np))^3}{n}\frac{M_4^4}{\overline{\sigma}^4}\Bigr)^{1/6}, \Bigl(\frac{(\log p)(\log(np))^2}{n^{1-2/q}\delta^{2/q}}\frac{B_n^2}{\overline{\sigma}^2}\Bigr)^{1/2}\biggr\} \tag{26}$$

when δ is smaller. Note that $\log(np/\delta) \asymp \log(np)$ as otherwise δ is extremely small so that the second bound is effective but also trivial due to small $n^{1-2/q}\delta^{2/q}$. For the third-moment match wild bootstrap, (ii) yields a slightly better result but the above bounds in (25) and (26) also apply. In Chernozhukov, Chetverikov and Kato (2017), the combination of Propositions 2.1 and 4.3 for the empirical bootstrap and the combination of Proposition 2.1 and Corollary 4.2 for the Gaussian wild bootstrap yield the Kolmogorov–Smirnov distance bound as

$$\begin{aligned}&\sup_t\left|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\{T_n^* \le t\}\right|\\ &\quad\le C_{q,K}\max\biggl\{\Bigl(\frac{\overline{B}_n^2(\log(np))^7}{n}\Bigr)^{1/6}, \Bigl(\frac{\overline{B}_n^2(\log(np))^3}{n^{1-2/q}\delta^{2/q}}\Bigr)^{1/3}\biggr\},\end{aligned}$$

with at least probability $1-\delta$, where $\overline{B}_n = \max\{M_4^2/\sigma_{(1)}^2, B_n/\sigma_{(1)}\}$ with the B_n in (E.3) and $\sigma_{(1)} = \min_j \sigma_j$. It is clear that the first component of the bound in (25) or (26) improves the first rate above by at least a factor of $(\overline{\sigma}/\sigma_{(1)}\log(np))^{1/3}$. As $q/(2(q+1)) > 1/3$ for all $q > 2$ and the bounds are trivial when $q \le 2$, the second components in (25) and (26) improves the second rate above by at least a factor of

$$\Bigl(\frac{n^{1-2/q}}{(\log p)(\log(np))^2}\frac{\overline{\sigma}^2}{B_n^2}\Bigr)^{\frac{q}{2(q+1)}-\frac{1}{3}} \quad \text{for } q > 2.$$

In linear regression, we observe $y_i = Z_i^T\beta + \varepsilon_i$. Suppose the design vectors are deterministic and normalized to $\sum_{i=1}^n Z_{i,j}^2 = n$. Suppose we want to control the spurious correlation in sure screening based on $\sum_{i=1}^n y_i Z_i/\sqrt{n}$ as in Fan and Lv (2008) and Fan and Zhou (2016). Let $X_i = y_i Z_i$. We have $X_i - \mathbb{E}X_i = \varepsilon_i Z_i$ and

$$T_n = \Bigl\|\sum_{i=1}^n y_i Z_i/\sqrt{n} - \mathbb{E}\sum_{i=1}^n y_i Z_i/\sqrt{n}\Bigr\|_\infty.$$

Suppose $\mathbb{E}\varepsilon_i = 0$ and $\mathbb{E}\varepsilon_i^2 = \sigma^2$. For $1 \le q \le \infty$, define

$$\left(\frac{1}{n}\sum_{i=1}^n \mathbb{E}|\varepsilon_i|^q\right)^{1/q} \le M_{\varepsilon,q}, \qquad \max_{j\le p}\left(\frac{1}{n}\sum_{i=1}^n |Z_{i,j}|^q\right)^{1/q} \le M_{Z,q}.$$

Then conditions (E.3) with $q = 4$, (E.4) and (E.5) can be fulfilled with

$$\mathfrak{M}_4 \le M_{\varepsilon,4} M_{Z,\infty}, \qquad \mathfrak{M}_{m,2} \le M_{\varepsilon,mq} M_{Z,mq/(1-q)}, \quad 1 \le q \le \infty,$$

where $m = 4, 6$ in (E.4), (E.5), respectively. Dezeure, Bühlmann and Zhang (2017) studied bootstrap simultaneous inference in high-dimensional linear regression under the sample size condition $n \ge (\log p)^7 + s^2(\log p)^3$ and the moment condition $M_{\varepsilon,4} + M_{Z,\infty} = O(1)$.

2.4. *Lévy–Prokhorov predistance and anticoncentration.* The Kolmogorov–Smirnov distance between two distribution functions can be bounded from the above by a sum of upper bounds for their Lévy–Prokhorov distance and the minimum of their modulus of continuity. For two random elements T_n and T_n^* living in a common metric space equipped with a probability measure $\mathbb{P}$, the Lévy–Prokhorov distance is the smallest $\epsilon > 0$ satisfying

$$\max\big[\mathbb{P}\{T_n \in A\} - \mathbb{P}\{T_n^* \in A(\epsilon)\}, \mathbb{P}\{T_n^* \in A\} - \mathbb{P}\{T_n \in A(\epsilon)\}\big] \le \epsilon \tag{27}$$

for all Borel sets A, where $A(\epsilon) = \{y : \min_{x\in A} d(x,y) < \epsilon\}$ is the ϵ-neighborhood of A. For comparison of the distributions of two maxima T_n and T_n^* for simultaneous testing, it is typically sufficient to consider one-sided intervals $A = (\infty, t]$ in (27). Choosing $A = (\infty, t]$ is also sufficient for studying the Kolmogorov–Smirnov distance between the distribution functions of T_n and T_n^*. Thus, our analysis focuses on the following quantity:

$$\eta_n(\epsilon) \equiv \eta_n^{(\mathbb{P})}(\epsilon; T_n, T_n^*) = \sup_{t\in\mathbb{R}} \eta_n^{(\mathbb{P})}(\epsilon, t; T_n, T_n^*) \tag{28}$$

with $\eta_n^{(\mathbb{P})}(\epsilon, t; T_n, T_n^*) = \max[\mathbb{P}\{T_n \le t-\epsilon\} - \mathbb{P}\{T_n^* < t\}, \mathbb{P}\{T_n^* \le t-\epsilon\} - \mathbb{P}\{T_n < t\}, 0]$. As the Lévy–Prokhorov distance over all one-sided intervals is the smallest ϵ satisfying $\eta_n(\epsilon) \le \epsilon$, we refer to the quantity $\eta_n(\epsilon)$ as Lévy–Prokhorov predistance for convenience. It does not define a distance between T_n and T_n^*, but satisfies a "pseudo-triangular inequality" in the sense that $\forall \widetilde{T}_n, \epsilon_1 + \epsilon_2 < \epsilon$ and $\epsilon_1, \epsilon_2 > 0$,

$$\eta_n^{(\mathbb{P})}(\epsilon; T_n, T_n^*) \le \eta_n^{(\mathbb{P})}(\epsilon_1; T_n, \widetilde{T}_n) + \eta_n^{(\mathbb{P})}(\epsilon_2; \widetilde{T}_n, T_n^*). \tag{29}$$

It is straightforward by the triangle inequality that the Kolmogorov–Smirnov distance between the cumulative distribution functions of T_n and T_n^*, equal to $\eta_n(0+)$, is bounded by

$$\begin{aligned} &\sup_{t\in\mathbb{R}}\big|\mathbb{P}\{T_n < t\} - \mathbb{P}\{T_n^* < t\}\big| \\ &\quad = \eta_n(0+) \le \eta_n(\epsilon) + \min\{\omega_n(\epsilon; T_n), \omega_n(\epsilon; T_n^*)\}, \quad \forall \epsilon > 0, \end{aligned} \tag{30}$$

where $\omega_n(\epsilon; T_n) = \omega_n^{(\mathbb{P})}(\epsilon; T_n) = \sup_{t\in\mathbb{R}} \mathbb{P}\{t-\epsilon < T_n < t\}$ and $\omega_n(\epsilon; T_n^*) = \omega_n^{(\mathbb{P})}(\epsilon; T_n^*)$ is defined in the same way with T_n replaced by T_n^*. The quantity $\omega_n(\epsilon; T_n)$, which is also called the Lévy concentration function, is the modulus of continuity of the cumulative distribution function of T_n.

The Lévy–Prokhorov predistance characterizes the convergence in distribution. When T_n has a fixed distribution function H_0, T_n^* converges in distribution to H_0 if and only if $\eta_n(\epsilon) \to 0$ $\forall \epsilon > 0$. On the other hand, $\lim_{\epsilon\to 0+} \omega_n(\epsilon; T_n) = 0$ if and only if H_0 is continuous. Of course, if T_n^* converges in distribution to a continuous H_0, then the distribution function of T_n^* converges to H_0 in the Kolmogorov–Smirnov distance. Moreover, as $\eta_n(\epsilon)$ is decreasing in ϵ, the condition $\eta_n(\epsilon) \to 0$ $\forall \epsilon > 0$ is necessary for the convergence $\eta_n(0+) \to 0$ in the Kolmogorov–Smirnov distance.

Inequality (30) asserts that the Kolmogorov–Smirnov distance is bounded by a sum of two quantities, the Lévy–Prokhorov predistance which allows a shift ϵ in the comparison of two distribution functions and the Lévy concentration as an upper bound for the error introduced by the shift. By allowing a shift, the Lévy–Prokhorov pre-distance can be further bounded by comparison of the expectations of smooth functions of T_n and T_n^* so that the Lindeberg interpolation can be applied as discussed in detail in Section 3. Upper bounds for the Lévy concentration, called the anticoncentration inequality, will be discussed in Section 4. The role of (30) is to explicitly spell out the roles of the comparison and anticoncentration theorems and to facilitate the notation in our analysis. We note that $\eta_n(\epsilon)$ is decreasing but $\min\{\omega_n(\epsilon; T_n), \omega_n(\epsilon; T_n^*)\}$ is increasing in ϵ. In our analysis, we pick an $\epsilon = 1/b_n$ to balance the rate of the two terms in (30). For example, as $\omega_n(1/b_n; T_n) \lesssim b_n^{-1}\sqrt{\log p}$ by Theorem 12 in Section 4, $b_n^{-1} \asymp ((\log p)^2/n)^{1/6}$ is used to achieve the rate $((\log p)^5/n)^{1/6}$ in Theorems 1 and 2.

In bootstrap, we are interested in approximating the distribution of T_n under the marginal probability $\mathbb{P}$ by the distribution of the bootstrap T_n^* under the conditional probability $\mathbb{P}^*$ given the original data. To streamline the notation, we write this comparison under a common probability measure by introducing a copy T_n^0 of T_n independent of the original data $\mathbf{X}$, so that $\mathbb{P}\{T_n \le t\} = \mathbb{P}\{T_n^0 \le t | \mathbf{X}\} = \mathbb{P}^*\{T_n^0 \le t\}$. This allows us to write

$$
\begin{aligned}
&\eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*) \\
&\quad = \max\big[\mathbb{P}^*\{T_n^0 \le t - \epsilon\} - \mathbb{P}^*\{T_n^* < t\}, \mathbb{P}^*\{T_n^* \le t - \epsilon\} - \mathbb{P}^*\{T_n^0 < t\}, 0\big] \\
&\quad = \max\big[\mathbb{P}\{T_n \le t - \epsilon\} - \mathbb{P}^*\{T_n^* < t\}, \mathbb{P}^*\{T_n^* \le t - \epsilon\} - \mathbb{P}\{T_n < t\}, 0\big].
\end{aligned}
$$

The following lemma connects the consistency of bootstrap to the tail probability of the random Lévy–Prokhorov predistance under $\mathbb{P}^*$ and Lévy concentration function $\omega_n(\epsilon; T_n)$.

LEMMA 1. *Let t_α^* be the $(1-\alpha)$-quantile of T_n^* under $\mathbb{P}^*$. Then, for all $\epsilon_n > 0$ and $\eta > 0$,*

$$
\big|\mathbb{P}\{T_n \le t_\alpha^*\} - (1-\alpha)\big| \le \sup_t \mathbb{P}\{\eta_n^{(\mathbb{P}^*)}(\epsilon_n, t; T_n^0, T_n^*) > \eta\} + \eta + \omega_n(\epsilon_n; T_n),
$$

and the Kolmogorov–Smirnov distance between $\mathbb{P}\{T_n \le t\}$ and $\mathbb{P}^*\{T_n^* < t\}$ is bounded by

$$
\sup_{t\in\mathbb{R}} \big|\mathbb{P}\{T_n < t\} - \mathbb{P}^*\{T_n^* < t\}\big| \le \eta + \omega_n^{(\mathbb{P})}(\epsilon_n; T_n)
$$

when $\eta_n^*(\epsilon_n) \le \eta$, where $\eta_n^*(\epsilon) \equiv \eta_n^{(\mathbb{P}^*)}(\epsilon; T_n^0, T_n^*) = \sup_{t\in\mathbb{R}} \eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*)$.

