IMPROVED CENTRAL LIMIT THEOREM AND BOOTSTRAP APPROXIMATIONS IN HIGH DIMENSIONS

VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, KENGO KATO, AND YUTA KOIKE

ABSTRACT. 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 econometric 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.,

$$T_n = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_{ij} - \mu_j)$$
 (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 ([11]), in carrying out reality checks for data snooping and testing superior predictive ability ([34, 21]), in constructing model confidence sets ([22]), in testing conditional and/or many unconditional moment inequalities ([2, 17, 15, 27]), 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 ([18, 19]), in carrying out inference on generalized instrumental variable models ([16]), and in constructing Lepski-type procedures for adaptive estimation and inference in nonparametric problems ([12]); more references can be found in [20] and especially

Date: December 24, 2019.

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

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 [11, 14]. 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 variance-covariance matrix $n^{-1} \sum_{i=1}^n \mathrm{E}[(X_i - \mu)(X_i - \mu)']$ and let the critical value $c_{1-\alpha}$ be the $(1-\alpha)$ th quantile of $\max_{1 \leq j \leq p} G_j$. Then under mild regularity conditions,

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

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 [14].

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 [36, 35, 15], U-statistics [17, 9, 10], Malliavin calculus [18], and homogeneous sums [25]. 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 [14] in 2014 on arXiv. The problem can be decomposed into two parts: (i) 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 [20] 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 [29]) multiplier bootstraps:

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

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 $\log p = o(n^{1/7})$. Their intuition for the improved bound is that the empirical and Mammen's bootstrap can approximately match the moments of the sample mean up to the third order, while the Gaussian approximation can only match the moments up to the second order. However, the recent preprint by the fourth

author [26] 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

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

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 the 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 multiplier bootstrap methods with Gaussian and Rademacher weights.

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

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

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.

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

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

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

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

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 infinitesimal factor η may substantially reduce over-rejection. 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 infinitesimal 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 :

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

where t is a vector in \mathbb{R}^p , which reduces to (1) if we set $t = 0_p$. 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 [31] 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

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

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

To prove (4), we develop a novel, 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 use a new exponential inequality for weighted sums of exchangeable random variables and 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 [20]. In addition, we sharpen the Gaussian approximation bounds for the multiplier processes developed in [26] 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 [28], using a recent result from the computer science literature on pseudo-random number generators by O'Donnell, Servedio, and Tan, derived in [30], which provides an anticoncentration inequality for maxima of Rademacher processes. Finally, to prove (6) and (7), we apply the original randomized Lindeberg method as developed in [20].

The rest of the paper is organized as followed. In the next section, we present our main results. In Section 3, we conduct a small simulation study confirming that all bootstrap schemes considered in this paper perform well in finite samples and comparing their performance in the high-dimensional regime. In Section 4, we develop the iterative randomized Lindeberg method. In Section 5, we derive new bounds for the Gaussian approximations using Stein's kernels. In Section 6, we give the proofs of the main results. In Section 7, we obtain a new exponential inequality for weighted

sums of exchangeable random variables as well as several other useful results. In Section 8, for ease of reference, we collect lemmas on maximal, deviation, and anti-concentration inequalities taken from the literature.

1.1. **Notation.** For any vectors $x, y \in \mathbb{R}^p$ and any scalar $c \in \mathbb{R}$, we write $x \leq y$ if $x_j \leq y_j$ for all $j = 1, \ldots, p$ and we write x + c to denote a vector in \mathbb{R}^p whose jth component is $x_j + c$ for all $j = 1, \ldots, p$. Also, for any sequences of scalars $\{a_n\}_{n\geq 1}$ and $\{b_n\}_{n\geq 1}$ we write $a_n \lesssim b_n$ if $a_n \leq Cb_n$ for all $n \geq 1$ and some constant C. Moreover, for any random variable T and a constant $\gamma \in (0,1)$, we define the γ th quantile of T as $\inf\{t \in \mathbb{R}: \gamma \leq P(T \leq t)\}$.

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

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

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

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

Second, define the bootstrap critical value $c_{1-\alpha}^B$ as the $(1-\alpha)$ th quantile of the conditional distribution of

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

given the data X_1, \ldots, X_n , where X_1^*, \ldots, X_n^* is a bootstrap sample. Depending on the bootstrap type, define X_1^*, \ldots, X_n^* according to one of the following schemes. In the case of the empirical bootstrap, let X_1^*, \ldots, X_n^* be a sequence of i.i.d. random variables sampled from the uniform distribution on $\{X_1 - \bar{X}_n, \ldots, X_n - \bar{X}_n\}$, where $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ denotes the sample mean of the data X_1, \ldots, X_n . In the case of the multiplier bootstrap, let e_1, \ldots, e_n be a sequence of i.i.d. random variables with mean zero and variance one, referred to as weights, which are independent of X_1, \ldots, X_n , and define $X_i^* = e_i(X_i - \bar{X}_n)$ for all $i = 1, \ldots, n$.

Except for imposing mean zero and variance one, we have a lot flexibility in terms of selecting the distribution of the weights e_1, \ldots, e_n but a particularly important case is when these weights are such that

$$E[e_i^3] = 1, \text{ for all } i = 1, \dots, n.$$
 (12)

If this condition is satisfied, we obtain the third-order matching multiplier bootstrap mentioned in the Introduction. Other popular choices include Rademacher weights, with e_i 's having uniform distribution on $\{-1,1\}$, and Gaussian weights, with e_i 's having N(0,1) distribution. Note that neither Rademacher nor Gaussian weights satisfy (12). In addition to the assumption that the weights e_1, \ldots, e_n have mean zero and unit variance, to simplify presentation, we will also assume throughout the paper, without further notice, that e_i 's are such that

$$E[\exp(e_i^2/4)] \le 2$$
, for all $i = 1, ..., n$, (13)

which means that the weights e_1, \ldots, e_n are sub-Gaussian random variables with the Orlicz ψ_2 -norm bounded by 2. An example of the weights e_1, \ldots, e_n satisfying both (12) and (13) is given in Lemma 7.3 (but keep in mind when we consider multiplier bootstraps, we do not assume (12); we only assume that the weights are i.i.d. with mean zero, unit variance, and sub-Gaussian, so that (13) is satisfied).

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

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

where $\hat{G} \sim N(0_p, \widehat{\Sigma}_n)$ and

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

is an estimator of the unknown Σ_n . We therefore, for brevity, sometimes refer to both quantities as the Gaussian critical values.

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, and so we assume that $n \geq 3$, so that, in particular, $\log(pn) \geq 1$. In addition, since we are primarily interested in the case with large p, we assume that $p \geq 2$.

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. Also, let \mathcal{R}^m be the class of all measurable sets in \mathbb{R}^p . We will use the following conditions:

Condition M: For all j = 1, ..., p, we have

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

Condition E: For all i = 1, ..., n and j = 1, ..., p, we have

$$\mathbb{E}[\exp(X_{ij}^2/B_n^2)] \le 2.$$

Condition S: For all i = 1, ..., n and $A \in \mathbb{R}^m$, we have

$$P(X_i \in A) = P(-X_i \in A).$$

The first part of Condition M means that each component of the random vectors X_i is scaled properly. Given the first part, the second part of Condition M holds if, for example, all random variables X_{ij} are bounded in absolute values by B_n . Condition E implies that the random variables X_{ij} are sub-Gaussian with the Orlicz ψ_2 -norm bounded by B_n . This condition is satisfied if the tails of the distribution of each X_{ij} are lighter than those of the $N(0, B_n^2/4)$ distribution; see [33] for details. Condition S means that the distribution of each X_i is symmetric around the mean. Importantly, none of these conditions restrict the covariance matrices $E[X_i X_i']$, and so our results do not follow from the classical results in the empirical process theory.

We will always impose Conditions M and E but we will use Condition S to deal with the multiplier bootstrap with Rademacher weights only. In particular, we will be able to improve the results for this type of the bootstrap if Condition S is satisfied. In addition, we note that Condition E is slightly stronger than the corresponding conditions we previously used in [14], but this additional restriction is not substantial: it is possible to relax Condition E in exchange for a more complicated statement of the main results.

2.3. Main Results. Here, we present our main results. The first result gives 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 M and E are satisfied. Then

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

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

This result improves upon the bound in [26], who obtained a similar result with the rate 1/6 instead of 1/4. Note that since $t \in \mathbb{R}^p$ in the definition of T_n in (8) is arbitrary, the bound (15) can be equivalently stated as

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

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

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

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

To prove Theorem 2.1, we first apply Theorem 4.1 below, which allows, among other things, to establish a result similar to that in (15) but for the third-order matching multiplier bootstrap. Next, we use the Stein kernel method to demonstrate that the maxima of the third-order matching multiplier bootstrap processes can be well approximated in distribution by the maxima of appropriate Gaussian processes. Finally, we obtain (15) by combining two steps via the triangle inequality.

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

Theorem 2.2 (Bootstrap Approximation). Suppose that Conditions M and E are satisfied and that $c_{1-\alpha}^B$ is obtained via either the empirical or the multiplier bootstrap methods. Then

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

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

This theorem improves upon the bounds in [20], 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, \ldots, e_n to satisfy (12). We do use slightly stronger conditions than those in [20] but we emphasize that it is not possible to obtain (16) using existing techniques. We therefore develop a novel technique, which we call the iterative randomized Lindeberg method.

To explain the intuition behind this method, recall that, for any smooth function $g: \mathbb{R}^p \to \mathbb{R}$ and any two sequences of independent random vectors X_1, \ldots, X_n and Y_1, \ldots, 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 [7] for example. The randomized Lindeberg method, introduced in [20], 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 $E[g^{(k)}(Z_1 + \cdots + Z_n)]$, up to some approximation error, where $g^{(k)}$ is a vector of the kth partial derivatives of g and Z_1, \ldots, 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 $E[g^{(k)}(Y_1 + \cdots + Y_n)]$, we apply the randomized Lindeberg method once again to approximate $E[g^{(k)}(Z_1 + \cdots + Z_n)]$ by $E[g^{(k)}(Y_1 + \cdots + Y_n)]$. 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.

The third 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 M, E, and S are satisfied and that $c_{1-\alpha}^B$ is obtained via the multiplier bootstrap with Rademacher weights. Then

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

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 $c_{1-\alpha}^R$ such that $P(T_n > c_{1-\alpha}^R) = \alpha$, 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 [30].

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

Theorem 2.4 (Bootstrap Approximation with Infinitesimal Factors). Suppose that Conditions M and E are satisfied and let $\eta > 0$ be a constant. Then there exists a constant depending only b_1 , b_2 , and η such that

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

if $c_{1-\alpha}^B$ is obtained via either the empirical or the multiplier bootstrap with weights satisfying (12) and

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

if $c_{1-\alpha}^B$ is obtained via the multiplier bootstrap and (12) does not hold.

To prove this theorem, we use the randomized Lindeberg method but with an important simplification that the infinitesimal 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 infinitesimal factor η to the critical value $c_{1-\alpha}^B$ may substantially reduce over-rejection, with a minimal effect on the power of the test.

To conclude this section, we note that in many applications requiring asymptotic linearization, the statistic T_n may not be exactly equal to but rather be asymptotically equal to that in (8). In addition, the vectors X_1, \ldots, X_n , often representing the values of the influence function, may not be directly observed but have to be estimated. We therefore emphasize that all our results can be extended to cover these cases via the so-called many approximate means framework using the same arguments as those used in [3] but we have opted not to carry out the extension here for brevity of the paper.

3. Simulation Results

In this section, we present results of a small-scale Monte Carlo simulation study. The purpose of the simulation study is two-fold. First, it confirms that all approximation methods discussed in the previous section work well in finite samples. Second, it compares the relative performance of different methods in the high-dimensional regime.

We generate random vectors X_1, \ldots, X_n by setting

$$X_{ij} = F^{-1}(\Phi(Y_{ij})), \text{ for all } i = 1, \dots, n \text{ and } j = 1, \dots, p,$$
 (18)

where random vectors Y_1, \ldots, Y_n are sampled independently from the centered Gaussian distribution with variance-covariance matrix Σ such that $\Sigma_{jk} = \rho^{|j-k|}$ for all $j,k=1,\ldots,p$, Φ is the cdf of the N(0,1) distribution, and, depending on the experiment, F^{-1} is the quantile function of either the Weibull or the Gamma distribution. For both distributions, we set the scale parameter to be one but we set the shape parameter k to be either 2, 3, or 4 in the case of the Weibull distribution and either 1, 3, or 5 in the case of the Gamma distribution. Depending on the experiment, we set the correlation parameter ρ to be either 0, 0.25, 0.5, or 0.75. Also, we set n=400 and p to be either 400 or 800.

We refer to (18) as the case of asymmetric distributions. In addition, since we obtain better bounds for the multiplier bootstrap with Rademacher weights if Condition S is satisfied, we also consider the case of symmetric distributions by setting

$$X_{ij} = F^{-1}(\Phi(Y_{ij}^1)) - F^{-1}(\Phi(Y_{ij}^2)), \text{ for all } i = 1, \dots, n \text{ and } j = 1, \dots, p,$$

where Y_1^1, \ldots, Y_n^1 and Y_1^2, \ldots, Y_n^2 are two independent copies of Y_1, \ldots, Y_n . Since approximations are better in this case, to differentiate between different types of approximations, we replace the sample size n = 400 by n = 100 and we keep the same choices for all other parameters.

For all types of the bootstrap, we calculate the critical value $c_{1-\alpha}^B$ using 500 bootstrap samples. To implement the third-order matching multiplier bootstrap, we sample the weights e_i from the distribution constructed in Lemma 7.3 with $\gamma = 0.2$. In all cases, we set the nominal level $\alpha = 0.1$. We estimate each rejection probability $P(T_n > c_{1-\alpha}^B)$ using 20,000 simulations.

The results of our simulations for the Weibull and the Gamma distributions are presented in Tables 1 and 2, respectively, and can be summarized as follows. First, we observe similar patterns in both tables. Second, all methods perform well in most cases even though we consider relatively small sample sizes, with the exception of the multiplier bootstrap with Rademacher weights, which tends to substantially over-reject in the case of the Gamma distributions, especially with small k. Third, in the case of the asymmetric distributions, the empirical and the third-order matching multiplier bootstrap methods clearly outperform the multiplier bootstrap methods with Gaussian and Rademacher weights. This is especially clear, for example, in the case of the Gamma distribution with k=3 and p=400, where the rejection probabilities $P(T_n > c_{1-\alpha}^B)$ are about 0.09-0.10 for the empirical and the third-order matching multiplier bootstrap methods but are about 0.13 - 0.15 for the multiplier bootstrap methods with Gaussian and Rademacher weights. Fourth, the multiplier bootstrap method with Gaussian weights improves and becomes comparable to the empirical and the third-order matching bootstrap methods in the case of symmetric distributions. However, the multiplier bootstrap method with Rademacher weights improves substantially more and in overall gives the best results among all methods in this case. An especially striking example of this conclusion is the case of the Gamma distribution with k=1 and p=800, where the rejection probabilities $P(T_n > c_{1-\alpha}^B)$ are about 0.10-0.11 for the multiplier bootstrap method with Rademacher weights but are about 0.05 - 0.07 for all other bootstrap methods.

4. Iterative Randomized Lindeberg Method

In this section, we develop the iterative randomized Lindeberg method and derive a distributional approximation result, Theorem 4.1, using this method. We will use this result in the next section to prove our main results on the bootstrap approximations in high dimensions, as stated in Section 2.

