Submitted to the Annals of Statistics

IMPROVED CENTRAL LIMIT THEOREM AND BOOTSTRAP APPROXIMATIONS IN HIGH DIMENSIONS

BY VICTOR CHERNOZHUOKOV[1], DENIS CHETVERIKOV[2,*] KENGO KATO[3,†] AND YUTA KOIKE[4,‡]

[1] *Department of Economics and Operations Research Center, MIT, vchern@mit.edu*

[2] *Department of Economics, UCLA, *chetverikov@econ.ucla.edu*

[3] *Department of Statistics and Data Science, Cornell University, †kk976@cornell.edu*

[4] *Department, Mathematics and Informatics Center and Graduate School of Mathematical Sciences, The University of Tokyo, ‡kyuta@ms.u-tokyo.ac.jp*

This paper deals with the Gaussian and bootstrap approximations to the distribution of the max statistic in high dimensions. This statistic takes the form of the maximum over components of the sum of independent random vectors and its distribution plays a key role in many high-dimensional estimation and testing problems. Using a novel iterative randomized Lindeberg method, the paper derives new bounds for the distributional approximation errors. These new bounds substantially improve upon existing ones and simultaneously allow for a larger class of bootstrap methods.

1. Introduction. Let $X_1, \ldots, X_n$ be independent random vectors in $\mathbb{R}^p$ such that $\mathrm{E}[X_{ij}] = \mu_j$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$, where X_{ij} denotes the jth component of the vector X_i. We are interested in approximating the distribution of the maximum coordinate of the centered sample mean of $X_1, \ldots, X_n$, i.e.,

$$T_n = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} - \mu_j). \tag{1}$$

The distribution of T_n plays a particularly important role in many high-dimensional settings, where p is potentially larger or much larger than n. For example, it appears in selecting the regularization parameters for the Lasso estimator and the Dantzig selector ([12]), in carrying out reality checks for data snooping and testing superior predictive ability ([42, 25]), in constructing model confidence sets ([26]), in testing conditional and/or many unconditional moment inequalities ([2, 19, 16, 31]), in multiple testing with the family-wise error rate control ([3]), in constructing simultaneous confidence intervals for high-dimensional parameters ([4]), in adaptive testing of regression and stochastic monotonicity ([20, 21]), in carrying out inference on generalized instrumental variable models ([18]), and in constructing Lepski-type procedures for adaptive estimation and inference in nonparametric problems ([13]); more references can be found in [22] and especially in [3]. It is therefore of great interest to develop methods for obtaining feasible and accurate approximations to the distribution of T_n, allowing for the high-dimensional $p \gg n$ case.

Toward this goal, the first three authors of this paper obtained the following Gaussian approximation result in [12, 15]. Let $G = (G_1, \ldots, G_p)'$ be a Gaussian random vector in $\mathbb{R}^p$ with mean $\mu = (\mu_1, \ldots, \mu_p)'$ and covariance matrix $n^{-1} \sum_{i=1}^{n} \mathrm{E}[(X_i - \mu)(X_i - \mu)']$ and let

MSC2020 subject classifications: 60F05, 62E17.

Keywords and phrases: bootstrap, central limit theorem, iterative randomized Lindeberg method, Stein kernel.

the critical value $c_{1-\alpha}$ be the $(1-\alpha)$th quantile of $\max_{1\le j\le p} G_j$. Then under mild regularity conditions,

$$\left|\mathrm{P}(T_n > c_{1-\alpha}) - \alpha\right| \le C\left(\frac{\log^7(pn)}{n}\right)^{1/6}, \tag{2}$$

where C is a constant that is independent of n and p. This result is important because the right-hand side of the bound (2) depends on p only via the logarithm of p, and hence it shows that the Gaussian approximation holds if $\log p = o(n^{1/7})$, which allows p to be much larger than n. Besides, building upon this result, the same authors have proved bounds similar to (2) for the critical values obtained by the Gaussian multiplier and empirical bootstraps in [15].

Gaussian approximation of the form (2) allows us to develop powerful inference methods for high-dimensional data in applications discussed above and has stimulated further developments into dependent data [44, 43, 16], U-statistics [19, 10, 11], Malliavin calculus [20], and homogeneous sums [29]. Despite such rapid developments, the literature has left much to be desired on coherent understanding of sharpness of the bound (2) for the Gaussian or bootstrap critical values since the first appearance of [15] in 2014 on arXiv. The problem can be decomposed into two parts: (i) sharpness of the bound in terms of dependence on n and (ii) sharpness of the bound in terms of dependence on p.

There are two important developments toward the question of sharpness of the bound (2) that should be mentioned. First, Deng and Zhang [22] considered direct bootstrap approximation without taking the root of Gaussian approximation, and proved the following bound for the critical value $c_{1-\alpha}$ obtained by the empirical or third-order matching (or Mammen's [36]) multiplier bootstraps:

$$\left|\mathrm{P}(T_n > c_{1-\alpha}) - \alpha\right| \le C\left(\frac{\log^5(pn)}{n}\right)^{1/6}. \tag{3}$$

Their bound improves the power of the logs in the previous bound (2), showing that the empirical and Mammen's bootstraps are consistent to approximate the distribution of T_n if $\log p = o(n^{1/5})$ instead of $\log p = o(n^{1/7})$. Second, the recent preprint by the fourth author [30] shows that the same bound (3) indeed holds for the Gaussian critical value as well.

In turn, in this paper, we show that in fact a much larger improvement is possible: under mild regularity conditions, we prove that

$$\left|\mathrm{P}(T_n > c_{1-\alpha}) - \alpha\right| \le C\left(\frac{\log^5(pn)}{n}\right)^{1/4}, \tag{4}$$

both for the Gaussian and bootstrap critical values $c_{1-\alpha}$. In comparison with the Gaussian approximation result (2), our new bound improves not only the power of the logs but also the power of the sample size n. Moreover, regarding the bootstrap types, we allow for not only the empirical and third-order matching multiplier bootstrap methods, but also for general multiplier bootstrap methods (with i.i.d weights), which match only two moments of the data, such as the multiplier bootstrap methods with Gaussian and Rademacher weights.

We remark that several authors have recently pointed out that an additional structural assumption on the covariance matrices of X_i's can improve the bound (4). In particular, Fang and Koike [23] showed that the right-hand side of (4) can be improved to $C(\log^4(pn)/n)^{1/3}$ when the covariance matrices are non-degenerate and can be further improved to $C(\log^3(p)/n)^{1/2}\log n$ when we additionally assume that X_i's have log-concave densities. The latter result is based on the fact that random vectors with log-concave densities admit Stein kernels with sub-Weibull entries, which is established by Fathi in [24]. Moreover, building on the important results by Lopes in [34] and Kuchibhotla and Rinaldo in [32], [17]

showed that the bound $C(\log^3(p)/n)^{1/2}\log n$ can be achieved even without the assumption of log-concave densities (non-degenerate covariance matrices are still required; [34] and [32] were the first to obtain dependence on n via $1/\sqrt{n}$ in (4) without requiring log-concave densities). In addition, Lopes, Lin and Müller [35] showed that the right-hand side of (4) can be improved to $Cn^{-1/2+\delta}$ for any $\delta > 0$ when the coordinates of X_i's have decaying variances. Compared to these results, our bound requires neither non-degenerate covariance matrices nor decaying variances.

In addition, we prove that if the distribution of the random vectors $X_1, \dots, X_n$ is symmetric around the mean, then even better approximation to the distribution of T_n is possible:

$$\left|\mathrm{P}(T_n > c_{1-\alpha}) - \alpha\right| \le C\left(\frac{\log^3(pn)}{n}\right)^{1/2} \tag{5}$$

as long as the critical value $c_{1-\alpha}$ is obtained via the multiplier bootstrap method with Rademacher weights. This new bound makes Rademacher weights particularly appealing in the high-dimensional settings, at least from a theoretical perspective.

We also consider bootstrap approximations with incremental factors, previously used by Andrews and Shi in [1] in the context of testing conditional moment inequalities. Specifically, for a small but fixed constant $\eta > 0$, called an incremental factor, we derive the following bounds:

$$\mathrm{P}(T_n > c_{1-\alpha} + \eta) - \alpha \le C\left(\frac{\log^3(pn)}{n}\right)^{1/2} \tag{6}$$

if $c_{1-\alpha}$ is obtained via either the empirical or the third-order matching multiplier bootstrap methods and

$$\mathrm{P}(T_n > c_{1-\alpha} + \eta) - \alpha \le C\left(\frac{\log^5(pn)}{n}\right)^{1/2} \tag{7}$$

if $c_{1-\alpha}$ is obtained via general multiplier bootstrap methods, where the constant C may depend on η. Even though these are one-sided bounds, they are useful because they show that in any test based on the statistic T_n, increasing the critical value $c_{1-\alpha}$ by an incremental factor η may substantially reduce the sample complexity for over-rejection. Namely, assuming $\log p \gtrsim \log n$ for simplicity, for the over-rejection probability to be less than or equal to a given level $0 < \Delta < 1 - \alpha$, the empirical bootstrap or multiplier bootstrap (without incremental factor) requires $n \gtrsim \Delta^{-4}\log^5 p$, while adding a constant incremental factor reduces the sample complexity to $n \gtrsim (\Delta^{-2}\log^3 p) \vee \log^5 p$ if we use the empirical or third-order matching bootstrap. It is worth noting that, given that in high-dimensional settings, where p is rapidly increasing together with n, $c_{1-\alpha}$ is typically also getting large as we increase n, adding an incremental factor η may not have a large impact on the power properties of the test.

In fact, all our results apply to a more general version of the statistic T_n:

$$T_n = \max_{1\le j\le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} - \mu_j + a_j), \tag{8}$$

where $a = (a_1, \dots, a_p)'$ is a vector in $\mathbb{R}^p$, which reduces to (1) if we set $a = 0_p$. In most applications mentioned above, the former version (1) is sufficient but there are some applications where the more general version (8) is required; for example, the latter was used by Bai, Shaikh, and Santos in [2] to extend the method of testing moment inequalities proposed in

[38] for the case of a small number of inequalities to the case of a large number of inequalities. For the rest of the paper, we will therefore work with the more general version (8) of the statistic T_n. In addition, we emphasize that our results can be equally applied with

$$T_n = \max_{1\le j\le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} - \mu_j + a_j) \right|$$

by replacing the p-dimensional vectors $X_i - \mu + a$ with the $2p$-dimensional vectors whose first p components are equal to $X_i - \mu + a$ and the last p components are equal to $-(X_i - \mu + a)$.

To prove (4), we develop a novel and iterative version of the randomized Lindeberg method. A key feature of our approach is that we carry out a careful analysis of the coefficients in the Taylor expansion underlying the Lindeberg method. In particular, we apply the Lindeberg method iteratively in combination with an anti-concentration inequality for maxima of Gaussian processes to bound these coefficients, which substantially improves upon the original randomized Lindeberg method proposed in [22]. In addition, we sharpen the Gaussian approximation bounds for the multiplier processes developed in [30] using Stein's kernels. In turn, to prove (5), we establish a new connection between the Rademacher bootstrap and the randomization tests, as discussed in [33], using a recent result from the computer science literature on pseudo-random number generators by O'Donnell, Servedio, and Tan [37], which provides an anti-concentration inequality for maxima of Rademacher processes. Finally, to prove error bounds (6) and (7), we apply the original randomized Lindeberg method as developed in [22].

