Robust Max Statistics for High-Dimensional Inference

Appendix S1 introduces preliminary material not covered in the main text. Appendix S2 outlines the proof of Theorem 1, and the main supporting arguments are given in Appendices S3-S5. Appendix S6 contains technical results on median-of-means estimators. Appendix S7 proves Proposition 1 in the main text. Lastly, Appendix S8 contains various background results.

S1 Preliminaries for supplementary material

Notation. The distribution of any random variable U is denoted as $\mathcal{L}(U)$. We write $\mathcal{L}(U|X)$ to refer to the conditional distribution of U given both the hold-out and non-hold-out sets of observations, whereas we write $\mathcal{L}(U|X')$ to refer to the conditional distribution of U given the hold-out set only. Similarly, we use $\mathbf{P}(\cdot|X)$ and $\mathbf{P}(\cdot|X')$ to denote conditional probabilities in the two cases just mentioned, and we use $\|\cdot\|_{L^q|X}$ and $\|\cdot\|_{L^q|X'}$ to denote the corresponding conditional L^q norms.

For any $d \in \{1, \dots, p\}$, recall that $J(d)$ denotes a set of indices corresponding to the d largest values among $\{\sigma_1, \dots, \sigma_p\}$. That is, $\{\sigma_{(1)}, \dots, \sigma_{(d)}\}$ $= \{\sigma_j | j \in J(d)\}$. Letting l_n be as in Assumption 1, define the integer

$$k_n = l_n^5 \wedge p,$$

which always satisfies $1 \leq l_n \leq k_n \leq p$. For each $d \in \{1, \ldots, p\}$, let

$$M_d(X) = \max_{j \in J(d)} \frac{1}{\sigma_j^\tau \sqrt{n}} \sum_{i=1}^{n} (X_{ij} - \mu_j). \tag{S1.1}$$

Letting $G(X) = (G_1(X), \ldots, G_p(X))$ be a centered Gaussian random vector with the same covariance matrix as X_1, the Gaussian counterpart of $M_d(X)$ is defined as

$$\tilde{M}_d(X) = \max_{j \in J(d)} \frac{G_j(X)}{\sigma_j^\tau}.$$

Next, for each $i \in \{1, \ldots, n\}$ and $j, d \in \{1, \ldots, p\}$, define

$$Y_{ij} = \varphi_{t_j}(X_{ij} - \mu_j) \quad \text{and} \quad M_d(Y) = \max_{j \in J(d)} \frac{1}{\sigma_j^\tau \sqrt{n}} \sum_{i=1}^{n} Y_{ij} - \mathbf{E}(Y_{ij}),$$

as well as

$$\hat{Y}_{ij} = \varphi_{\hat{t}_j}(X_{ij} - \mu_j) \quad \text{and} \quad M_d(\hat{Y}) = \max_{j \in J(d)} \frac{1}{\hat{\sigma}_j^\tau \sqrt{n}} \sum_{i=1}^{n} \hat{Y}_{ij}.$$

Let $\xi_1, \ldots, \xi_n$ be i.i.d. standard Gaussian random variables, generated independently of $X_1, \ldots, X_{n+m_n}$, and define

$$M_d^*(X) = \max_{j \in J(d)} \frac{1}{\sigma_j^\tau \sqrt{n}} \sum_{i=1}^{n} \xi_i (X_{ij} - \bar{X}_j). \tag{S1.2}$$

Let $\tilde{X}_j$ denote the median-of-means estimator for μ_j described in Section 2 and define

$$\hat{Z}_{ij} = \varphi_{\hat{t}_j}(X_{ij} - \tilde{X}_j) \quad \text{and} \quad M_d^*(\hat{Z}) = \max_{j \in J(d)} \frac{1}{\hat{\sigma}_j^\tau \sqrt{n}} \sum_{i=1}^{n} \xi_i (\hat{Z}_{ij} - \hat{Z}_j).$$

where we let $\hat{Z}_j = \frac{1}{n}\sum_{i=1}^n \hat{Z}_{ij}$.

Frequently-used inequalities. As a shorthand for the Kolmogorov metric between generic random variables U and V we write

$$d_{\mathrm{K}}(\mathcal{L}(U), \mathcal{L}(V)) = \sup_{t\in\mathbb{R}} |\mathbf{P}(U \le t) - \mathbf{P}(V \le t)|.$$

We will often use the following two basic inequalities that hold for any random variables U and V, and any number $s > 0$,

$$d_{\mathrm{K}}(\mathcal{L}(U), \mathcal{L}(V)) \;\le\; \sup_{t\in\mathbb{R}} \mathbf{P}(|V - t| \le s) \,+\, \mathbf{P}(|U - V| > s), \tag{S1.3}$$

and

$$\sup_{t\in\mathbb{R}} \mathbf{P}(|U - t| \le s) \;\le\; \sup_{t\in\mathbb{R}} \mathbf{P}(|V - t| \le s) + 2d_{\mathrm{K}}(\mathcal{L}(U), \mathcal{L}(V)). \tag{S1.4}$$