Let $V_1, \ldots, V_n, Z_1, \ldots, Z_n$ be a sequence of independent random vectors in \mathbb{R}^p such that $\mathrm{E}[V_{ij}] = \mathrm{E}[Z_{ij}] = 0$ for all $i = 1, \ldots, n$ and $j = 1, \ldots, p$, 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, ..., p, we have

$$\frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[V_{ij}^{2} + Z_{ij}^{2}\right] \leq C_{v} \text{ and } \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left[V_{ij}^{4} + Z_{ij}^{4}\right] \leq C_{v} B_{n}^{2}.$$

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

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

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

$$E[||V_i||_{\infty}^8 + ||Z_i||_{\infty}^8] \le C_b B_n^8 \log^4(pn).$$

Condition A: There exists a constant $C_a > 0$ such that for all $(y,t) \in \mathbb{R}^p \times \mathbb{R}_+$, we have

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

We can now state the main result of this section:

Theorem 4.1 (Distributional Approximation via Iterative Randomized Lindeberg Method). Suppose that Conditions V, P, B, and A are satisfied. In addition, suppose that

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

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbb{E}[V_{ij}V_{ik}V_{il}] - \mathbb{E}[Z_{ij}Z_{ik}Z_{il}]) \right| \le C_m B_n^2 \sqrt{\log^3(pn)}$$
 (20)

for some constant C_m . Then

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

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

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

$$\varrho_{\epsilon} = \sup_{y \in \mathbb{R}^p} \left| P\left(S_{n,\epsilon}^V \le y \right) - P\left(S_n^Z \le y \right) \right|, \tag{21}$$

where

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

In addition, denote $\epsilon^0=(1,\ldots,1)'\in\mathbb{R}^n$ and for $D=[4\log n]+1$, define random vectors $\epsilon^1,\ldots,\epsilon^D\in\{0,1\}^n$ such that for all $d=1,\ldots,D$, the random vector ϵ^d obeys the following conditions: (i) for all $i=1,\ldots,n$, $\epsilon^d_i=0$ if $\epsilon^{d-1}_i=0$ and (ii) for $I_{d-1}=\{i=1,\ldots,n\colon\epsilon^{d-1}_i=1\}$, the random variables $\{\epsilon^d_i\}_{i\in I_{d-1}}$ are exchangeable conditional on ϵ^{d-1} and satisfy

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

This uniquely determines the joint distribution of $\epsilon^1, \ldots, \epsilon^D$. We will also assume that $\epsilon^1, \ldots, \epsilon^D$ are independent of $V_1, \ldots, V_n, Z_1, \ldots, Z_n$. Moreover, for all $i = 1, \ldots, n$ and $j, k, l = 1, \ldots, p$, denote

$$\begin{split} \mathcal{E}_{i,jk}^V &= \mathrm{E}[V_{ij}V_{ik}], \ \mathcal{E}_{i,jkl}^V = \mathrm{E}[V_{ij}V_{ik}V_{il}], \\ \mathcal{E}_{i,jk}^Z &= \mathrm{E}[Z_{ij}Z_{ik}], \ \mathcal{E}_{i,jkl}^Z = \mathrm{E}[Z_{ij}Z_{ik}Z_{il}]. \end{split}$$

Finally, for all $n \geq 1$ and $d = 0, \ldots, D$, let $\mathcal{B}_{n,1,d}$ and $\mathcal{B}_{n,2,d}$ be some strictly positive constants and let \mathcal{A}_d be the event that

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^d (\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z) \right| \le \mathcal{B}_{n,1,d}$$

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^d (\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z) \right| \le \mathcal{B}_{n,2,d}.$$

We then have the following results:

Lemma 4.1. Suppose that Conditions V, P, B, and A are satisfied. Then for any d = 0, ..., D-1 and any constant $\phi > 0$ such that

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

on the event A_d , we have

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

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

Proof. Since we assume throughout the paper that $p \geq 2$, the asserted claim 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

$$C_p B_n \log^2(pn) \le \sqrt{n}. \tag{23}$$

This condition will be useful in the proof.

Fix d = 0, ..., D-1 and $e^d \in \{0,1\}^n$ such that if $\epsilon^d = e^d$, then \mathcal{A}_d holds. All arguments in this proof will be conditional on $\epsilon^d = e^d$. For brevity of notation, however, we make this conditioning implicit and write $P(\cdot)$ and $E[\cdot]$ instead of $P(\cdot \mid \epsilon^d = e^d)$ and $E[\cdot \mid \epsilon^d = e^d]$, respectively.

Fix any five-times continuously differentiable and decreasing function $g_0 \colon \mathbb{R} \to \mathbb{R}$ such that (i) $g_0(t) \geq 0$ for all $t \in \mathbb{R}$, (ii) $g_0(t) = 0$ for all $t \geq 1$, and (iii) $g_0(t) = 1$ for all $t \leq 0$. For this function, there exists a constant $C_g > 0$ such that

$$\sup_{t \in \mathbb{R}} \left(|g_0^{(1)}(t)| \vee |g_0^{(2)}(t)| \vee |g_0^{(3)}(t)| \vee |g_0^{(4)}(t)| \vee |g_0^{(5)}(t)| \right) \le C_g.$$

In this proof, we will use the symbol \lesssim to denote inequalities that hold up to a constant depending only on C_v , C_p , C_b , C_a , and C_g . Since g_0 can be chosen to be universal, we say that the inequality for ϱ_{ϵ^d} in the statement of the lemma holds up to a constant depending only on C_v , C_p , C_b , and C_a .

Fix $\phi > 0$ and set $\beta = \phi \log p$. Define functions $g: \mathbb{R} \to \mathbb{R}$ and $F: \mathbb{R}^p \to \mathbb{R}$ by $g(t) = g_0(\phi t)$ for all $t \in \mathbb{R}$ and

$$F(w) = \beta^{-1} \log \left(\sum_{j=1}^{p} \exp(\beta w_j) \right), \text{ for all } w \in \mathbb{R}^p.$$

It is immediate that the function g satisfies

$$g(t) = \begin{cases} 1 & \text{if } t \le 0, \\ 0 & \text{if } t \ge \phi^{-1}. \end{cases}$$
 (24)

It is also straightforward to check that the function F has the following property:

$$\max_{1 \le j \le p} w_j \le F(w) \le \max_{1 \le j \le p} w_j + \phi^{-1}, \text{ for all } w \in \mathbb{R}^p;$$
 (25)

see [11] for details. Also, for all $y \in \mathbb{R}^p$, define the function $m^y \colon \mathbb{R}^p \to \mathbb{R}$ by

$$m^y(w) = q(F(w - y)), \text{ for all } w \in \mathbb{R}^p.$$

Below, we will need partial derivatives of m^y up to the fifth order. For brevity of notation, we will use indices to denote these derivatives. For example, for any $j, k, l, r, h = 1, \ldots, p$, we will write

$$m_{jklrh}^y(w) = \frac{\partial^5 m^y(w)}{\partial w_i \partial w_k \partial w_l \partial w_r \partial w_h}, \quad \text{for all } w \in \mathbb{R}^p.$$

Using straightforward but lengthy algebra, we can show that the function m^y has the following property: for all j, k, l, r, h = 1, ..., p, there exist functions $U^y_{jk} \colon \mathbb{R}^p \to \mathbb{R}, \ U^y_{jkl} \colon \mathbb{R}^p \to \mathbb{R}, \ U^y_{jklr} \colon \mathbb{R}^p \to \mathbb{R}, \ \text{and} \ U^y_{jklrh} \colon \mathbb{R}^p \to \mathbb{R} \ \text{such that}$ (i) for all $w \in \mathbb{R}^p$, we have

$$|m_{jk}^{y}(w)| \le U_{jk}^{y}(w), \quad |m_{jkl}^{y}(w)| \le U_{jkl}^{y}(w),$$
 (26)

$$|m_{iklr}^y(w)| \le U_{iklr}^y(w), \quad |m_{iklrh}^y(w)| \le U_{iklrh}^y(w), \tag{27}$$

(ii) for all $w_1 \in \mathbb{R}^p$ and $w_2 \in \mathbb{R}^p$ such that $\beta ||w_2||_{\infty} \leq 1$, we have

$$U_{jklr}^{y}(w_1 + w_2) \lesssim U_{jklr}^{y}(w_1), \quad U_{jklrh}^{y}(w_1 + w_2) \lesssim U_{jklrh}^{y}(w_2),$$
 (28)

and (iii) for all $w \in \mathbb{R}^p$,

$$\sum_{j,k=1}^{p} U_{jk}^{y}(w) \lesssim \phi^{2} \log p, \quad \sum_{j,k,l=1}^{p} U_{jkl}^{y}(w) \lesssim \phi^{3} \log^{2} p, \tag{29}$$

$$\sum_{j,k,l,r=1}^{p} U_{jklr}^{y}(w) \lesssim \phi^{4} \log^{3} p, \quad \sum_{j,k,l,r,h=1}^{p} U_{jklrh}^{y}(w) \lesssim \phi^{5} \log^{4} p.$$
 (30)

For example, we can set

$$U_{jk}^{y}(w) = C_{g}(\phi^{2} + \phi\beta) \frac{\exp(\beta(w_{j} - y_{j})) \exp(\beta(w_{k} - y_{k}))}{\left(\sum_{i=1}^{p} \exp(\beta(w_{i} - y_{i}))\right)^{2}} + C_{g}\phi\beta 1\{j = k\} \frac{\exp(\beta(w_{j} - y_{j}))}{\sum_{i=1}^{p} \exp(\beta(w_{i} - i))}, \text{ for all } w \in \mathbb{R}^{p};$$

see [11] and [14] for more details.

Further, for all $y \in \mathbb{R}^p$, define

$$\mathcal{I}^y = m^y (S_{n \epsilon^d}^V) - m^y (S_n^Z) \tag{31}$$

and

$$h^{y}(Y;x) = 1 \left\{ -x < \max_{1 \le j \le p} (Y_{j} - y_{j}) \le x \right\}, \text{ for all } x \ge 0 \text{ and } Y \in \mathbb{R}^{p}.$$
 (32)

Also, denote

$$W = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\epsilon_i^{d+1} V_i + (1 - \epsilon_i^{d+1}) Z_i \right), \tag{33}$$

for the random vector e^{d+1} introduced in front of the lemma.

For the rest of the proof, we proceed in five steps. In the first step, we show that

$$\sup_{y \in \mathbb{R}^p} |\mathcal{E}[\mathcal{I}^y]| \lesssim \frac{B_n^2 \phi^4 \log^4(pn)}{n^2} + \left(\mathcal{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \times \left(\frac{\mathcal{B}_{n,1,d} \phi^2 \log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2,d} \phi^3 \log^2 p}{n} + \frac{B_n^2 \phi^4 \log^3(pn)}{n} \right). \tag{34}$$

In the second step, we show that

$$\varrho_{\epsilon^d} \lesssim \frac{\sqrt{\log p}}{\phi} + \sup_{y \in \mathbb{R}^p} |\mathbf{E}[\mathcal{I}^y]|.$$
(35)

Combining two steps, we obtain the asserted claim. In Steps 3, 4, and 5, we provide some auxiliary calculations.

Step 1. Here, we prove (34). Recalling that $I_d = \{i = 1, ..., n : \epsilon_i^d = 1\}$, let \mathcal{S}_n be the set of all one-to-one functions mapping $\{1, ..., |I_d|\}$ to I_d , and

let σ be a random function with uniform distribution on S_n such that σ is independent of $V_1, \ldots, V_n, Z_1, \ldots, Z_n$, and ϵ^{d+1} .

Denote

$$W_i^{\sigma} = \frac{1}{\sqrt{n}} \sum_{j=1}^{i-1} V_{\sigma(j)} + \frac{1}{\sqrt{n}} \sum_{j=i+1}^{|I_d|} Z_{\sigma(j)} + \frac{1}{\sqrt{n}} \sum_{j \notin I_d} Z_j, \text{ for all } i = 1, \dots, |I_d|.$$

Note that for any function $m: \mathbb{R}^p \to \mathbb{R}$ and any $i \in I_d$, it follows from Lemma 7.2 that

$$E\left[\frac{\sigma^{-1}(i)}{|I_d|+1} m \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{V_i}{\sqrt{n}}\right) + \left(1 - \frac{\sigma^{-1}(i)}{|I_d|+1}\right) m \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{Z_i}{\sqrt{n}}\right)\right]$$

is equal to E[m(W)]. We will use this property extensively below without explicit mentioning.

Now, fix $y \in \mathbb{R}^p$ and observe that

$$\mathcal{I}^{y} = \sum_{i=1}^{|I_d|} \left(m^y \left(W_i^{\sigma} + \frac{V_{\sigma(i)}}{\sqrt{n}} \right) - m^y \left(W_i^{\sigma} + \frac{Z_{\sigma(i)}}{\sqrt{n}} \right) \right).$$

Hence, letting $f: [0,1] \to \mathbb{R}$ be a function defined by

$$f(t) = \sum_{i=1}^{|I_d|} \mathbb{E}\left[m^y \left(W_i^{\sigma} + \frac{tV_{\sigma(i)}}{\sqrt{n}}\right) - m^y \left(W_i^{\sigma} + \frac{tZ_{\sigma(i)}}{\sqrt{n}}\right)\right], \text{ for all } t \in [0, 1],$$

it follows that $E[\mathcal{I}^y] = f(1)$ and by Taylor's expansion,

$$f(1) = f(0) + f^{(1)}(0) + \frac{f^{(2)}(0)}{2} + \frac{f^{(3)}(0)}{6} + \frac{f^{(4)}(\tilde{t})}{24}$$
, where $\tilde{t} \in (0, 1)$.

Here, f(0) = 0 by construction and $f^{(1)}(0) = 0$ because $E[V_{ij}] = E[Z_{ij}] = 0$ for all $i \in I_d$ and $j = 1, \ldots, p$. We thus need to bound $|f^{(2)}(0)|$, $|f^{(3)}(0)|$, and $|f^{(4)}(\tilde{t})|$. To this end, we show in Steps 3, 4, and 5 that

$$|f^{(2)}(0)| \lesssim \frac{B_n^2 \phi^4 \log^4(pn)}{n^2} + \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \left(\frac{\mathcal{B}_{n,1,d} \phi^2 \log p}{\sqrt{n}} + \frac{B_n^2 \phi^4 \log^3(pn)}{n} \right), \quad (36)$$

$$|f^{(3)}(0)| \lesssim \frac{B_n^3 \phi^5 \log^5(pn)}{n^{5/2}} + \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \left(\frac{\mathcal{B}_{n,2,d} \phi^3 \log^2 p}{n} + \frac{B_n^3 \phi^5 \log^5(pn)}{n^{3/2}} \right), \quad (37)$$

and

$$|f^{(4)}(\tilde{t})| \lesssim \frac{B_n^2 \phi^4 \log^3 p}{n^2} + \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \frac{B_n^2 \phi^4 \log^3 p}{n},$$
 (38)

respectively. Combining these inequalities and using (22) gives (34) and completes Step 1.

Step 2. Here, we prove (35). Fix $y \in \mathbb{R}^p$ and observe that

$$\begin{split} & \mathbf{P}(S_{n,\epsilon^d}^V \leq y) \leq \mathbf{P}(F(S_{n,\epsilon^d}^V - y - \phi^{-1}) \leq 0) \leq \mathbf{E}[m^{y+\phi^{-1}}(S_{n,\epsilon^d}^V)] \\ & \leq \mathbf{E}[m^{y+\phi^{-1}}(S_n^Z)] + |\mathbf{E}[\mathcal{I}^{y+\phi^{-1}}]| \leq \mathbf{P}(S_n^Z \leq y + 2\phi^{-1}) + |\mathbf{E}[\mathcal{I}^{y+\phi^{-1}}]| \\ & \leq \mathbf{P}(S_n^Z \leq y) + 2C_a\phi^{-1}\sqrt{\log p} + |\mathbf{E}[\mathcal{I}^{y+\phi^{-1}}]|, \end{split}$$

where the first inequality follows from (25), the second from $m^{y+\phi^{-1}}(\cdot) = g(F(\cdot - y - \phi^{-1}))$ and (24), the third from (31), the fourth from (24) and (25), and the fifth from Condition A. Similarly,

$$\begin{split} & \mathrm{P}(S_{n,\epsilon^d}^V \leq y) = \mathrm{P}(S_{n,\epsilon^d}^V - y \leq 0) \\ & \geq \mathrm{P}(F(S_{n,\epsilon^d}^V - y + \phi^{-1}) \leq \phi^{-1}) \geq \mathrm{E}[m^{y-\phi^{-1}}(S_{n,\epsilon^d}^V)] \\ & \geq \mathrm{E}[m^{y-\phi^{-1}}(S_n^Z)] - |\mathrm{E}[\mathcal{I}^{y-\phi^{-1}}]| \geq \mathrm{P}(S_n^Z \leq y - 2\phi^{-1}) - |\mathrm{E}[\mathcal{I}^{y-\phi^{-1}}]| \\ & \geq \mathrm{P}(S_n^Z \leq y) - 2C_a\phi^{-1}\sqrt{\log p} - |\mathrm{E}[\mathcal{I}^{y-\phi^{-1}}]|. \end{split}$$

Combining the presented bounds gives (35) and completes Step 2.