Finally, we conduct a small scale simulation study. Our simulation study shows that (i) all bootstrap methods considered in this paper perform reasonably well in high dimensions; (ii) for asymmetric distributions, the empirical and the third-order matching multiplier bootstrap methods outperform the multiplier bootstrap methods with Gaussian and Rademacher weights; and (iii) for symmetric distributions, the multiplier bootstrap with Rademacher weights performs the best, which is consistent with Theorem 2.3 ahead. See the Supplementary Material for details.

The rest of the paper is organized as follows. In the next section, we present our main results. In Section 3, we develop the iterative randomized Lindeberg method, which is the first key component in deriving our main results. In Section 4, we provide new bounds for the Gaussian approximations using Stein's kernels, which is the second key component in deriving our main results. In Section 5, we give proofs of the main results. In the Supplemental Material, we collect additional derivations and conduct a small simulation study.

1.1. *Notation.* For any vectors $x, y \in \mathbb{R}^p$ and any scalar $c \in \mathbb{R}$, we write $x \le y$ if $x_j \le y_j$ for all $j = 1, \dots, p$ and write $x + c$ to denote the vector in $\mathbb{R}^p$ whose jth component is $x_j + c$ for all $j = 1, \dots, p$. Also, for any sequences of scalars $\{a_n\}_{n\ge1}$ and $\{b_n\}_{n\ge1}$ we write $a_n \lesssim b_n$ if $a_n \le C b_n$ for all $n \ge 1$ for some constant C. Recall that, for any random variable T and a constant $\gamma \in (0,1)$, the γth quantile of T is defined as $\inf\{t \in \mathbb{R} \colon \mathrm{P}(T \le t) \ge \gamma\}$. Finally, we use the notation $X_{1:n} = (X_1, \dots, X_n)$.

2. Main Results. In this section, we present our main results. We first formally define all the critical values $c_{1-\alpha}$ to be used throughout the paper. We then discuss the required regularity conditions and present the results.

2.1. *Gaussian and Bootstrap Critical Values.* First, define the Gaussian critical value $c_{1-\alpha}^G$ as the $(1-\alpha)$th quantile of

$$T_n^G = \max_{1\le j\le p}(G_j + a_j), \tag{9}$$

where G is a centered Gaussian random vector in $\mathbb{R}^p$ with the covariance matrix

$$\Sigma_n = \frac{1}{n}\sum_{i=1}^n \mathrm{E}[(X_i-\mu)(X_i-\mu)'], \tag{10}$$

which coincides with the covariance matrix of $\sqrt{n}(\bar{X}_n - \mu)$. Second, define the bootstrap critical value $c_{1-\alpha}^B$ as the $(1-\alpha)$th quantile of the conditional distribution of

$$T_n^* = \max_{1\le j\le p}\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_{ij}^* + a_j) \tag{11}$$

given the data $X_1,\dots,X_n$, where $X_1^*,\dots,X_n^*$ is a (not necessarily empirical) bootstrap sample. We consider the following types of the bootstrap:

- Empirical bootstrap: let $X_1^*,\dots,X_n^*$ be a sequence of i.i.d. random variables sampled from the uniform distribution on $\{X_1-\bar{X}_n,\dots,X_n-\bar{X}_n\}$, where $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ denotes the sample mean of the data $X_1,\dots,X_n$.
- Multiplier bootstrap: let $e_1,\dots,e_n$ be a sequence of i.i.d. random variables with mean zero and unit variance, referred to as weights, which are independent of $X_1,\dots,X_n$. Define $X_i^* = e_i(X_i-\bar{X}_n)$ for all $i=1,\dots,n$.

For the multiplier bootstrap, we will assume throughout the paper that the weights $e_1,\dots,e_n$ are such that

$$\begin{aligned}&e_i = e_{i,1}+e_{i,2}\text{, where } e_{i,1} \text{ and } e_{i,2} \text{ are independent, } e_{i,1} \text{ has}\\ &\text{the } N(0,\sigma_e^2) \text{ distribution for some } \sigma_e\ge 0, \text{ and } |e_{i,2}|\le 3.\end{aligned} \tag{12}$$

Condition (12) is mild and covers many commonly used weights, such as:

- Gaussian weights: $e_{i,1}\sim N(0,1)$ and $e_{i,2}=0$.
- Rademacher weights: $e_{i,1}=0$ (i.e., $\sigma_e=0$) and $\mathrm{P}(e_{i,2}=\pm 1)=1/2$.
- Mammen's weights [36]: $e_{i,1}=0$ and

$$\mathrm{P}\left(e_{i,2}=\frac{1\pm\sqrt{5}}{2}\right)=\frac{\sqrt{5}\mp 1}{2\sqrt{5}}.$$

See Remark 2.1 for further discussion on Condition (12).

Occasionally, we will also consider the weights with unit third moment, namely,

$$\mathrm{E}[e_i^3]=1, \quad \text{for all } i=1,\dots,n. \tag{13}$$

The weights satisfying Condition (13) correspond to the third-order matching multiplier bootstrap mentioned in the Introduction. We note that Mammen's weights satisfy both Conditions (12) and (13), but neither Rademacher nor Gaussian weights satisfy Condition (13). See Lemma I.3 in the Supplemental Material, where we provide a more general class of distributions for the weights satisfying both Conditions (12) and (13).

Before proceeding to the regularity conditions, we also note that the multiplier bootstrap critical value $c_{1-\alpha}^B$ with Gaussian weights can be regarded as a feasible version of the Gaussian critical value $c_{1-\alpha}^G$. Indeed, it is easy to see that the former can be alternatively defined as the $(1-\alpha)$th quantile of the distribution of

$$T_n^{\hat{G}} = \max_{1\le j\le p}(\hat{G}_j + a_j),$$

where $\hat{G} \sim N(0_p, \widehat{\Sigma}_n)$ and $\widehat{\Sigma}_n$ is the empirical covariance matrix

$$\widehat{\Sigma}_n = n^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)(X_i - \bar{X}_n)'. \tag{14}$$

For brevity, we sometimes refer to both quantities as the Gaussian critical values.

REMARK 2.1 (On Condition (12)). Condition (12) is technical and can be weakened depending on the moment conditions on X_i. A key step in the proof of Theorem 2.2 is to apply Theorem 3.1 ahead to approximate the conditional distribution of T_n^* with that of the multiplier bootstrap statistic with weights following a Beta distribution that matches the moments of e_i up to the third order (to be precise, we first replace the Gaussian components $e_{i,1}$ by bounded weights in the proof of Theorem 2.2). Condition (12) will be used to verify Conditions V, P, and B when we apply Theorem 3.1 there. If, e.g., X_i are bounded by B_n, then the conclusion of Theorem 2.2 continues to hold for sub-exponential weights. Since current Condition (12) already covers many commonly used bootstrap weights, however, we do not pursue this generality of the weights to keep our presentation reasonable concise.

2.2. *Regularity Conditions.* First, observe that given the construction of the statistic T_n in (8) and its Gaussian and bootstrap analogs in (9) and (11), it is without loss of generality to assume that $\mu_j = 0$ for all $j = 1, \ldots, p$, which is what we do for the rest of the paper. Also, all our results follow immediately if $n = 2$, so we assume $n \geq 3$, which in particular implies $\log(pn) \geq 1$. In addition, since we are primarily interested in the case with large p, we assume $p \geq 2$.

Second, let b_1 and b_2 be some strictly positive constants such that $b_1 \leq b_2$ and let $\{B_n\}_{n \geq 1}$ be a sequence of constants such that $B_n \geq 1$ for all $n \geq 1$. Here, the sequence $\{B_n\}_{n \geq 1}$ can diverge to infinity as the sample size n increases.

Condition E: *For all $i = 1, \ldots, n$ and $j = 1, \ldots, p$, we have*

$$\mathrm{E}[\exp(|X_{ij}|/B_n)] \leq 2.$$

Condition M: *For all $j = 1, \ldots, p$, we have*

$$b_1^2 \leq \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[X_{ij}^2] \quad \textit{and} \quad \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[X_{ij}^4] \leq B_n^2 b_2^2.$$

Condition S: *For all $i = 1, \ldots, n$, the distribution of X_i is symmetric in the sense that X_i and $-X_i$ are identically distributed.*

Condition E implies that the random variables X_{ij} are sub-exponential with the Orlicz ψ_1-norm bounded by B_n; see [40] for details. The same sub-exponential condition was assumed in e.g. [15] and [22]; see Condition (E.1) in [15] and (E.1) in [22]. The first part of Condition M, which we refer to as the variance lower bound condition, requires that each component of the random vectors X_i is scaled properly. The variance lower bound condition is needed to apply the anti-concentration inequalities (cf. Lemmas J.3 and J.4 in the Supplemental Material) but can be dropped in Theorem 2.4 ahead. Also, at least for Theorems 2.1 and 2.2, it can be relaxed by using Theorem 10 in [22]. However, to consistently state all the results, we work with the present assumption. Given the first part, the second part of Condition M holds if, for example, all random variables X_{ij} are bounded by B_n and $n^{-1} \sum_{i=1}^{n} \mathrm{E}[X_{ij}^2] \leq b_2^2$ for all $j = 1, \ldots, p$. Condition S means that the distribution of each X_i is symmetric around

the mean. Importantly, none of these conditions restrict the correlation matrices of X_i, and so our results do not follow from the classical results in empirical process theory.

In what follows, we will always maintain Conditions E and M and will assume Condition S only in Theorem 2.3, which shows that imposing the symmetric distributions improves the approximation bound for the multiplier bootstrap with Rademacher weights.

2.3. *Main Results.* We first present a non-asymptotic bound on the error of the Gaussian approximation to the distribution of the statistic T_n:

THEOREM 2.1 (Gaussian Approximation). *Suppose that Conditions E and M are satisfied. Then*

$$\left|\mathrm{P}\left(T_n > c^G_{1-\alpha}\right) - \alpha\right| \leq C\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}, \tag{15}$$

where C is a constant depending only on b_1 and b_2.

This result improves upon the bound in [30], who obtained a similar result with the rate $1/6$ instead of $1/4$. Since $a \in \mathbb{R}^p$ in the definition of T_n in (8) is arbitrary, the bound (15) can be equivalently stated as

$$\sup_{A\in\mathcal{A}}\left|\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \in A\right) - \mathrm{P}(G \in A)\right| \leq C\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4},$$

where $G \sim N(0_p, \Sigma_n)$ and $\mathcal{A}$ is the class of all hyper-rectangles in $\mathbb{R}^p$, i.e. sets of the form

$$A = \Big\{w = (w_1, \ldots, w_p)' \in \mathbb{R}^p \colon a_{lj} \leq w_j \leq a_{rj} \text{ for all } j = 1, \ldots, p\Big\},$$

for some constants $-\infty \leq a_{lj} \leq a_{rj} \leq \infty$ with $j = 1, \ldots, p$. This gives a quantitative Central Limit Theorem (CLT) over the hyper-rectangles in high dimensions.

The proof of Theorem 2.1, which is deferred to Section 5, is fairly complicated and goes somewhat backward: (i) we first compare the conditional distribution of a third-order matching bootstrap statistic T_n^* with that of the Gaussian multiplier bootstrap statistic $T_n^{\hat{G}}$, and then compare the conditional distribution of $T_n^{\hat{G}}$ with the distribution of T_n^G. These two comparisons rely on the Gaussian approximation via Stein kernel (Theorem 4.1). Then, (ii) we use the preceding comparison between T_n^* and T_n^G to verify the anti-concentration for T_n^* to invoke Theorem 3.1 and compare the conditional distribution of T_n^* with the distribution of T_n. The proof of Theorem 3.1 relies on a novel technique which we call the *iterative randomized Lindeberg method*. The conclusion of Theorem 2.1 follows from combining the results in Steps (i) and (ii) and the triangle inequality.