If g is a centered Gaussian random variable and $q \ge 1$, then there is an absolute constant $c > 0$ such that

$$\|g\|_{L^q} \;\le\; c\sqrt{q}\|g\|_{L^2}, \tag{S1.5}$$

as recorded in (Vershynin, 2012, Eqn. 2.11). When referring to Chebyshev's inequality, we will typically use it in the following form for a generic random variable U,

$$\mathbf{P}(|U| \ge e\|U\|_{L^q}) \le e^{-q}. \tag{S1.6}$$

For any random variables $U_1, \ldots, U_p$, we have

$$\Big\| \max_{1 \le j \le p} U_j \Big\|_{L^q} \le p^{1/q} \max_{1 \le j \le p} \|U_j\|_{L^q}. \tag{S1.7}$$

Lastly, for any two real vectors $(a_1, \ldots, a_p)$ and $(b_1, \ldots, b_p)$, we have

$$\Big| \max_{1 \le j \le p} a_j - \max_{1 \le j \le p} b_j \Big| \le \max_{1 \le j \le p} |a_j - b_j|. \tag{S1.8}$$

S2 Proof of Theorem 1

Observe that the left side of the bound in Theorem 1 is given by

$$\sup_{s \in \mathbb{R}} \Big| \mathbf{P}(\mathcal{M}_n \le s) - \mathbf{P}(\mathcal{M}_n^* \le s | X) \Big| = d_{\mathrm{K}}\big(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_p^*(\hat{Z})|X)\big). \tag{S2.9}$$

We will bound this distance in three main parts

$$\begin{aligned} d_{\mathrm{K}}\big(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_p^*(\hat{Z})|X)\big) \;\; \le \;\; & d_{\mathrm{K}}\big(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(\tilde{M}_{k_n}(X))\big) & \text{(S2.10)} \\ & + \; d_{\mathrm{K}}\Big(\mathcal{L}(\tilde{M}_{k_n}(X)), \mathcal{L}(M_{k_n}^*(X)|X)\Big) & \text{(S2.11)} \\ & + \; d_{\mathrm{K}}\big(\mathcal{L}(M_{k_n}^*(X)|X), \mathcal{L}(M_p^*(\hat{Z})|X)\big). & \text{(S2.12)} \end{aligned}$$

The three terms on the right side respectively correspond to a Gaussian approximation, a Gaussian comparison, and a bootstrap approximation. These terms are respectively addressed in Proposition 2 of Appendix S3, Proposition 6 of Appendix S4 and Proposition 7 of Appendix S5, which show that all the terms are $\mathcal{O}(n^{-\frac{1}{2}+\epsilon})$ with probability at least $1 - \mathcal{O}(n^{-\delta/4})$. $\square$

Comments on proof techniques. Although decompositions similar to the one in (S2.10)-(S2.12) have become a well-established approach to analyzing max statistics, our work in handling each of the three parts involves a number of technical challenges that differ from those that occur in the more commonly studied setting of light-tailed data (e.g. Chernozhukov et al., 2023; Lopes, 2022, and references therein). At a high level, these extra challenges arise from two main sources, which are (1) that the robust max statistic $M_p(\hat{Y})$ is constructed with ingredients for increasing robustness that are absent from conventional max statistics, and (2) that the presence of heavy-tailed data precludes the use of certain probabilistic tools that are more suited to light-tailed data. For example, in analyzing the term (S2.10), we must account for the bias that is introduced by the truncation functions $\varphi_{\hat{t}_1}, \dots, \varphi_{\hat{t}_p}$ (see the proofs of Lemmas 2 and 10), as well the fluctuations of the robust MOM estimates $\hat{\sigma}_1, \dots, \hat{\sigma}_p$ from their population counterparts $\sigma_1, \dots, \sigma_p$ (see the proofs of Lemmas 3 and 13). Next, the analysis of the term (S2.11) involves a Gaussian comparison that hinges on the entrywise accuracy of a sample covariance matrix in the proof of Lemma 6. Whereas this step would be rather straightforward in the setting of light-tailed data, our analysis requires a more specialized argument based on the Fug-Nagaev inequality (Lemma 21), which can succeed even when the data have only a limited number of moments. The third term (S2.12) also involves accounting for the effect of centering the coordinates of the original observations with the robust MOM estimates $\tilde{X}_1, \ldots, \tilde{X}_p$ for the population means $\mu_1, \ldots, \mu_p$ (see the proofs of Lemmas 7 and 12), as well as the effects of the truncation functions and scale estimates mentioned in connection with the term (S2.10). Lastly, it is worth noting that existing results on the MOM estimates $\tilde{X}_1, \ldots, \tilde{X}_p$ and $\hat{\sigma}_1, \ldots, \hat{\sigma}_p$ were not adequate for purposes, and consequently, we developed Lemmas 12-15 to quantify the performance of the MOM estimates in a way that is specifically adapted to our setting of high dimensionality with variance decay structure.