Step 3. Here, we prove (36). We have

$$f^{(2)}(0) = \frac{1}{n} \sum_{i=1}^{|I_d|} \sum_{j,k=1}^p \mathbb{E} \left[m_{jk}^y(W_i^{\sigma}) (V_{\sigma(i)j} V_{\sigma(i)k} - Z_{\sigma(i)j} Z_{\sigma(i)k}) \right]$$

$$= \frac{1}{n} \sum_{i \in I_d} \sum_{j,k=1}^p \mathbb{E} \left[m_{jk}^y(W_{\sigma^{-1}(i)}^{\sigma}) (V_{ij} V_{ik} - Z_{ij} Z_{ik}) \right]$$

$$= \frac{1}{n} \sum_{i \in I_d} \sum_{j,k=1}^p \mathbb{E} \left[m_{jk}^y(W_{\sigma^{-1}(i)}^{\sigma}) \right] (\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z), \tag{39}$$

where the third line follows from observing that conditional on σ , $W^{\sigma}_{\sigma^{-1}(i)}$ is independent of $V_{ij}V_{ik} - Z_{ij}Z_{ik}$. Thus, denoting

$$\begin{split} R_{i,jk}^{\sigma} &= m_{jk}^{y}(W_{\sigma^{-1}(i)}^{\sigma}) - \frac{\sigma^{-1}(i)}{|I_{d}|+1} m_{jk}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{V_{i}}{\sqrt{n}}\right) \\ &- \left(1 - \frac{\sigma^{-1}(i)}{|I_{d}|+1}\right) m_{jk}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{Z_{i}}{\sqrt{n}}\right), \end{split}$$

for all $i \in I_d$ and $j, k = 1, \ldots, p$, we have $f^{(2)}(0) = \mathcal{I}_{2,1} + \mathcal{I}_{2,2}$, where

$$\mathcal{I}_{2,1} = \frac{1}{n} \sum_{i \in I} \sum_{j,k=1}^{p} E[m_{jk}^{y}(W)] (\mathcal{E}_{i,jk}^{V} - \mathcal{E}_{i,jk}^{Z})$$

and

$$\mathcal{I}_{2,2} = \frac{1}{n} \sum_{i \in I_d} \sum_{j,k=1}^{p} \mathbb{E}[R_{i,jk}^{\sigma}] (\mathcal{E}_{i,jk}^{V} - \mathcal{E}_{i,jk}^{Z})$$

by our discussion in the beginning of Step 1. To bound $\mathcal{I}_{2,1}$, we have

$$\begin{aligned} |\mathcal{I}_{2,1}| &\leq \sum_{j,k=1}^{p} \mathrm{E}[|m_{jk}^{y}(W)|] \max_{1 \leq j,k \leq p} \left| \frac{1}{n} \sum_{i=1}^{n} \epsilon_{i}^{d} (\mathcal{E}_{i,jk}^{V} - \mathcal{E}_{i,jk}^{Z}) \right| \\ &\leq \frac{\mathcal{B}_{n,1,d}}{\sqrt{n}} \sum_{j,k=1}^{p} \mathrm{E}[|m_{jk}^{y}(W)|] \end{aligned}$$

by the definition of \mathcal{A}_d in front of the lemma. In addition, by the definition of m^y , we have $m_{ik}^y(W) = 0$ if

$$\max_{1 \le j \le p} (W_j - y_j) \le -\phi^{-1} \quad \text{or} \quad \max_{1 \le j \le p} (W_j - y_j) > \phi^{-1},$$

which means that

$$m_{jk}^{y}(W) = h^{y}(W; \phi^{-1}) m_{jk}^{y}(W)$$
 (40)

by the definition of h^y in (32). Thus, since

$$\mathcal{P} = P\left(-\phi^{-1} < \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Z_{ij} - y_j) \le \phi^{-1}\right) \le \frac{2C_a \sqrt{\log p}}{\phi}$$

by Condition A, it follows that $\sum_{j,k=1}^{p} \mathrm{E}[|m_{jk}^{y}(W)|]$ is equal to

$$\sum_{j,k=1}^{p} \operatorname{E}[h(W;\phi^{-1})|m_{jk}^{y}(W)|] \leq \sum_{j,k=1}^{p} \operatorname{E}[h(W;\phi^{-1})U_{jk}(W)]$$

$$\lesssim (\phi^{2}\log p)\operatorname{P}\left(-\phi^{-1} < \max_{1\leq j\leq p}(W_{j} - y_{j}) \leq \phi^{-1}\right)$$

$$\leq (\phi^{2}\log p)\left(2\operatorname{E}[\varrho_{\epsilon^{d+1}}] + \mathcal{P}\right) \lesssim (\phi^{2}\log p)\left(\operatorname{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi}\right), \quad (41)$$

where the first inequality follows from (26), the second from (29), and the third from the definition of $\varrho_{\epsilon^{d+1}}$ in (21) and (33). Hence,

$$|\mathcal{I}_{2,1}| \lesssim \frac{\mathcal{B}_{n,1,d}\phi^2 \log p}{\sqrt{n}} \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right).$$

To bound $\mathcal{I}_{2,2}$, by another Taylor's expansion, for all $i \in I_d$ and $j, k = 1, \ldots, p$, we have

$$|\mathbf{E}[R_{i,jk}^{\sigma}]| \leq \sum_{l,r=1}^{p} \mathbf{E}\left[\left|m_{jklr}^{y}\left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_{i}}{\sqrt{n}}\right) \frac{V_{il}V_{ir}}{n}\right|\right] + \sum_{l,r=1}^{p} \mathbf{E}\left[\left|m_{jklr}^{y}\left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}Z_{i}}{\sqrt{n}}\right) \frac{Z_{il}Z_{ir}}{n}\right|\right]$$

for some $\hat{t} \in (0,1)$, possibly depending on i, j, and k, and so $|\mathcal{I}_{2,2}| \leq \mathcal{I}_{2,2,1} + \mathcal{I}_{2,2,2}$, where

$$\mathcal{I}_{2,2,1} = \frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \mathbb{E}\left[\left| m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\hat{t}V_i}{\sqrt{n}} \right) V_{il} V_{ir} \right|\right] \times |\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z|$$

and

$$\mathcal{I}_{2,2,2} = \frac{1}{n^2} \sum_{i \in I_d} \sum_{i,k,l,r=1}^p \mathbb{E}\left[\left| m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\hat{t}Z_i}{\sqrt{n}} \right) Z_{il} Z_{ir} \right|\right] \times |\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z|.$$

Below, we bound $\mathcal{I}_{2,2,1}$ and note that the same argument also applies to $\mathcal{I}_{2,2,2}$. Denote $x = C_p B_n \log(pn) / \sqrt{n} + \phi^{-1}$ and $\tilde{V}_i = 1\{||V_i||_{\infty} \leq C_p B_n \log(pn)\}$ for all $i \in I_d$. Then for all $i \in I_d$ and $j, k = 1, \ldots, p$, we have

$$\sum_{l,r=1}^{p} \operatorname{E}\left[\tilde{V}_{i} \left| m_{jklr}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_{i}}{\sqrt{n}} \right) V_{il} V_{ir} \right| \right] \\
= \sum_{l,r=1}^{p} \operatorname{E}\left[\tilde{V}_{i} h^{y} (W_{\sigma^{-1}(i)}^{\sigma}; x) \left| m_{jklr}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_{i}}{\sqrt{n}} \right) V_{il} V_{ir} \right| \right] \\
\leq \sum_{l,r=1}^{p} \operatorname{E}\left[\tilde{V}_{i} h^{y} (W_{\sigma^{-1}(i)}^{\sigma}; x) U_{jklr}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_{i}}{\sqrt{n}} \right) |V_{il} V_{ir}| \right] \\
\lesssim \sum_{l,r=1}^{p} \operatorname{E}\left[h^{y} (W_{\sigma^{-1}(i)}^{\sigma}; x) U_{jklr}^{y} (W_{\sigma^{-1}(i)}^{\sigma}) |V_{il} V_{ir}| \right], \tag{42}$$

where the equality follows from the same argument as that leading to (40), the first inequality follows from (27) and the second from (22) and (28). In turn, denoting $\tilde{Z}_i = \{\|Z_i\|_{\infty} \leq C_p B_n \log(pn)\}$, it follows that for all $l, r = 1, \ldots, p$, the expectation in (42) is equal to

$$\begin{split}
& \mathbf{E}\Big[h^{y}(W_{\sigma^{-1}(i)}^{\sigma};x)U_{jklr}^{y}(W_{\sigma^{-1}(i)}^{\sigma})\Big]\mathbf{E}[|V_{il}V_{ir}|] \\
& \lesssim \mathbf{E}\Big[\tilde{V}_{i}\tilde{Z}_{i}h^{y}(W_{\sigma^{-1}(i)}^{\sigma};x)U_{jklr}^{y}(W_{\sigma^{-1}(i)}^{\sigma})\Big]\mathbf{E}[|V_{il}V_{ir}|] \\
& \lesssim \mathbf{E}\Big[\tilde{V}_{i}\tilde{Z}_{i}h^{y}(W;2x)U_{jklr}^{y}(W_{\sigma^{-1}(i)}^{\sigma})\Big]\mathbf{E}[|V_{il}V_{ir}|] \\
& \lesssim \mathbf{E}\Big[h^{y}(W;2x)U_{jklr}^{y}(W)\Big]\mathbf{E}[|V_{il}V_{ir}|],
\end{split} \tag{43}$$

where the first inequality follows from Condition P, the second from the definitions of h^y in (32), W in (33), and $W^{\sigma}_{\sigma^{-1}(i)}$ in the beginning of Step 1,

and the third from (22) and (28). Thus,

$$\begin{split} &\frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \operatorname{E} \left[\tilde{V}_i \left| m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\hat{t}V_i}{\sqrt{n}} \right) V_{il} V_{ir} \right| \right] \times |\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z| \\ &\lesssim \frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \operatorname{E} \left[h^y(W;2x) U_{jklr}^y(W) \right] \operatorname{E}[|V_{il} V_{ir}|] \times |\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z| \\ &\lesssim \frac{B_n^2}{n} \sum_{j,k,l,r=1}^p \operatorname{E} \left[h^y(W;2x) U_{jklr}^y(W) \right] \lesssim \frac{B_n^2 \phi^4 \log^3 p}{n} \left(\operatorname{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \end{split}$$

where the second inequality follows from Condition V since by Hölder's inequality,

$$\max_{1 \le j,k,l,r \le p} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[|V_{il}V_{ir}|] \times |\mathcal{E}_{i,jk}^{V} - \mathcal{E}_{i,jk}^{Z}|$$

$$\lesssim \max_{1 \le j,k \le p} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[|V_{ij}V_{ik}|^{2}] + \max_{1 \le j,k \le p} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[|Z_{ij}Z_{ik}|^{2}] \lesssim B_{n}^{2},$$

and the third inequality follows from (22) and the same arguments as those leading to (41). In addition,

$$\frac{1}{n^{2}} \sum_{i \in I_{d}} \sum_{j,k,l,r=1}^{p} E\left[(1 - \tilde{V}_{i}) \middle| m_{jklr}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_{i}}{\sqrt{n}} \right) V_{il} V_{ir} \middle| \right] \times |\mathcal{E}_{i,jk}^{V} - \mathcal{E}_{i,jk}^{Z}| \\
\lesssim \frac{\phi^{4} \log^{3} p}{n} \sum_{i=1}^{n} E\left[(1 - \tilde{V}_{i}) ||V_{i}||_{\infty}^{2} \right] \lesssim \frac{B_{n}^{2} \phi^{4} \log^{4}(pn)}{n^{2}}, \tag{44}$$

where the first inequality follows from (27) and (30) and Condition V and the second from noting that $\mathrm{E}[(1-\tilde{V}_i)\|V_i\|_\infty^2] \leq (\mathrm{E}[1-\tilde{V}_i])^{1/2}(\mathrm{E}[\|V_i\|_\infty^4])^{1/2}$ by Hölder's inequality and using Conditions P and B. This shows that

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

and since the same bound holds for $\mathcal{I}_{2,2,2}$ as well, it follows that

$$|\mathcal{I}_{2,2}| \lesssim \frac{B_n^2 \phi^4 \log^3 p}{n} \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) + \frac{B_n^2 \phi^4 \log^4(pn)}{n^2}.$$

Combining the bounds on $\mathcal{I}_{2,1}$ and $\mathcal{I}_{2,2}$ gives (36) and completes Step 3.

Step 4. Here, we prove (37). We have

$$f^{(3)}(0) = \frac{1}{n^{3/2}} \sum_{i \in I, j} \sum_{k, l=1}^{p} E[m_{jkl}^{y}(W_{\sigma^{-1}(i)}^{\sigma})] (\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z})$$

by the same argument as that in (39). Further, denoting

$$R_{i,jkl}^{\sigma} = m_{jkl}^{y}(W_{\sigma^{-1}(i)}^{\sigma}) - \frac{\sigma^{-1}(i)}{|I_{d}| + 1} m_{jkl}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{V_{i}}{\sqrt{n}}\right) - \left(1 - \frac{\sigma^{-1}(i)}{|I_{d}| + 1}\right) m_{jkl}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{Z_{i}}{\sqrt{n}}\right),$$

for all $i \in I_d$ and j, k, l = 1, ..., p, we have $f^{(3)}(0) = \mathcal{I}_{3,1} + \mathcal{I}_{3,2}$, where

$$\mathcal{I}_{3,1} = \frac{1}{n^{3/2}} \sum_{i \in I_d} \sum_{j,k,l=1}^p E[m_{jkl}^y(W)] (\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z)$$

and

$$\mathcal{I}_{3,2} = \frac{1}{n^{3/2}} \sum_{i \in I_d} \sum_{j,k,l=1}^p \mathbb{E}[R_{i,jkl}^{\sigma}] (\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z}).$$

Here, $|\mathcal{I}_{3,1}|$ can be bounded using the same arguments as those used to bound $|\mathcal{I}_{2,1}|$ in the previous step. This gives

$$|\mathcal{I}_{3,1}| \lesssim \frac{\mathcal{B}_{n,2,d}\phi^3 \log^2 p}{n} \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right).$$

To bound $|\mathcal{I}_{3,2}|$, we have like in the case of $|\mathcal{I}_{2,2}|$ in the previous step that $|\mathcal{I}_{3,2}| \leq \mathcal{I}_{3,2,1} + \mathcal{I}_{3,2,2}$, where

$$\mathcal{I}_{3,2,1} = \frac{1}{n^{5/2}} \sum_{i \in I_d} \sum_{i,k,l,r,h=1}^p \mathbb{E}\left[\left| m_{jklrh}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\hat{t}V_i}{\sqrt{n}} \right) V_{ir} V_{ih} \right| \right] \times |\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z|$$

and

$$\mathcal{I}_{3,2,2} = \frac{1}{n^{5/2}} \sum_{i \in I_d} \sum_{j,k,l,r,h=1}^p \mathbb{E}\left[\left| m_{jklrh}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\hat{t}Z_i}{\sqrt{n}} \right) Z_{ir} Z_{ih} \right| \right] \times |\mathcal{E}_{i,jkl}^V - \mathcal{E}_{i,jkl}^Z|.$$

Further, since

$$|\mathcal{E}_{i,jkl}^{V}| \leq \mathrm{E}[|V_{ij}V_{ik}V_{il}|] = \mathrm{E}\left[\tilde{V}_{i}|V_{ij}V_{ik}V_{il}|\right] + \mathrm{E}\left[(1-\tilde{V}_{i})|V_{ij}V_{ik}V_{il}|\right]$$

$$\lesssim B_{n}\log(pn)\mathrm{E}[|V_{ij}V_{ik}|] + (\mathrm{E}[1-\tilde{V}_{i}])^{1/2}(\mathrm{E}[||V_{i}||_{\infty}^{6}])^{1/2}$$

$$\lesssim B_{n}\log(pn)\mathrm{E}[|V_{ij}V_{ik}|] + B_{n}^{3}\log^{3/2}(pn)/n^{2}$$
(45)

and similarly

$$|\mathcal{E}_{i,jkl}^Z| \lesssim B_n \log(pn) \mathbb{E}[|Z_{ij}Z_{ik}|] + B_n^3 \log^{3/2}(pn)/n^2$$

by Conditions P and B, we have by the same argument as in the previous step that

$$\frac{1}{n^{5/2}} \sum_{i \in I_d} \sum_{j,k,l,r,h=1}^{p} \operatorname{E}\left[\tilde{V}_i \left| m_{jklrh}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_i}{\sqrt{n}} \right) V_{ir} V_{ih} \right| \right] \times |\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z}| \\
\lesssim \frac{1}{n^{5/2}} \sum_{i \in I_d} \sum_{j,k,l,r,h=1}^{p} \operatorname{E}\left[h^{y}(W;2x) U_{jklrh}^{y}(W) \right] \operatorname{E}[|V_{ir} V_{ih}|] \times |\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z}| \\
\lesssim \left(\frac{B_n^3 \phi^5 \log^5(pn)}{n^{3/2}} + \frac{B_n^3 \phi^5 \log^{11/2}(pn)}{n^{7/2}} \right) \left(\operatorname{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) \\
\lesssim \frac{B_n^3 \phi^5 \log^5(pn)}{n^{3/2}} \left(\operatorname{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right),$$

where the last inequality follows from (23) since $B_n \geq 1$. Also, again like in the previous step,

$$\frac{1}{n^{5/2}} \sum_{i \in I_d} \sum_{j,k,l,r,h=1}^{p} \mathbb{E}\left[(1 - \tilde{V}_i) \left| m_{jklrh}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\hat{t}V_i}{\sqrt{n}} \right) V_{ir} V_{ih} \right| \right] \\
\times |\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z}| \lesssim \frac{B_n \phi^5 \log^4 p}{n^{3/2}} \sum_{i=1}^{n} \mathbb{E}\left[(1 - \tilde{V}_i) ||V_i||_{\infty}^2 \right] \lesssim \frac{B_n^3 \phi^5 \log^5(pn)}{n^{5/2}}$$

since by Condition V and Hölder's inequality,

$$|\mathcal{E}_{i,jkl}^{V}| \le \mathbb{E}[|V_{ij}V_{ik}V_{il}|] \le \left(\mathbb{E}[|V_{ij}V_{ik}|^2] \times \mathbb{E}[V_{il}^2]\right)^{1/2} \lesssim nB_n$$

and similarly $|\mathcal{E}_{i,jkl}^Z| \lesssim nB_n$. Thus,

$$\mathcal{I}_{3,2,1} \lesssim \frac{B_n^3 \phi^5 \log^5(pn)}{n^{3/2}} \Big(\mathrm{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \Big) + \frac{B_n^3 \phi^5 \log^5(pn)}{n^{5/2}}$$

and since the same bound holds for $\mathcal{I}_{3,2,2}$, it follows that

$$\mathcal{I}_{3,2} \lesssim \frac{B_n^3 \phi^5 \log^5(pn)}{n^{3/2}} \Big(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \Big) + \frac{B_n^5 \phi^5 \log^5(pn)}{n^{5/2}}.$$

Combining these bounds gives (37) and completes Step 4.