We derive in the next two sections upper bounds for the Lévy–Prokhorov predistances $\eta_n(\epsilon)$ and $\eta_n^*(\epsilon)$ and the Lévy concentration function $\omega_n(\epsilon; T_n)$, respectively.

3. Comparison theorems. Let h_0 be a smooth decreasing function taking value 1 in $(-\infty, -1]$ and 0 in $[0, \infty)$. As we will explicitly explain at the beginning of the proof of Theorem 5, it follows directly from the definition of the Lévy–Prokhorov predistance in (28) that

$$
\eta_n(1/b_n) \le \sup_{t\in\mathbb{R}} \big|\mathbb{E}h_t(b_n T_n) - \mathbb{E}h_t(b_n T_n^*)\big|, \quad \forall b_n > 0,
$$

where $h_t(\cdot) = h_0(\cdot - t)$ is the location shift of h_0. In this section, we develop comparison theorems which provide expansions and bounds for

$$
\mathbb{E}f(X_1, \dots, X_n) - \mathbb{E}^* f(X_1^*, \dots, X_n^*)
$$

in terms of average moments of $\{X_i, i \le n\}$ and $\{X_i^*, i \le n\}$. Here, $f(x_1, \dots, x_n)$ is a smooth function of n vectors $x_i \in \mathbb{R}^p$ and $\mathbb{E}$ and $\mathbb{E}^*$ may represent two arbitrary measures. The bootstrap is treated as a special case where $\mathbb{E}^*$ is the conditional expectation given $\mathbf{X}$ under $\mathbb{E}$.

To make a connection between quantities of the form $\mathbb{E}h_t(b_n T_n)$, which is Lipschitz smooth in X_i at the best, and $\mathbb{E}f(X_1, \dots, X_n)$, which is required to be more smooth in our analysis, we approximate the maximum function $T_n = \max_j \sum_{i=1}^n X_{i,j}/\sqrt{n}$ of $\{X_i\}$ by the smooth max function $F_\beta(Z_n)$ as in Chernozhukov, Chetverikov and Kato (2013), where $Z_n = (X_1 + \cdots + X_n)/n^{1/2}$ and

$$F_\beta(z) = \frac{1}{\beta}\log\left(\sum_{j=1}^p e^{\beta z_j}\right), \quad \forall z = (z_1, \dots, z_p)^T. \tag{31}$$

For $\beta > 0$, the function $F_\beta(z)$ is infinitely differentiable and

$$\max(z_1, \dots, z_p) \le F_\beta(z) \le \max(z_1, \dots, z_p) + \beta^{-1}\log p.$$

It follows that (cf. Proof of Theorem 5 in the Supplementary Material, Deng and Zhang (2020)) for $\beta_n = 2b_n \log p$,

$$\eta_n(1/b_n) \le \sup_{t\in\mathbb{R}}\left|\mathbb{E}h_t\left(2b_n F_{\beta_n}(Z_n)\right) - \mathbb{E}h_t\left(2b_n F_{\beta_n}(Z_n^*)\right)\right|, \tag{32}$$

where $Z_n^* = (X_1^* + \cdots + X_n^*)/n^{1/2}$. In the Supplementary Material, we provide upper bounds for the derivatives of $F_\beta(z)$ and $f = h \circ (b_n F_\beta)$ via the Faa di Bruno formula.

We shall put $\mathbf{X}$ and $\mathbf{X}^*$ in the same probability space to better present our analysis. For this purpose, we use slightly different notation between the general and bootstrap cases. In the general case where both $\mathbb{E}$ and $\mathbb{E}^*$ are treated as deterministic, the problem does not involve the joint distribution between $\{X_i\}$ and $\{X_i^*\}$. This allows us to assume without loss of generality that $(X_i, X_i^*) \in \mathbb{R}^{p\times 2}$, $1 \le i \le n$, are independent matrices under $\mathbb{E}$, so that the problem concerns

$$\Delta_n(f) = \mathbb{E}\left\{f(X_1, \dots, X_n) - f\left(X_1^*, \dots, X_n^*\right)\right\}.$$

In the bootstrap case, $\mathbb{E}^*$ is the conditional expectation given $\mathbf{X}$ and we consider

$$\begin{aligned}\Delta_n^*(f) =& \mathbb{E}^*\left\{f\left(X_1^0, \dots, X_n^0\right) - f\left(X_1^*, \dots, X_n^*\right)\right\} \\ =& \mathbb{E}f(X_1, \dots, X_n) - \mathbb{E}^* f\left(X_1^*, \dots, X_n^*\right),\end{aligned} \tag{33}$$

where $\mathbf{X}^0 = (X_1^0, \dots, X_n^0)^T$ is an independent copy of $\mathbf{X}$. As (X_i^0, X_i^*) are still independent random matrices under $\mathbb{E}^*$, we can conveniently write the mean squared approximation error as

$$\mathbb{E}\left[\mathbb{E}^*\left\{f\left(X_1^0, \dots, X_n^0\right) - f\left(X_1^*, \dots, X_n^*\right)\right\}\right]^2.$$

In either cases, we assume throughout this section that $\mathbb{E}X_i = \mathbb{E}^* X_i^* = 0$, so that the average centered moments are

$$\mu^{(m)} = \frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i^{\otimes m}, \qquad \nu^{(m)} = \frac{1}{n}\sum_{i=1}^n \mathbb{E}^*\left(X_i^*\right)^{\otimes m}. \tag{34}$$

We consider in separate sections the Lindeberg method and comparison bounds for two general measures, the maxima, the empirical bootstrap and the wild bootstrap.

3.1. *A coherent Lindeberg interpolation.* Let $(X_i, X_i^*) \in \mathbb{R}^{p\times 2}$ be independent random matrices under $\mathbb{E}$, $\mathbf{U}_i = (X_1, \dots, X_{i-1}, 0, X_{i+1}^*, \dots, X_n^*)$, and $\mathbf{V}_i = (X_1, \dots, X_i, X_{i+1}^*, \dots, X_n^*)$. The original Lindeberg (1922) proof of the central limit theorem begins with the decomposition

$$\Delta_n(f) = \mathbb{E}\{f(\mathbf{V}_n) - f(\mathbf{V}_0)\} = \sum_{i=1}^n \mathbb{E}\{f(\mathbf{V}_i) - f(\mathbf{V}_{i-1})\},$$

followed by a Taylor expansion of the increments $f(\mathbf{V}_i) - f(\mathbf{V}_{i-1})$ at $\mathbf{U}_i$, so that

$$\begin{aligned} \Delta_n(f) &= \sum_{m=1}^{m^*-1} \Delta_{n,m} + \mathrm{Rem}, \\ \Delta_{n,m} &= \frac{1}{m!}\sum_{i=1}^n \langle \mathbb{E} f_i^{(m)}(\mathbf{U}_i), \mathbb{E}X_i^{\otimes m} - \mathbb{E}(X_i^*)^{\otimes m}\rangle, \end{aligned} \tag{35}$$

where $f_i^{(m)}(x_1, \dots, x_n) = (\partial/\partial x_i)^{\otimes m} f(x_1, \dots, x_n)$. To prove the central limit theorem, Lindeberg (1922) took $m^* = 3$ and Gaussian X_i^* with the same first two moments as X_i, so that $\Delta_n(f) = \mathrm{Rem}$. In this approach, $f(\mathbf{V}_i)$ can be viewed as an interpolation between $f(\mathbf{V}_n) = f(\mathbf{X})$ and $f(\mathbf{V}_0) = f(\mathbf{X}^*)$. The ideal has found much broader applications recently; see, for example, Chatterjee (2006). However, the decomposition (35) may not yield the best bounds for $\Delta_n(f)$ when $\mathbb{E}X_i^{\otimes m} - \mathbb{E}(X_i^*)^{\otimes m}$ are heterogeneous, for example, in the case of the empirical bootstrap with heteroscedastic X_i.

We further develop the Lindeberg approach (35) as follows to bound the quantity $\Delta_n(f)$ in terms of the difference of the average moments of $\{X_i\}$ and $\{X_i^*\}$,

$$\frac{1}{n}\sum_{i=1}^n \mathbb{E}X_i^{\otimes m} - \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i^*)^{\otimes m}, \tag{36}$$

instead of the difference in the moments of individual X_i and X_i^* as in a direct application of (35). This improvement, which can be viewed as a "coherent" Lindeberg interpolation and facilitates our analyses of the bootstrap for the maxima of the sums of X_i, is achieved by taking the average of the Lindeberg interpolation over all permutations of the index i.

Consider permutation invariant functions $f(x_1, \dots, x_n)$ of $x_i \in \mathbb{R}^p$, $1 \le i \le n$, satisfying

$$f(x_1, \dots, x_n) = f(x_{\sigma_1}, \dots, x_{\sigma_n})$$

for all permutations $\sigma = (\sigma_1, \dots, \sigma_n)$ of $\{1, \dots, n\}$. While $\Delta_n(f)$ of (35) is invariant in the permutation σ, the individuals components $\Delta_{n,m}$ and the remainder term on the right-hand side are not. Thus, the worst scenario bounds for $|\Delta_{n,m}|$ and $|\mathrm{Rem}|$ may not yield optimal results compared with the coherent Lindeberg interpolation, which we formally describe as follows.

Suppose $\mathbb{E}X_i = \mathbb{E}X_i^* = 0$. For permutations $\sigma = (\sigma_1, \dots, \sigma_n)$ of $\{1, \dots, n\}$, let

$$\mathbf{U}_{\sigma,k} = (X_{\sigma_1}, \dots, X_{\sigma_{k-1}}, X_{\sigma_{k+1}}^*, \dots, X_{\sigma_n}^*).$$

As $\Delta_n(f)$ invariant under permutation of the index i, for each permutation σ (35) yields

$$\Delta_n(f) = \sum_{m=2}^{m^*-1} \Delta_{n,m,\sigma} + \mathrm{Rem}_\sigma,$$

with $\Delta_{n,m,\sigma} = (m!)^{-1}\sum_{k=1}^{n}\langle \mathbb{E} f_{\sigma_k}^{(m)}(\mathbf{U}_{\sigma,k}, 0), \mathbb{E}X_{\sigma_k}^{\otimes m} - \mathbb{E}(X_{\sigma_k}^*)^{\otimes m}\rangle$. This leads to the expansion

$$\begin{aligned}\Delta_n(f) &= \mathbb{E}\{f(X_1,\ldots,X_n) - f(X_1^*,\ldots,X_n^*)\}\\ &= \sum_{m=2}^{m^*-1}\mathbb{A}_\sigma(\Delta_{n,m,\sigma}) + \mathbb{A}_\sigma(\mathrm{Rem}_\sigma),\end{aligned}\tag{37}$$

where $\mathbb{A}_\sigma$ is the operator of averaging over all permutations σ of $\{1,\ldots,n\}$. The expansion in (37) can be viewed as a coherent version of the original one in (35) as the fluctuation with respect to the choice of σ is removed by taking average over all permutations. The following lemma will be used to approximate $\mathbb{A}_\sigma(\Delta_{n,m,\sigma})$ and $\mathbb{A}_\sigma(\mathrm{Rem}_\sigma)$ by quantities of the same form with the difference of the average moments (36) in place of $\mathbb{E}X_i^{\otimes m} - \mathbb{E}(X_i^*)^{\otimes m}$. Define

$$\zeta_{k,i} = \delta_k X_i + (1-\delta_k)X_i^*,$$

where $\{\delta_k\}$ are Bernoulli variables independent of $\{X_i, X_i^*, i \le n\}$ under $\mathbb{E}$ with $\mathbb{P}\{\delta_k = 1\} = k/(n+1)$. Let $\mathbb{A}_{\sigma,k}$ be the operator of taking the average over all permutations σ and all $k = 1,\ldots,n$ and the expectation with respect to δ_k, conditionally on $\{X_i, X_i^*, i \le n\}$,

$$\mathbb{A}_{\sigma,k}h(\sigma,k,\zeta_{k,\sigma_k},\mathbf{X},\mathbf{X}^*) = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{n!}\sum_{\sigma}\left\{\frac{kh(\cdot, X_{\sigma_k})}{n+1} + \frac{(n+1-k)h(\cdot, X_{\sigma_k}^*)}{n+1}\right\},\tag{38}$$

for all Borel functions h, where $h(\cdot,\xi) = h(\sigma,k,\mathbf{X}_{(\sigma)},\mathbf{X}_{(\sigma)}^*,\xi)$ and $\mathbf{X}_{(\sigma)}$ is the permutation over rows of $\mathbf{X}$.