Conventions. In the appendices supporting the proof of Theorem 1, we may assume without loss of generality that ϵ satisfies $\epsilon < 1/2$ and that $n \geq c$ for any fixed constant $c > 0$, for otherwise the result is trivially true. Also, we will often use c to denote a generic positive constant not depending on n, whose value may differ at each appearance. Lastly, we may assume without loss of generality that $(\mu_1, \ldots, \mu_p) = \mathbf{E}(X_1) = 0$, because the conditions in Assumption 1 are shift invariant, and that $\max_{1\leq j\leq p} \sigma_j^2 = 1$, because the Kolmogorov metric is scale invariant. To avoid repetitiveness, these conventions will not be stated explicitly in most of the results presented in the appendices.

S3 Gaussian approximation

Proposition 2. *If the conditions of Theorem 1 hold, then*

$$d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(\tilde{M}_{k_n}(X))) \lesssim n^{-\frac{1}{2}+\epsilon}.$$

Proof. The proof is based on the decomposition

$$d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(\tilde{M}_{k_n}(X))) \leq \mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n,$$

where we define

$$\mathrm{I}_n = d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_{k_n}(\hat{Y}))),$$

$$\mathrm{II}_n = d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(\hat{Y})), \mathcal{L}(M_{k_n}(Y))),$$

$$\mathrm{III}_n = d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(Y)), \mathcal{L}(\tilde{M}_{k_n}(X))).$$

Below, the terms I_n, II_n, and III_n are shown to be at most of order $n^{-\frac{1}{2}+\epsilon}$ in Propositions 3, 4, and 5 respectively. $\square$

Proposition 3. *If the conditions of Theorem 1 hold, then*

$$d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_{k_n}(\hat{Y}))) \lesssim n^{-\frac{1}{2}+\epsilon}.$$

Proof. For any $t \in \mathbb{R}$, define the events

$$A(t) = \Big\{ \max_{j \in J(k_n)} \frac{\sum_{i=1}^n \hat{Y}_{ij}}{\sqrt{n}\hat{\sigma}_j^\tau} \leq t \Big\} \quad \text{and} \quad B(t) = \Big\{ \max_{j \in J(k_n)^c} \frac{\sum_{i=1}^n \hat{Y}_{ij}}{\sqrt{n}\hat{\sigma}_j^\tau} > t \Big\}.$$

It is straightforward to check that for any $t \in \mathbb{R}$, we have

$$\Big| \mathbf{P}(M_p(\hat{Y}) \leq t) - \mathbf{P}(M_{k_n}(\hat{Y}) \leq t) \Big| \;=\; \mathbf{P}(A(t) \cap B(t)).$$

Next, it can be checked that for any real numbers $s_{1,n}$ and $s_{2,n}$ satisfying $s_{1,n} \leq s_{2,n}$, the inclusion

$$A(t) \cap B(t) \ \subset \ A(s_{2,n}) \cup B(s_{1,n})$$

holds simultaneously for all $t \in \mathbb{R}$. Therefore, after taking the supremum over $t \in \mathbb{R}$, we have

$$d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_{k_n}(\hat{Y}))) \ \leq \ \mathbf{P}(A(s_{2,n})) + \mathbf{P}(B(s_{1,n})). \tag{S3.13}$$

Let

$$\omega = \frac{\epsilon}{24(\beta \vee 1)C}, \tag{S3.14}$$

$$d_n = \left\lfloor \tfrac{\omega^2}{4} \mathtt{r}(R(l_n)) \vee 2 \right\rfloor, \tag{S3.15}$$

where $\mathtt{r}(R(l_n)) := \operatorname{tr}(R(l_n))^2 / \|R(l_n)\|_F^2 = \frac{l_n^2}{\|R(l_n)\|_F^2}$ is the stable rank of $R(l_n)$. We will choose $s_{1,n}$ and $s_{2,n}$ according to

$$\begin{aligned} s_{1,n} &= c_1 k_n^{-\beta(1-\tau)/2} (\log(n) \vee 2), \\ s_{2,n} &= c_2 l_n^{-\beta(1-\tau)} \sqrt{\log(d_n)}, \end{aligned} \tag{S3.16}$$

for some constants $c_1, c_2 > 0$ not depending on n. It can also be checked that for any fixed choices of c_1 and c_2, the inequality $s_{1,n} \leq s_{2,n}$ holds for all large n due to the definitions of k_n, l_n and d_n.