Step 5. Here, we prove (38). We have $f^{(4)}(\tilde{t}) = \mathcal{I}_{4,1} - \mathcal{I}_{4,2}$, where

$$\mathcal{I}_{4,1} = \frac{1}{n^2} \sum_{i \in I_d} \sum_{i,k,l,r=1}^p \mathbb{E}\left[m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\tilde{t}V_i}{\sqrt{n}}\right) V_{ij} V_{ik} V_{il} V_{ir}\right]$$

and

$$\mathcal{I}_{4,2} = \frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \mathbb{E}\left[m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\tilde{t}Z_i}{\sqrt{n}}\right) Z_{ij} Z_{ik} Z_{il} Z_{ir}\right].$$

Here, again denoting $x = C_p B_n \log(pn) / \sqrt{n} + \phi^{-1}$, we have

$$\frac{1}{n^{2}} \sum_{i \in I_{d}} \sum_{j,k,l,r=1}^{p} \operatorname{E}\left[\tilde{V}_{i} \left| m_{jklr}^{y} \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\tilde{t}V_{i}}{\sqrt{n}} \right) V_{ij} V_{ik} V_{il} V_{ir} \right| \right] \\
\lesssim \frac{1}{n^{2}} \sum_{i \in I_{d}} \sum_{j,k,l,r=1}^{p} \operatorname{E}\left[\tilde{V}_{i} h^{y} (W_{\sigma^{-1}(i)}^{\sigma}; x) U_{jklr}^{y} (W_{\sigma^{-1}(i)}^{\sigma}) | V_{ij} V_{ik} V_{il} V_{ir} | \right] \\
\leq \frac{1}{n^{2}} \sum_{i \in I_{d}} \sum_{j,k,l,r=1}^{p} \operatorname{E}\left[h^{y} (W_{\sigma^{-1}(i)}^{\sigma}; x) U_{jklr}^{y} (W_{\sigma^{-1}(i)}^{\sigma}) \right] \operatorname{E}[|V_{ij} V_{ik} V_{il} V_{ir} |],$$

where the first inequality follows from the same argument as that leading to (42). In addition, for all $i \in I_d$ and j, k, l, r = 1, ..., p, we have

$$\mathrm{E}\Big[h^{y}(W^{\sigma}_{\sigma^{-1}(i)};x)U^{y}_{jklr}(W^{\sigma}_{\sigma^{-1}(i)})\Big]\lesssim \mathrm{E}\Big[h^{y}(W;2x)U^{y}_{jklr}(W)\Big]$$

by the same argument as that leading to (43). Hence,

$$\begin{split} &\frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \mathbf{E} \left[\tilde{V}_i \left| m_{jklr}^y \left(W_{\sigma^{-1}(i)}^\sigma + \frac{\tilde{t}V_i}{\sqrt{n}} \right) V_{ij} V_{ik} V_{il} V_{ir} \right| \right] \\ &\lesssim \frac{1}{n^2} \sum_{j,k,l,r=1}^p \mathbf{E} [h^y(W;2x) U_{jklr}^y(W)] \max_{1 \leq j,k,l,r \leq p} \sum_{i=1}^n \mathbf{E} [|V_{ij} V_{ik} V_{il} V_{ir}|] \\ &\lesssim \frac{B_n^2 \phi^4 \log^3 p}{n} \left(\mathbf{E} [\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right), \end{split}$$

where the second inequality follows from (22), (30), Condition V, and the same arguments as those leading to (41). In addition,

$$\frac{1}{n^2} \sum_{i \in I_d} \sum_{j,k,l,r=1}^p \mathbb{E}\left[(1 - \tilde{V}_i) \left| m_{jklr}^y \left(W_{\sigma^{-1}(i)}^{\sigma} + \frac{\tilde{t}V_i}{\sqrt{n}} \right) V_{ij} V_{ik} V_{il} V_{ir} \right| \right] \\
\lesssim \frac{\phi^4 \log^3 p}{n^2} \sum_{i=1}^n \mathbb{E}[(1 - \tilde{V}_i) ||V_i||_{\infty}^4] \lesssim \frac{B_n^2 \phi^4 \log^3 p}{n^2}$$

by (23) and the arguments similar to those leading to (44). Therefore,

$$|\mathcal{I}_{4,1}| \lesssim \frac{B_n^2 \phi^4 \log^3 p}{n} \left(\mathbb{E}[\varrho_{\epsilon^{d+1}}] + \frac{\sqrt{\log p}}{\phi} \right) + \frac{B_n^2 \phi^4 \log^3 p}{n^2},$$

and since the same bound holds for $|\mathcal{I}_{4,2}|$ as well, it follows that (38) holds, which completes Step 5 and the proof of the lemma.

Corollary 4.1. Suppose that all assumptions of Lemma 4.1 are satisfied. Then there exists a constant K > 0 depending only C_v , C_p , and C_b such that for all d = 0, ..., 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

$$E[\varrho_{\epsilon^d}1\{\mathcal{A}_d\}] \lesssim \frac{\sqrt{\log p}}{\phi} + \frac{B_n^2 \phi^4 \log^4(pn)}{n^2} + \left(E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + \frac{\sqrt{\log p}}{\phi}\right) \\
\times \left(\frac{\mathcal{B}_{n,1,d} \phi^2 \log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2,d} \phi^3 \log^2 p}{n} + \frac{B_n^2 \phi^4 \log^3(pn)}{n}\right) \tag{46}$$

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

Proof. Like in the proof of Lemma 4.1, we can assume, without loss of generality, that $\phi \geq 1$, and so (22) implies (23), which we use below.

Fix d = 0, ..., D - 1 and $\phi > 0$ such that (22) holds. Then, given that \mathcal{A}_d depends only on ϵ^d , we have by Lemma 4.1 that

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

up to a constant depending only on C_v , C_p , C_b , and C_a . Thus, given (22) and the fact that $B_n \geq 1$, the asserted claim will follow if we can show that

$$E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_d\}] \le E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + 4/n. \tag{47}$$

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

$$E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d}\}]$$

$$= E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d}\}1\{\mathcal{A}_{d+1}\}] + E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d}\}(1 - 1\{\mathcal{A}_{d+1}\})]$$

$$\leq E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + E[1\{\mathcal{A}_{d}\}(1 - 1\{\mathcal{A}_{d+1}\})]$$

$$\leq E[\varrho_{\epsilon^{d+1}}1\{\mathcal{A}_{d+1}\}] + 1 - P(\mathcal{A}_{d+1} \mid \mathcal{A}_{d}). \tag{48}$$

Moreover, by Lemma 7.1, for all j, k = 1, ..., p and t > 0, we have

$$P\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d+1}(\mathcal{E}_{i,jk}^{V}-\mathcal{E}_{i,jk}^{Z})\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d}(\mathcal{E}_{i,jk}^{V}-\mathcal{E}_{i,jk}^{Z})\right| + t \mid \epsilon^{d}\right)$$

$$\leq 2\exp\left(-\frac{nt^{2}}{32\sum_{i=1}^{n}(\mathcal{E}_{i,jk}^{V}-\mathcal{E}_{i,jk}^{Z})^{2}}\right) \leq 2\exp\left(-\frac{t^{2}}{128C_{v}B_{n}^{2}}\right),$$

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

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i^d (\mathcal{E}_{i,jk}^V - \mathcal{E}_{i,jk}^Z) \right| \le \mathcal{B}_{n,1,d}$$

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

$$P\left(\max_{1\leq j,k\leq p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d+1}(\mathcal{E}_{i,jk}^{V}-\mathcal{E}_{i,jk}^{Z})\right|>\mathcal{B}_{n,1,d+1}\mid\mathcal{A}_{d}\right)\leq\frac{2p^{2}}{(pn)^{3}}\leq\frac{2}{n}.$$

In addition, as in Step 4 of the proof of Lemma 4.1, we have for all i = 1, ..., n and j, k, l = 1, ..., p that

$$\mathcal{E}_{i,jkl}^{V} \le C_p B_n \log(pn) \mathbb{E}[|V_{ij}V_{ik}|] + C_b B_n^3 \log^{3/2}(pn)/n^2$$

and

$$\mathcal{E}_{i,ikl}^{Z} \le C_p B_n \log(pn) \mathbb{E}[|Z_{ij}Z_{ik}|] + C_b B_n^3 \log^{3/2}(pn)/n^2$$

by Conditions P and B. Hence, by Condition V and (23), there exists a constant C depending only on C_v , C_p , and C_b such that

$$\frac{32}{n} \sum_{i=1}^{n} (\mathcal{E}_{i,jkl}^{V} - \mathcal{E}_{i,jkl}^{Z})^{2} \le CB_{n}^{4} \log^{2}(pn).$$

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

$$\begin{split} \mathbf{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d+1}(\mathcal{E}_{i,jkl}^{V}-\mathcal{E}_{i,jkl}^{Z})\right| > \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d}(\mathcal{E}_{i,jkl}^{V}-\mathcal{E}_{i,jkl}^{Z})\right| + t\mid\epsilon^{d}\right) \\ \leq 2\exp\left(-\frac{nt^{2}}{32\sum_{i=1}^{n}(\mathcal{E}_{i,jkl}^{V}-\mathcal{E}_{i,jkl}^{Z})^{2}}\right) \leq 2\exp\left(-\frac{t^{2}}{CB_{n}^{4}\log^{2}(pn)}\right). \end{split}$$

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

$$P\left(\max_{1\leq j,k,l\leq p}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_{i}^{d+1}(\mathcal{E}_{i,jkl}^{V}-\mathcal{E}_{i,jkl}^{Z})\right|>\mathcal{B}_{n,2,d+1}\mid\mathcal{A}_{d}\right)\leq\frac{2p^{3}}{(pn)^{3}}\leq\frac{2}{n}.$$

Thus, $1 - P(A_{d+1} \mid A_d) \le 4/n$, which in combination with (48) implies (47) and completes the proof.

Lemma 4.2. For any constant $\phi > 0$ such that (22) holds, we have

$$E[\varrho_{\epsilon^D}1\{\mathcal{A}_D\}] \le 1/n.$$

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,

$$P(\epsilon^{D} \neq (0, \dots, 0)') = P\left(\sum_{i=1}^{n} \epsilon_{i}^{D} \ge 1\right) \le E\left[\sum_{i=1}^{n} \epsilon_{i}^{D}\right]$$

$$= E\left[E\left[\sum_{i=1}^{n} \epsilon_{i}^{D} \mid \sum_{i=1}^{n} \epsilon_{i}^{D-1}\right]\right] = E\left[\frac{1}{2}\sum_{i=1}^{n} \epsilon_{i}^{D-1}\right]$$

$$= \dots = E\left[\frac{1}{2^{D}}\sum_{i=1}^{n} \epsilon_{i}^{0}\right] = \frac{n}{2^{D}} \le \frac{n}{2^{4 \log n}} \le \frac{1}{n}.$$

Hence,

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

which gives the asserted claim and completes the proof.

Proof of Theorem 4.1. Throughout the proof, we will assume that

$$C_p^4 B_n^2 \log^5(pn) \le n \tag{49}$$

since otherwise the asserted claim is trivial.

Let K be the constant from Corollary 4.1 and for all $d = 0, \dots, D$, define

$$\mathcal{B}_{n,1,d} = C_1(d+1)B_n \log^{1/2}(pn)$$
 and $\mathcal{B}_{n,2,d} = C_1(d+1)B_n^2 \log^{3/2}(pn)$, (50)

where $C_1 = C_m + K$, so that \mathcal{A}_0 holds by (19) and (20) and, in addition, the requirements of Corollary 4.1 on $\mathcal{B}_{n,1,d}$ and $\mathcal{B}_{n,2,d}$ also hold.

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

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

and for all d = 0, ..., D - 1, apply Corollary 4.1 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 (49). Given that

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

and

$$\frac{\mathcal{B}_{n,2,d}\phi_d^3 \log^2 p}{n} \vee \frac{B_n^2 \phi_d^4 \log^3(pn)}{n} \le \frac{C_1 \vee 1}{f_{d+1}},$$

this leads to

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

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

$$f_d \le C_2 \left(f_{d+1}^{2/3} + (d+1)^{2/3} + 1 \right), \text{ for all } d = 0, \dots, D-1.$$

Here, we have $f_D = 1$ by Lemma 4.2 since $B_n \ge 1$ by assumption. Therefore, by a simple induction argument, we conclude that there exists a constant $C \ge 1$ depending only on C_2 such that

$$f_d \leq C(d+1)$$
, for all $d = 0, \dots, D$.

In particular, it follows that

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

Since A_0 holds by construction, so that $1\{A_0\} = 1$, the asserted claim follows by combining this inequality and the definition of ϱ_{ϵ^0} .

5. 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

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

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 variance-covariance matrix Σ .

Theorem 5.1 (Gaussian Approximation via Stein Kernels). If $\Sigma_{jj} \geq c$ for all j = 1, ..., p and some constant c > 0, then

$$\sup_{y \in \mathbb{R}^p} \left| \mathbf{P}(V \le y) - \mathbf{P}(Z \le y) \right| \le C \left(\Delta \log^2 p \right)^{1/2},$$

where C is a constant depending only on c and

$$\Delta = E \left[\max_{1 \le j,k \le p} |\tau_{jk}(V) - \Sigma_{jk}| \right].$$

Remark 5.1. This theorem improves upon Proposition 4.1 in [25], which shows that

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

under the same conditions.

Proof. In this proof, we will use the same notation as that used in the proof of Lemma 4.1. In particular, we will use the constant $\phi > 0$ and the functions m^y and h^y . Moreover, we will use indices to denote partial derivatives, e.g. $m_{jk}^y(w) = \frac{\partial^2 m^y(w)}{\partial w_j \partial w_k}$. Throughout the proof, we will assume, without loss of generality, that V and Z are independent. We proceed in two steps.

Step 1. Denote

$$\mathcal{I}^y = m^y(V) - m^y(Z)$$
, for all $y \in \mathbb{R}^p$.

It then follows from exactly the same arguments as those in Step 2 of the proof of Lemma 4.1, with Lemma 8.3 playing the role of Condition A, that

$$\sup_{y \in \mathbb{R}^p} \left| P(V \le y) - P(Z \le y) \right| \le C_1 \left(\frac{\sqrt{\log p}}{\phi} + \sup_{y \in \mathbb{R}^p} |E[\mathcal{I}^y]| \right),$$

where C_1 is a constant depending only on c. In addition, we will prove in Step 2 below that

$$\sup_{y \in \mathbb{R}^p} |\mathcal{E}[\mathcal{I}^y]| \le C_2 \phi \Delta \log^{3/2} p, \tag{51}$$

where C_2 is another constant depending only on c. Combining these inequalities and substituting $\phi = 1/(\Delta \log p)^{1/2}$ gives the asserted claim.