LEMMA 2. *For all permutation invariant functions $f(x_1,\ldots,x_n)$,*

$$\mathbb{A}_{\sigma,k}\big(I_{\{\sigma_k=i\}}f(\mathbf{U}_{\sigma,k},\zeta_{k,i})\big)$$

does not depend on i. Consequently, for any function $g_i(\cdot,\cdot)$, $1 \le i \le n$,

$$\mathbb{A}_{\sigma,k}\langle f(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k}), g_{\sigma_k}(\mathbf{X},\mathbf{X}^*)\rangle = \left\langle \mathbb{A}_{\sigma,k}\big(f(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k})\big), \frac{1}{n}\sum_{i=1}^{n}g_i(\mathbf{X},\mathbf{X}^*)\right\rangle.$$

We consider smooth functions with slightly stronger permutation invariance properties. Suppose that for certain permutation invariant functions $f^{(m,0)}(x_1,\ldots,x_n)$,

$$f_n^{(m)}(x_1,\ldots,x_{n-1},0) = f^{(m,0)}(x_1,\ldots,x_{n-1},0), \quad m = 0,2,\ldots,m^*-1,\tag{39}$$

where $f_n^{(m)}(x_1,\ldots,x_n) = (\partial/\partial x_n)^{\otimes m} f(x_1,\ldots,x_n)$ is as in (35). Such $f^{(m,0)}$ exist if $f(x_1,\ldots,x_n) = f_0(x_1,\ldots,x_n,0)$ for a permutation invariant function $f_0(x_1,\ldots,x_n,x_{n+1})$ involving $n+1$ vectors, for example, a function of the sum $x_1+\cdots+x_n$. In this case, we may pick

$$f^{(m,0)}(x_1,\ldots,x_n) = (\partial/\partial x_{n+1})^{\otimes m} f_0(x_1,\ldots,x_n,x_{n+1})|_{x_{n+1}=0}.$$

It follows from (37), Lemma 2 and (39) that

$$\begin{aligned}&\mathbb{A}_\sigma(\Delta_{n,m,\sigma})\\ &\quad = n\mathbb{A}_{\sigma,k}\big((m!)^{-1}\langle \mathbb{E}f^{(m,0)}(\mathbf{U}_{\sigma,k},0), \mathbb{E}X_{\sigma_k}^{\otimes m} - \mathbb{E}(X_{\sigma_k}^*)^{\otimes m}\rangle\big)\\ &\quad \approx n\mathbb{A}_{\sigma,k}\big((m!)^{-1}\langle \mathbb{E}f^{(m,0)}(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k}), \mathbb{E}X_{\sigma_k}^{\otimes m} - \mathbb{E}(X_{\sigma_k}^*)^{\otimes m}\rangle\big)\\ &\quad = \left\langle \frac{n}{m!}\mathbb{A}_{\sigma,k}\big(\mathbb{E}f^{(m,0)}(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k})\big), \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}X_i^{\otimes m} - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i^*)^{\otimes m}\right\rangle,\end{aligned}\tag{40}$$

so that $\mathbb{A}_\sigma(\Delta_{n,m,\sigma})$ is small when the average moments between $\{X_i\}$ and $\{X_i^*\}$ are close to each other. Interestingly, a combination of Slepian's (1962) smart path interpolation and Stein's (1981) leave-one-out method also allows comparison of the average of the second moment, but not the third moment and beyond. The Edgeworth expansion, a classical tool for high-order analysis of the bootstrap, is not available in our analysis as we are interested in the regime where the Gaussian approximation may fail to begin with.

3.2. *A general comparison theorem.* In this subsection, we present upper bounds for the absolute value of $\Delta_n(f)$ in (37) for smooth permutation invariant functions $f(x_1,\dots,x_n)$ in a general setting, where $(X_i, X_i^*) \in \mathbb{R}^{p\times 2}$, $1 \le i \le n$, are assumed to be independent random matrices under $\mathbb{E}$. Conditions up to the m^*th moment will be imposed, for example, $m^* = 4$ in (37).

In addition to invariance condition (39), we assume the following stability condition on derivatives of order m^*. For integers $m_1 \ge 2$ and $m_2 \ge 0$ with $m_1 + m_2 \le m^*$, define

$$f^{(m_1,m_2)}(x_1,\dots,x_{n-1},x_n) = \big((\partial/\partial x_n)^{\otimes m_2}\big) \otimes f^{(m_1,0)}(x_1,\dots,x_{n-1},x_n).$$

Here, $((\partial/\partial x_n)^{\otimes m_2}) \otimes f^{(m_1,0)}$, a product of two tensors, is treated as an $m = m_1 + m_2$ dimensional tensor with elements $(\frac{\partial}{\partial x_{n,j_1}})\cdots(\frac{\partial}{\partial x_{n,j_{m_2}}}) f^{(m_1,0)}_{j_{m_2+1},\dots,j_{m_2+m_1}}$. Suppose that for $m_1 \ge 2$ and $m_2 = m^* - m_1$, for example, $(m_1, m_2) = (2, 2)$ or $(3, 1)$ for $m^* = 4$,

$$\mathbb{P}\left\{\begin{aligned} &\big|f^{(m_1,m_2)}_{j_1,\dots,j_{m^*}}(x_1,\dots,x_{n-1},t\xi_i)\big| \le g\left(\frac{\|\xi_i\|}{u_n}\right)\bar f^{(m^*)}_{j_1,\dots,j_{m^*}}(x_1,\dots,x_{n-1},0),\\ &\big|f^{(m^*)}_{j_1,\dots,j_{m^*}}(x_1,\dots,x_{n-1},t\xi_i)\big| \le g\left(\frac{\|\xi_i\|}{u_n}\right)\bar f^{(m^*)}_{j_1,\dots,j_{m^*}}(x_1,\dots,x_{n-1},0)\end{aligned}\right\} = 1 \tag{41}$$

for all $0 \le t \le 1$ and $1 \le i \le n$, where ξ_i is either X_i or X_i^*. Suppose further that for some permutation invariant $f^{(m^*)}_{\max}(x_1,\dots,x_n)$,

$$\mathbb{P}\left\{\bar f^{(m^*)}_{j_1,\dots,j_{m^*}}(x_1,\dots,x_{n-1},0) \le g\left(\frac{\|\xi_i\|}{u_n}\right)\big(f^{(m^*)}_{\max}(x_1,\dots,x_{n-1},\xi_i)\big)_{j_1,\dots,j_{m^*}}\right\} = 1 \tag{42}$$

for the same ξ_i. Define $G_k = (\mathbb{E}\{1/g(\|X_k\|/u_n)\}) \wedge (\mathbb{E}\{1/g(\|X_k^*\|/u_n)\})$ and

$$\begin{aligned}\mu^{(m)}_{\max} = \Bigg(\Bigg[\max\Bigg\{&\sum_{k=1}^n \frac{\mathbb{E}|X_k|^m g(\frac{\|X_k\|}{u_n})}{nG_k}, \sum_{k=1}^n \frac{\mathbb{E}|X_k^*|^m g(\frac{\|X_k^*\|}{u_n})}{nG_k},\\ &\sum_{k=1}^n \frac{\mathbb{E}|X_k|^m \mathbb{E}g(\frac{\|X_k^*\|}{u_n})}{nG_k}, \sum_{k=1}^n \frac{\mathbb{E}|X_k^*|^m \mathbb{E}g(\frac{\|X_k\|}{u_n})}{nG_k}\Bigg\}\Bigg]^{1/m}\Bigg)^{\otimes m}.\end{aligned} \tag{43}$$

When $g(t)$ is increasing in t and $\mathbb{P}\{\max_{1\le i\le n}(\|X_i\| \vee \|X_i^*\|) \le cu_n\} = 1$ for a constant c,

$$\mu^{(m)}_{\max} \le g^2(c)\left(\left(\max\left\{\sum_{k=1}^n \frac{\mathbb{E}|X_k|^m}{n}, \sum_{k=1}^n \frac{\mathbb{E}|X_k^*|^m}{n}\right\}\right)^{1/m}\right)^{\otimes m}.$$

Let $\mathbf{U}_{\sigma,k}$ and $\zeta_{k,i}$ be as in Lemma 2 and define

$$F^{(m)} = \frac{n}{m!}\mathbb{A}_{\sigma,k}\big(\mathbb{E}f^{(m,0)}(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k})\big) = \sum_{k=1}^n \frac{1}{m!n!}\sum_{i=1}^n \sum_{\sigma,\sigma_k=i} \mathbb{E}f^{(m,0)}(\mathbf{U}_{\sigma,k},\zeta_{k,i}),$$

where $\mathbb{A}_{\sigma,k}$ is the operator defined in (38). Similarly, define

$$F^{(m)}_{\max} = \frac{n}{m!}\mathbb{A}_{\sigma,k}\big(\mathbb{E}f^{(m)}_{\max}(\mathbf{U}_{\sigma,k},\zeta_{k,\sigma_k})\big) = \sum_{k=1}^n \frac{1}{m!n!}\sum_{i=1}^n \sum_{\sigma,\sigma_k=i} \mathbb{E}f^{(m)}_{\max}(\mathbf{U}_{\sigma,k},\zeta_{k,i}).$$

THEOREM 4. *Let* $(X_i, X_i^*) \in \mathbb{R}^{p\times 2}$, $1 \le i \le n$, *be independent random matrices under expectation* $\mathbb{E}$. *Let* $m^* \in \{3, 4\}$. *Suppose* (41) *and* (42) *hold. Then*

$$\mathbb{E} f(X_1, \dots, X_n) - \mathbb{E} f\big(X_1^*, \dots, X_n^*\big) = \sum_{m=2}^{m^*-1} \big\langle F^{(m)}, \mu^{(m)} - \nu^{(m)}\big\rangle + \mathrm{Rem},$$

where $\mu^{(m)} = n^{-1}\sum_{i=1}^n \mathbb{E}X_i^{\otimes m}$ *and* $\nu^{(m)} = n^{-1}\sum_{i=1}^n \mathbb{E}(X_i^*)^{\otimes m}$ *as in* (34), *and*

$$|\,\mathrm{Rem}\,| \le \left\{2 + 4\sum_{m=2}^{m^*-1}\binom{m^*}{m}\right\}\big\langle F_{\max}^{(m^*)}, \mu_{\max}^{(m^*)}\big\rangle.$$

We may apply Theorem 4 directly to $\{X_i\}$ and $\{X_i^*\}$ or their truncated versions as we will show in Theorems 5 and 6 in the next two subsections.

In Theorem 4, the difference between the left- and right-hand sides of (40) is absorbed in the remainder term, which itself is expressed in terms of the average of moment-like quantities in (43), under conditions (41) and (42).

3.3. *Comparison theorem for the maxima of sums.* As in (1) and (8), let

$$T_n = \left\|\sum_{i=1}^n X_i/\sqrt{n}\right\|_\infty, \qquad T_n^* = \left\|\sum_{i=1}^n X_i^*/\sqrt{n}\right\|_\infty.$$

For random matrices $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \dots, \widetilde{X}_n)$ and $\widetilde{\mathbf{X}}^* = (\widetilde{X}_1^*, \dots, \widetilde{X}_n^*)$ and $b_n > 0$, define

$$\Omega_0 = \left\{\left\|\sum_{i=1}^n \frac{X_i - \widetilde{X}_i}{n^{1/2}}\right\|_\infty > \frac{1}{4b_n}\right\}, \qquad \Omega_0^* = \left\{\left\|\sum_{i=1}^n \frac{X_i^* - \widetilde{X}_i^*}{n^{1/2}}\right\|_\infty > \frac{1}{4b_n}\right\}. \tag{44}$$

THEOREM 5. *Let* $(X_i, X_i^*) \in \mathbb{R}^{p\times 2}$, $1 \le i \le n$, *be independent random matrices under expectation* $\mathbb{E}$, $m^* \in \{3, 4\}$, $\eta_n(\epsilon)$ *be the Lévy–Prokhorov pre-distance in* (28), *and* $u_n = \sqrt{n}/(2b_n \log p)$.