Comparison of the the Gaussian multiplier bootstrap statistic $T_n^{\hat{G}}$ with T_n^G relies on the following Gaussian-to-Gaussian comparison inequality, which can be of independent interest and whose proof is presented in Section 4 as a consequence of Theorem 4.1:

PROPOSITION 2.1 (Gaussian-to-Gaussian Comparison). *If Z_1 and Z_2 are centered Gaussian random vectors in $\mathbb{R}^p$ with covariance matrices Σ^1 and Σ^2, respectively, and Σ^2 is such that $\Sigma^2_{jj} \geq c$ for all $j = 1, \ldots, p$ for some constant $c > 0$, then*

$$\sup_{y\in\mathbb{R}^p}\Big|\mathrm{P}(Z_1 \leq y) - \mathrm{P}(Z_2 \leq y)\Big| \leq C\Big(\Delta \log^2 p\Big)^{1/2},$$

where C is a constant depending only on c and $\Delta = \max_{1\leq j,k\leq p}|\Sigma^1_{jk} - \Sigma^2_{jk}|$.

REMARK 2.2. Two comments on Proposition 2.1 are warranted. First, Proposition 2.1 improves upon Theorem 2 in [14], which shows that

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\max_{1\le j\le p} Z_{1j}\le x\right)-\mathrm{P}\left(\max_{1\le j\le p} Z_{2j}\le x\right)\right|\le C\Big(\Delta\log^2 p\Big)^{1/3},$$

under the same conditions. Second, the bound in this proposition is sharp in the sense that there exists a constant $c>0$ such that for infinitely many values of p, there exist centered Gaussian random vectors Z_1 and Z_2 in $\mathbb{R}^p$ such that the covariance matrix Σ^2 of Z_2 satisfies $\Sigma^2_{jj}=1$ for all $j=1,\dots,p$ and

$$\sup_{y\in\mathbb{R}^p}\Big|\mathrm{P}(Z_1\le y)-\mathrm{P}(Z_2\le y)\Big|\ge c\Big(\Delta\log^2 p\Big)^{1/2}.$$

The latter claim is proven in Appendix C of the Supplemental Material. ∎

Comparison of the conditional distribution of the third-order matching bootstrap statistic T_n^* with that of T_n (Theorem 3.1) relies on the iterative randomized Lindeberg method. An intuition behind the iterative randomized Lindeberg method goes as follows. Recall that, for any smooth function $g\colon\mathbb{R}^p\to\mathbb{R}$ and any two sequences of independent random vectors $X_1,\dots,X_n$ and $Y_1,\dots,Y_n$ in $\mathbb{R}^p$, in order to approximate $\mathrm{E}[g(X_1+\cdots+X_n)]$ by $\mathrm{E}[g(Y_1+\cdots+Y_n)]$, the original Lindeberg method constructs an interpolation path from $\mathrm{E}[g(X_1+\cdots+X_n)]$ to $\mathrm{E}[g(Y_1+\cdots+Y_n)]$ by replacing X_i's with Y_i's one-by-one in a given order and uses Taylor's expansion to show that the change in the expectation at each step is sufficiently small; see [8] for example. The randomized Lindeberg method, introduced in [22], is similar to the original Lindeberg method but it replaces X_i's with Y_i's in a randomly selected order. It turns out that this randomization may bring substantial benefits to the final bound. In turn, to improve upon this version of the randomized Lindeberg method, we carry out a careful analysis of the coefficients in the Taylor's expansions underlying the method. In particular, given that kth order coefficients take the form of $\mathrm{E}[g^{(k)}(Z_1+\cdots+Z_n)]$, up to some approximation error, where $g^{(k)}$ is a vector of the kth partial derivatives of g and $Z_1,\dots,Z_n$ is a sequence such that some of its elements are given by X_i's and others by Y_i, and using the fact that it is easier in our setting to bound $\mathrm{E}[g^{(k)}(Y_1+\cdots+Y_n)]$, we apply the randomized Lindeberg method once again to approximate $\mathrm{E}[g^{(k)}(Z_1+\cdots+Z_n)]$ by $\mathrm{E}[g^{(k)}(Y_1+\cdots+Y_n)]$. Here, since a new application of the method will bring new Taylor's coefficients, we apply the same method over and over again until the approximation error becomes sufficiently small. We demonstrate that this iterative use of the randomized Lindeberg method gives further substantial benefits to the final bound. See also the discussion before Lemma 3.1 concerning comparisons of the iterative randomized Lindeberg method with the randomized Lindeberg method used in [22] and the related Slepian-Stein method used in our earlier work [12, 15].

Our second main result gives a non-asymptotic bound on the deviation of the bootstrap rejection probabilities $\mathrm{P}(T_n>c^B_{1-\alpha})$ from the nominal level α for the empirical and the multiplier bootstrap methods:

THEOREM 2.2 (Bootstrap Approximation). *Suppose that Conditions E and M are satisfied and that $c^B_{1-\alpha}$ is obtained via either the empirical bootstrap or the multiplier bootstrap with weights satisfying (12). Then*

$$\left|\mathrm{P}\left(T_n>c^B_{1-\alpha}\right)-\alpha\right|\le C\left(\frac{B_n^2\log^5(pn)}{n}\right)^{1/4}, \tag{16}$$

where C is a constant depending only on b_1 and b_2.

This theorem improves upon the bounds in [22], who obtained a similar result with the rate $1/6$ instead of $1/4$. In addition, we allow for a larger class of multiplier bootstrap methods. In particular, we do not require the weights $e_1, \dots, e_n$ to satisfy (13). The proof of this theorem is given in Section 5.

Our third main result gives a non-asymptotic bound on the deviation of the bootstrap rejection probabilities from the nominal level for the multiplier bootstrap method with Rademacher weights in the case of symmetric distributions:

THEOREM 2.3 (Rademacher Bootstrap Approximation in Symmetric Case). *Suppose that Conditions E, M, and S are satisfied and that $c^B_{1-\alpha}$ is obtained via the multiplier bootstrap with Rademacher weights. Then*

$$\left|\mathrm{P}\left(T_n > c^B_{1-\alpha}\right) - \alpha\right| \le C\left(\frac{B_n^2 \log^3(pn)}{n}\right)^{1/2}, \tag{17}$$

where C is a constant depending only on b_1 and b_2.

This theorem implies that the multiplier bootstrap with Rademacher weights is very accurate in the symmetric case. To prove it, we note that under the assumption of symmetric distributions, one can construct the randomization critical value $c^R_{1-\alpha}$ such that $\mathrm{P}(T_n > c^R_{1-\alpha}) = \alpha$, up to possible mass points in the distribution of T_n. Thus, given that the critical value based on the multiplier bootstrap with Rademacher weights turns out to be a feasible version of this randomization critical value and the two are close to each other, (17) follows if we can show that the distribution of T_n is not too concentrated. To this end, we use an anti-concentration inequality for maxima of Rademacher processes derived in [37]. The proof of Theorem 2.3 is given in Appendix G of the Supplemental Material.

Our fourth and final result shows that one-sided bounds in the bootstrap approximation can be substantially improved if we allow for incremental factors:

THEOREM 2.4 (Bootstrap Approximation with Incremental Factors). *Suppose that Conditions E and M are satisfied and let $\eta > 0$ be a constant that may depend on n and p. Then there exists a constant C depending only b_1 and b_2 such that the following hold.*

(i) *If $B_n^2 \log^5(pn) \le n$ and $c^B_{1-\alpha}$ is obtained via either the empirical bootstrap or the multiplier bootstrap with weights satisfying (12) and (13), then we have*

$$\mathrm{P}(T_n > c^B_{1-\alpha} + \eta) \le \alpha + C(1 \vee \eta^{-4})\left(\frac{B_n^2 \log^3(pn)}{n}\right)^{1/2}.$$

(ii) *If $n^{-1}\sum_{i=1}^n \mathrm{E}[X_{ij}^2] \le b_2^2$ for all $j = 1, \dots, p$ and $c^B_{1-\alpha}$ is obtained via the multiplier bootstrap with weights satisfying (12), then*

$$\mathrm{P}(T_n > c^B_{1-\alpha} + \eta) \le \alpha + C(1 \vee \eta^{-4})\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/2}.$$

Theorem 2.4 allows η to (slowly) decrease with n and/or p. For example, if we choose $\eta \sim (\log n)^{-1}$, then the over-rejection probability is of order $n^{-1/2}$ in n up to log factors, while only requiring p to be $\log p = o\big(n^{1/3}/\mathrm{polylog}(n)\big)$ in (i) and $\log p = o\big(n^{1/5}/\mathrm{polylog}(n)\big)$ in (ii) provided that B_n is bounded in n.

To prove this theorem, we use the randomized Lindeberg method but with an important simplification that the incremental factor η now absorbs all the terms arising from smoothing the functions of the form $x \mapsto 1\{\max_{1\le j\le p} x_j > c\}$, which is used in the Lindeberg method.

As discussed in the Introduction, Theorem 2.4 is useful if one is concerned with the finite-sample over-rejection of tests based on the statistic T_n as it says that adding an incremental factor η to the critical value $c^B_{1-\alpha}$ may substantially reduce over-rejection, with a minimal effect on the power of the test. The proof of Theorem 2.4 is given in Appendix H of the Supplemental Material.

We conclude this section with a few remarks on cases with polynomial moment conditions and approximate sample means. First, while we maintain the sub-exponential condition (Condition E) for X_i throughout the main text, combining Theorems 2.1 and 2.2, and a simple truncation argument, we are able to derive analogs of Gaussian and bootstrap approximation results under polynomial moment conditions similar to e.g. Condition (E.1) in [15]. Those additional results can be found in Appendix A of the Supplemental Material. Second, in many applications (such as simultaneous inference for high-dimensional statistical models; cf. [5]), the statistic T_n can only be asymptotically approximated by the maximum coordinate of the sample mean of independent random vectors. Also, those random vectors, often corresponding to the influence functions, may not be directly observable but have to be estimated. We emphasize here that all our results can be extended to such approximate sample mean cases using the same arguments as those used in [3]; however, we have opted not to carry out the extension here for brevity of the paper.

3. Iterative Randomized Lindeberg Method. In this section, we derive a distributional approximation result, Theorem 3.1, using a novel proof technique, which we call the iterative randomized Lindeberg method. We will use this result in Section 5 to prove our main results on the Gaussian and bootstrap approximations in high dimensions, as stated in Section 2.

Let $V_1, \dots, V_n, Z_1, \dots, Z_n$ be a sequence of independent random vectors in $\mathbb{R}^p$ such that $\mathrm{E}[V_{ij}] = \mathrm{E}[Z_{ij}] = 0$ for all $i = 1, \dots, n$ and $j = 1, \dots, p$, where V_{ij} and Z_{ij} denote the jth components of V_{ij} and Z_{ij}, respectively. We will assume that these vectors obey the following conditions:

Condition V: *There exists a constant $C_v > 0$ such that for all $j = 1, \dots, p$, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left[V_{ij}^4 + Z_{ij}^4\right] \le C_v B_n^2.$$

Condition P: *There exists a constant $C_p \ge 1$ such that for all $i = 1, \dots, n$, we have*

$$\mathrm{P}\Big(\|V_i\|_\infty \vee \|Z_i\|_\infty > C_p B_n \log(pn)\Big) \le 1/n^4.$$

Condition B: *There exists a constant $C_b > 0$ such that for all $i = 1, \dots, n$, we have*

$$\mathrm{E}\Big[\|V_i\|_\infty^8 + \|Z_i\|_\infty^8\Big] \le C_b B_n^8 \log^8(pn).$$

Condition A: *There exist constants $C_a > 0$ and $\delta \ge 0$ such that for all $(y, t) \in \mathbb{R}^p \times (0, \infty)$, we have*

$$\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \le y + t\right) - \mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i \le y\right) \le C_a\left(t\sqrt{\log p} + \delta\right).$$

Note that the constants C_v, C_p, C_b, and C_a appearing in these conditions are not supposed to be dependent on their indices, e.g. C_p here is not allowed to change with p; the indices are introduced with the only goal to differentiate between the constants.