Step 2. Here, we prove (51). Fix $y \in \mathbb{R}^p$ and denote

$$\Psi(t) = \mathbb{E}[m^y(\sqrt{t}V + \sqrt{1-t}Z)], \text{ for all } t \in [0,1].$$

Then

$$E[\mathcal{I}^y] = \Psi(1) - \Psi(0) = \int_0^1 \Psi'(t)dt,$$

where

$$\Psi'(t) = \frac{1}{2} \sum_{j=1}^{p} \mathbb{E} \left[m_j^y (\sqrt{t}V + \sqrt{1-t}Z) \left(\frac{V_j}{\sqrt{t}} - \frac{Z_j}{\sqrt{1-t}} \right) \right].$$

Also, since τ is the Stein kernel for V,

$$\sum_{j=1}^{p} \mathbb{E}\left[m_{j}^{y}(\sqrt{t}V + \sqrt{1-t}Z)\frac{V_{j}}{\sqrt{t}}\right] = \sum_{j,k=1}^{p} \mathbb{E}\left[m_{jk}^{y}(\sqrt{t}V + \sqrt{1-t}Z)\tau_{jk}(V)\right]$$

and by the multivariate Stein identity,

$$\sum_{j=1}^{p} \mathbb{E}\left[m_{j}^{y}(\sqrt{t}V + \sqrt{1-t}Z)\frac{Z_{j}}{\sqrt{1-t}}\right] = \sum_{j,k=1}^{p} \mathbb{E}[m_{jk}^{y}(\sqrt{t}V + \sqrt{1-t}Z)\Sigma_{jk}].$$

Therefore,

$$\Psi'(t) = \frac{1}{2} \sum_{j,k=1}^{p} E\left[m_{jk}^{y}(\sqrt{t}V + \sqrt{1-t}Z)(\tau_{jk}(V) - \Sigma_{jk})\right].$$

In addition, by (40),

$$m_{jk}^{y}(w) = h^{y}(w; \phi^{-1})m_{jk}^{y}(w)$$

for all $w \in \mathbb{R}^p$. Substituting here $w = \sqrt{tV} + \sqrt{1-tZ}$ and using the definition of h^y in (32), we obtain

$$m_{jk}^{y}(\sqrt{t}V + \sqrt{1-t}Z) = h^{y(V,t)}\left(Z, \frac{1}{\phi\sqrt{1-t}}\right)m_{jk}^{y}(\sqrt{t}V + \sqrt{1-t}Z),$$

where

$$y(V,t) = \frac{y - \sqrt{t}V}{\sqrt{1-t}}.$$

Hence, using (26) and (29), we have

$$|\Psi'(t)| \leq K_1 \phi^2(\log p) \mathbb{E}\left[h^{y(V,t)}\left(Z, \frac{1}{\phi\sqrt{1-t}}\right) \times \max_{1 \leq j,k \leq p} |\tau_{jk}(V) - \Sigma_{jk}|\right]$$

$$= K_1 \phi^2(\log p) \mathbb{E}\left[\mathbb{E}\left[h^{y(V,t)}\left(Z, \frac{1}{\phi\sqrt{1-t}}\right) \mid V\right] \times \max_{1 \leq j,k \leq p} |\tau_{jk}(V) - \Sigma_{jk}|\right]$$

$$\leq \frac{K_2 \phi \log^{3/2} p}{\sqrt{1-t}} \mathbb{E}\left[\max_{1 \leq j,k \leq p} |\tau_{jk}(V) - \Sigma_{jk}|\right]$$

by the law of iterated expectations and Lemma 8.3, where K_1 is a universal constant and K_2 is a constant depending only on c. Conclude that

$$|\mathrm{E}[\mathcal{I}^y]| \le \int_0^1 \Psi'(t)dt \le 2K_2 \phi \Delta \log^{3/2} p,$$

which gives (51) and completes Step 2 and the proof of the theorem.

Theorem 5.1 has two important corollaries. First, note that if V is a centered Gaussian random vector, then by the multivariate Stein identity, its Stein kernel coincides with its variance-covariance matrix. Hence, Theorem 5.1 immediately implies the following result:

Corollary 5.1 (Gaussian-to-Gaussian Comparison). If Z_1 and Z_2 are centered Gaussian random vectors in \mathbb{R}^p with variance-covariance matrices Σ^1 and Σ^2 , respectively, and Σ^2 is such that $\Sigma_{jj}^2 \geq c$ for all $j = 1, \ldots, p$ and some constant c > 0, then

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

where C is a constant depending only on c and

$$\Delta = \max_{1 \le j,k \le p} |\Sigma_{jk}^1 - \Sigma_{jk}^2|.$$

Remark 5.2. This corollary improves upon Theorem 2 in [13], which shows that

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

under the same conditions.

Second, combining Theorem 5.1 with Lemma 4.6 in [26] gives the following result:

Corollary 5.2 (Multiplier-Bootstrap-to-Gaussian Comparison). Let there be vectors a_1, \ldots, a_n in \mathbb{R}^p such that

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

for some constants c, B > 0. Also, let $\varepsilon_1, \ldots, \varepsilon_n$ be independent N(0, 1) random variables and for some constants $\alpha, \beta > 0$ and $v \in [0, 1]$, let e_1, \ldots, e_n be

independent random variables sampled from the distribution of $\zeta+\sqrt{1-v}((\alpha+\beta)\eta-\alpha)((\alpha+\beta+1)/(\alpha\beta))^{1/2}$, where ζ and η are independent random variables such that ζ has the N(0,v) distribution and η has the Beta (α,β) distribution. Then

$$E[e_i] = 0$$
 and $E[e_i^2] = 1$, for all $i = 1, ..., n$, (52)

and, moreover, the random vectors

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

satisfy

$$\sup_{y \in \mathbb{R}^p} \left| P(V \le y) - P(Z \le y) \right| \le C \left(\frac{B^2 \log^5 p}{n} \right)^{1/4}, \tag{53}$$

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

Proof. Recall that the Beta (α, β) distribution has support [0, 1], pdf $f_{\alpha,\beta}(x) = x^{\alpha-1}(1-x)^{\beta-1}/\mathrm{B}(\alpha,\beta)$ for $x \in [0,1]$, mean $\mu = \alpha/(\alpha+\beta)$, and variance $\sigma^2 = \alpha\beta/((\alpha+\beta)^2(\alpha+\beta+1))$. Hence, the random variables e_1, \ldots, e_n have distribution of $\zeta + \sqrt{1-v}(\eta-\mu)/\sigma$ and so satisfy (52).

Further, define

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

where

$$f(x) = \sigma f_{\alpha,\beta}(\sigma x + \mu), \text{ for all } x \in \left(-\frac{\mu}{\sigma}, \frac{1-\mu}{\sigma}\right),$$

is the pdf of $(\eta - \mu)/\sigma$. It is then easy to check, using L'Hospital's rule, that there exists a constant C_1 depending only on α and β such that

$$|\tau(x)| \le C_1$$
, for all $x \in \left(-\frac{\mu}{\sigma}, \frac{1-\mu}{\sigma}\right)$

and also, using integration by parts, that

$$E[e_1\varphi(e_1)] = E[\varphi'(e_1)\{v + (1-v)\tau(e_1)\}]$$

for any continuously differentiable function $\varphi \colon \mathbb{R} \to \mathbb{R}$. Then, by Lemma 4.6 in [26], a Stein kernel τ^V for the random vector V satisfies

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

for some constant C_2 depending only on C_1 . Combining this bound with Theorem 5.1 and observing that

$$E[Z_j Z_k] = \frac{1}{n} \sum_{i=1}^n a_{ij} a_{ik}, \text{ for all } j, k = 1, \dots, p$$

gives (53) and completes the proof.

6. Proofs of Main Theorems

In this section, we provide proofs of our main results stated in Section 2. We will use the following six lemmas:

Lemma 6.1. Suppose that Conditions M and E are satisfied. Then

$$\max_{1 \le j \le p} ||X_i||_{\infty} \le B_n \sqrt{5 \log(pn)} \tag{54}$$

with probability at least $1 - 1/(2n^4)$.

Proof. By the union bound and Markov's inequality, we have for any x > 0 that

$$P\left(\max_{1 \le i \le n} \max_{1 \le j \le p} |X_{ij}| > x\right) \le pn \max_{1 \le i \le n} \max_{1 \le j \le p} P(|X_{ij}| > x)$$

$$\le pn \max_{1 \le i \le n} \max_{1 \le j \le p} \frac{E[\exp(|X_{ij}|^2/B_n^2)]}{\exp(x^2/B_n^2)} \le 2pn \exp(-x^2/B_n^2).$$

Substituting here $x = B_n \sqrt{5 \log(pn)}$ gives the asserted claim.

Lemma 6.2. Suppose that Conditions M and E are satisfied and denote $\tilde{X}_i = X_i - \bar{X}_n$ for all i = 1, ..., 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 $n \geq n_0$, the inequality

$$B_n^2 \log^5(pn) \le cn \tag{55}$$

implies that the inequalities

$$\frac{b_1}{2} \le \frac{1}{n} \sum_{i=1}^n \tilde{X}_{ij}^2 \le 2b_2 \text{ and } \frac{1}{n} \sum_{i=1}^n \tilde{X}_{ij}^4 \le 2B_n^2 b_2, \text{ for all } j = 1, \dots, p,$$
 (56)

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_{ij} \tilde{X}_{ik} - \mathbb{E}[X_{ij} X_{ik}]) \right| \le C B_n \sqrt{\log(pn)}, \tag{57}$$

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_{ij} \tilde{X}_{ik} \tilde{X}_{il} - \mathbb{E}[X_{ij} X_{ik} X_{il}]) \right| \le C B_n^2 \log(pn)$$
 (58)

hold jointly with probability at least 1 - 1/n.

Proof. Fix $m \in \{1, 2, 3, 4\}$ and let $\mathcal{P} = \{1, \dots, p\}^m$. Also, for any $y = (y_1, \dots, y_p)' \in \mathbb{R}^p$ and $h = (h_1, \dots, h_m)' \in \mathcal{P}$, denote $y^h = y_{h_1} \cdots y_{h_m}$. Then note that there exists a constant $A_1 \geq 1$ depending only on b_2 such that

$$\max_{h \in \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[(X_{i}^{h} - \mathrm{E}[X_{i}^{h}])^{2}] \le \max_{h \in \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[(X_{i}^{h})^{2}]$$
$$\le A_{1}^{2} B_{n}^{2(m-1)} \log^{(m-2) \vee 0}(pn)$$

by Condition M, Lemma 6.1, and calculations similar to those in (45). Also, by standard calculations (see Lemma 2.2.2 and discussion on page 95 of [32]), for some universal constant $A_2 \ge 1$,

$$\mathbb{E}\left[\max_{1 \le i \le n} \max_{h \in \mathcal{P}} (X_i^h - \mathbb{E}[X_i^h])^2\right] \le A_2^2 B_n^{2m} \log^m(pn)$$

by Condition E. Hence, by Lemma 8.1,

$$\mathbb{E}\left[\max_{h\in\mathcal{P}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_{i}^{h}-\mathbb{E}[X_{i}^{h}])\right|\right] \\
\leq K_{1}\left(A_{1}B_{n}^{m-1}\log^{(m/2-1/2)\vee(1/2)}(pn)+\frac{A_{2}B_{n}^{m}\log^{m/2+1}(pn)}{\sqrt{n}}\right) \\
\leq K_{1}B_{n}^{m-1}(A_{1}+A_{2})\log^{(m/2-1/2)\vee(1/2)}(pn)$$

for some universal constant $K_1 \ge 1$, where the second inequality follows from (55). Thus, applying Lemma 8.2 with $\eta = 1$, $\beta = 1/2$, and $t = 3K_1B_n^{m-1}(A_1 + A_2)\log^{(m/2-1/2)\vee(1/2)}(pn)$ shows that

$$\max_{h \in \mathcal{P}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i^h - \mathbb{E}[X_i^h]) \right| > 5K_1 B_n^{m-1} (A_1 + A_2) \log^{(m/2 - 1/2) \vee (1/2)} (pn)$$

with probability at most

$$\exp(-3\log(pn)) + 3\exp\left(-K_2\left(\frac{B_n^{m-1}\log^{(m/2-1/2)\vee(1/2)}(pn)}{B_n^m\log^{m/2}(pn)/\sqrt{n}}\right)^{1/2}\right)$$

$$\leq (pn)^{-3} + 3\exp\left(-K_2\log(pn)/c^{1/4}\right) \leq 4/(pn)^3 \leq 1/(4n),$$

for some universal constant $K_2 > 0$, where the first inequality follows from (55) and the second holds if we set $c = (1 \wedge (K_2/3))^4$. Thus, for $A_3 = 5K_1(A_1 + A_2)$, letting \mathcal{A} be the event that the inequalities

$$\max_{1 \le j \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_{ij} \right| \le A_3 \sqrt{\log(pn)},$$

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} X_{ik} - \mathbb{E}[X_{ij} X_{ik}]) \right| \le A_3 B_n \sqrt{\log(pn)},$$

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} X_{ik} X_{il} - \mathbb{E}[X_{ij} X_{ik} X_{il}]) \right| \le A_3 B_n^2 \log(pn),$$

$$\max_{1 \le j,k,l,r \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} X_{ik} X_{il} X_{ir} - \mathbb{E}[X_{ij} X_{ik} X_{il} X_{ir}]) \right| \le A_3 B_n^3 \log^{3/2}(pn)$$

hold jointly, we have that the probability of A is at least 1 - 1/n. On the other hand, given that

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

$$\leq \max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} X_{ik} - E[X_{ij} X_{ik}]) \right| + \sqrt{n} \max_{1 \le j \le p} |\bar{X}_{ij}|^{2}$$

and

$$\max_{1 \leq j,k,l \leq p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\tilde{X}_{ij} \tilde{X}_{ik} \tilde{X}_{il} - E[X_{ij} X_{ik} X_{il}]) \right| \\
\leq \max_{1 \leq j,k,l \leq p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_{ij} X_{ik} X_{il} - E[X_{ij} X_{ik} X_{il}]) \right| \\
+ 2\sqrt{n} \max_{1 \leq j \leq p} |\bar{X}_{nj}|^3 + \max_{1 \leq j,k,l \leq p} |\bar{X}_{nl}| \times \left| \frac{3}{\sqrt{n}} \sum_{i=1}^{n} X_{ij} X_{ik} \right|,$$

it follows that the inequalities (57) and (58) with some constant C depending only on b_2 hold on A.

In addition, it follows from Condition M that the first part of (56) holds as long as

$$\max_{1 \le j \le p} \left| \frac{1}{n} \sum_{i=1}^{n} (\tilde{X}_{ij}^{2} - \mathbb{E}[X_{ij}^{2}]) \right| \le (b_{1}/2) \wedge b_{2}.$$

However, on the event (57), we have

$$\max_{1 \le j \le p} \left| \frac{1}{n} \sum_{i=1}^{n} (\tilde{X}_{ij}^{2} - E[X_{ij}^{2}]) \right| \le \max_{1 \le j,k \le p} \left| \frac{1}{n} \sum_{i=1}^{n} (\tilde{X}_{ij} \tilde{X}_{ik} - E[X_{ij} X_{ik}]) \right| \\
\le \frac{C B_{n} \sqrt{\log(pn)}}{\sqrt{n}} \le \frac{C}{\log^{2}(pn)} \le (b_{1}/2) \wedge b_{2},$$

where the last inequality holds as long as $n \geq n_0$ for some constant n_0 depending only on b_1 , b_2 , and C.

Finally, it follows from Condition M that the second part of (57) holds as long as

$$\max_{1 \le j \le p} \left| \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_{ij}^{4} - \mathbb{E}[X_{ij}^{4}] \right| \le B_{n}^{2} b_{2},$$

which holds on \mathcal{A} for all $n \geq n_0$ and some n_0 depending only on b_2 by the same arguments as those used above. The asserted claim follows.

Lemma 6.3. Suppose that Conditions M and E are satisfied and that the random variables X_1^*, \ldots, X_n^* are obtained via either the empirical or multiplier bootstrap with weights satisfying (12). Then with probability at least

1-2/n, we have

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

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

Proof. Let Y_1, \ldots, Y_n be vectors in \mathbb{R}^p such that

$$||Y_i||_{\infty} \le 2B_n \sqrt{5\log(pn)} \quad \text{for all } i = 1, \dots, n,$$
 (59)

$$b_1/2 \le \frac{1}{n} \sum_{i=1}^n Y_{ij}^2 \le 2b_2$$
 and $\frac{1}{n} \sum_{i=1}^n Y_{ij}^4 \le 2B_n^2 b_2$, for all $j = 1, \dots, p$, (60)

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{ij} Y_{ik} - \mathbb{E}[X_{ij} X_{ik}]) \right| \le C_m B_n \sqrt{\log(pn)}, \tag{61}$$

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_{ij} Y_{ik} Y_{il} - E[X_{ij} X_{ik} X_{il}]) \right| \le C_m B_n^2 \log(pn), \tag{62}$$

where C_m is the constant C from Lemma 6.2. Also, let Y_1^*, \ldots, Y_n^* be independent random vectors with each Y_i^* having uniform distribution on $\{Y_1, \ldots, Y_n\}$. In addition, let $\tilde{e}_1, \ldots, \tilde{e}_n$ be a sequence of independent random variables with distribution constructed in Lemma 7.3. This distribution is such that for all $i = 1, \ldots, n$, we have $\tilde{e}_i = \tilde{e}_{i,1} + \tilde{e}_{i,2}$, where $\tilde{e}_{i,1}$ and $\tilde{e}_{i,2}$ are independent, $\tilde{e}_{i,1}$ has the $N(0, \sigma^2)$ distribution, $\tilde{e}_{i,2}$ has a two-point distribution, and $\sigma > 0$ can be chosen to be a universal constant. Thus, using (60) and applying Lemma 8.3 conditional on $\tilde{e}_{1,2}, \ldots, \tilde{e}_{n,2}$, it follows that for all $y \in \mathbb{R}^p$ and t > 0,

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{e}_{i}Y_{i} \leq y+t\right) - P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{e}_{i}Y_{i} \leq y\right) \leq K_{1}t\sqrt{\log p}, \quad (63)$$

where K_1 is a constant depending only on b_1 and σ .