(*i*) *Let* $\mu_{\max}^{(m)}$ *be given in* (43) *with* $g(t) = e^{2m^* t}$. *Then*

$$\begin{aligned}\eta_n(1/b_n) \le C_{m^*}\Bigg(&\sum_{m=2}^{m^*-1} \frac{b_n^m (\log p)^{m-1}}{n^{m/2-1}}\big\|\mu^{(m)} - \nu^{(m)}\big\|_{\max}\\ &+ \frac{b_n^{m^*}(\log p)^{m^*-1}}{n^{m^*/2-1}}\big\|\mu_{\max}^{(m^*)}\big\|_{\max}\Bigg),\end{aligned} \tag{45}$$

where $\mu^{(m)}$ *and* $\nu^{(m)}$ *are as in Theorem* 4.

(*ii*) *Let* $\widetilde{\mathbf{X}} = (\widetilde{X}_1, \dots, \widetilde{X}_n)$ *and* $\widetilde{\mathbf{X}}^* = (\widetilde{X}_1^*, \dots, \widetilde{X}_n^*)$. *Suppose that* $(\widetilde{X}_i, \widetilde{X}_i^*)$ *are independent matrices under* $\mathbb{P}$, $\mathbb{E}\widetilde{X}_i = \mathbb{E}\widetilde{X}_i^* = 0$, *and* $\mathbb{P}\{\|\widetilde{\mathbf{X}}\|_{\max} \vee \|\widetilde{\mathbf{X}}^*\|_{\max} \le c_1 u_n\} = 1$ *for a constant* c_1. *Then*

$$\begin{aligned}\eta_n(1/b_n) \le{}& C_{m^*,c_1}\sum_{m=2}^{m^*} \frac{b_n^m(\log p)^{m-1}}{n^{m/2-1}}\big\|\widetilde{\mu}^{(m)} - \widetilde{\nu}^{(m)}\big\|_{\max}\\ &+ C_{m^*,c_1}\frac{b_n^{m^*}(\log p)^{m^*-1}}{n^{m^*/2-1}}\big\|\widetilde{\mu}^{(m^*)}\big\|_{\max} + \mathbb{P}\{\Omega_0\} + \mathbb{P}\{\Omega_0^*\},\end{aligned} \tag{46}$$

where $\widetilde{\mu}^{(m)} = n^{-1}\sum_{i=1}^n \mathbb{E}\widetilde{X}_i^{\otimes m}$, $\widetilde{\nu}^{(m)} = n^{-1}\sum_{i=1}^n \mathbb{E}(\widetilde{X}_i^*)^{\otimes m}$, *and* Ω_0 *and* Ω_0^* *are as in* (44).

We may consider $\widetilde{\mathbf{X}} = (\widetilde{X}_{i,j})_{n\times p} = (\widetilde{X}_1, \ldots, \widetilde{X}_n)$ as a truncated version of $\mathbf{X}$ given by

$$\widetilde{X}_{i,j} = X_{i,j} I_{\{|X_{i,j}|\le a_n\}} - \mathbb{E} X_{i,j} I_{\{|X_{i,j}|\le a_n\}}. \tag{47}$$

In this case, the following lemma can be used to bound $\mathbb{P}\{\Omega_0\}$.

LEMMA 3. *Let M_m be as in* (5) *with $m > 2$, $\widetilde{\mathbf{X}}$ as in* (47) *with a_n satisfying $M_m\{n/\log(p/\epsilon_n)\}^{1/m} \le a_n \le \tilde{a}_n = \{c_1 n^{1/2}/(b_n \log(p/\epsilon_n))\}$ with $c_1 > 0$, and Ω_0 as in* (44). *Then, for sufficiently large constant C_{m,c_1}, it implies by $b_n^m(\log(p/\epsilon_n))^{m-1} M_m^m/n^{m/2-1} \le 1/C_{m,c_1}$ that*

$$\mathbb{P}\{\Omega_0\} \le \epsilon_n + \mathbb{P}\{\widetilde{\Omega}_0\} \le \epsilon_n + C_{m,c_1} \frac{b_n^m(\log(p/\epsilon_n))^{m-1}}{n^{m/2-1}} \mathfrak{M}_{m,2}^m, \tag{48}$$

where $\widetilde{\Omega}_0 = \{\max_{1\le j\le p} |n^{-1/2}\sum_{i=1}^n X_{i,j} I_{\{|X_{i,j}|>\tilde{a}_n\}}| > 1/(8b_n)\}$ and $\mathfrak{M}_{m,2}$ is as in (6).

We note that the upper bound for a_n is no smaller than the lower bound due to the condition

$$b_n^m\big(\log(p/\epsilon_n)\big)^{m-1} M_m^m/n^{m/2-1} \le 1/C_{m,c_1}.$$

3.4. *Efron's empirical bootstrap.* We have already obtained upper bounds for the Lévy–Prokhorov predistance (28) in terms of the average moments of X_i and X_i^* in Theorem 5. In bootstrap, the Lévy–Prokhorov predistance is a random variable due to the involvement of $\mathbb{P}^*$,

$$\eta_n^*(\epsilon) \equiv \eta_n^{(\mathbb{P}^*)}(\epsilon; T_n^0, T_n^*) = \sup_{t\in\mathbb{R}} \eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*), \tag{49}$$

where $\eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*) = \max[\mathbb{P}^*\{T_n^0 \le t-\epsilon\} - \mathbb{P}^*\{T_n^* < t\}, \mathbb{P}^*\{T_n^* \le t-\epsilon\} - \mathbb{P}^*\{T_n^0 < t\}, 0]$ as in Lemma 1, and T_n^* is the bootstrapped T_n. Recall that $\mathbb{P}^*\{T_n^0 \le t\} = \mathbb{P}\{T_n \le t\}$ as T_n^0 is an independent copy of T_n. In this subsection, we derive more explicit bounds for $\eta_n^*(\epsilon)$ in terms of the average moments of $\{X_i\}$ for Efron's empirical bootstrap.

For the empirical bootstrap, the difference of the average moments between X_i and X_i^* is

$$\begin{aligned}\nu^{(m)} - \mu^{(m)} &= \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^{\otimes m} - \frac{1}{n}\sum_{i=1}^n \mathbb{E} X_i^{\otimes m} \\ &= \frac{1}{n}\sum_{i=1}^n (X_i^{\otimes m} - \mu^{(m)}) + \sum_{k=1}^m \binom{m}{k} \mathrm{Sym}\Bigg((-\overline{X})^{\otimes k} \sum_{i=1}^n \frac{X_i^{\otimes(m-k)}}{n}\Bigg),\end{aligned}$$

where $\nu^{(m)}$ and $\mu^{(m)}$ are as in (34) with the assumption $\mu^{(1)} = 0$ and $\mathrm{Sym}(A)$ denotes the symmetrization of tensor A by taking the average over all permutations of the index of its elements. It can be seen from the above expression that the quantities $\|\mu^{(m)} - \nu^{(m)}\|_{\max}$ in the right-hand side of (45), and $\|\mu^{(4)}_{\max}\|_{\max}$ as well, can be bounded by empirical process methods. However, as high moments are involved, some level of truncation may still be needed to obtain sharp results when $\|\mathbf{X}\|_{\max}$ is unbounded. Therefore, a direct application of the error bound (46) with truncation is natural. This approach is taken here.

THEOREM 6. *Let $X_i \in \mathbb{R}^p$ be independent random vectors and X_i^* generated by the empirical bootstrap. Let $b_n > 0$, M_4 be as in* (5), *$\{c_1, c_2\}$ be fixed positive constants, and $\tilde{a}_n = c_1\sqrt{n}/(b_n \log(p/\epsilon_n))$. Suppose $\log(p/\epsilon_n) \le c_2 n$. Then*

$$\eta_n^*(1/b_n) \le C_{c_1,c_2} b_n^2\big(\log(p/\epsilon_n)\big)^{3/2} M_4^2/n^{1/2} + 2\epsilon_n + \mathbb{P}\big\{\|\mathbf{X}\|_{\max} > \tilde{a}_n\big\} \tag{50}$$

with at least probability $1 - (\mathbb{P}\{\|\mathbf{X}\|_{\max} > \tilde{a}_n\} + 2\epsilon_n)$, and with $\mathfrak{M}_4$ as in (5),

$$\mathbb{P}\big\{\eta_n^*(1/b_n) > C_{c_1,c_2}\big(\epsilon_n + b_n^4\big(\log(p/\epsilon_n)\big)^3 \mathfrak{M}_4^4/(\epsilon_n n)\big)\big\} \le \epsilon_n. \tag{51}$$

3.5. *Wild bootstrap*. Let $\{W_i\}$ be a sequence of i.i.d. variables independent of $\mathbf{X}$ and satisfying $\mathbb{E}W_i = 0$ and $\mathbb{E}W_i^2 = 1$. The wild bootstrap (Liu (1988), Wu (1986), Mammen (1993)) is defined in (13). Recall that we assume $\mathbb{E}X_i = 0$ without loss of generality in our analysis. As $\|\sum_{i=1}^n W_i\overline{X}/\sqrt{n}\|_\infty = O_P(1)\|\overline{X}\|_\infty$ is typically negligible in the analysis of the maxima of the sum of X_i^* under mild conditions, for simplicity we may study

$$X_i^* = W_i X_i. \tag{52}$$

Suppose the moments of individual X_i^* matches that of X_i under the joint expectation $\mathbb{E}$,

$$\mathbb{E}X_i^{\otimes m} = \mathbb{E}(W_i X_i)^{\otimes m}, \quad m = 1, \dots, m^* - 1, \tag{53}$$

where m^* represents the highest order of expansion involved in the comparison theorem. Condition (53) holds for $m^* = 4$ when $\mathbb{E}W_i^3 = 1$ (Liu (1988), Mammen (1993)), and all m^* for the Rademacher wild bootstrap when $\mathbb{E}X_i^{\otimes m} = 0$ for all positive odd m smaller than m^*,

$$\begin{aligned} &\{\mathbb{E}W_i^3 = 1 \text{ and } m^* = 4\} \quad \text{or} \\ &\{\mathbb{E}W_i^4 = 1 \text{ and } \mathbb{E}X_i^{\otimes m} = 0\ \forall \text{ odd } m \in [1, m^*)\}. \end{aligned} \tag{54}$$

We note that (53) always holds for $m^* = 3$ due to the default conditions $\mathbb{E}W_i = 0$ and $\mathbb{E}W_i^2 = 1$. Under this moment condition and the sub-Gaussian condition (17) on W_i, a modification of the proof of Theorem 6 yields the following result.

THEOREM 7. *Let $X_i \in \mathbb{R}^p$ be independent random vectors and X_i^* generated by the wild bootstrap as in* (13). *Suppose* (53) *holds with $m^* \in \{3, 4\}$ and* (17) *holds with $\tau_0 < \infty$. Let M_{m^*} and $\mathfrak{M}_{m^*,2}$ be as in* (5) *and* (6), *respectively. Let $b_n > 0$, $\epsilon_n \le \overline{\epsilon}_n$ and $\tilde{a}_n = c_1\sqrt{n}/\{(b_n(\log(p/\overline{\epsilon}_n))^{1/2}(\log(p/\epsilon_n))^{1/2}\}$. Suppose $\log p \le c_2 n$ with a constant $c_2 > 0$ and $M = M_{m^*}(n/\log(p/\overline{\epsilon}_n))^{1/m^*-1/4}$. Then, for a sufficiently large constant C_{m^*,τ_0,c_1,c_2},*

$$\begin{aligned} \eta_n^*(1/b_n) \le{}& C_{m^*,\tau_0,c_1,c_2} b_n^2 (\log(p/\overline{\epsilon}_n))^{1/2}(\log(p/\epsilon_n))/n^{1/2}M^2 + \overline{\epsilon}_n \\ &+ \mathbb{P}\left\{\max_{1\le j\le p}\left|\sum_{i=1}^n \frac{X_{i,j}I_{\{|X_{i,j}|>\tilde{a}_n\}}}{\sqrt{n}}\right| > 1/(8b_n)\right\} \end{aligned} \tag{55}$$

with at least probability $1 - (\mathbb{P}\{C_0\tau_0^2 b_n^2 \log(p/\overline{\epsilon}_n)\max_{1\le j\le p}\sum_{i=1}^n X_{i,j}^2 I_{\{|X_{i,j}|>\tilde{a}_n\}} > n\} + 2\epsilon_n)$ *and*

$$\eta_n^*(1/b_n) \le C'_{m^*,\tau_0,c_1,c_2}\left(\overline{\epsilon}_n + \frac{b_n^{m^*}(\log(p/\overline{\epsilon}_n))^{\frac{m^*}{2}-1}(\log(p/\epsilon_n))^{\frac{m^*}{2}}}{\epsilon_n \cdot n^{\frac{m^*}{2}-1}}\mathfrak{M}_{m^*,2}^{m^*}\right) \tag{56}$$

with at least probability $1 - \epsilon_n$.