The following is the main result of this section:

THEOREM 3.1 (Distributional Approximation via Iterative Randomized Lindeberg Method). *Suppose that Conditions V, P, B, and A are satisfied. In addition, suppose that*

$$\max_{1\le j,k\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathrm{E}[V_{ij}V_{ik}]-\mathrm{E}[Z_{ij}Z_{ik}])\right|\le C_m B_n\sqrt{\log(pn)} \tag{18}$$

and

$$\max_{1\le j,k,l\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathrm{E}[V_{ij}V_{ik}V_{il}]-\mathrm{E}[Z_{ij}Z_{ik}Z_{il}])\right|\le C_m B_n^2\sqrt{\log^3(pn)} \tag{19}$$

for some constant C_m. Then

$$\sup_{y\in\mathbb{R}^p}\left|\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i\le y\right)-\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\le y\right)\right|\le C\left(\left(\frac{B_n^2\log^5(pn)}{n}\right)^{1/4}+\delta\right),$$

where C is a constant depending only on C_v, C_p, C_b, C_a, and C_m.

REMARK 3.1 (On Sharpness of Theorem 3.1). We do not claim sharpness of Theorem 3.1 in the high-dimensional case $p\gg n$ (when p is fixed, the theorem is not sharp in view of the classical Berry-Esseen bound). On one hand, classical Edgeworth expansions in the low-dimensional case suggest that conditions like (19) should lead to better distributional approximation results than the corresponding Gaussian approximation results, which we do not observe in Theorem 3.1 since Theorem 2.1 gives the same dependence on both n and p for the Gaussian approximation. On the other hand, to the best of our knowledge, there exist no analogs of Edgeworth expansions in high dimensions. The question whether conditions like (19) can be used to improve distributional approximations (relative to the Gaussian approximations) thus remains open. ∎

To prove this result, we will need additional notation. For all $\epsilon\in\{0,1\}^n$, define

$$\varrho_\epsilon=\sup_{y\in\mathbb{R}^p}\left|\mathrm{P}\left(S_{n,\epsilon}^V\le y\right)-\mathrm{P}\left(S_n^Z\le y\right)\right|, \tag{20}$$

where

$$S_{n,\epsilon}^V=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_i V_i+(1-\epsilon_i)Z_i)\quad\text{and}\quad S_n^Z=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i.$$

We will replace ϵ with a certain sequence of random vectors $\epsilon^0,\dots,\epsilon^D\in\{0,1\}^n$, independent of $V_1,\dots,V_n,Z_1,\dots,Z_n$, and derive recursive bounds for ρ_{ϵ^d} for $d=0,\dots,D$, which lead to the desired bound in Theorem 3.1. Such a sequence of random vectors $\epsilon^0,\dots,\epsilon^D\in\{0,1\}^n$ is constructed as follow:

- Set $D=\lceil 4\log n\rceil+1$ and initialize $\epsilon^0=(1,\dots,1)$.
- Let $U_1,\dots,U_D$ be a sequence of independent uniform $[0,1]$ random variables that are independent of $V_1,\dots,V_n,Z_1,\dots,Z_n$.
- For $d=1,\dots,D$: conditionally on ϵ^{d-1} and $U_1,\dots,U_D$, set $\epsilon_i^d=0$ if $\epsilon_i^{d-1}=0$, and generate $\{\epsilon_i^d\}_{i\in I_{d-1}}$ with $I_{d-1}=\{i=1,\dots,n:\epsilon_i^{d-1}=1\}$ as i.i.d. Bernoulli(U_d) random variables.

It is not difficult to see that for each $d=1,\dots,D$, the random vector ϵ^d satisfies the following properties:

(i) for all $i=1,\dots,n$, $\epsilon_i^d=0$ if $\epsilon_i^{d-1}=0$, and

(ii) for $I_{d-1} = \{i = 1, \dots, n : \epsilon_i^{d-1} = 1\}$, the random variables $\{\epsilon_i^d\}_{i \in I_{d-1}}$ are exchangeable conditional on ϵ^{d-1} and satisfy

$$\mathrm{P}\left(\sum_{i \in I_{d-1}} \epsilon_i^d = s \mid \epsilon^{d-1} \right) = \frac{1}{|I_{d-1}| + 1}, \quad \text{for all } s = 0, \dots, |I_{d-1}|. \tag{21}$$

Indeed, to see that (21) holds, observe that, conditional on ϵ^{d-1} and U_d, $\sum_{i \in I_{d-1}} \epsilon_i^d$ follows the binomial distribution with parameters $|I_{d-1}|$ and (success probability) U_d, so that

$$\begin{aligned} \mathrm{P}\left(\sum_{i \in I_{d-1}} \epsilon_i^d = s \mid \epsilon^{d-1} \right) &= \binom{|I_{d-1}|}{s} \int_0^1 u^s (1-u)^{|I_{d-1}|-s} du \\ &= \binom{|I_{d-1}|}{s} \frac{s!(|I_{d-1}| - s)!}{(|I_{d-1}| + 1)!} = \frac{1}{|I_{d-1}| + 1}. \end{aligned}$$

Also, two properties (i) and (ii) ensure that $S^V_{n,\epsilon^d} - n^{-1/2} \sum_{i \notin I_{d-1}} Z_i$ is the randomized Lindeberg interpolant between $n^{-1/2} \sum_{i \in I_{d-1}} V_i$ and $n^{-1/2} \sum_{i \in I_{d-1}} Z_i$; see Lemma I.2 and the discussion at the beginning of Step 1 of the proof of Lemma 3.1.

Further, for all $i = 1, \dots, n$ and $j, k, l = 1, \dots, p$, define

$$\mathcal{E}^V_{i,jk} = \mathrm{E}[V_{ij} V_{ik}], \ \mathcal{E}^V_{i,jkl} = \mathrm{E}[V_{ij} V_{ik} V_{il}],$$

$$\mathcal{E}^Z_{i,jk} = \mathrm{E}[Z_{ij} Z_{ik}], \ \mathcal{E}^Z_{i,jkl} = \mathrm{E}[Z_{ij} Z_{ik} Z_{il}].$$

For all $n \geq 1$ and $d = 0, \dots, D$, let $\mathcal{B}_{n,1,d}$ and $\mathcal{B}_{n,2,d}$ be some strictly positive constants, and define the event $\mathcal{A}_d$ by

$$\begin{aligned} \mathcal{A}_d = &\left\{ \max_{1 \leq j,k \leq p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i^d (\mathcal{E}^V_{i,jk} - \mathcal{E}^Z_{i,jk}) \right| \leq \mathcal{B}_{n,1,d} \right\} \\ &\bigcap \left\{ \max_{1 \leq j,k,l \leq p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \epsilon_i^d (\mathcal{E}^V_{i,jkl} - \mathcal{E}^Z_{i,jkl}) \right| \leq \mathcal{B}_{n,2,d} \right\}. \end{aligned}$$

The proof of Theorem 3.1 proceeds as follows. In Lemma 3.1 and Corollary 3.1, we establish a recursive inequality for $\mathrm{E}[\varrho_{\epsilon^d} 1\{\mathcal{A}_d\}]$, $d = 0, \dots, D$. Next, we show in Lemma 3.2 that $\mathrm{E}[\varrho_{\epsilon^D} 1\{\mathcal{A}_D\}]$ is bounded by $1/n$. Then, we use an induction argument backward to derive a bound for $\mathrm{E}[\varrho_{\epsilon^0} 1\{\mathcal{A}_0\}]$. Since $\epsilon_i^0 = 1$ for all i, this gives the claim of the theorem once we appropriately choose the constants $\mathcal{B}_{n,1,d}$ and $\mathcal{B}_{n,2,d}$. The proof of Lemma 3.1 is long and is given in Appendix D of the Supplemental Material.

The derivation of the recursive inequality is based on connecting S^V_{n,ϵ^d} with S^Z_n by the randomized Lindeberg method originally developed by [22]. A similar approach was used in [12, 15] to connect S^V_{n,ϵ^0} with G, where the Slepian–Stein method was applied instead. Unlike the latter approach, the randomized Lindeberg method allows us to match the moments of S^V_{n,ϵ^d} and S^Z_n up to the third order rather than the second order. This leads to improvement on the power of $\log(pn)$ factors. In addition, we incorporate a smoothing effect induced by Z_i via Condition A into our argument. This along with the higher-order moment matching lead to improvement on the power of the sample size n.

LEMMA 3.1. *Suppose that Conditions V, P, B, and A are satisfied. Then for any $d = 0, \dots, D-1$ and any constant $\phi > 0$ such that*

$$C_p B_n \phi \log^2(pn) \leq \sqrt{n}, \tag{22}$$

we have on the event $\mathcal{A}_d$*,*

$$\varrho_{\epsilon^d} \lesssim \frac{\sqrt{\log p}}{\phi} + \delta + \frac{B_n^2\phi^4\log^5(pn)}{n^2} + \left(\mathrm{E}[\varrho_{\epsilon^{d+1}} \mid \epsilon^d] + \frac{\sqrt{\log p}}{\phi} + \delta\right) \times \left(\frac{\mathcal{B}_{n,1,d}\phi^2\log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2,d}\phi^3\log^2 p}{n} + \frac{B_n^2\phi^4\log^3(pn)}{n}\right)$$

up to a constant depending only on C_v*,* C_p*,* C_b*, and* C_a*.*

REMARK 3.2 (Choice of ϕ). We will choose ϕ to depend on n via $n^{1/4}$ when applying this lemma. ∎

COROLLARY 3.1. *Suppose that all assumptions of Lemma 3.1 are satisfied. Then there exists a constant* $K > 0$ *depending only on* C_v*,* C_p*, and* C_b *such that for all* $d = 0, \dots, D-1$*, if* $\mathcal{B}_{n,1,d+1} \geq \mathcal{B}_{n,1,d} + KB_n\log^{1/2}(pn)$ *and* $\mathcal{B}_{n,2,d+1} \geq \mathcal{B}_{n,2,d} + KB_n^2\log^{3/2}(pn)$*, then for any constant* $\phi > 0$ *satisfying (22), we have*

$$\mathrm{E}[\varrho_{\epsilon^d}1\{\mathcal{A}_d\}] \lesssim \frac{\sqrt{\log p}}{\phi} + \delta + \frac{B_n^2\phi^4\log^5(pn)}{n^2} + \left(\mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + \frac{\sqrt{\log p}}{\phi} + \delta\right) \times \left(\frac{\mathcal{B}_{n,1,d}\phi^2\log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2,d}\phi^3\log^2 p}{n} + \frac{B_n^2\phi^4\log^3(pn)}{n}\right) \tag{23}$$

up to a constant depending only on C_v*,* C_p*,* C_b*, and* C_a*.*

PROOF. Since we assume throughout the paper that $p \geq 2$, the conclusion is trivial if $\phi < 1$. We will therefore assume in the proof that $\phi \geq 1$. In turn, $\phi \geq 1$ together with (22) imply that

$$C_pB_n\log^2(pn) \leq \sqrt{n}. \tag{24}$$

This condition will be useful in the proof.