Now, to prove the asserted claim, we will apply Theorem 4.1 several times, where V_i 's will vary and Z_i 's will always be given by $Z_i = \tilde{e}_i Y_i$ for all $i=1,\ldots,n$. In particular, we will consider the following cases: (i) $V_i=X_i$, (ii) $V_i=Y_i^*$, and (iii) $V_i=e_i Y_i$ with e_i 's satisfying (12) and (13). In all these cases, Conditions V, P, and B with C_v , C_p , and C_b depending only on b_1 and b_2 follow immediately from Conditions M and E and the inequalities in (59) and (60). Also, Condition A with $C_a=K_1$ follows from (63). Hence, an application of Theorem 4.1 is justified if we can verify that

$$\max_{1 \le j,k \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbb{E}[V_{ij}V_{ik}] - Y_{ij}Y_{ik}) \right| \le C_m B_n \sqrt{\log(pn)}$$
 (64)

and

$$\max_{1 \le j,k,l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\mathbb{E}[V_{ij} V_{ik} V_{il}] - Y_{ij} Y_{ik} Y_{il}) \right| \le C_m B_n^2 \sqrt{\log^3(pn)}.$$
 (65)

In turn, in the case (i), (64) and (65) follow from (61) and (62). In the case (ii), (64) and (65) follow from noting that the left-hand sides of (64) and (65) are both zero by the construction of Y_i^* 's. In the case (iii), (64) and (65) follow from noting that the left-hand sides of (64) and (65) are both zero by the fact that $E[e_i^2] = E[e_i^3] = 1$.

Thus, applying Theorem 4.1 shows that in all cases, for all $y \in \mathbb{R}^p$, we have

$$\left| P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i \le y\right) - P\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{e}_i Y_i \le y\right) \right| \le K_2 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}$$

for some constant K_2 depending only on b_1 , b_2 , K_1 , and C_m . Now, noting that if we take $Y_i = X_i - \bar{X}_n$ for all i = 1, ..., n, then (59) holds with probability at least $1 - 1/(2n^4)$ by Lemma 6.1 and (60), (61), and (62) hold jointly with probability at least 1 - 1/n by Lemma 6.2 (in the application of Lemma 6.2, we can assume that the required condition (55) is satisfied since otherwise the asserted claim is trivial), it follows from the triangle inequality that

$$\sup_{x \in \mathbb{R}} |P(T_n \le x) - P(T_n^* \le x \mid X)| \le 2K_2 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}$$

with probability at least 1 - 2/n. The asserted claim follows.

Lemma 6.4. Suppose that Conditions M and E are satisfied. Then

$$\sup_{x \in \mathbb{R}} |P(T_n \le x) - P(T_n^G \le x)| \le C \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}, \tag{66}$$

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

Proof. Without loss of generality, assume that (55) holds since otherwise the asserted claim is trivial. Define the random variables e_1, \ldots, e_n as in Corollary 5.2 with v = 0, $\alpha = 1/2$ and $\beta = 3/2$. It is then easy to check that $E[e_i^3] = 1$ for all $i = 1, \ldots, n$. Thus, Lemma 6.3 implies that we have

$$\sup_{x \in \mathbb{R}} |P(T_n \le x) - P(T_n^* \le x \mid X)| \le C_1 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}$$

with probability at least 1-2/n, where T_n^* is the multiplier bootstrap statistic with weights e_1, \ldots, e_n and C_1 is a constant depending only on b_1 and b_2 .

Next, let A_n be the event that (56) and (57) hold jointly. By Lemma 6.2, we have $P(A_n) \ge 1 - 1/n$. Moreover, on A_n we have

$$\sup_{x \in \mathbb{R}} |P(T_n^* \le x \mid X) - P(T_n^{\hat{G}} \le x \mid X)| \le C_2 \left(\frac{B_n^2 \log^5 p}{n}\right)^{1/4}$$

by Corollary 5.2 and

$$\sup_{x \in \mathbb{R}} |P(T_n^{\hat{G}} \le x \mid X) - P(T_n^G \le x)| \le C_3 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}$$

by Corollary 5.1, where C_2 and C_3 are constants 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 (66) is non-stochastic, so that if (66) holds with strictly positive probability, then it holds with probability one.

Lemma 6.5. Suppose that Conditions M and E are satisfied and that the random variables X_1^*, \ldots, X_n^* are obtained via the multiplier bootstrap with weights e_1, \ldots, e_n violating (12). Then with probability at least 1 - 2/n, we have

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

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

Remark 6.1. The constant C in this result depends on $E[e_1^3]$ continuously, and so, given that $|E[e_1^3]| \leq \sqrt{E[e_1^6]} \leq \sqrt{4^3 \cdot 3! \cdot E[\exp(e_1^2/4)]}$, we can take C independent of $E[e_1^3]$ under the implicitly maintained assumption that (13) holds.

Proof. Without loss of generality, assume that (55) holds since otherwise the asserted claim is trivial. Let \mathcal{A}_n be the event that $\max_{1 \leq j \leq p} \|X_i\|_{\infty} \leq B_n \sqrt{5 \log(pn)}$ and (56)–(58) hold jointly. By Lemmas 6.1 and 6.2, we have $P(\mathcal{A}_n) \geq 1 - 2/n$. Moreover, by Corollary 5.1,

$$\sup_{x \in \mathbb{R}} |P(T_n^{\hat{G}} \le x \mid X) - P(T_n^G \le x)| \le C_1 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4}$$

on \mathcal{A}_n , where C_1 is a constant depending only on b_1 and b_2 . Next, define the function $f:(0,1)\to\mathbb{R}$ by

$$f(\alpha) = \frac{2\sqrt{2}(1-2\alpha)}{3\sqrt{\alpha(1-\alpha)}},$$
 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 $f(\alpha^*) = 2^{3/2} \mathrm{E}[e_1^3]$. Now, let ζ_1, \ldots, ζ_n and η_1, \ldots, η_n be respectively i.i.d. N(0,1/2) and $\mathrm{Beta}(\alpha^*, 1-\alpha^*)$ variables such that $\zeta_1, \ldots, \zeta_n, \eta_1, \ldots, \eta_n, X_1, \ldots, X_n$ are independent. We

set $\tilde{e}_i = \zeta_i + \sqrt{2}(\eta_i - \alpha^*)/\sqrt{\alpha^*(1 - \alpha^*)}$ for i = 1, ..., n, so that the distribution of \tilde{e}_i 's is equal to the distribution of e_i 's in Corollary 5.2 with v = 1/2, $\alpha = \alpha^*$ and $\beta = 1 - \alpha^*$. It is then easy to check that $E[\tilde{e}_i] = 0$, $E[\tilde{e}_i^2] = 1$, and $E[\tilde{e}_i^3] = E[e_i^3]$ for all i = 1, ..., n. Also, set

$$\tilde{T}_n^* = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{e}_i \tilde{X}_i + t_j),$$

where $\tilde{X}_i := X_i - \bar{X}_n$. Then applying Corollary 5.2, we have on A_n that

$$\sup_{x \in \mathbb{R}} |P(\tilde{T}_n^* \le x \mid X) - P(T_n^{\hat{G}} \le x \mid X)| \le C_2 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4},$$

where C_2 is a constant depending only on α^* , b_1 and b_2 .

Meanwhile, using (56) and applying Lemma 8.3 conditional on η_1, \ldots, η_n , it follows that for all $y \in \mathbb{R}^p$ and t > 0,

$$P_{\tilde{e}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{e}_{i}\tilde{X}_{i} \leq y+t\right) - P_{\tilde{e}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{e}_{i}\tilde{X}_{i} \leq y\right) \leq K_{1}t\sqrt{\log p}, \quad (67)$$

where K_1 is a constant depending only on b_1 . Thus, conditionally on X_1, \ldots, X_n , we can apply Theorem 4.1 with $V_i = X_i^* = e_i \tilde{X}_i$ and $Z_i = \tilde{e}_i \tilde{X}_i$ on the event \mathcal{A}_n . In fact, 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 inequality $\max_{1 \leq j \leq p} \|X_i\|_{\infty} \leq B_n \sqrt{5 \log(pn)}$, (56) and the sub-Gaussianity of the multiplier variables. Also, Condition A with $C_a = K_1$ follows from (67). Moreover, (19) and (20) are evident by construction. Consequently, we have on \mathcal{A}_n that

$$\sup_{x \in \mathbb{R}} |P(T_n^* \le x \mid X) - P(\tilde{T}_n^* \le x \mid X)| \le C_3 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4},$$

where C_3 is a constant depending only on α^* , b_1 and b_2 . The asserted claim now follows from combining these bounds via the triangle inequality and using Lemma 6.4.

Lemma 6.6 (Anti-Concentration of T_n). Suppose that Conditions M and E are satisfied. Then for any $x \in \mathbb{R}$ and t > 0,

$$P(T_n \le x + t) - P(T_n \le x) \le C \left(t \sqrt{\log p} + \left(\frac{B_n^2 \log^5(pn)}{n} \right)^{1/4} \right),$$

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

Proof. Fix $y \in \mathbb{R}^p$ and t > 0 and let the random variables e_1, \ldots, e_n be independently sampled from the distribution defined in Lemma 7.3 (and independently of X_1, \ldots, X_n). Then for some constant C depending only

on b_1 and b_2 , with probability at least 1 - 3/n,

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{i} \leq y + t\right)$$

$$\leq P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_{i}(X_{i} - \bar{X}_{n}) \leq y + t \mid X\right) + C\left(\frac{B_{n}^{2}\log^{5}(pn)}{n}\right)^{1/4}$$

$$\leq P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}e_{i}(X_{i} - \bar{X}_{n}) \leq y \mid X\right) + Ct\sqrt{\log p} + C\left(\frac{B_{n}^{2}\log^{5}(pn)}{n}\right)^{1/4}$$

$$\leq P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_{i} \leq y\right) + Ct\sqrt{\log p} + 2C\left(\frac{B_{n}^{2}\log^{5}(pn)}{n}\right)^{1/4},$$

where the first and the third inequalities follow from Lemma 6.3 and the second from Lemmas 6.2 and 8.3. Here, both the left-hand side and the right-hand side are non-stochastic, and so if the right-hand side exceeds the left-hand side with strictly positive probability, it must do so with probability one. This gives the asserted claim.

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 6.4 by applying (66) with $x = c_{1-\alpha}^G$.

Proof of Theorem 2.2. Let C_1 , C_2 , and C_3 be the constant C in Lemmas 6.3, 6.5, and 6.6, respectively. Denote

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

By Lemmas 6.3 and 6.5, we have

$$\sup_{x \in \mathbb{R}} |P(T_n \le x) - P(T_n^* \le x \mid X)| \le \beta_n$$

with probability at least 1 - 2/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

$$P(T_n^* \le c_{1-\alpha+\beta_n} \mid X) \ge P(T_n \le c_{1-\alpha+\beta_n}) - \beta_n \ge 1 - \alpha$$

and

$$P(T_n^* \le c_{1-\alpha-3\beta_n} \mid X) \le P(T_n \le c_{1-\alpha-3\beta_n}) + \beta_n$$

$$\le 1 - \alpha - 2\beta_n + C_2 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4} < 1 - \alpha,$$

where the second inequality follows from Lemma 6.6. Therefore,

$$P(c_{1-\alpha-3\beta_n} < c_{1-\alpha}^B \le c_{1-\alpha+\beta_n}) \ge 1 - 2/n,$$

and so

$$P(T_n > c_{1-\alpha}^B) \le P(T_n > c_{1-\alpha-3\beta_n}) + 2/n \le \alpha + 3\beta_n + 2/n \le \alpha + 5\beta_n$$

and

$$P(T_n > c_{1-\alpha}^B) \ge P(T_n > c_{1-\alpha+\beta_n}) - 2/n$$

 $\ge \alpha - \beta_n - C_2 \left(\frac{B_n^2 \log^5(pn)}{n}\right)^{1/4} - 2/n \ge \alpha - 4\beta_n,$

where the second inequality follows from Lemma 6.6. Combining these inequalities gives the asserted claim.

Proof of Theorem 2.3. Let e_1, \ldots, e_n be Rademacher weights and assume that $B_n^2 \log(pn) \leq cn$ and $n \geq n_0$ with the same constants c and n_0 as those in Lemma 6.2 since otherwise the asserted claim is trivial. Then by the proof of Lemma 6.2, there exists a constant $C_1 \geq 1$ depending only on b_1 and b_2 such that

$$\|\bar{X}_n\|_{\infty} \le C_1 \sqrt{\frac{\log(pn)}{n}} \text{ and } \frac{b_1}{2} \le \frac{1}{n} \sum_{i=1}^n X_{ij}^2 \le 2b_2, \text{ for all } j = 1, \dots, p$$
 (68)

with probability at least 1 - 1/n.

Further, for all $\gamma \in (0,1)$, let $c_{1-\gamma}^{B,0}$ be the $(1-\gamma)$ th quantile of the conditional distribution of

$$T_n^{*,0} = \max_{1 \le j \le p} \frac{1}{\sqrt{n}} \sum_{i=1}^n (e_i X_{ij} + t_j)$$

given X_1, \ldots, X_n and observe that by Lemma 8.4, there exists a constant $C_2 \geq 1$ depending only on b_1 and b_2 such that on the event that (54) and (68) hold jointly, we have

$$\sup_{x \in \mathbb{R}} P(x \le T_n^{*,0} \le x + t \mid X) \le C_2 \left(t \sqrt{\log p} + \sqrt{\frac{B_n^2 \log^2(pn)}{n}} \right), \text{ for all } t > 0.$$

Hence, given that (54) holds with probability at least 1 - 1/n by Lemma 6.1, applying Lemma 7.4, which is justified by Condition S, we obtain

$$\sup_{\gamma \in (0,1)} |P(T_n > c_{1-\gamma}^{B,0}) - \gamma| \le C_2 \sqrt{\frac{B_n^2 \log^2(pn)}{n}} + \frac{2}{n}.$$
 (69)

In addition, for

$$\beta_n = C_2 \sqrt{\frac{B_n^2 \log^2(pn)}{n}} + C_1 C_2 \sqrt{\frac{2 \log^2(pn) \log n}{n}},$$

we have on the event that (54) and (68) hold jointly that

$$c_{1-\gamma+\beta_n}^{B,0} - c_{1-\gamma}^{B,0} \ge C_1 \sqrt{\frac{2\log(pn)\log n}{n}}, \text{ for all } \gamma \in (\beta_n, 1)$$

since otherwise we would have

$$P(c_{1-\gamma}^{B,0} \le T_n^{*,0} \le c_{1-\gamma+\beta_n}^{B,0} \mid X) < C_2 \sqrt{\frac{B_n^2 \log^2(pn)}{n}} + C_1 C_2 \sqrt{\frac{2 \log^2(pn) \log n}{n}}$$

and simultaneously

$$P(c_{1-\gamma}^{B,0} \le T_n^{*,0} \le c_{1-\gamma+\beta_n}^{B,0} \mid X)$$

$$= P(T_n^{*,0} \le c_{1-\gamma+\beta_n}^{B,0} \mid X) - P(T_n^{*,0} < c_{1-\gamma}^{B,0} \mid X)$$

$$\ge 1 - \gamma + \beta_n - (1 - \gamma) = \beta_n,$$

which is a contradiction.

Thus, on the event that (54) and (68) hold jointly, we have

$$\begin{split} & \mathbf{P}(T_n^* \leq c_{1-\alpha+2\beta_n}^{B,0} \mid X) \\ & \geq \mathbf{P}\left(T_n^{*,0} + C_1 \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \right| \sqrt{\frac{\log(pn)}{n}} \leq c_{1-\alpha+2\beta_n}^{B,0} \mid X\right) \\ & \geq \mathbf{P}\left(T_n^{*,0} + C_1 \sqrt{\frac{2\log(pn)\log n}{n}} \leq c_{1-\alpha+2\beta_n}^{B,0} \mid X\right) - 2/n \\ & \geq \mathbf{P}(T_n^{*,0} \leq c_{1-\alpha+\beta_n}^{B,0} \mid X) - 2/n \geq 1 - \alpha + \beta_n - 2/n > 1 - \alpha, \end{split}$$

where the first inequality follows from (68) and the second from the Hoeffding inequality. In addition, by the same arguments, again on the event that (54) and (68) hold jointly, we have

$$P(T_n^* \le c_{1-\alpha-2\beta_n}^{B,0} \mid X) \le P(T_n^{*,0} \le c_{1-\alpha-\beta_n}^{B,0} \mid X) + 2/n$$

$$\le 1 - \alpha - \beta_n + 2/n + C_2 \sqrt{\frac{B_n^2 \log^2(pn)}{n}} < 1 - \alpha.$$

Hence.

$$P(c_{1-\alpha-2\beta_n}^{B,0} < c_{1-\alpha}^{B} \le c_{1-\alpha+2\beta_n}^{B,0}) \ge 1 - 2/n.$$

Conclude that

$$P(T_n > c_{1-\alpha}^B) \le P(T_n > c_{1-\alpha-2\beta_n}^{B,0}) + 2/n$$

 $\le \alpha + 2\beta_n + 2/n + \beta_n \le \alpha + 4\beta_n$

and

$$P(T_n > c_{1-\alpha}^B) \ge P(T_n > c_{1-\alpha+2\beta_n}^{B,0}) - 2/n$$

 $\ge \alpha - 2\beta_n - 2/n - \beta_n \ge \alpha - 4\beta_n,$

where the second lines follow from (69). The asserted claim follows.