While (55) is comparable with (50) in Theorem 6, (56) requires the weaker moment $\mathfrak{M}_{m^*,2}$ than the $\mathfrak{M}_{m^*}$ in (51).

In the rest of the subsection, we study the implication of a martingale structure in the original Lindeberg expansion (35) for wild bootstrap. This would lead to a comparison theory more useful for the high order $m^* > 4$. Let

$$\mathbf{U}_i^0 = (X_1^0, \dots, X_{i-1}^0, 0, X_{i+1}^*, \dots, X_n^*), \qquad \mathbf{V}_i^0 = (X_1^0, \dots, X_i^0, X_{i+1}^*, \dots, X_n^*),$$

where $\mathbf{X}^0 = (X_1^0, \dots, X_n^0)^T$ is an independent copy of $\mathbf{X}$. Let $f_i^{(m)} = (\partial/\partial x_i)^m f$ and $\Delta_n^*(f)$ be as in (33). The bootstrap version of the Lindeberg expansion (35) is

$$\Delta_n^*(f) = \sum_{m=2}^{m^*-1}\Delta_{n,m}^* + \text{Rem} \tag{57}$$

with $\Delta_{n,m}^* = (m!)^{-1}\sum_{i=1}^n \langle \mathbb{E}^* f_i^{(m)}(\mathbf{U}_i^0), \mathbb{E}^*(X_i^0)^{\otimes m} - \mathbb{E}^*(X_i^*)^{\otimes m}\rangle$.

Consider the case where X_i^* are defined as in (52). By (53), $\mathbb{E}\{\mathbb{E}^*(X_i^0)^{\otimes m} - \mathbb{E}^*(X_i^*)^{\otimes m}\} = 0$. As $\mathbb{E}^* f_i^{(m)}(\mathbf{U}_i^0)$ is a function of $(X_{i+1}, \dots, X_n)$, $\Delta_{n,m}^*$ is a sum of martingale differences. This directly leads to the comparison inequalities in Proposition 1 below. Consider functions f satisfying

$$\mathbb{P}\left\{\begin{array}{c} |f_i^{(m^*)}(x_1, \dots, x_{i-1}, t\xi_i, x_{i+1}, \dots, x_n)| \\ \le g\left(\dfrac{\|\xi_i\|}{u_n}\right) f_{\max}^{(m^*)}(x_1, \dots, x_{i-1}, 0, x_{i+1}, \dots, x_n) \end{array}\right\} = 1, \quad 0 \le t \le 1, \tag{58}$$

for $\xi_i = X_i$ or X_i^*, with real-valued $g(t)$ and m^*-tensor-valued $f_{\max}^{(m^*)}$. Let

$$s_{n,m,i}^2 = \langle (\mathbb{E}^* f_i^{(m)}(\mathbf{U}_i^0))^{\otimes 2}, \mathbb{E}(X_i^{\otimes m} - \mathbb{E}X_i^{\otimes m})^{\otimes 2}(\mathbb{E}W_i^m)^2 \rangle, \quad 2 \le m < m^*,$$

$$s_{n,m^*,i}^2 = \left\langle (\mathbb{E}^* f_{\max}^{(m^*)}(\mathbf{U}_i^0))^{\otimes 2}, \mathbb{E}\left[\mathbb{E}^* g\left(\frac{\|X_i^*\|}{u_n}\right)|X_i^*|^{\otimes m^*} - \mathbb{E}g\left(\frac{\|X_i^*\|}{u_n}\right)|X_i^*|^{\otimes m^*}\right]^{\otimes 2}\right\rangle,$$

$$\overline{\mathrm{Rem}} = \frac{1}{m^*!}\sum_{i=1}^n \left\langle \mathbb{E}^* f_{\max}^{(m^*)}(\mathbf{U}_i^0), \mathbb{E}g\left(\frac{\|X_i^0\|}{u_n}\right)|X_i^0|^{\otimes m^*} + \mathbb{E}g\left(\frac{\|X_i^*\|}{u_n}\right)|X_i^*|^{\otimes m^*}\right\rangle.$$

PROPOSITION 1. *Let X_i and X_i^* be as in (52) and $\Delta_n^*(f)$ as in (57). Suppose (53) and (58). Then*

$$\begin{aligned} \mathbb{E}|\Delta_n^*(f)| &\le \sum_{m=2}^{m^*-1} \frac{1}{m!}\left(\sum_{i=1}^n \mathbb{E}s_{n,m,i}^2\right)^{1/2} + \mathbb{E}(\overline{\mathrm{Rem}}), \\ (\mathbb{E}|\Delta_n^*(f)|^2)^{1/2} &\le \sum_{m=2}^{m^*} \frac{1}{m!}\left(\sum_{i=1}^n \mathbb{E}s_{n,m,i}^2\right)^{1/2} + (\mathbb{E}(\overline{\mathrm{Rem}})^2)^{1/2}. \end{aligned} \tag{59}$$

For Efron's empirical bootstrap,

$$\mathbb{E}^*(X_i^*)^{\otimes m} = n^{-1}\sum_{k=1}^n (X_k - \overline{X})^{\otimes m} \tag{60}$$

involves all data points, so that the martingale argument does not directly apply. An application of the martingale Bernstein inequality (Steiger (1969), Freedman (1975)) leads to the following theorem.

THEOREM 8. *Theorem 7 is still valid for general $m^* > 2$ when ϵ_n is defined by*

$$\epsilon_n = b_n^2(\log p)\{\log(1/\epsilon_n)/n\}^{1/2}M^2 + \kappa_{n,m^*}(\mathfrak{M}_{m^*,1}/M_{m^*})^{m^*} \tag{61}$$

provided that $M \ge \mathfrak{M}_{4,1}$ with the $\mathfrak{M}_{m,1}$ in (6).

Consider $m^* = 6$. When $M^6 \asymp \mathfrak{M}_{6,1}$ and $Mb_n \asymp \sqrt{\log p}$, the second term in (61) is of no greater order than $\{b_n^2(\log p)n^{-1/2}M^2\}^4$, so that by Theorem 8

$$(\log p)^4/n \to 0 \quad \Rightarrow \quad \epsilon_n \to 0.$$

In this case, taking $m^* > 6$ does not improve the order of ϵ_n in Theorem 8.

Next, we derive upper bounds for

$$\eta_n^{(q)}(\epsilon) = \sup_{t\in\mathbb{R}}[\mathbb{E}\{\eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*)\}^q]^{1/q} \tag{62}$$

with the $\eta_n^{(\mathbb{P}^*)}(\epsilon, t; T_n^0, T_n^*)$ in (49). The quantity $\eta_n^{(q)}(\epsilon)$ can be viewed as a weak Lévy–Prokhorov predistance, as the supreme is taken outside the expectation. However, this weak version of the Lévy–Prokhorov predistance is still stronger than the unconditional one. In fact, we have

$$\eta_n(\epsilon) \equiv \sup_{t\in\mathbb{R}} \eta_n^{(\mathbb{P})}(\epsilon, t; T_n, T_n^*) \le \eta_n^{(q)}(\epsilon) \le \big\|\eta_n^*(\epsilon)\big\|_{L_q(\mathbb{P})}, \quad q \ge 1,$$

where $\eta_n^{(\mathbb{P})}(\epsilon, t; T_n, T_n^*)$ is as in (28). See (49) and the discussion below (28).

In addition to the average moments M_m defined in (5), we use quantities

$$\begin{aligned} M_{m,1} &= \left\| \sum_{i=1}^n \frac{\mathbb{E}\exp(2m\|W_i X_i\|_\infty/u_n)|W_i X_i|^m}{n\mathbb{E}\exp(-2m\|X_i\|_\infty/u_n)} \right\|_\infty^{1/m}, \\ M_{m,2} &= \left\| \frac{1}{n} \sum_{i=1}^n \frac{\mathbb{E}|W_i X_i|^m}{\mathbb{E}\exp(-2m\|X_i\|_\infty/u_n)} \right\|_\infty^{1/m}, \end{aligned} \tag{63}$$

to bound the $\eta_n^{(q)}(\epsilon)$ in (62). When $\mathbb{P}\{\|X_i\|_\infty \le a_n\} = 1$,

$$M_{m,1} \le e^{\frac{2a_n}{u_n}} \big(\mathbb{E}|W_1|^m \mathbb{E}e^{2m|W_1|\frac{a_n}{u_n}}\big)^{1/m} M_m, \qquad M_{m,2} \le e^{\frac{2a_n}{u_n}} \big(\mathbb{E}|W_1|^m\big)^{1/m} M_m.$$

In any case, controlling $M_{m,1}$ requires W_1 to have a finite moment generating function in the interval $[0, 2m^* a_n/u_n]$.

THEOREM 9. *Let $a_n = c_1\sqrt{n}/(b_n \log p)$, $m^* \ge 3$ and $\eta_n^{(q)}(\cdot)$ be as in* (62).

(*i*) *Let X_i^* be as in* (52). *Suppose* (53) *holds. Let $b_n > 0$ and $u_n = \sqrt{n}/(2b_n \log p)$ in* (63). *Then*

$$\eta_n^{(1)}(1/b_n) \le C_{m^*}\left(\sum_{m=2}^{m^*-1} \big|\mathbb{E}W_1^m\big| \frac{b_n^m (\log p)^{m-1}}{n^{m/2-1/2}} M_{2m,2}^m + \frac{b_n^{m^*}(\log p)^{m^*-1}}{n^{m^*/2-1}} M_{m^*,1}^{m^*}\right). \tag{64}$$

(*ii*) *Let X_i^* be as in* (13). *Suppose* (54) *and* (17) *hold. Then, for $1 \le q \le 2$,*

$$\begin{aligned} &\eta_n^{(q)}(1/b_n) \\ &\quad \le C_{m^*,\tau_0,c_1}\left(\frac{b_n^2 \log p}{n^{1/2}} M_4^2 + \kappa_{n,m^*}^{1/q}\right) \\ &\qquad + \left[\mathbb{E}\min\left\{2, C_{\tau_0}\frac{b_n^2 \log p}{n} \max_{1\le j\le p} \sum_{i=1}^n X_{i,j}^2 I_{\{|X_{i,j}|>a_n\}}\right\}\right]^{1/q} \\ &\quad \le C_{m^*,\tau_0,c_1}\left(\frac{b_n^2 \log p}{n^{1/2}} M_4^2 + \kappa_{n,m^*}^{1/q}\right) \\ &\qquad + \left[\mathbb{E}\min\left\{2, C_{m^*,\tau_0,c_1}\frac{b_n^{m^*}(\log p)^{m^*-1}}{n^{m^*/2}} \max_{1\le j\le p} \sum_{i=1}^n |X_{i,j}|^{m^*} I_{\{|X_{i,j}|>a_n\}}\right\}\right]^{1/q}, \end{aligned} \tag{65}$$

where $\kappa_{n,m} = b_n^m (\log p)^{m-1} n^{1-m/2} M_m^m$. Moreover,

$$\big(\mathbb{E}|\eta_n^*(1/b_n)|^q\big)^{1/q} \le (1+q)\big\{q^{-1}2^{1/q}\eta_n^{(q)}(1/b_n)\big\}^{q/(q+1)}. \tag{66}$$

Compared with the first term on the right-hand side of (50), the first term on the right-hand side of (65) is of smaller order by at least a factor $\sqrt{\log(p/\kappa_{n,4})}$.