Fix $d = 0, \dots, D-1$. Then, given that $\mathcal{A}_d$ depends only on ϵ^d, we have by Lemma 3.1 that

$$\mathrm{E}[\varrho_{\epsilon^d}1\{\mathcal{A}_d\}] \lesssim \frac{\sqrt{\log p}}{\phi} + \delta + \frac{B_n^2\phi^4\log^5(pn)}{n^2} + \left(\mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}] + \frac{\sqrt{\log p}}{\phi} + \delta\right) \times \left(\frac{\mathcal{B}_{n,1,d}\phi^2\log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2,d}\phi^3\log^2 p}{n} + \frac{B_n^2\phi^4\log^3(pn)}{n}\right)$$

up to a constant depending only on C_v, C_p, C_b, and C_a. Thus, given that (22) implies that $\sqrt{\log p}/\phi \geq 1/n$, the conclusion of the corollary will follow if we can show that

$$\mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}] \leq \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + 4/n. \tag{25}$$

To this end, we first observe that , as $\varrho_{\epsilon^{d+1}} \in [0,1]$,

$$\begin{aligned}
\mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}] &= \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}1\{\mathcal{A}_{d+1}\}] + \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}(1-1\{\mathcal{A}_{d+1}\})] \\
&\leq \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + \mathrm{E}[1\{\mathcal{A}_d\}(1-1\{\mathcal{A}_{d+1}\})] \\
&= \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + \underbrace{\mathrm{P}(\mathcal{A}_d) - \mathrm{P}(\mathcal{A}_d \cap \mathcal{A}_{d+1})}_{=\mathrm{P}(\mathcal{A}_d)(1-\mathrm{P}(\mathcal{A}_{d+1}|\mathcal{A}_d))} \\
&\leq \mathrm{E}[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + 1 - \mathrm{P}(\mathcal{A}_{d+1} \mid \mathcal{A}_d).
\end{aligned} \tag{26}$$

Moreover, by Lemma I.1 in the Supplemental Material, for all $j,k=1,\dots,p$ and $t>0$, we have

$$\mathrm{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d+1}(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d}(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\right| + t \mid \epsilon^d\right)$$
$$\le 2\exp\left(-\frac{nt^2}{32\sum_{i=1}^n(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})^2}\right) \le 2\exp\left(-\frac{t^2}{128C_vB_n^2}\right),$$

where the second inequality follows from Condition V. Applying this inequality with $t = 8B_n\sqrt{6C_v\log(pn)}$ and using the fact that

$$\max_{1\le j,k\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d}(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\right| \le \mathcal{B}_{n,1,d} \quad \text{on } \mathcal{A}_d,$$

we have by the union bound that for any $\mathcal{B}_{n,1,d+1} \ge \mathcal{B}_{n,1,d} + t$,

$$\mathrm{P}\left(\max_{1\le j,k\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d+1}(\mathcal{E}^V_{i,jk}-\mathcal{E}^Z_{i,jk})\right| > \mathcal{B}_{n,1,d+1} \mid \mathcal{A}_d\right) \le \frac{2p^2}{(pn)^3} \le \frac{2}{n}.$$

In addition, for all $i=1,\dots,n$ and $j,k,l=1,\dots,p$, setting $\tilde{V}_i = 1\{\|V_i\|_\infty \le C_pB_n\log(pn)\}$, we have that

$$\begin{aligned}|\mathcal{E}^V_{i,jkl}| &\le \mathrm{E}[|V_{ij}V_{ik}V_{il}|] = \mathrm{E}\left[\tilde{V}_i|V_{ij}V_{ik}V_{il}|\right] + \mathrm{E}\left[(1-\tilde{V}_i)|V_{ij}V_{ik}V_{il}|\right] \\ &\le C_pB_n\log(pn)\mathrm{E}[|V_{ij}V_{ik}|] + (\mathrm{E}[1-\tilde{V}_i])^{1/2}(\mathrm{E}[\|V_i\|_\infty^6])^{1/2} \\ &\le C_pB_n\log(pn)\mathrm{E}[|V_{ij}V_{ik}|] + C_b^{3/8}B_n^3\log^3(pn)/n^2\end{aligned} \tag{27}$$

and similarly

$$|\mathcal{E}^Z_{i,jkl}| \le C_pB_n\log(pn)\mathrm{E}[|Z_{ij}Z_{ik}|] + C_b^{3/8}B_n^3\log^3(pn)/n^2$$

by Conditions P and B. Hence, by Condition V and (24), there exists a constant C depending only on C_v, C_p, and C_b such that

$$\frac{32}{n}\sum_{i=1}^n(\mathcal{E}^V_{i,jkl}-\mathcal{E}^Z_{i,jkl})^2 \le CB_n^4\log^2(pn).$$

Thus, by the same argument as above, for all $j,k,l=1,\dots,p$ and $t>0$,

$$\mathrm{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d+1}(\mathcal{E}^V_{i,jkl}-\mathcal{E}^Z_{i,jkl})\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d}(\mathcal{E}^V_{i,jkl}-\mathcal{E}^Z_{i,jkl})\right| + t \mid \epsilon^d\right)$$
$$\le 2\exp\left(-\frac{nt^2}{32\sum_{i=1}^n(\mathcal{E}^V_{i,jkl}-\mathcal{E}^Z_{i,jkl})^2}\right) \le 2\exp\left(-\frac{t^2}{CB_n^4\log^2(pn)}\right).$$

Applying this inequality with $t = \sqrt{3C}B_n^2\log^{3/2}(pn)$ shows that for any $\mathcal{B}_{n,2,d+1} \ge \mathcal{B}_{n,2,d} + t$, we have

$$\mathrm{P}\left(\max_{1\le j,k,l\le p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \epsilon_i^{d+1}(\mathcal{E}^V_{i,jkl}-\mathcal{E}^Z_{i,jkl})\right| > \mathcal{B}_{n,2,d+1} \mid \mathcal{A}_d\right) \le \frac{2p^3}{(pn)^3} \le \frac{2}{n}.$$

Thus, $1-\mathrm{P}(\mathcal{A}_{d+1}\mid\mathcal{A}_d) \le 4/n$, which in combination with (26) implies (25) and completes the proof. ∎

LEMMA 3.2. *For any constant $\phi > 0$ such that (22) holds, we have* $\mathrm{E}[\varrho_{\epsilon^D} 1\{\mathcal{A}_D\}] \le 1/n$.

PROOF. Recall that $D = [4\log n] + 1$ and note that $\varrho_{\epsilon^D} = 0$ if $\epsilon^D = (0,\dots,0)'$. Moreover, by Markov's inequality,

$$\mathrm{P}(\epsilon^D \neq (0,\dots,0)') = \mathrm{P}\left(\sum_{i=1}^n \epsilon_i^D \ge 1\right) \le \mathrm{E}\left[\sum_{i=1}^n \epsilon_i^D\right] = \mathrm{E}\left[\mathrm{E}\left[\sum_{i=1}^n \epsilon_i^D \mid \sum_{i=1}^n \epsilon_i^{D-1}\right]\right]$$
$$= \mathrm{E}\left[\frac{1}{2}\sum_{i=1}^n \epsilon_i^{D-1}\right] = \cdots = \mathrm{E}\left[\frac{1}{2^D}\sum_{i=1}^n \epsilon_i^0\right] = \frac{n}{2^D} \le \frac{n}{2^{4\log n}} \le \frac{1}{n},$$

where the equalities on the second line follow from (21). Hence,

$$\mathrm{E}[\varrho_{\epsilon^D} 1\{\mathcal{A}_D\}] \le \mathrm{E}[\varrho_{\epsilon^D}] \le \mathrm{P}(\epsilon^D \neq (0,\dots,0)') \le 1/n,$$

as desired. ∎

PROOF OF THEOREM 3.1. Throughout the proof, we will assume that

$$C_p^4 B_n^2 \log^5(pn) \le n \tag{28}$$

since otherwise the conclusion of the theorem is trivial.

Let K be the constant from Corollary 3.1 and for all $d = 0,\dots,D$, define

$$\mathcal{B}_{n,1,d} = C_1(d+1)B_n \log^{1/2}(pn) \quad \text{and} \quad \mathcal{B}_{n,2,d} = C_1(d+1)B_n^2 \log^{3/2}(pn), \tag{29}$$

where $C_1 = C_m + K$, so that $\mathcal{A}_0$ holds by (18) and (19) and, in addition, the requirements of Corollary 3.1 on $\mathcal{B}_{n,1,d}$ and $\mathcal{B}_{n,2,d}$ also hold.

Now, for all $d = 0,\dots,D$, define

$$f_d = \inf\left\{x \ge 1 \colon \mathrm{E}[\varrho_{\epsilon^d} 1\{\mathcal{A}_d\}] \le x\left(\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4} + \delta\right)\right\}.$$

Note that $f_d < \infty$ because $\varrho_{\epsilon^d} \le 1$. Then, for all $d = 0,\dots,D-1$, apply Corollary 3.1 with

$$\phi = \phi_d = \frac{n^{1/4}}{B_n^{1/2}\log^{3/4}(pn)((d+1)f_{d+1})^{1/3}},$$

which satisfies the required condition (22) since we assume (28). Since

$$\frac{B_n^2\phi_d^4\log^5(pn)}{n^2} \le \frac{\log^2(pn)}{n} \le \frac{\log^{1/4}(pn)}{n^{1/4}} \le \frac{C_p B_n^{1/2}\log^{1/4}(pn)}{n^{1/4}}$$
$$\le \frac{C_p\sqrt{\log p}}{\phi_d} \le C_p((d+1)f_{d+1})^{1/3}\left(\frac{B_n^2\log^5(pn)}{n}\right)^{1/4},$$
$$\frac{\mathcal{B}_{n,1,d}\phi_d^2\log p}{\sqrt{n}} \le \frac{C_1(d+1)}{((d+1)f_{d+1})^{2/3}}, \quad \text{and}$$
$$\frac{\mathcal{B}_{n,2,d}\phi_d^3\log^2 p}{n} \bigvee \frac{B_n^2\phi_d^4\log^3(pn)}{n} \le \frac{C_1 \vee 1}{f_{d+1}},$$

we have by Corollary 3.1

$$\mathrm{E}[\rho_{\epsilon^d} 1\{\mathcal{A}_d\}] \le C_2\left(f_{d+1}^{2/3} + (d+1)^{2/3} + 1\right)\left(\left(\frac{B_n^2\log^5(pn)}{n}\right)^{1/4} + \delta\right)$$

for some constant $C_2 \geq 1$ depending only on C_v, C_p, C_b, C_a, and C_m. Hence,

$$f_d \leq C_2\Big(f_{d+1}^{2/3} + (d+1)^{2/3} + 1\Big), \quad \text{for all } d = 0, \ldots, D-1.$$

Here, we have $f_D = 1$ by Lemma 3.2 since $B_n \geq 1$ by assumption. Therefore, by a simple induction argument, we conclude that there exists a constant $C \geq 1$ depending only on C_2 such that

$$f_d \leq C(d+1), \quad \text{for all } d = 0, \ldots, D.$$

In particular, it follows that

$$\varrho_{\epsilon^0} 1\{\mathcal{A}_0\} = \mathrm{E}[\varrho_{\epsilon^0} 1\{\mathcal{A}_0\}] \leq C\left(\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4} + \delta\right).$$

Since $\mathcal{A}_0$ holds by construction, so that $1\{\mathcal{A}_0\} = 1$, the desired bound follows by combining this inequality and the definition of ϱ_{ϵ^0}. ∎

4. Stein Kernels and Gaussian Approximation. Let $C_b^2(\mathbb{R}^p)$ be the class of twice continuously differentiable functions φ on $\mathbb{R}^p$ such that φ and all its partial derivatives up to the second order are bounded where $p \geq 2$. Let V be a centered random vector in $\mathbb{R}^p$ and assume that there exists a measurable function $\tau : \mathbb{R}^p \to \mathbb{R}^{p \times p}$ such that

$$\sum_{j=1}^{p} \mathrm{E}[\partial_j \varphi(V) V_j] = \sum_{j,k=1}^{p} \mathrm{E}[\partial_{jk} \varphi(V) \tau_{jk}(V)]$$

for all $\varphi \in C_b^2(\mathbb{R}^p)$. This function τ is called a *Stein kernel* for the random vector V. Also, let Z be a centered Gaussian random vector in $\mathbb{R}^p$ with covariance matrix Σ.