Proof of Theorem 2.4. We proceed in three steps.

Step 1. Here, we show that in the setting of Section 4, if Conditions V, P, and B are satisfied, then for all $y \in \mathbb{R}^p$, we have

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{i} \leq y\right) - P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{i} \leq y + \eta/2\right)$$

$$\lesssim \frac{\mathcal{B}_{n,1}\log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2}\log^{2}p}{n} + \frac{B_{n}^{2}\log^{3}(pn)}{n} + \frac{B_{n}^{2}\log^{4}(pn)}{n^{2}}$$

up to a constant depending only on C_v , C_p , C_b , and η , where $\mathcal{B}_{n,1}$ and $\mathcal{B}_{n,2}$ are the left-hand sides of (19) and (20), respectively. To show this result, note that for all $y \in \mathbb{R}^p$, we have

$$P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{i} \leq y\right) \leq P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_{i} \leq y + 2/\phi\right) + |E[\mathcal{I}^{y+\phi}]|$$

by Step 2 of the proof of Lemma 4.1, where all the notations are the same as in Lemma 4.1. So, we set $\phi = 4/\eta$ and bound $|\mathbf{E}[\mathcal{I}^{y+\phi}]|$ using Steps 1, 3, 4, and 5 of the proof of Lemma 4.1 with the only difference that all terms like

$$P\left(-\phi^{-1} < \max_{1 \le j \le p} (W_j - y_j) \le \phi^{-1}\right)$$

are now upper bounded by one rather than by $2E[\varrho_{\epsilon^1}] + \mathcal{P}$. This gives the claim of this step.

Step 2. Here, we show that there exists a constant $C \geq 1$ depending only on b_1 , b_2 , and η such that with probability at least 1 - 2/n, we have for all $x \in \mathbb{R}$ that

$$\mathcal{P}_x = P(T_n^* \le x \mid X) - P(T_n \le x + \eta/2) \le C\left(\frac{B_n^2 \log^s(pn)}{n}\right)^{1/2}$$
 (70)

where s = 3 if $c_{1-\alpha}^B$ is obtained via either the empirical or the multiplier bootstrap with weights satisfying (12) and s = 5 if $c_{1-\alpha}^B$ is obtained via the multiplier bootstrap with weights violating (12). To do so, we proceed as in the proof of Lemma 6.3, with the following differences: (i) we now use Step 1 instead of Theorem 4.1, and (ii) in the case of the multiplier bootstrap with weights violating (12), instead of (62), we use the bound

$$\max_{1 \le j, k, l \le p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_{ij} Y_{ik} Y_{il} \right| \le 4b_2 B_n \sqrt{5n \log(pn)},$$

which follows from (59) and (60). This leads to the following inequality: with probability at least 1 - 1/n, for all $x \in \mathbb{R}^p$,

$$\mathcal{P}_x \lesssim \frac{\mathcal{B}_{n,1} \log p}{\sqrt{n}} + \frac{\mathcal{B}_{n,2} \log^2 p}{n} + \frac{B_n^2 \log^3(pn)}{n} + \frac{B_n^2 \log^4(pn)}{n^2}$$

up to a constant depending only b_1 , b_2 , and η , where $\mathcal{B}_{n,1} = B_n \sqrt{\log(pn)}$ in all cases and $\mathcal{B}_{n,2} = B_n^2 \log(pn)$ in the case of the empirical and the multiplier

bootstrap with weights satisfying (12) and $\mathcal{B}_{n,2} = B_n \sqrt{n \log(pn)}$ in the case of the multiplier bootstrap with weights violating (12). The asserted claim of this step follows.

Step 3. Here, we complete the proof. Let β_n be the right-hand side of (70) and for all $\gamma \in (0,1)$, let $c_{1-\gamma}$ be the $(1-\gamma)$ th quantile of T_n . Then by Step 2, with probability at least 1-2/n, we have

$$P(T_n^* \le c_{1-\alpha-2\beta_n} - \eta \mid X) \le P(T_n < c_{1-\alpha-2\beta_n}) + \beta_n < 1 - \alpha.$$

Therefore,

$$P(c_{1-\alpha}^B \ge c_{1-\alpha-2\beta_n} - \eta) \ge 1 - 2/n,$$

and so

$$P(T_n > c_{1-\alpha}^B + \eta) \le P(T_n > c_{1-\alpha-2\beta_n}) + 2/n \le \alpha + 4\beta_n.$$

The asserted claim follows.

7. Technical Lemmas

Lemma 7.1 (Exponential Inequality for Weighted Sums of Exchangeable Random Variables). Let a_1, \ldots, a_n be some constants in \mathbb{R} and let X_1, \ldots, X_n be exchangeable random variables such that $|X_i| \leq 1$ almost surely for all $i = 1, \ldots, n$. Then

$$P\left(\left|\sum_{i=1}^{n} a_i X_i\right| > \left|\sum_{i=1}^{n} a_i\right| + t\right) \le 2 \exp\left(-\frac{t^2}{32 \sum_{i=1}^{n} a_i^2}\right), \quad \text{for all } t > 0.$$

Proof. Since the random variables X_i are exchangeable, we can and will, without loss of generality, assume that

$$|a_1| \ge |a_2| \ge \dots \ge |a_n|. \tag{71}$$

Next, define the sample mean $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ and observe that $|\bar{X}_n| \leq 1$. Hence, denoting

$$Y_i = X_i - \bar{X}_n$$
, for all $i = 1, \dots, n$,

we have by the triangle inequality that

$$\left| \sum_{i=1}^{n} a_i X_i \right| = \left| \sum_{i=1}^{n} a_i (X_i - \bar{X}_n) + \sum_{i=1}^{n} a_i \bar{X}_n \right| \le \left| \sum_{i=1}^{n} a_i Y_i \right| + \left| \sum_{i=1}^{n} a_i \right|. \tag{72}$$

Now, observe that Y_1, \ldots, Y_n are exchangeable random variables, and so for all $i = 1, \ldots, n$,

$$E[Y_i \mid Y_1, \dots, Y_{i-1}] = E\left[\frac{1}{n-i+1} \sum_{j=i}^n Y_j \mid Y_1, \dots, Y_{i-1}\right].$$

Hence, denoting

$$R_i = Y_i + \frac{1}{n-i+1} \sum_{j=1}^{i-1} Y_j$$
, for all $i = 1, \dots, n$, (73)

it follows that for all $i = 1, \ldots, n$,

$$E[R_i \mid R_1, \dots, R_{i-1}] = E[R_i \mid Y_1, \dots, Y_{i-1}]$$

$$= E\left[Y_i + \frac{1}{n-i+1} \sum_{j=1}^{i-1} Y_j \mid Y_1, \dots, Y_{i-1}\right]$$

$$= E\left[\frac{1}{n-i+1} \sum_{j=1}^{n} Y_j \mid Y_1, \dots, Y_{i-1}\right] = 0.$$

Thus, $(R_i, \mathcal{F}_i)_{i=1}^n$, where $\mathcal{F}_i = \{R_1, \dots, R_i\}$ for all $i = 1, \dots, n$, is a martingale difference sequence. In addition, for all $i = 1, \dots, n$,

$$|R_i| = \left| Y_i + \frac{1}{n-i+1} \sum_{j=1}^{i-1} Y_j \right| = \left| Y_i - \frac{1}{n-j+1} \sum_{j=i}^n Y_j \right| \le \max_{1 \le j \le n} |X_j| \le 2.$$

Moreover, using an induction argument, it follows from (73) that for all i = 1, ..., n,

$$Y_i = R_i - \sum_{j=1}^{i-1} \frac{R_j}{n-j},\tag{74}$$

Indeed, (74) holds trivially for i=1. Hence, assuming that (74) holds for all $i=1,\ldots,k-1$ for some $k=2,\ldots,n$, we have that

$$Y_k = R_k - \frac{1}{n-k+1} \sum_{j=1}^{k-1} Y_j = R_k - \frac{1}{n-k+1} \sum_{j=1}^{k-1} \left(R_j - \sum_{l=1}^{j-1} \frac{R_l}{n-l} \right)$$
$$= R_k - \frac{1}{n-k+1} \sum_{j=1}^{k-1} R_j \left(1 - \frac{k-1-j}{n-j} \right) = R_k - \sum_{j=1}^{k-1} \frac{R_j}{n-j},$$

meaning that (74) holds for i = k as well, and thus for all i = 1, ..., n by induction. In turn, it follows from (74) that

$$\sum_{i=1}^{n} a_i Y_i = \sum_{i=1}^{n} c_i R_i,$$

where

$$c_i = a_i - \frac{1}{n-i} \sum_{j=i+1}^{n} a_j$$
, for all $i = 1, \dots, n$.

Here, we have

$$|c_i| \le 2|a_i|$$
, for all $i = 1, \ldots, n$,

by (71), and so

$$\sum_{i=1}^{n} c_i^2 \le 4 \sum_{i=1}^{n} a_i^2.$$

Hence, by the Azuma-Hoeffding inequality, for any t > 0,

$$P\left(\left|\sum_{i=1}^{n} a_i Y_i\right| > t\right) = P\left(\left|\sum_{i=1}^{n} c_i R_i\right| > t\right)$$

$$\leq 2 \exp\left(-\frac{t^2}{8 \sum_{i=1}^{n} c_i^2}\right) \leq 2 \exp\left(-\frac{t^2}{32 \sum_{i=1}^{n} a_i^2}\right).$$

Combining this bound with (72) gives the asserted claim of the lemma.

Lemma 7.2 (Randomized Lindeberg Interpolation). Let S_n be the set of all one-to-one functions mapping $\{1, \ldots, n\}$ to $\{1, \ldots, n\}$. Also, let X_1, \ldots, X_n and Y_1, \ldots, Y_n be sequences of vectors in \mathbb{R}^p , U be a random variable with uniform distribution on [0,1], and σ be a random function with uniform distribution on S_n . Assume that U is independent of σ , and for all $k = 1, \ldots, n$, denote

$$W_k^{\sigma} = \sum_{j=1}^{k-1} X_{\sigma(j)} + \sum_{j=k+1}^n Y_{\sigma(j)}$$

and

$$W_k = \begin{cases} W_{\sigma^{-1}(k)}^{\sigma} + X_k & \text{if } U \le \frac{\sigma^{-1}(k)}{n+1}, \\ W_{\sigma^{-1}(k)}^{\sigma} + Y_k & \text{if } U > \frac{\sigma^{-1}(k)}{n+1}. \end{cases}$$

Then the distribution of W_k is independent of k, i.e. there exists a random vector $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ with values in $\{0,1\}^n$ such that for all $k = 1, \ldots, n$, the distribution of W_k is equal to that of

$$\sum_{i=1}^{n} \left(\epsilon_i X_i + (1 - \epsilon_i) Y_i \right). \tag{75}$$

Moreover, the random variables $\epsilon_1, \ldots, \epsilon_n$ are exchangeable and are such that $P(\sum_{i=1}^n \epsilon_i = s) = 1/(n+1)$ for all $s = 0, \ldots, n$. In particular, $E[\sum_{i=1}^n \epsilon_i] = n/2$.

Remark 7.1. The first asserted claim of this lemma is the same as Lemma 2 in [20]. We present a self-contained proof of this claim below for reader's convenience.

Proof. Fix k = 1, ..., n. To show that the distribution of W_k is independent of k, it suffices to show that for any subset S of $\{1, ..., n\}$,

$$P\left(W_k = \sum_{i \in S} X_i + \sum_{i \notin S} Y_i\right) \tag{76}$$

is independent of k. To do so, fix any $S \subset \{1, \ldots, n\}$ and denote s = |S|. If $k \notin S$, then

$$P\left(W_{k} = \sum_{i \in S} X_{i} + \sum_{i \notin S} Y_{i}\right)$$

$$= P\left(\left\{\sigma^{-1}(i) \leq s, \forall i \in S\right\} \cap \left\{\sigma^{-1}(k) = s + 1\right\} \cap \left\{U > \frac{s + 1}{n + 1}\right\}\right)$$

$$= P\left(\left\{\sigma^{-1}(i) \leq s, \forall i \in S\right\} \cap \left\{\sigma^{-1}(k) = s + 1\right\}\right) \times P\left(U > \frac{s + 1}{n + 1}\right)$$

$$= \frac{1}{n} \frac{1}{\binom{n-1}{s}} \left(1 - \frac{s + 1}{n + 1}\right) = \frac{s!(n - s)!}{(n + 1)!},$$

where we used the fact that σ^{-1} is also uniformly distributed on S_n . Similarly, if $k \in S$, then

$$P\left(W_k = \sum_{i \in S} X_i + \sum_{i \notin S} Y_i\right)$$

$$= P\left(\left\{\sigma^{-1}(i) \le s, \forall i \in S\right\} \cap \left\{\sigma^{-1}(k) = s\right\} \cap \left\{U \le \frac{s}{n+1}\right\}\right)$$

$$= P\left(\left\{\sigma^{-1}(i) \le s, \forall i \in S\right\} \cap \left\{\sigma^{-1}(k) = s\right\}\right) \times P\left(U \le \frac{s}{n+1}\right)$$

$$= \frac{1}{n} \frac{1}{\binom{n-1}{s-1}} \frac{s}{n+1} = \frac{s!(n-s)!}{(n+1)!}.$$

Hence, the probability in (76) is independent of k, and so is the distribution of V_k .

Further, since W_k can only take values of the form $\sum_{i \in S} X_i + \sum_{i \notin S} Y_i$, where S is a subset of $\{1, \ldots, n\}$, it follows that there exists a random vector $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$ with values in $\{0, 1\}^n$ such that the distribution of W_k is equal to that of (75). To see that the random variables $\epsilon_1, \ldots, \epsilon_n$ are exchangeable, note that for any subset S of $\{1, \ldots, n\}$ with s = |S| elements,

$$P(\epsilon_i = 1 \ \forall i \in S \text{ and } \epsilon_i = 0 \ \forall i \notin S)$$
$$= P\left(W_k = \sum_{i \in S} X_i + \sum_{\notin S} Y_i\right) = \frac{s!(n-s)!}{(n+1)!},$$

which depends on the set S only via s. Thus, permuting the random variables $\epsilon_1, \ldots, \epsilon_n$ in the vector ϵ creates a vector with the same distribution, which means that these random variables are exchangeable.

Finally, for any $s = 0, \ldots, n$,

$$P\left(\sum_{i=1}^{n} \epsilon_{i} = s\right) = \binom{n}{s} \frac{s!(n-s)!}{(n+1)!} = \frac{1}{n+1}$$

and

$$E\left[\sum_{i=1}^{n} \epsilon_{i}\right] = \sum_{s=0}^{n} \frac{s}{n+1} = \frac{n(n+1)}{2(n+1)} = \frac{n}{2}.$$

This completes the proof of the lemma.

Lemma 7.3 (Third-Order Matching Multipliers with Gaussian Component). Let $\gamma \in (0; 1/2 - 1/(2\sqrt{5}))$ be a constant. Then

$$\sigma = \left(1 - \frac{(1 - \gamma)^{1/3} \gamma^{1/3}}{(1 - 2\gamma)^{2/3}}\right)^{1/2}$$

is a real number satisfying $\sigma > 0$. Further, denote

$$a = \frac{(1-\gamma)^{2/3}}{\gamma^{1/3}(1-2\gamma)^{1/3}}$$
 and $b = -\frac{\gamma^{2/3}}{(1-\gamma)^{1/3}(1-2\gamma)^{1/3}}$

and let e_1 and e_2 be independent random variables such that e_1 has the $N(0, \sigma^2)$ distribution and e_2 takes values a and b with probabilities γ and $1 - \gamma$, respectively. Then the random variable $e = e_1 + e_2$ has the following properties:

$$E[e] = 0, \quad E[e^2] = 1, \quad and \quad E[e^3] = 1.$$
 (77)

Remark 7.2. The distribution of the random variable *e* constructed in this lemma is different from that used in [20]. It appears that our construction is easier to work with.