The proof of Theorem 9, given in the Supplementary Material (Deng and Zhang (2020)), involves two issues. The first one is to relate the maxima T_n in (1) and T_n^* in (8) to smooth functions $f(x_1, \dots, x_n)$ in Proposition 1. This is done via the smooth max function in (31) as discussed at the beginning of this section. The second issue involves heterogeneity among X_i. When $\mathbb{P}\{\|\mathbf{X}\|_{\max} \le u_n\} = 1$, the quantities in (63) are bounded under the condition $M_{m^*} = O(1)$ on the average moments. However, a direct application of (59) requires the stronger condition

$$\frac{1}{n}\sum_{i=1}^{n} \max_{1\le j\le p} \mathbb{E}|X_{i,j}|^{m^*} = O(1)$$

as in Theorem 8. This issue is again resolved through Lemma 2.

4. Anticoncentration of the maxima. As we have discussed at the end of Section 2, the Kolmogorov–Smirnov distance between two distribution functions can be bounded from the above by the sum of the Lévy–Prokhorov predistance and the minimum of the Lévy concentration of the two distribution functions,

$$\sup_{t\in\mathbb{R}}\bigl|\mathbb{P}\{T_n \le t\} - \mathbb{P}\bigl\{T_n^* < t\bigr\}\bigr| \le \eta_n(\epsilon) + \min\bigl\{\omega_n(\epsilon; T_n), \omega_n\bigl(\epsilon; T_n^*\bigr)\bigr\} \tag{67}$$

as in (30). The above two terms are also required if one wants to use Lemma 1 to derive an upper bound for $|\mathbb{P}\{T_n \le t_\alpha^*\} - (1-\alpha)|$. As upper bounds for the Lévy–Prokhorov predistance $\eta_n(\epsilon)$ and its bootstrap version $\eta_n^*(\epsilon)$ have already been established in Section 3, the aim of this section is to develop anticoncentration inequalities to bound the Lévy concentration function $\omega_n(\epsilon; T_n)$ from the above. We note that once a comparison theorem becomes available as an upper bound for $\eta_n(\epsilon)$, an anticoncentration inequality for T_n can be established from one for T_n^*, as

$$\omega_n(\epsilon; T_n) \le \omega_n\bigl(\epsilon; T_n^*\bigr) + 2\sup_{t\in\mathbb{R}}\bigl|\mathbb{P}\{T_n \le t\} - \mathbb{P}\bigl\{T_n^* < t\bigr\}\bigr| \le 3\omega_n\bigl(\epsilon; T_n^*\bigr) + 2\eta_n(\epsilon) \tag{68}$$

by the triangle inequality and (67), and vice versa.

To study the consistency of the Gaussian wild bootstrap, say $T_n^{*,\mathrm{Gauss}}$ for the approximation of the distribution of T_n, the Kolmogorov–Smirnov distance of interest is bounded by

$$\begin{aligned}&\sup_{t\in\mathbb{R}}\bigl|\mathbb{P}\{T_n \le t\} - \mathbb{P}^*\bigl\{T_n^{*,\mathrm{Gauss}} < t\bigr\}\bigr| \\ &\quad \le \eta_n^*(\epsilon) + \min\bigl\{\omega_n^{(\mathbb{P})}(\epsilon; T_n), \omega_n^{(\mathbb{P}^*)}\bigl(\epsilon; T_n^{*,\mathrm{Gauss}}\bigr)\bigr\},\end{aligned}$$

where $\eta_n^*(\epsilon) = \eta_n^{(\mathbb{P}^*)}(\epsilon; T_n^0, T_n^{*,\mathrm{Gauss}})$ and $\omega_n^{(\mathbb{P})}(\epsilon; T_n)$ are as in (49) and (30) respectively. Thus, an anticoncentration inequality for the Gaussian maxima $T_n^{*,\mathrm{Gauss}}$ under $\mathbb{P}^*$ suffices (Chernozhukov, Chetverikov and Kato (2015)). This approach has been taken in Chernozhukov, Chetverikov and Kato (2013, 2017) among others. However, the inequality (68) with $T_n^* = T_n^{*,\mathrm{Gauss}}$, which requires a small Lévy–Prokhorov predistance $\eta_n(\epsilon) = \eta_n^{(\mathbb{P})}(\epsilon; T_n, T_n^{*,\mathrm{Gauss}})$, is not helpful in our study as we are interested in scenarios where the Gaussian approximation may not hold.

Our idea is to derive anticoncentration inequalities for the maxima T_n of sums of possibly skewed independent random vectors through a mixed wild bootstrap which has a Gaussian component and also provides the third moment match as Liu (1988) and Mammen (1993) advocated. Compared with the Gaussian wild bootstrap, such a mixed wild bootstrap enjoys both the anticoncentration properties of the Gaussian component through conditioning and sharper approximation of the distribution of T_n through the fourth-order comparison theorems developed in Section 3.

The multiplier of the above mixed wild, bootstrap can be defined as

$$W_i^{**} = a_0\delta_i Z_i + b_0(1-\delta_i)W_i^0, \tag{69}$$

where δ_i, Z_i, W_i^0, $i = 1, \dots, n$, are independent random variables, δ_i are Bernoulli variables with $\mathbb{P}\{\delta_i = 1\} = p_0 = 1 - \mathbb{P}\{\delta_i = 0\}$, $Z_i \sim N(0, 1)$, and W_i^0 can be taken as Mammen's bootstrap multiplier in (16). In this mixed wild bootstrap, a_0, b_0 and p_0 are positive constants satisfying

$$\begin{aligned} &0 < p_0 < 1, \quad \mathbb{E}(W_i^{**})^2 = a_0^2 p_0 + b_0^2(1-p_0) = 1, \\ &\mathbb{E}(W_i^{**})^3 = b_0^3(1-p_0) = 1. \end{aligned} \tag{70}$$

For any $p_0 \in (0, 1)$, the values of a_0 and b_0 are determined by

$$b_0 = (1-p_0)^{-1/3}, \qquad a_0 = \sqrt{p_0^{-1}\big(1-(1-p_0)^{1/3}\big)}.$$

For example, $a_0 = 0.6423387$ and $b_0 = 1.259921$ for $p_0 = 1/2$.

Suppose $\mathbb{E}X_i = 0$ as in Section 3. Given the multiplier (69) and the original data $X_i = (X_{i,1}, \dots, X_{i,p})^T$, $i = 1, \dots, n$, the mixed wild bootstrap for T_n is defined through

$$X_i^{**} = W_i^{**} X_i, \qquad Z_{n,j}^{**} = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{i,j}^{**}, \quad \text{and} \quad T_n^{**} = \max_{1 \le j \le p} Z_{n,j}^{**}. \tag{71}$$

We conveniently avoid the complication of subtracting the sample mean from X_i as the primary purpose of this mixed wild bootstrap is to provide a vehicle to derive anticoncentration inequalities for the maxima T_n for the original data. Once an upper bound for $\omega_n^{(\mathbb{P})}(\epsilon; T_n)$ is established, the consistency of the bootstrap can be studied through (67) and Lemma 1.

Let $\mathbb{P}^{**}$ be the conditional probability given $\{X_i, \delta_i, W_i^0, i = 1, \dots, n\}$. We find that under $\mathbb{P}^{**}$, Z_n^{**} is a Gaussian vector with individual mean and standard deviation

$$\begin{aligned} \mu_j^{**} &= \mathbb{E}^{**}(Z_{n,j}^{**}) = \frac{b_0}{\sqrt{n}} \sum_{i=1}^n (1-\delta_i) W_i^0 X_{i,j}, \\ \sigma_j^{**} &= \sqrt{\mathrm{Var}^{**}(Z_{n,j}^{**})} = \left(\frac{a_0^2}{n} \sum_{i=1}^n \delta_i X_{i,j}^2 \right)^{1/2}. \end{aligned} \tag{72}$$

Anticoncentration inequalities for T_n^{**} under the marginal probability $\mathbb{P}$ can be derived from the conditional one under $\mathbb{P}^{**}$ via

$$\omega_n^{(\mathbb{P})}(\epsilon; T_n^{**}) \le \mathbb{E}\big[\omega_n^{(\mathbb{P}^{**})}(\epsilon; T_n^{**})\big], \tag{73}$$

where $\omega_n^{(\mathbb{P}^{**})}(\epsilon; T_n^{**})$, a function of the random vector $(\mu_j^{**}, \sigma_j^{**}, 1 \le j \le p)$, is the Lévy concentration function of T_n^{**} under the conditional probability $\mathbb{P}^{**}$ as in (30).

In what follows, we present anticoncentration inequalities for the maxima of Gaussian vectors, sums in the mixed wild bootstrap and sums of general independent vectors with zero mean.

THEOREM 10. *Let $\xi = (\xi_1, \dots, \xi_p)^T$ be a multivariate Gaussian vector with marginal distributions $\xi_j \sim N(\mu_j, \sigma_j^2)$, $\sigma_{(1)} \le \dots \le \sigma_{(p)}$ be the ordered values of $\sigma_1, \dots, \sigma_p$. Then, for all $x_m \ge 1$,*

$$\sup_x \frac{d}{dx} \mathbb{P}\Big\{ \max_{1 \le j \le p} \xi_j \le x \Big\} \le \max_{1 \le m \le p} \left\{ \frac{x_m}{\sigma_{(m)}} + \sum_{k=1}^{m-1} \frac{\varphi(x_k - 1/x_k)}{\sigma_{(k)}} \right\}. \tag{74}$$

Consequently, with $\overline{\sigma} = (2+\sqrt{2\log p})/\{1/\sigma_{(1)} + \max_{1\le m\le p}(1+\sqrt{2\log m})/\sigma_{(m)}\} \ge \sigma_{(1)}$,

$$\mathbb{P}\Big\{a < \max_{1\le j\le p} \xi_j \le a+\epsilon\Big\} \le \frac{\epsilon}{\overline{\sigma}}\big(2+\sqrt{2\log p}\big), \quad \forall \epsilon > 0, a\in\mathbb{R}. \tag{75}$$

Given $\{\sigma_j\}$, *there exist certain* $\xi_j \sim N(0,\sigma_j)$ *and constants* $a>0$ *and* $C_0 \le 2^7/(1-1/4)$ *such that*

$$\mathbb{P}\Big\{a \le \max_{1\le j\le p} \xi_j \le a+\epsilon\Big\} \ge \frac{\epsilon}{\overline{\sigma}}\left(\frac{2+\sqrt{2\log p}}{C_0}\right) \tag{76}$$

for all ϵ *satisfying* $0 \le (\epsilon/\overline{\sigma})(2+\sqrt{2\log p}) \le 1/8$. *Moreover,* (76) *also holds for certain independent* $\xi_j \sim N(\mu_j,\sigma_j)$ *with possibly different nonzero* μ_j *and the same* $\{a, C_0\}$.

Anticoncentration of the maxima of Gaussian vectors have been considered in the literature; For example, Nazarov (2003), Klivans, O'Donnell and Servedio (2008) and Chernozhukov, Chetverikov and Kato (2015). These results provides $C_0(2+\sqrt{2\log p})/\sigma_{(1)}$ as an upper bound for (74) or $C_0(2+\sqrt{2\log p})\epsilon/\sigma_{(1)}$ for (75). A main advantage of Theorem 10 is the use of potentially much large $\overline{\sigma}$ instead of $\sigma_{(1)}$. For example, when $1/\sigma_{(1)} \ge (1+\sqrt{2\log p})/\sigma_{(m)}$ for all $1\le m\le p$, we have $\overline{\sigma} = (2+\sqrt{2\log p})(\sigma_{(1)}/2)$ and therefore the right-hand side of (75) becomes $2\epsilon/\sigma_{(1)}$. Moreover, Theorem 10 is sharp up to the constant factor C_0. The anticoncentration inequality for general $\xi_j \sim N(\mu_j,\sigma_j^2)$ is needed to study the mixed wild bootstrap T_n^{**} under the conditional probability $\mathbb{P}^{**}$, in view of (72).