THEOREM 4.1 (Gaussian Approximation via Stein Kernels). *If $\Sigma_{jj} \geq c$ for all $j = 1, \ldots, p$ and some constant $c > 0$, then*

$$\sup_{y \in \mathbb{R}^p} \Big|\mathrm{P}(V \leq y) - \mathrm{P}(Z \leq y)\Big| \leq C\Big(\Delta \log^2 p\Big)^{1/2},$$

where C is a constant depending only on c and $\Delta = \mathrm{E}\left[\max_{1 \leq j,k \leq p} |\tau_{jk}(V) - \Sigma_{jk}|\right]$.

REMARK 4.1. This theorem improves upon Proposition 4.1 in [29], which shows that

$$\sup_{y \in \mathbb{R}^p} \Big|\mathrm{P}(V \leq y) - \mathrm{P}(Z \leq y)\Big| \leq C\Big(\Delta \log^2 p\Big)^{1/3}$$

under the same conditions. ∎

Theorem 4.1 is proven in Appendix E of the Supplemental Material. It has two important corollaries. The first is Proposition 2.1, a sharp Gaussian-to-Gaussian comparison inequality stated in Section 2:

PROOF OF PROPOSITION 2.1. If V is a centered Gaussian random vector, then by the multivariate Stein identity, its Stein kernel coincides with its covariance matrix. Hence, Theorem 4.1 immediately implies the conclusion of Proposition 2.1. ∎

Second, combining Theorem 4.1 with Lemma 4.6 in [30] gives the following result:

COROLLARY 4.1 (Multiplier-Bootstrap-to-Gaussian Comparison). *Let $a_1,\dots,a_n$ be vectors in $\mathbb{R}^p$ such that*

$$\min_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n a_{ij}^2\ge c \quad \textit{and} \quad \max_{1\le j\le p}\frac{1}{n}\sum_{i=1}^n a_{ij}^4\le B^2$$

for some constants $c,B>0$. Also, let $\varepsilon_1,\dots,\varepsilon_n$ be independent $N(0,1)$ random variables. Moreover, for some constants $\alpha,\beta>0$, let $e_1,\dots,e_n$ be independent standardized Beta(α,β) random variables so that

$$\mathrm{E}[e_i]=0 \quad \textit{and} \quad \mathrm{E}[e_i^2]=1, \quad \textit{for all } i=1,\dots,n. \tag{30}$$

Then, for the random vectors

$$V=\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i a_i \quad \textit{and} \quad Z=\frac{1}{\sqrt{n}}\sum_{i=1}^n \varepsilon_i a_i$$

we have

$$\sup_{y\in\mathbb{R}^p}\bigl|\mathrm{P}(V\le y)-\mathrm{P}(Z\le y)\bigr|\le C\left(\frac{B^2\log^5 p}{n}\right)^{1/4}, \tag{31}$$

where C is a constant depending only on c, α and β.

PROOF. Recall that $\eta\sim\mathrm{Beta}(\alpha,\beta)$ has density function $f_{\alpha,\beta}(x)\propto x^{\alpha-1}(1-x)^{\beta-1}$ for $x\in[0,1]$, mean $\mu=\alpha/(\alpha+\beta)$, and variance $\sigma^2=\alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$. By definition, the common distribution of the random variables $e_1,\dots,e_n$ equals that of $(\eta-\mu)/\sigma$.

Define

$$\tau(x)=-\frac{\int_{-\mu/\sigma}^x sf(s)ds}{f(x)}=\frac{\int_x^{(1-\mu)/\sigma} sf(s)ds}{f(x)} \quad \text{for } x\in\left(-\frac{\mu}{\sigma},\frac{1-\mu}{\sigma}\right),$$

where $f(x)=\sigma f_{\alpha,\beta}(\sigma x+\mu)$ for $x\in\big(-\frac{\mu}{\sigma},\frac{1-\mu}{\sigma}\big)$ is the density function of $(\eta-\mu)/\sigma$. From L'Hospital's rule, there exists a constant C_1 depending only on α and β such that $|\tau(x)|\le C_1$ for all $x\in\big(-\frac{\mu}{\sigma},\frac{1-\mu}{\sigma}\big)$. Also, by integration by parts, $\mathrm{E}[e_1\varphi(e_1)]=\mathrm{E}[\varphi'(e_1)\tau(e_1)]$ for any continuously differentiable function $\varphi\colon\mathbb{R}\to\mathbb{R}$. Then, by Lemma 4.6 in [30], a Stein kernel τ^V for the random vector V satisfies

$$\mathrm{E}\left[\max_{1\le j,k\le p}\left|\tau_{jk}^V(V)-\frac{1}{n}\sum_{i=1}^n a_{ij}a_{ik}\right|\right]\le C_2\sqrt{\frac{\log p}{n}}\times\max_{1\le j\le p}\sqrt{\frac{1}{n}\sum_{i=1}^n a_{ij}^4}$$

for some constant C_2 depending only on C_1. The desired conclusion (31) follows from combining this bound with Theorem 4.1 and observing that $\mathrm{E}[Z_jZ_k]=n^{-1}\sum_{i=1}^n a_{ij}a_{ik}$ for all $j,k=1,\dots,p$. ∎

5. Proofs of Theorems 2.1 and 2.2. In this section, we provide proofs of Theorems 2.1 and 2.2. Proofs of Theorems 2.3 and 2.4 will be given in Appendices G and H of the Supplemental Material. To simplify notation, we write

$$\delta_n=\left(\frac{B_n^2\log^5(pn)}{n}\right)^{1/4} \quad \text{and} \quad \upsilon_n=\sqrt{\frac{B_n^2\log^3(pn)}{n}}.$$

Our proof strategy for Theorems 2.1 and 2.2 is summarized as follows. First, we consider the multiplier bootstrap statistic T_n^* with the weights e_i constructed from the standardized

Beta(α,β) distribution and parameters α and β chosen so that $\mathrm{E}[e_i^3] = 1$. Thanks to Corollary 4.1 and Proposition 2.1, we have Gaussian approximation to this statistic with the rate δ_n. This implies that Condition A in Section 3 is satisfied with $Z_i = e_i(X_i - \bar{X}_n)$ and $\delta = \delta_n$ due to the Gaussian anti-concentration inequality in Lemma J.3 of the Supplemental Material. In turn, the latter allows us to invoke Theorem 3.1, which gives the approximation to T_n by T_n^* with the rate δ_n. (Note that having $\mathrm{E}[e_i^3] = 1$ is important here since otherwise Theorem 3.1 would give a slower approximation rate.) Combining this result with the aforementioned Gaussian approximation for T_n^*, we obtain the Gaussian approximation for T_n with the rate δ_n. This is done in Lemma 5.3 and gives Theorem 2.1.

Second, we consider the empirical bootstrap statistic T_n^*. Since we now have the Gaussian approximation for T_n with the rate δ_n, it follows that Condition A is satisfied with $Z_i = X_i$ and $\delta = \delta_n$. Hence, applying Theorem 3.1 with $V_i = X_i^*$ and $Z_i = X_i$, we can verify the empirical bootstrap approximation for T_n with the rate δ_n. This is done in Lemma 5.5 and gives one part of Theorem 2.2.

Third, we consider the multiplier bootstrap statistic T_n^* with arbitrary weights e_i satisfying (12). By choosing parameters α and β appropriately, we can match the first three moments of these weights by weights constructed from the standardized Beta(α,β) distribution. Thus, yet another application of Theorem 3.1 allows us to link the distribution of any multiplier bootstrap statistic to the distribution of the multiplier bootstrap statistic with weights constructed from the standardized Beta(α,β) distribution and further, via Corollary 4.1 and Proposition 2.1, to the Gaussian distribution. This leads to the Gaussian approximation for the multiplier bootstrap statistic T_n^* with the rate δ_n. This is done in Lemma 5.6 and gives the other part of Theorem 2.2.

Before proceeding to the main body of the proofs, we present a few auxiliary results.

LEMMA 5.1. *Suppose that Condition E is satisfied. Then*

$$\max_{1\le i\le n} \|X_i\|_\infty \le 5B_n \log(pn) \tag{32}$$

with probability at least $1 - 1/(2n^4)$. *In addition,*

$$\max_{1\le i\le n} \mathrm{E}\left[\|X_i\|_\infty^8\right] \le CB_n^8 \log^8(pn),$$

where C *is a universal constant.*

PROOF. By the union bound, Markov's inequality, and Condition E, we have for any $x > 0$ that

$$\begin{aligned}\mathrm{P}\left(\max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}| > x\right) &\le pn \max_{1\le i\le n}\max_{1\le j\le p}\mathrm{P}(|X_{ij}| > x)\\ &\le pn \max_{1\le i\le n}\max_{1\le j\le p}\frac{\mathrm{E}[\exp(|X_{ij}|/B_n)]}{\exp(x/B_n)} \le 2pn\exp(-x/B_n).\end{aligned}$$

Substituting here $x = 5B_n\log(pn)$ gives the first asserted claim. The second asserted claim follows from combining Condition E, inequalities on page 95 in [39], and Lemma 2.2.2 in [39]. ∎

LEMMA 5.2. *Suppose that Conditions E and M are satisfied and set* $\tilde{X}_i = X_i - \bar{X}_n$ *for all* $i = 1,\dots,n$. *Then there exist a universal constant* $c \in (0,1]$ *and constants* $C > 0$ *and* $n_0 \in \mathbb{N}$ *depending only on* b_1 *and* b_2 *such that for all* $n \ge n_0$, *if the inequality*

$$B_n^2 \log^5(pn) \le cn \tag{33}$$

holds, then the following events hold jointly with probability at least $1 - 1/n - 3\upsilon_n$*:*

$$\frac{b_1^2}{2} \le \frac{1}{n}\sum_{i=1}^n \tilde{X}_{ij}^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n \tilde{X}_{ij}^4 \le 2B_n^2 b_2^2, \quad \text{for all } j = 1, \dots, p, \tag{34}$$

$$\max_{1\le j,k\le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{X}_{ij}\tilde{X}_{ik} - \mathrm{E}[X_{ij}X_{ik}]) \right| \le CB_n\sqrt{\log(pn)}, \tag{35}$$

$$\max_{1\le j,k,l\le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{X}_{ij}\tilde{X}_{ik}\tilde{X}_{il} - \mathrm{E}[X_{ij}X_{ik}X_{il}]) \right| \le CB_n^2\sqrt{\log^3(pn)}. \tag{36}$$

The proof of this lemma is rather standard but long, and so is deferred to Appendix F of the Supplemental Material.

LEMMA 5.3. *Suppose that Conditions E and M are satisfied. Then*

$$\sup_{x\in\mathbb{R}} |\mathrm{P}(T_n \le x) - \mathrm{P}(T_n^G \le x)| \le C\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}, \tag{37}$$

where C *is a constant depending only on* b_1 *and* b_2*.*

PROOF. Without loss of generality, we may assume that (33) holds and that n is large enough so that $n \ge n_0$ for n_0 from Lemma 5.2, since otherwise the conclusion of the lemma is trivial by taking C large enough. This will justify an application of Lemma 5.2 when needed. In addition, by again taking C large enough, we may assume that $1/n^4 + 2/n + 3\upsilon_n < 1$.