Proof. To show that σ is a real number satisfying $\sigma > 0$, it suffices to show that

$$(1 - 2\gamma)^2 > (1 - \gamma)\gamma,$$

which in turn is equivalent to

$$5\gamma^2 - 5\gamma + 1 > 0,$$

which holds by the choice of γ . Further, it is straightforward to check that

$$E[e_2] = 0$$
, $E[e_2^2] = 1 - \sigma^2$, and $E[e_2^3] = 1$.

Thus, given that

$$E[e_1] = 0$$
, $E[e_1^2] = \sigma^2$, and $E[e_1^3] = 0$,

the equalities in (77) follow from

$$E[e] = E[e_1] + E[e_2], E[e^2] = E[e_1^2] + E[e_2^2], \text{ and } E[e^3] = E[e_1^3] + E[e_2^3].$$

This completes the proof of the lemma.

Lemma 7.4 (Randomization Tests with Mass Points). Let \mathcal{X} and X be a set and a random variable taking values in this set. Also, let G be a set of M one-to-one functions mapping \mathcal{X} onto \mathcal{X} such that (i) for all $g \in G$, the distribution of g(X) is equal to that of X, (ii) for all $g \in G$, we have

 $g^{-1} \in G$, and (iii) for all $g_1, g_2 \in G$, we have $g_2 \circ g_1 \in G$. Further, let T be a function mapping \mathcal{X} to \mathbb{R} and for $\alpha \in (0,1)$, define $\phi \colon \mathcal{X} \to \{0,1\}$ by

$$\phi(x) = \begin{cases} 1, & \text{if } \sum_{g \in G} 1\{T(x) > T(g(x))\} \ge M(1 - \alpha), \\ 0, & \text{if } \sum_{g \in G} 1\{T(x) > T(g(x))\} < M(1 - \alpha), \end{cases} \text{ for all } x \in \mathcal{X}.$$

Finally, define $\chi \colon \mathcal{X} \to \mathbb{R}$ by

$$\chi(x) = \max_{t \in \mathbb{R}} |\{g \in G \colon T(g(x)) = t\}| / M, \quad \text{for all } x \in \mathcal{X}.$$

Then

$$\alpha - \mathrm{E}[\chi(X)] \le \mathrm{E}[\phi(X)] \le \alpha.$$

Remark 7.3. If X is observable data and T(X) is a statistic, we can think of $\phi(X)$ as a level α randomization test that exploits symmetries of X with respect to a set of transformations G. The result presented here is then similar to Theorem 15.2.1 in [28], with the difference coming from the fact that we do not allow the function ϕ to take values in (0,1) and instead quantify how much the test can under-reject because of the mass points.

Proof. Define $\phi \colon \mathcal{X} \times \mathcal{X} \to \{0,1\}$ by

$$\phi(x,y) = \begin{cases} 1, & \text{if } \sum_{g \in G} 1\{T(x) > T(g(y))\} \ge M(1-\alpha), \\ 0, & \text{if } \sum_{g \in G} 1\{T(x) > T(g(y))\} < M(1-\alpha), \end{cases} \text{ for all } x, y \in \mathcal{X},$$

so that $\phi(x) = \phi(x, x)$ for all $x \in \mathcal{X}$. Observe that for any $x \in \mathcal{X}$, we have

$$\frac{1}{M} \sum_{g \in G} \phi(g(X), X) \leq \alpha \quad \text{and} \quad \frac{1}{M} \sum_{g \in G} \phi(g(X), X) \geq \alpha - \chi(X)$$

by construction of the function ϕ . Hence,

$$\alpha \ge \frac{1}{M} \sum_{g \in G} \mathbb{E}[\phi(g(X), X)] = \frac{1}{M} \sum_{g \in G} \mathbb{E}[\phi(g(X), g(X))]$$
$$= \frac{1}{M} \sum_{g \in G} \mathbb{E}[\phi(X, X)] = \mathbb{E}[\phi(X, X)] = \mathbb{E}[\phi(X)],$$

where the first equality follows from noting that for all $g_2 \in G$, we have $\{T(g_1(X))\}_{g_1 \in G} = \{T(g_1(g_2(X)))\}_{g_1 \in X}$, and the second from noting that g(X) is equal in distribution to X for all $g \in G$. Similarly, we also have

$$\alpha - \mathrm{E}[\chi(X)] \le \frac{1}{M} \sum_{g \in G} \mathrm{E}[\phi(g(X), X)] = \mathrm{E}[\phi(X)].$$

Combining these bounds gives the asserted claim.

8. Other Useful Lemmas

In this section, we collect maximal, deviation, and anti-concentration inequalities that are useful for our analysis. Lemmas 8.1–8.3 are taken from [14]. Lemma 8.4 is essentially taken from [30].

Lemma 8.1 (Maximal Inequality for Centered Sums). Let X_1, \ldots, X_n be independent centered random vectors in \mathbb{R}^p with $p \geq 2$. Define Z, M, and σ^2 by $Z = \max_{1 \leq j \leq p} |\sum_{i=1}^n X_{ij}|$, $M = \max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |X_{ij}|$ and $\sigma^2 = \max_{1 \leq j \leq p} \sum_{i=1}^n \mathrm{E}[X_{ij}^2]$. Then

$$E[Z] \le K(\sigma \sqrt{\log p} + \sqrt{E[M^2]} \log p).$$

where K is a universal constant.

Lemma 8.2 (Deviation Inequality for Centered Sums). Assume the setting of Lemma 8.1. For every $\eta > 0, \beta \in (0,1]$ and t > 0, we have

$$P\{Z \ge (1+\eta) E[Z] + t\} \le \exp\{-t^2/(3\sigma^2)\} + 3 \exp\{-(t/(K\|M\|_{\psi_\beta}))^\beta\},$$

where K is a constant depending only on η and β .

Lemma 8.3 (Gaussian Anti-Concentration Inequality). Let $Y = (Y_1, \ldots, Y_n)'$ be a centered Gaussian random vector in \mathbb{R}^p with $p \geq 2$ such that $\mathrm{E}[Y_j^2] \geq b$ for all $j = 1, \ldots, p$ and some constant b > 0. Then for every $y \in \mathbb{R}^p$ and t > 0,

$$P(Y \le y + t) - P(Y \le y) \le Ct\sqrt{\log p},$$

where C is a constant depending only on b.

Lemma 8.4 (Rademacher Anti-Concentration Inequality). Let Z_1, \ldots, Z_n be vectors in \mathbb{R}^p with $p \geq 2$ and let e_1, \ldots, e_n be independent Rademacher random variables. Define $Y = n^{-1/2} \sum_{i=1}^n e_i Z_{ij}$ and assume that for some constants $b_1, b_2, B > 0$, (i) $b_1 n \leq \sum_{i=1}^n Z_{ij}^2 \leq b_2 n$ for all $j = 1, \ldots, p$ and (ii) $\|Z_i\|_{\infty} \leq B$ for all $i = 1, \ldots, n$. Then for every $y \in \mathbb{R}^p$ and $t \geq B/\sqrt{n}$,

$$P(Y \le y + t) - P(Y \le y) \le Ct\sqrt{\log p},$$

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

Proof. By the proof of Theorem 7.1 in [30], there exists a constant K depending only on b_1 and b_2 such that for all $y \in \mathbb{R}^p$ and $t \geq B/\sqrt{n}$, we have

$$P(Y \le y + t) - P(Y \le y) \le Kt\sqrt{\log p} + \exp(\log p - K/t^2). \tag{78}$$

Here, since the asserted claim is trivial if $2t^2(\log(1/t) + \log p) > K$, we can assume that $2t^2(\log(1/t) + \log p) \leq K$, in which case the right-hand side of (78) is bounded from above by

$$Kt\sqrt{\log p} + \exp(-K/(2t^2)) \le Kt\sqrt{\log p} + t.$$

The asserted claim follows.

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] Bentkus, V. (1985). Lower bounds for the rate of convergence in the central limit theorem in Banach spaces. *Litovskii Matematicheskii Sbornik* **25** 10-21.
- [6] Canay, I., Romano, J. and Shaikh, A. (2016). Randomization tests under an approximate symmetry assumption. *Econometrica* **85** 1013-1030.
- [7] Chatterjee, S. (2006). A generalization of the Lindeberg principle. *The Annals of Probability* **34** 2061-2076.
- [8] Chen, X. (2018). Gaussian and bootstrap approximations for highdimensional U-statistics and their applications. *Annals of Statistics* **46** 642-678.
- [9] Chen, X. and Kato, K. (2019a). Randomized incomplete *U*-statistics in high dimensions. *Annals of Statistics*, to appear
- [10] Chen, X. and Kato, K. (2019b). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probability Theory and Related Fields*, to appear.
- [11] 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.
- [12] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Anticoncentration and honest, adaptive confidence bands. Annals of Statistics 42 1787-1818.
- [13] 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.
- [14] Chernozhukov, V., Chetverikov, D. and Kato, K. (2017). Central limit theorems and bootstrap in high dimensions. *Annals of Probability* **45** 2309-2352.
- [15] 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.

- [16] Chesher, A. and Rosen, A. (2019). Generalized instrumental variable models, methods, and applications. *Cemmap working paper CWP41/19*.
- [17] Chetverikov, D. (2018). Adaptive tests of conditional moment inequalities. *Econometric Theory* **34** 186-227.
- [18] Chetverikov, D. (2019). Testing regression monotonicity in econometric models. *Econometric Theory* **35** 729-776.
- [19] Chetverikov, D., Wilhelm, D., and Kim, D. (2018). An adaptive test of stochastic monotonicity. Cemmap working paper CWP24/18.
- [20] Deng, H. and Zhang, C.-H. (2017). Beyond Gaussian approximation: bootstrap for maxima of sums of independent random vectors. arXiv:1705.09528.
- [21] Hansen, P. (2005). A test for superior predictive ability. *Journal of Business and Economic Statistics* **23** 365-380.
- [22] Hansen, P., Lunde, A. and Nason, J. (2011). The model confidence set. *Econometrica* **79** 453-497.
- [23] Horowitz, J. (2001). The bootstrap. *Handbook of Econometrics, Volume* 5 3159-3228.
- [24] Koike, Y. (2019a). Gaussian approximation of maxima of Wiener functionals and its application to high-frequency data. Annals of Statistics 47 1663–1687.
- [25] Koike, Y. (2019b). High-dimensional central limit theorems for homogeneous sums. *arXiv:1902.03809*.
- [26] Koike, Y. (2019c). Notes on the dimension dependence in high-dimensional central limit theorems for hyperrectangles. arXiv:1911.00160.
- [27] Koning, N. and Bekker P. (2019). Exact testing of many moment inequalities against multiple violations. arXiv:1904.12775.
- [28] Lehmann, E. and Romano, J. (2005). Testing Statistical Hypotheses. Springer Texts in Statistics.
- [29] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* **21** 255-285.
- [30] O'Donnell, R., Servedio, R., and Tan, L. (2018). Fooling polytopes. arXiv:1808.04035.
- [31] Romano, J., Shaikh, A. and Wolf, M. (2014). A practical two-step method for testing moment inequalities. *Econometrica* 82 1979-2002.
- [32] van der Vaart, A.W. and Wellner, J.A. (1996). Weak Convergence and Empirical Processes: With Applications to Statistics. Springer.
- [33] Vershynin, R. (2011). Introduction to the non-asymptotic analysis of random matrices. arXiv:1011.3027.
- [34] White, H. (2000). A reality check for data snooping. *Econometrica* 5 1007-1126.
- [35] Zhang, X. and Cheng, G. (2017). Gaussian approximation for high dimensional vector under physical dependence. *Bernoulli* **24** 2640–2675.
- [36] Zhang, D. and Wu, W.-B. (2017). Gaussian approximation for high dimensional time series. *Annals of Statistics* **45** 1895-1919.

TABLE 1. Results of Monte Carlo experiments for bootstrap rejection probabilities $P(T_n > c_{1-\alpha}^B)$ with $\alpha = 10\%$ and 4 types of bootstrap: multiplier bootstrap with Gaussian weights (GB), empirical bootstrap (EB), multiplier bootstrap with Rademacher weights (RB), and third-order matching multiplier bootstrap (MB). The case of the Weibull distributions.

Design 1: Asymmetric Distributions, n = 400

k	0		p =	400		p = 800			
	ρ	GB	EB	RB	MB	GB	EB	RB	MB
2	.00	.117	.098	.125	.099	.125	.102	.133	.102
	.25	.121	.100	.126	.099	.121	.097	.129	.097
	.50	.114	.095	.122	.096	.124	.100	.133	.102
	.75	.117	.098	.122	.099	.121	.099	.128	.099
	.00	.110	.105	.115	.105	.106	.100	.114	.101
3	.25	.105	.101	.110	.100	.107	.102	.114	.099
9	.50	.103	.098	.108	.098	.107	.101	.113	.100
	.75	.106	.103	.112	.101	.104	.099	.112	.098
	.00	.096	.099	.101	.097	.095	.099	.102	.098
4	.25	.096	.099	.102	.098	.098	.102	.105	.103
	.50	.093	.095	.097	.095	.100	.102	.107	.103
	.75	.099	.101	.103	.101	.098	.102	.104	.100

Design 2: Symmetric Distributions, n = 100

k	ρ		p =	400		p = 800			
		GB	EB	RB	MB	GB	EB	RB	MB
2	.00	.088	.087	.110	.087	.082	.083	.108	.081
	.25	.083	.082	.104	.082	.082	.083	.108	.081
	.50	.089	.088	.109	.087	.082	.082	.109	.081
	.75	.090	.090	.108	.089	.085	.084	.108	.084
	.00	.088	.090	.109	.088	.086	.086	.109	.084
3	.25	.086	.088	.108	.087	.085	.086	.109	.085
	.50	.090	.090	.110	.089	.087	.088	.110	.086
	.75	.093	.095	.109	.093	.089	.089	.111	.089
4	.00	.086	.090	.108	.086	.085	.086	.108	.081
	.25	.085	.087	.105	.084	.082	.081	.104	.080
	.50	.090	.091	.109	.089	.088	.088	.111	.085
	.75	.092	.092	.107	.090	.093	.092	.113	.091

TABLE 2. Results of Monte Carlo experiments for bootstrap rejection probabilities $P(T_n > c_{1-\alpha}^B)$ with $\alpha = 10\%$ and 4 types of bootstrap: multiplier bootstrap with Gaussian weights (GB), empirical bootstrap (EB), multiplier bootstrap with Rademacher weights (RB), and third-order matching multiplier bootstrap (MB). The case of the Gamma distributions.

Design 1: Asymmetric Distributions, n = 400

k			p =	400		p = 800			
	ρ	GB	EB	RB	MB	GB	EB	RB	MB
1	.00	.143	.081	.166	.087	.157	.084	.190	.092
	.25	.151	.085	.171	.093	.156	.081	.190	.091
	.50	.142	.081	.167	.087	.155	.078	.185	.087
	.75	.143	.082	.164	.088	.150	.080	.179	.088
	.00	.135	.096	.147	.098	.136	.092	.152	.096
3	.25	.131	.092	.143	.095	.140	.092	.155	.095
3	.50	.130	.092	.142	.092	.134	.092	.151	.096
	.75	.129	.096	.140	.097	.130	.090	.144	.093
	.00	.123	.094	.134	.096	.126	.093	.136	.093
5	.25	.124	.095	.133	.096	.130	.094	.144	.097
	.50	.118	.094	.130	.095	.130	.094	.142	.098
	.75	.123	.094	.132	.096	.125	.092	.135	.093

Design 2: Symmetric Distributions, n = 100

k	ρ		p =	400		p = 800			
		GB	EB	RB	MB	GB	EB	RB	MB
1	.00	.070	.061	.107	.068	.064	.053	.110	.061
	.25	.066	.059	.103	.064	.062	.053	.108	.062
	.50	.071	.063	.108	.069	.063	.053	.108	.062
	.75	.074	.066	.107	.072	.065	.055	.104	.062
	.00	.081	.078	.109	.079	.073	.070	.107	.071
3	.25	.080	.077	.107	.079	.076	.072	.109	.074
9	.50	.081	.077	.109	.080	.076	.074	.109	.076
	.75	.087	.085	.111	.086	.082	.076	.112	.079
	.00	.081	.080	.105	.081	.077	.076	.107	.076
5	.25	.081	.079	.105	.079	.077	.075	.106	.076
	.50	.083	.080	.107	.083	.082	.079	.111	.081
	.75	.090	.088	.112	.090	.086	.084	.113	.084

(V. Chernozhukov) Department of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02142, USA.

E-mail address: vchern@mit.edu

(D. Chetverikov) Department of Economics, UCLA, Bunche Hall, 8283, 315 Portola Plaza, Los Angeles, CA 90095, USA.

 $E\text{-}mail\ address: \verb|chetverikov@econ.ucla.edu||$

(K. Kato) Department of Statistics and Data Science, Cornell University, 1194 Comstock Hall, Ithaca 14853, USA.

 $E ext{-}mail\ address: kk976@cornell.edu}$

(Y. Koike) Mathematics and Informatics Center and Graduate School of Mathematical Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.

E-mail address: kyuta@ms.u-tokyo.ac.jp