THEOREM 11. *Let* $X_i = (X_{i,1},\dots,X_{i,p})^T \in \mathbb{R}^p$ *be independent centered random vectors with* $p>1$ *and* T_n^{**} *the mixed wild bootstrap given by* (69) *and* (71). *Let* $\sigma_j^2 = \sum_{i=1}^n \mathbb{E}X_{i,j}^2/n$ *and* $\{\sigma_{(j)}, 1\le j\le p, \overline{\sigma}\}$ *be as in* (4). *Suppose* $\mathbb{P}\{\|\mathbf{X}\|_{\max} \le a_n\} = 1$ *for certain constants* a_n *satisfying*

$$\max_{1\le j\le p} \frac{\log(j^2\overline{\sigma}/(\epsilon\sqrt{\log p}))}{\sigma_{(j)}^2} \le p_0 n/\big(8a_n^2\big). \tag{77}$$

Then, with the (a_0, b_0, p_0) *in* (69),

$$\omega_n^{(\mathbb{P})}(\epsilon; T_n^{**}) = \sup_{t\in\mathbb{R}} \mathbb{P}\{t \le T_n^{**} \le t+\epsilon\} \le C_{a_0,b_0,p_0}\frac{\epsilon}{\overline{\sigma}}\sqrt{\log p}. \tag{78}$$

If we use the mixed wild bootstrap (71) to approximate the distribution of T_n, Theorem 11 and the comparison theorems in Section 3 can be directly applied to establish the consistency of the bootstrap via (49). However, for studying the consistency of bootstrap methods in general through (67), we desire an anticoncentration inequality for the original data. This can be done by comparing the distributions of T_n^{**} and T_n, resulting in the following theorem.

THEOREM 12. *Let* $X_i \in \mathbb{R}^p$ *be independent with* $p>1$, $\mathbb{E}X_i = 0$, M_m *and* $\overline{\sigma}$ *be as in* (5) *and* (4), *respectively,* $b_n > 0$ *and* $\omega_n^{(\mathbb{P})}(\epsilon; T_n)$ *be as in* (30) *with the* T_n *in* (1). *Let* $a_n = c_1\sqrt{n}/(b_n\log p)$ *for some constant* $c_1>0$. *Then, for a certain positive constant* C_{c_1},

$$\begin{aligned} &\omega_n^{(\mathbb{P})}(1/b_n; T_n) \\ &\quad \le \frac{C_0}{b_n\overline{\sigma}}\sqrt{\log p} + C_{c_1}\kappa_{n,4} + 2\mathbb{P}\left\{\max_{1\le j\le p}\left|\sum_{i=1}^n \frac{X_{i,j}I_{\{|X_{i,j}|>a_n\}}}{\sqrt{n}}\right| > \frac{1}{8b_n}\right\}. \end{aligned} \tag{79}$$

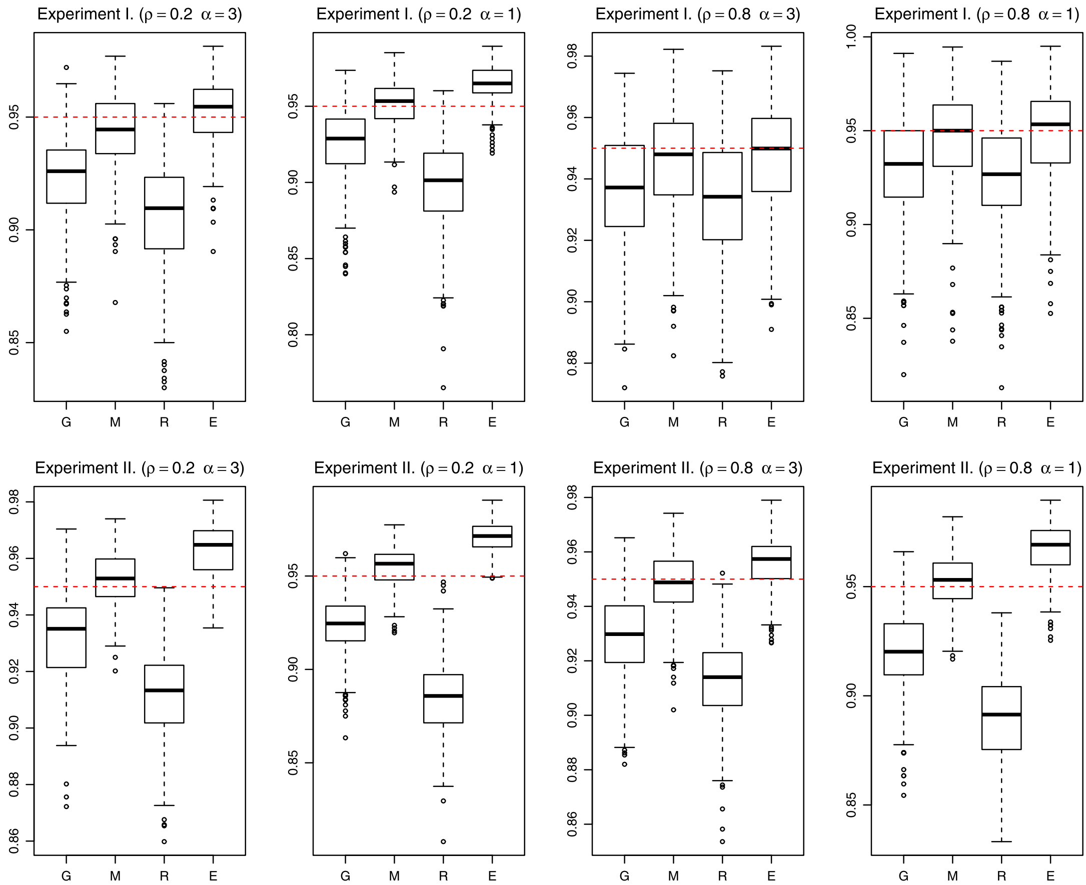

FIG. 1. *Simulated relative frequency of the simultaneous coverage of* 500 95% *simultaneous confidence intervals for each bootstrap scheme*: *G*, *M and R*, *respectively*, *represent the Gaussian*, *Mammen and Rademacher wild bootstrap*, *while E represents Efron's empirical bootstrap*.

We have derived comparison theorems up to a general order $m^* \geq 3$ under the moment matching condition (53). This includes $m^* > 4$ for the Rademacher wild bootstrap for symmetric X_i. However, as the Rademacher multiplier does not have a Gaussian component, we settle for $m^* = 4$ in the above theorem. If the Gaussian wild bootstrap is used as a vehicle to prove Theorem 12, (53) holds only for $m^* = 3$ and the term $\kappa_{n,4} = b_n^4 (\log p)^3 M_4^4/n$ will have to be replaced by $\kappa_{n,3} = b_n^3 (\log p)^2 M_3^3/\sqrt{n}$, leading to the condition $\log p \ll n^{1/7}$ for $b_n \gtrsim \sqrt{\log p}$ as in Chernozhukov, Chetverikov and Kato (2015).

5. Simulation results. We study the performance of different bootstrap procedures in two experiments. In both experiments, we generate vectors $X_i = (X_{i,1}, \dots, X_{i,p})^T$ in a Gaussian copula model, where $F(X_{i,j}) = \Phi(Y_{i,j})$ and $Y_i = (Y_{i,1}, \dots, Y_{i,p})^T$ are i.i.d. $N(0, \Sigma)$ with $N(0,1)$ marginal distributions, $n = 200$, $p = 400$ and F represents the gamma distribution with unit scale and shape parameter $\alpha = \mathbb{E}X_{i,j} \in \{1, 3\}$. We pick $\Sigma_{j,k} = \operatorname{Cov}(Y_{i,j}, Y_{i,k}) = \rho + (1-\rho)I_{\{j=k\}}$ in Experiment I, and $\Sigma_{j,k} = \rho^{|j-k|}$ in Experiment II, with $\rho \in \{0.2, 0.8\}$. Four bootstrap methods are considered: the Gaussian wild bootstrap with $W_i \sim N(0,1)$, Mammen's wild bootstrap, the Rademacher wild bootstrap with $\mathbb{P}\{W_i = \pm 1\} = 1/2$ and Efron's empirical bootstrap. Note that the skewness for $X_{i,j}$ is $2/\sqrt{\alpha}$, for example, 2 for $\alpha = 1$ and $2/\sqrt{3}$ for $\alpha = 3$. Thus, in this setting, the Gaussian multiplier and Rademacher wild bootstrap methods do not match the third moment of the

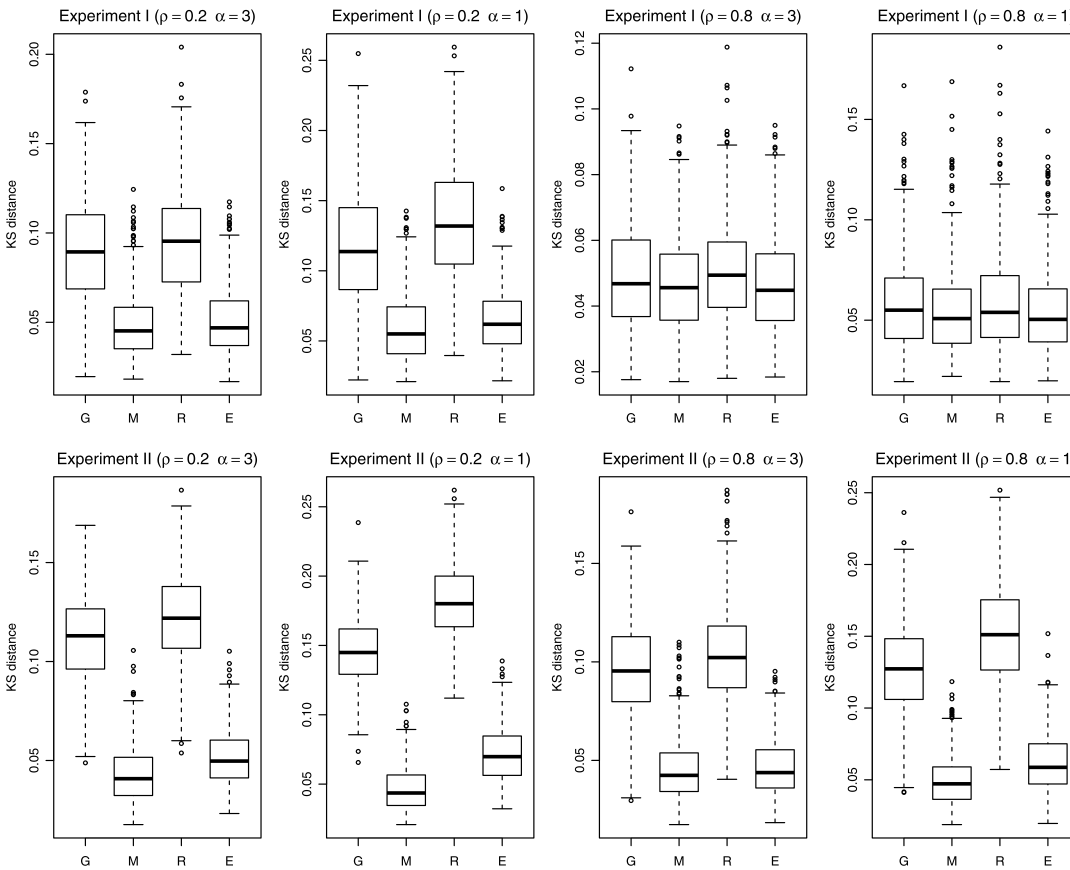

FIG. 2. *The Kolmogorov–Smirnov distances of* 500 *runs for each bootstrap scheme*: *G*, *M*, *R and E*, *respectively*, *represent the Gaussian*, *Mammen*, *Rademacher and empirical bootstrap schemes*.

original data. Our theorems in Section 2 therefore assert that Mammen's wild bootstrap and empirical bootstrap have better approximation properties. This theoretical claim is supported by our simulation results.

Since $\mathbb{E}X_i$ is unknown, the wild bootstrap is defined as $X_i^* = W_i(X_i - \overline{X})$. We compare the distribution of $T_n = \max_j \sum_{i=1}^n (X_{i,j} - \mathbb{E}X_{i,j})/\sqrt{n}$ against their bootstrapped versions. The true distribution of T_n is evaluated based on 5000 simulations. The results for the four bootstrap schemes are based on 500 copies of $\mathbf{X}$, and 500 copies of $\mathbf{X}^*$ for each observation of $\mathbf{X}$.