Let $\mathcal{A}_n$ be the event that (32) and (34)–(36) hold jointly. By Lemmas 5.1 and 5.2, $\mathrm{P}(\mathcal{A}_n) \ge 1 - 1/(2n^4) - 1/n - 3\upsilon_n > 0$. Further, let $e_1, \dots, e_n$ be independent standardized Beta$(1/2, 3/2)$ random variables, standardized in such a way that they have mean zero and unit variance (cf. Corollary 4.1), that are independent of $X_{1:n} = (X_1, \dots, X_n)$. It is not difficult to check that $\mathrm{E}[e_i^3] = 1$ for all $i = 1, \dots, n$.

Let T_n^* be the multiplier bootstrap statistic with weights $e_1, \dots, e_n$. Condition on $X_{1:n}$ such that $\mathcal{A}_n$ holds. Then, by Corollary 4.1 and the definition of $\mathcal{A}_n$, we have

$$\sup_{y\in\mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(X_i - \bar{X}_n) \le y \mid X_{1:n} \right) - \mathrm{P}(\hat{G} \le y \mid X_{1:n}) \right| \le C_1\delta_n, \tag{38}$$

while by Proposition 2.1, we have

$$\sup_{y\in\mathbb{R}^p} |\mathrm{P}(\hat{G} \le y \mid X_{1:n}) - \mathrm{P}(G \le y)| \le C_2\delta_n,$$

where C_1 and C_2 are constants depending only on b_1 and b_2.

Next, we shall invoke Theorem 3.1 to compare the distribution of T_n with the conditional distribution of T_n^*. Formally, let $Y_1, \dots, Y_n$ be independent copies of $X_1, \dots, X_n$ that are independent of $X_{1:n}$, and define T_n' by T_n with X_i's replaced by Y_i's. Then, $\mathrm{P}(T_n \le x) = \mathrm{P}(T_n' \le x \mid X_{1:n})$. Condition on $X_{1:n}$ such that $\mathcal{A}_n$ holds and apply Theorem 3.1 with $V_i = Y_i$ and $Z_i = e_i\tilde{X}_i$ for all $i = 1, \dots, n$. Since $\mathrm{E}[e_i] = 0$ and $\mathrm{E}[e_i^2] = \mathrm{E}[e_i^3] = 1$ for all $i = 1, \dots, n$, it is not difficult to see from the definition of $\mathcal{A}_n$ that Conditions V, P, and B, as well as inequalities (18) and (19) of Theorem 3.1 are satisfied with appropriate constants C_v, C_p, C_b, and C_m that depend only on b_1, b_2. It remains to verify Condition A in Theorem 3.1. Observe

that for any $y \in \mathbb{R}^p$ and $t > 0$,

(39)
$$\begin{aligned}
\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(X_i - \bar{X}_n) \le y + t \mid X_{1:n}\right) &\le \mathrm{P}\left(\hat{G} \le y + t \mid X_{1:n}\right) + C_1\delta_n \quad \text{(by (38))} \\
&\le \mathrm{P}\left(\hat{G} \le y \mid X_{1:n}\right) + K_1 t\sqrt{\log p} + C_1\delta_n \quad \text{(by Lemma J.3 and (34))} \\
&\le \mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n e_i(X_i - \bar{X}_n) \le y \mid X_{1:n}\right) + K_1 t\sqrt{\log p} + 2C_1\delta_n, \quad \text{(by (38))}
\end{aligned}$$

where $K_1 > 0$ is a constant depending only on b_1. Thus, applying Theorem 3.1, we conclude that

$$\sup_{x\in\mathbb{R}} |\mathrm{P}(T_n \le x) - \mathrm{P}(T_n^* \le x \mid X_{1:n})| = \sup_{x\in\mathbb{R}} |\mathrm{P}(T_n' \le x \mid X_{1:n}) - \mathrm{P}(T_n^* \le x \mid X_{1:n})| \le C_3\delta_n$$

for some constant C_3 depending only on b_1 and b_2. The asserted claim follows from these bounds via the triangle inequality by noting that the left-hand side of (37) is non-stochastic, so that if (37) holds with strictly positive probability (recall that $\mathrm{P}(\mathcal{A}_n) > 0$), then it holds with probability one. ∎

LEMMA 5.4. *Suppose that Conditions E and M are satisfied. Then for any $y \in \mathbb{R}^p$ and $t > 0$,*

$$\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le y + t\right) - \mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le y\right) \le C\left(t\sqrt{\log p} + \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}\right),$$

where C is a constant depending only on b_1 and b_2.

PROOF. Fix $y \in \mathbb{R}^p$ and $t > 0$. Then for some constant C depending only on b_1 and b_2,

$$\begin{aligned}
\mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le y + t\right) &\le \mathrm{P}(G \le y + t) + C\delta_n \le \mathrm{P}(G \le y) + Ct\sqrt{\log p} + C\delta_n \\
&\le \mathrm{P}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i \le y\right) + Ct\sqrt{\log p} + 2C\delta_n,
\end{aligned}$$

where the first and the third inequalities follow from Lemma 5.3 and the second from Lemma J.3 of the Supplemental Material. This gives the asserted claim. ∎

LEMMA 5.5. *Suppose that Conditions E and M are satisfied and that the random variables $X_1^*, \ldots, X_n^*$ are obtained via the empirical bootstrap. Then with probability at least $1 - 2/n - 3\upsilon_n$, we have*

$$\sup_{x\in\mathbb{R}} |\mathrm{P}\left(T_n \le x\right) - \mathrm{P}\left(T_n^* \le x \mid X_{1:n}\right)| \le C\left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4},$$

where C is a constant depending only on b_1 and b_2.

PROOF. As before, we may assume that (33) holds and that n is large enough so that $n \ge n_0$ for n_0 from Lemma 5.2, since otherwise the conclusion of the lemma is trivial by taking C large enough. This will justify an application of Lemma 5.2 when needed.

Let $Y_1, \dots, Y_n$ be vectors in $\mathbb{R}^p$ such that

$$\|Y_i\|_\infty \le 10 B_n \log(pn) \quad \text{for all } i = 1, \dots, n, \tag{40}$$

$$b_1^2/2 \le \frac{1}{n}\sum_{i=1}^n Y_{ij}^2 \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^n Y_{ij}^4 \le 2B_n^2 b_2^2, \quad \text{for all } j = 1, \dots, p, \tag{41}$$

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{ij}Y_{ik} - \mathrm{E}[X_{ij}X_{ik}]) \right| \le C_m B_n \sqrt{\log(pn)}, \tag{42}$$

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{ij}Y_{ik}Y_{il} - \mathrm{E}[X_{ij}X_{ik}X_{il}]) \right| \le C_m B_n^2 \sqrt{\log^3(pn)}, \tag{43}$$

where C_m is the constant C from Lemma 5.2. Also, let $Y_1^*, \dots, Y_n^*$ be independent random vectors with each Y_i^* having uniform distribution on $\{Y_1, \dots, Y_n\}$.

To prove the asserted claim, we will apply Theorem 3.1 with $V_i = Y_i^*$ and $Z_i = X_i$ for all $i = 1, \dots, n$. Conditions V, P, and B with constants C_v, C_p, and C_b depending only on b_1 and b_2 follow immediately from Conditions E and M, Lemma 5.1, and the inequalities in (40) and (41). Also, Condition A with $\delta = \delta_n$ and C_a depending only on b_1 and b_2 follows from Lemma 5.4. Hence, an application of Theorem 3.1 is justified if we can verify (18) and (19) but these inequalities follow from (42) and (43) by noting that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n (\mathrm{E}[V_{ij}V_{ik}] - Y_{ij}Y_{ik}) = 0 \quad \text{and} \quad \frac{1}{\sqrt{n}}\sum_{i=1}^n (\mathrm{E}[V_{ij}V_{ik}V_{il}] - Y_{ij}Y_{ik}Y_{il}) = 0$$

for all $j, k, l = 1, \dots, p$. Now, applying Theorem 3.1 shows that for all $y \in \mathbb{R}^p$, we have

$$\left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n V_i \le y \right) - \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \le y \right) \right| \le K_1 \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4}$$

for some constant K_1 depending only on b_1, b_2, and C_m. The asserted claim follows from this bound by setting $Y_i = X_i - \bar{X}_n$ for all $i = 1, \dots, n$, and noting that in this case (40) holds with probability at least $1 - 1/(2n^4)$ by Lemma 5.1 and (41), (42), and (43) hold jointly with probability at least $1 - 1/n - 3\upsilon_n$ by Lemma 5.2. ∎

LEMMA 5.6. *Suppose that Conditions E and M are satisfied and that the random variables $X_1^*, \dots, X_n^*$ are obtained via the multiplier bootstrap with weights $e_1, \dots, e_n$ satisfying (12). Then with probability at least $1 - 2/n - 3\upsilon_n$, we have*

$$\sup_{x \in \mathbb{R}} |\mathrm{P}(T_n \le x) - \mathrm{P}(T_n^* \le x \mid X_{1:n})| \le C \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4},$$

where C is a constant depending only on $\mathrm{E}[e_1^3]$, b_1 and b_2.

REMARK 5.1. The constant C in this result depends on $\mathrm{E}[e_1^3]$ continuously, and so we can take C independent of $\mathrm{E}[e_1^3]$ under the implicitly maintained assumption that (12) holds.

PROOF. As before, we may assume that (33) holds and that n is large enough so that $n \ge n_0$ for n_0 from Lemma 5.2, since otherwise the conclusion of the lemma is trivial by taking C large enough. This will justify an application of Lemma 5.2 when needed.

Let $\mathcal{A}_n$ be the event that (32) and (34)–(36) hold jointly. By Lemmas 5.1 and 5.2, we have $\mathrm{P}(\mathcal{A}_n) \geq 1 - 2/n - 3\upsilon_n$. Moreover, by Proposition 2.1,

$$\sup_{y \in \mathbb{R}^p} |\mathrm{P}(\hat{G} \leq y \mid X_{1:n}) - \mathrm{P}(G \leq y)| \leq C_1 \delta_n \tag{44}$$

on the event $\mathcal{A}_n$, where C_1 is a constant depending only on b_1 and b_2.

Next, we claim that the case with $\sigma_e > 0$ can be reduced to the case with $\sigma_e = 0$ (and the constant 3 appearing in (12) replaced by some other universal constant). To prove this claim, define random variables $e'_1, \ldots, e'_n$ as in Corollary 4.1 with $\alpha = \beta = 1$ such that they are independent of everything else. Then on the event $\mathcal{A}_n$, by Corollary 4.1, we have that

$$\sup_{y \in \mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e'_i \tilde{X}_i \leq y \mid X_{1:n} \right) - \mathrm{P}\left(\frac{1}{\sigma_e \sqrt{n}} \sum_{i=1}^n e_{i,1} \tilde{X}_i \leq y \mid X_{1:n} \right) \right| \leq C_2 \delta_n,$$

where $\tilde{X}_i = X_i - \bar{X}_n$ for all $i = 1, \ldots, n$ and C_2 is a constant depending only on b_1 and b_2. Therefore, noting that the sequences $\{e_{i,1}\}_{i=1}^n$, $\{e_{i,2}\}_{i=1}^n$, and $\{e'_i\}_{i=1}^n$ are independent, we have on $\mathcal{A}_n$ that

$$\begin{aligned}
&\sup_{y \in \mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \tilde{X}_i \leq y \mid X_{1:n} \right) - \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (\sigma_e e'_i + e_{i,2}) \tilde{X}_i \leq y \mid X_{1:n} \right) \right| \\
&\leq \mathrm{E}\left[\sup_{y \in \mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e_{i,1} \tilde{X}_i \leq y \mid X_{1:n}, \{e_{i,2}\}_{i=1}^n \right) \right.\right. \\
&\qquad \left.\left. - \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \sigma_e e'_i \tilde{X}_i \leq y \mid X_{1:n}, \{e_{i,2}\}_{i=1}^n \right) \right| \mid X_{1:n} \right] \leq C_2 \delta_n.
\end{aligned}$$

Thus, it suffices to prove the asserted claim with e_i's replaced by $\sigma_e e'_i + e_{i,2}$'s, which are bounded by a universal constant (note that $\sigma_e \leq 1$ since e_i has unit variance).