Figure 1 plots the simulated relative frequency of the simultaneous coverage of 95% bootstrap simultaneous confidence intervals for each bootstrap scheme in the four combinations of (ρ, α) in Experiments I and II. This is closely related to the risk $|\mathbb{P}\{T_n > t_\alpha^*\} - \alpha|$. The results for the Kolmogorov–Smirnov distance are shown in Figure 2 which contains 8 boxplots of the Kolmogorov–Smirnov distances between the true T_n and bootstrapped T_n^*.

Corresponding to our theoretical results, this simulation study demonstrates that Mammen's wild bootstrap is the best among all four schemes, empirical bootstrap is a close second, while Gaussian and Rademacher wild bootstrap methods are clearly worse. Because of the skewness of the Gamma distribution, an explanation of the poor performance of the Gaussian and Rademacher wild bootstrap methods is the lack of the third moment match as our theoretical results indicate. We would like to mention that the difference among bootstrap procedures in two settings (Experiment I, $\rho = 0.8$, $\alpha = 3$ or 1) are not as significant as

TABLE 1
The Kolmogorov–Smirnov distances between the bootstrapped T_n^* and true T_n

Setting	Gaussian		Mammen		Rademacher		Empirical	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
I, $\rho = 0.2, \alpha = 3$	0.08996	0.02907	0.04893	0.01883	0.09484	0.02916	0.05088	0.01873
I, $\rho = 0.2, \alpha = 1$	0.11660	0.03958	0.05964	0.02377	0.13428	0.04088	0.06457	0.02231
I, $\rho = 0.8, \alpha = 3$	0.04910	0.01610	0.04699	0.01510	0.05091	0.01587	0.04690	0.01503
I, $\rho = 0.8, \alpha = 1$	0.05861	0.02364	0.05443	0.02198	0.05880	0.02432	0.05452	0.02107
II, $\rho = 0.2, \alpha = 3$	0.11106	0.02299	0.04324	0.01443	0.12176	0.02254	0.05105	0.01397
II, $\rho = 0.2, \alpha = 1$	0.14542	0.02451	0.04677	0.01622	0.18143	0.02654	0.07190	0.02053
II, $\rho = 0.8, \alpha = 3$	0.09558	0.02485	0.04575	0.01629	0.10335	0.02493	0.04667	0.01488
II, $\rho = 0.8, \alpha = 1$	0.12780	0.03229	0.04998	0.01839	0.15043	0.03404	0.06249	0.02055

the others, possibly due to the smaller effective dimensionality caused by high correlation. Nevertheless, Mammen's wild bootstrap and empirical bootstrap still perform slightly better.

In addition to the plots, Table 1 provides the mean and standard deviation of the Kolmogorov–Smirnov distance between the bootstrap estimates and the true cumulative distribution function of T_n, and Table 2 provides the mean and standard deviation of the coverage probabilities of 95% simultaneous confidence intervals with each bootstrap scheme. These tables depicts the same picture as the plots.

It is worth mentioning that the empirical bootstrap does not always perform worse than Mammen's wild bootstrap (Figure 1, Experiment I, $\rho = 0.2, \alpha = 3$). Recall that we discuss in Section 2 that the empirical bootstrap doesn't offer exact moments match, and the fluctuation of the difference between true moments and empirically bootstrapped ones leads to a slightly weaker consistency statement in Theorem 1. However, the difference between the fourth moments,

$$\mu^{(4)} - \nu^{(4)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i^{\otimes 4} - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}^*(X_i^*)^{\otimes 4},$$

for the empirical bootstrap can be much smaller than that for Mammen's. This may provide an explanation of the performance of the Mammen and empirical bootstraps in these two settings.

TABLE 2
Relative frequency of bootstrap coverage of 95% simultaneous confidence intervals

Setting	Gaussian		Mammen		Rademacher		Empirical	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
I, $\rho = 0.2, \alpha = 3$	0.9232	0.01938	0.9446	0.01544	0.9072	0.2199	0.9527	0.01422
I, $\rho = 0.2, \alpha = 1$	0.9251	0.02308	0.9517	0.01422	0.8975	0.02938	0.9646	0.01131
I, $\rho = 0.8, \alpha = 3$	0.9364	0.01876	0.9457	0.01706	0.9331	0.01912	0.9471	0.01649
I, $\rho = 0.8, \alpha = 1$	0.9303	0.02671	0.9458	0.02447	0.9251	0.02785	0.9486	0.02357
II, $\rho = 0.2, \alpha = 3$	0.9323	0.01513	0.9527	0.00970	0.9124	0.01563	0.9628	0.00876
II, $\rho = 0.2, \alpha = 1$	0.9230	0.01613	0.9545	0.00955	0.8853	0.01890	0.9707	0.00721
II, $\rho = 0.8, \alpha = 3$	0.9291	0.01456	0.9479	0.01061	0.9129	0.01540	0.9562	0.00872
II, $\rho = 0.8, \alpha = 1$	0.9196	0.01850	0.9524	0.01172	0.8894	0.02079	0.9673	0.01116

Acknowledgments. The first author was supported is part by NSF Grant DMS-1454817. The second author was supported is part by NSF Grants DMS-1513378, IIS-1407939, DMS-1721495, IIS-1741390 and CCF-1934924.

SUPPLEMENTARY MATERIAL

Supplement to "Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors" (DOI: 10.1214/20-AOS1946SUPP; .pdf). This supplement contains proofs of all the theoretical results stated in the main body of the paper.

REFERENCES

BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Rev. Econ. Stud.* **81** 608–650. MR3207983 https://doi.org/10.1093/restud/rdt044

BELLONI, A., CHERNOZHUKOV, V. and KATO, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika* **102** 77–94. MR3335097 https://doi.org/10.1093/biomet/asu056

BENTKUS, V. Y. (1986). Dependence of the Berry–Esseen estimate on the dimension. *Litovsk. Mat. Sb.* **26** 205–210. MR0862741

BENTKUS, V. (2003). On the dependence of the Berry–Esseen bound on dimension. *J. Statist. Plann. Inference* **113** 385–402. MR1965117 https://doi.org/10.1016/S0378-3758(02)00094-0

BLANCHET, J., KANG, Y. and MURTHY, K. (2019). Robust Wasserstein profile inference and applications to machine learning. *J. Appl. Probab.* **56** 830–857. MR4015639 https://doi.org/10.1017/jpr.2019.49

CAI, T., LIU, W. and XIA, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *J. Amer. Statist. Assoc.* **108** 265–277. MR3174618 https://doi.org/10.1080/01621459.2012.758041

CHANG, J., ZHOU, W., ZHOU, W.-X. and WANG, L. (2017). Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering. *Biometrics* **73** 31–41. MR3632349 https://doi.org/10.1111/biom.12552

CHATTERJEE, S. (2006). A generalization of the Lindeberg principle. *Ann. Probab.* **34** 2061–2076. MR2294976 https://doi.org/10.1214/009117906000000575

CHEN, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Ann. Statist.* **46** 642–678. MR3782380 https://doi.org/10.1214/17-AOS1563

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *Ann. Statist.* **43** 1896–1928. MR3375871 https://doi.org/10.1214/15-AOS1329

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *J. Amer. Statist. Assoc.* **112** 1684–1696. MR3750891 https://doi.org/10.1080/01621459.2016.1228536

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.* **41** 2786–2819. MR3161448 https://doi.org/10.1214/13-AOS1161

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probab. Theory Related Fields* **162** 47–70. MR3350040 https://doi.org/10.1007/s00440-014-0565-9

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2017). Central limit theorems and bootstrap in high dimensions. *Ann. Probab.* **45** 2309–2352. MR3693963 https://doi.org/10.1214/16-AOP1113

DENG, H. and ZHANG, C.-H. (2020). Supplement to "Beyond Gaussian approximation: Bootstrap for maxima of sums of independent random vectors." https://doi.org/10.1214/20-AOS1946SUPP

DEZEURE, R., BÜHLMANN, P. and ZHANG, C.-H. (2017). High-dimensional simultaneous inference with the bootstrap. *TEST* **26** 685–719. MR3713586 https://doi.org/10.1007/s11749-017-0554-2

EFRON, B. (1979). Bootstrap methods: Another look at the jackknife. *Ann. Statist.* **7** 1–26. MR0515681

FAN, J. and LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 849–911. MR2530322 https://doi.org/10.1111/j.1467-9868.2008.00674.x

FAN, J. and ZHOU, W.-X. (2016). Guarding against spurious discoveries in high dimensions. *J. Mach. Learn. Res.* **17** 1–34. MR3580356

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3** 100–118. MR0380971 https://doi.org/10.1214/aop/1176996452

GINÉ, E. and ZINN, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.* **18** 851–869. MR1055437

GÖTZE, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724–739. MR1106283

HALL, P. (1988). Theoretical comparison of bootstrap confidence intervals. *Ann. Statist.* **16** 927–985. MR0959185 https://doi.org/10.1214/aos/1176350933

HALL, P. and PRESNELL, B. (1999). Intentionally biased bootstrap methods. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 143–158. MR1664116 https://doi.org/10.1111/1467-9868.00168

HOROWITZ, J. L. (2019). Bootstrap methods in econometrics. *Ann. Rev. Econ.* **11**.

KLIVANS, A. R., O'DONNELL, R. and SERVEDIO, R. A. (2008). Learning geometric concepts via Gaussian surface area. In *Foundations of Computer Science*, 2008. *FOCS'*08. *IEEE* 49*th Annual IEEE Symposium on* 541–550. IEEE Press, New York.

LINDEBERG, J. W. (1922). Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211–225. MR1544569 https://doi.org/10.1007/BF01494395

LIU, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models. *Ann. Statist.* **16** 1696–1708. MR0964947 https://doi.org/10.1214/aos/1176351062

MAMMEN, E. (1993). Bootstrap and wild bootstrap for high-dimensional linear models. *Ann. Statist.* **21** 255–285. MR1212176 https://doi.org/10.1214/aos/1176349025

NAGAEV, S. V. (1976). An estimate of the remainder term in the multidimensional central limit theorem. In *Proceedings of the Third Japan-USSR Symposium on Probability Theory* (*Tashkent*, 1975). *Lecture Notes in Math.* **550** 419–438. MR0443043

NAZAROV, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis. Lecture Notes in Math.* **1807** 169–187. Springer, Berlin. MR2083397 https://doi.org/10.1007/978-3-540-36428-3_15

NING, Y. and LIU, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist.* **45** 158–195. MR3611489 https://doi.org/10.1214/16-AOS1448

PRÆSTGAARD, J. and WELLNER, J. A. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086. MR1245301

SAZONOV, V. V. (1981). *Normal Approximation—Some Recent Advances. Lecture Notes in Math.* **879**. Springer, Berlin. MR0643968

SENATOV, V. V. (1980). Some uniform estimates of the convergence rate in the multidimensional central limit theorem. *Teor. Veroyatn. Primen.* **25** 757–770. MR0595137

SHAO, J. and TU, D. S. (1995). *The Jackknife and Bootstrap. Springer Series in Statistics*. Springer, New York. MR1351010 https://doi.org/10.1007/978-1-4612-0795-5

SINGH, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Ann. Statist.* **9** 1187–1195. MR0630102

SLEPIAN, D. (1962). The one-sided barrier problem for Gaussian noise. *Bell Syst. Tech. J.* **41** 463–501. MR0133183 https://doi.org/10.1002/j.1538-7305.1962.tb02419.x

STEIGER, W. L. (1969). A best possible Kolmogoroff-type inequality for martingales and a characteristic property. *Ann. Math. Stat.* **40** 764–769. MR0240843 https://doi.org/10.1214/aoms/1177697586

STEIN, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151. MR0630098

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14** 1261–1350. MR0868303 https://doi.org/10.1214/aos/1176350142

ZHANG, X. and CHENG, G. (2017). Simultaneous inference for high-dimensional linear models. *J. Amer. Statist. Assoc.* **112** 757–768. MR3671768 https://doi.org/10.1080/01621459.2016.1166114

ZHANG, D. and WU, W. B. (2017). Gaussian approximation for high dimensional time series. *Ann. Statist.* **45** 1895–1919. MR3718156 https://doi.org/10.1214/16-AOS1512

ZHANG, C.-H. and ZHANG, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 217–242. MR3153940 https://doi.org/10.1111/rssb.12026

ZHILOVA, M. (2016). Non-classical Berry–Esseen inequality and accuracy of the weighted bootstrap. Preprint. Available at arXiv:1611.02686.