Further, define the function $f \colon (0,1) \to \mathbb{R}$ by

$$f(\alpha) = \frac{2\sqrt{2}(1-2\alpha)}{3\sqrt{\alpha(1-\alpha)}}, \qquad \text{for all } \alpha \in (0,1).$$

One can directly check that $f(\alpha)$ is the skewness of the Beta$(\alpha, 1-\alpha)$ distribution for all $\alpha \in (0,1)$. Since $\lim_{\alpha \to 0} f(\alpha) = \infty$, $\lim_{\alpha \to 1} f(\alpha) = -\infty$ and f is continuous, there is an $\alpha^* \in (0,1)$ satisfying $f(\alpha^*) = \mathrm{E}[e_1^3]$. We define random variables $\tilde{e}_1, \ldots, \tilde{e}_n$ as in Corollary 4.1 with $\alpha = \alpha^*$ and $\beta = 1 - \alpha^*$ such that they are independent of everything else. It is then easy to check that $\mathrm{E}[\tilde{e}_i] = 0$, $\mathrm{E}[\tilde{e}_i^2] = 1$, and $\mathrm{E}[\tilde{e}_i^3] = \mathrm{E}[e_i^3]$ for all $i = 1, \ldots, n$. Also, applying Corollary 4.1, we have on $\mathcal{A}_n$ that

$$\sup_{y \in \mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{e}_i \tilde{X}_i \leq y \mid X_{1:n} \right) - \mathrm{P}\left(\hat{G} \leq y \mid X_{1:n} \right) \right| \leq C_3 \delta_n, \tag{45}$$

where C_3 is a constant depending only on α^*, b_1 and b_2.

We now apply Theorem 3.1 with $V_i = e_i \tilde{X}_i$ and $Z_i = \tilde{e}_i \tilde{X}_i$ for all $i = 1, \ldots, n$ conditional on $X_{1:n}$ on the event $\mathcal{A}_n$. Conditions V, P, and B with C_v, C_p, and C_b depending only on α^*, b_1 and b_2 follow immediately from the inequalities (32) and (34) and the boundedness of e_i's and $\tilde{e}_i$'s. Condition A with $\delta = \delta_n$ follows from (45) and the derivation in (39). Moreover, (18) and (19) are evident by construction. Thus, by Theorem 3.1, we have on $\mathcal{A}_n$ that

$$\sup_{y \in \mathbb{R}^p} \left| \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{e}_i \tilde{X}_i \leq y \mid X_{1:n} \right) - \mathrm{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \tilde{X}_i \leq y \mid X_{1:n} \right) \right| \leq C_4 \delta_n, \tag{46}$$

where C_4 is a constant depending only on α^*, b_1 and b_2. The asserted claim now follows from combining (44), (45), and (46) via the triangle inequality and using Lemma 5.3. ∎

We are now in the position to prove the main results from Section 2.

PROOF OF THEOREM 2.1. The asserted claim follows immediately from Lemma 5.3 by applying (37) with $x = c^G_{1-\alpha}$. ∎

PROOF OF THEOREM 2.2. Let C_1, C_2, and C_3 be the constants C in Lemmas 5.4, 5.5, and 5.6, respectively. Set

$$\beta_n = (1 \vee C_1 \vee C_2 \vee C_3) \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4}.$$

By Lemmas 5.5 and 5.6, we have $\sup_{x \in \mathbb{R}} |\mathrm{P}(T_n \le x) - \mathrm{P}(T_n^* \le x \mid X_{1:n})| \le \beta_n$ with probability at least $1 - 2/n - 3\upsilon_n$. Hence, letting $c_{1-\gamma}$ be the $(1-\gamma)$th quantile of T_n for all $\gamma \in (0,1)$, we have with the same probability that

$$\mathrm{P}(T_n^* \le c_{1-\alpha+\beta_n} \mid X_{1:n}) \ge \mathrm{P}(T_n \le c_{1-\alpha+\beta_n}) - \beta_n \ge 1 - \alpha, \quad \text{and}$$

$$\begin{aligned} \mathrm{P}(T_n^* \le c_{1-\alpha-3\beta_n} \mid X_{1:n}) &\le \mathrm{P}(T_n \le c_{1-\alpha-3\beta_n}) + \beta_n \\ &\le 1 - \alpha - 2\beta_n + C_1 \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4} < 1 - \alpha, \end{aligned}$$

where the second inequality follows from Lemma 5.4. Therefore,

$$\mathrm{P}(c_{1-\alpha-3\beta_n} < c^B_{1-\alpha} \le c_{1-\alpha+\beta_n}) \ge 1 - 2/n - 3\upsilon_n \ge 1 - 5\upsilon_n,$$

so that

$$\mathrm{P}(T_n > c^B_{1-\alpha}) \le \mathrm{P}(T_n > c_{1-\alpha-3\beta_n}) + 5\upsilon_n \le \alpha + 3\beta_n + 5\upsilon_n \le \alpha + 8\beta_n \quad \text{and}$$

$$\begin{aligned} \mathrm{P}(T_n > c^B_{1-\alpha}) &\ge \mathrm{P}(T_n > c_{1-\alpha+\beta_n}) - 5\upsilon_n \\ &\ge \alpha - \beta_n - C_1 \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4} - 5\upsilon_n \ge \alpha - 7\beta_n, \end{aligned}$$

where the second inequality follows from Lemma 5.4. Combining these inequalities gives the asserted claim. ∎

Acknowledgments. We are grateful to Tim Armstrong, Matias Cattaneo, Xiaohong Chen, and Tengyuan Liang for helpful discussions. We also thank seminar participants at the University of Pennsylvania and Yale University.

Funding. K. Kato is suport by the NSF DMS-1952306 and DMS-2014636.

SUPPLEMENTARY MATERIAL

The Supplementary Material contains proofs omitted in the main text as well as several technical tools, and the simulation results.

REFERENCES

[1] Andrews, D. and Shi, X. (2013). Inference based on conditional moment inequalities. *Econometrica* **81** 609-666.

[2] Bai, Y., Santos, A., and Shaikh, A. (2019). A practical method for testing many moment inequalities. *Becker Friedman Institute for Economics Working Paper*.

[3] Belloni, A., Chernozhukov, V., Chetverikov, D., Hansen, C., and Kato, K. (2018). High-dimensional econometrics and regularized GMM. *arXiv:1806.01888*.

[4] Belloni, A., Chernozhukov, V., Chetverikov, D., and Wei, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in Z-estimation framework. *Annals of Statistics* **46** 3643-3675.

[5] 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.

[6] Bentkus, V. (1985). Lower bounds for the rate of convergence in the central limit theorem in Banach spaces. *Litovskii Matematicheskii Sbornik* **25** 10-21.

[7] Canay, I., Romano, J. and Shaikh, A. (2016). Randomization tests under an approximate symmetry assumption. *Econometrica* **85** 1013-1030.

[8] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *Annals of Probability* **34** 2061-2076.

[9] Chen, X. (2018). Gaussian and bootstrap approximations for high-dimensional U-statistics and their applications. *Annals of Statistics* **46** 642-678.

[10] Chen, X. and Kato, K. (2019). Randomized incomplete U-statistics in high dimensions. *Annals of Statistics* **47** 3127-3156.

[11] Chen, X. and Kato, K. (2020). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probability Theory and Related Fields* **176** 1097-1163.

[12] Chernozhukov, V., Chetverikov, D. and Kato, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* **41** 2786-2819.

[13] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics* **42** 1787-1818.

[14] Chernozhukov, V., Chetverikov, D. and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* **162** 47-70.

[15] Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability* **45** 2309-2352.

[16] Chernozhukov, V., Chetverikov, D. and Kato, K. (2019). Inference on causal and structural parameters using many moment inequalities. *Review of Economic Studies* **86** 1867-1900.

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

[18] Chesher, A. and Rosen, A. (2019). Generalized instrumental variable models, methods, and applications. *Cemmap working paper CWP41/19*.

[19] Chetverikov, D. (2018). Adaptive tests of conditional moment inequalities. *Econometric Theory* **34** 186-227.

[20] Chetverikov, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory* **35** 729-776.

[21] Chetverikov, D., Wilhelm, D., and Kim, D. (2018). An adaptive test of stochastic monotonicity. *Cemmap working paper CWP24/18*.

[22] Deng, H. and Zhang, C.-H. (2020). Beyond Gaussian approximation: bootstrap for maxima of sums of independent random vectors. *Annals of Statistics* **48** 3643-3671.

[23] Fang, X. and Koike, Y. (2020). High-dimensional Central Limit Theorems by Stein's Method. To appear in *Annals of Applied Probability*.

[24] Fathi, M. (2020). Stein kernels and moment maps. *Annals of Probability* **47** 2172–2185.

[25] Hansen, P. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics* **23** 365-380.

[26] Hansen, P., Lunde, A. and Nason, J. (2011). The model confidence set. *Econometrica* **79** 453-497.

[27] Horowitz, J. (2001). The bootstrap. *Handbook of Econometrics, Volume 5* 3159-3228.

[28] Koike, Y. (2019a). Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. *Annals of Statistics* **47** 1663–1687.

[29] Koike, Y. (2019b). High-dimensional central limit theorems for homogeneous sums. *arXiv:1902.03809*.

[30] Koike, Y. (2021). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. *Japanese Journal of Statistics and Data Science* **1** 257–297.

[31] Koning, N. and Bekker, P. (2019). Exact testing of many moment inequalities against multiple violations. *arXiv:1904.12775*.

[32] Kuchibhotla, A. and Rinaldo, A. (2020). High-dimensional CLT for sums of non-degenerate random vectors: $n^{-1/2}$ rate. *arXiv:2009.13673*.

[33] Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics.

[34] Lopes, M. (2020). Central limit theorem and bootstrap approximation in high dimensions with near $1/\sqrt{n}$ rates. *arXiv:2009.06004*.

[35] Lopes, M., Lin, Z., and Müller, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Annals of Statistics* **48** 1214–1229.

[36] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* **21** 255-285.

[37] O'Donnell, R., Servedio, R., and Tan, L. (2019). Fooling polytopes. *STOC 2019: Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 614–625

[38] Romano, J., Shaikh, A. and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica* **82** 1979-2002.

[39] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

[40] Vershynin, R. (2018). *High-Dimensional Probability*. Cambridge University Press.

[41] Wainwright, M. J. (2019). *High-Dimensional Statistics*. Cambridge University Press.

[42] White, H. (2000). A reality check for data snooping. *Econometrica* **5** 1007-1126.

[43] Zhang, X. and Cheng, G. (2017). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675.

[44] Zhang, D. and Wu, W.-B. (2017). Gaussian approximation for high dimensional time series. *Annals of Statistics* **45** 1895-1919.