To bound $\mathbf{P}(A(s_{2,n}))$, we have

$$\mathbf{P}\big(A(s_{2,n})\big) \ \leq \mathbf{P}(\tilde{M}_{k_n}(X) \leq s_{2,n}) + d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(\hat{Y})), \mathcal{L}(\tilde{M}_{k_n}(X))),$$

where the first term on the right hand side is of order $n^{-1/2}$ by Lemma 1, and the second term is of order $n^{-\frac{1}{2}+\epsilon}$ by Propositions 4 and 5. Lastly, Lemma 2 shows that $\mathbf{P}(B(s_{1,n}))$ is of order $\frac{1}{n}$, which completes the proof. $\square$

Lemma 1. *Suppose conditions of Theorem 1 hold. Then, there is a constant $c_2 > 0$, not depending on n, such that the following bound holds when $s_{2,n} = c_2 l_n^{-\beta(1-\tau)}\sqrt{\log(d_n)}$,*

$$\mathbf{P}\Big(\tilde{M}_{k_n}(X) \leq s_{2,n}\Big) \lesssim n^{-1/2}.$$

Proof. Observe that for any $t \in \mathbb{R}$,

$$\mathbf{P}(\tilde{M}_{k_n}(X) \leq t) \leq \mathbf{P}(\tilde{M}_{l_n}(X) \leq t). \tag{S3.17}$$

Let $(a_j)_{j\in J(l_n)}$ and b be positive numbers with $\max_{j\in J(l_n)} a_j \leq b$. For any sequence of random variables $(U_j)_{j\in J(l_n)}$ and $t \geq 0$, it is straightforward to check that

$$\mathbf{P}\Big(\max_{j\in J(l_n)} U_j \leq t\Big) \leq \mathbf{P}\Big(\max_{j\in J(l_n)} a_j U_j \leq bt\Big).$$

Consider the choice $a_j = \sigma_j^{-(1-\tau)}$ and note that under Assumption 1(iii), there is a constant $c_0 > 0$ not depending on n such that the bound $a_j \leq c_0 l_n^{(1-\tau)\beta} =: b$ holds for all $j \in J(l_n)$. So, if we let $U_j = G_j(X)/\sigma_j^{\tau}$, then the previous two displays imply

$$\mathbf{P}\big(\tilde{M}_{k_n}(X) \leq t\big) \ \leq \ \mathbf{P}\Big(\max_{j\in J(l_n)} \frac{G_j(X)}{\sigma_j} \leq c_0 l_n^{(1-\tau)\beta} t\Big).$$

Consequently, if we let ω be as defined in (S3.14), and let $c_2 = \frac{1}{c_0}\omega\sqrt{2(1-\omega)}$ in the definition (S3.16) of $s_{2,n}$, then choosing $t = s_{2,n}$ in the previous display gives

$$\mathbf{P}\Big(\tilde{M}_{k_n}(X) \leq s_{2,n}\Big) \;\leq\; \mathbf{P}\Big(\max_{j\in J(l_n)} \frac{G_j(X)}{\sigma_j} \;\leq \omega\sqrt{2(1-\omega)\log(d_n)}\Big).$$

To bound the probability on the right, we apply Lemma 17 with $(l_n, d_n, \omega, \omega)$ playing the roles of (d, k, a, b) in the statement of that result, which yields

$$\mathbf{P}\Big(\tilde{M}_{k_n}(X) \leq s_{2,n}\Big) \;\lesssim\; d_n^{-(1-\omega)^3/\omega}(\log(d_n))^{\frac{1-\omega(2-\omega)-\omega}{2\omega}}.$$

(Note that the conditions of Lemma 17 are applicable because $2 \leq d_n \leq \frac{\omega^2}{4}\mathtt{r}(R(l_n))$ when n is sufficiently large.) Furthermore, by Assumption 1(iv), we have

$$d_n \asymp \mathtt{r}(R(l_n)) = \frac{l_n^2}{\|R(l_n)\|_F^2} \;\gtrsim\; l_n^{\frac{1}{C}} \;\gtrsim\; n^{\omega}.$$

Hence, there is a constant $c > 0$ not depending on n such that

$$\begin{aligned}
\mathbf{P}\Big(\tilde{M}_{k_n}(X) \leq s_{2,n}\Big) \;&\lesssim\; d_n^{-\frac{(1-\omega)^3}{\omega}}\log(d_n)^c \\
&\lesssim\; n^{-(1-\omega)^3}\log(n)^c \\
&\lesssim\; n^{-1/2}
\end{aligned}$$

as needed. □

Lemma 2. *If the conditions of Theorem 1 hold, then there is a constant $c_1 > 0$ not depending on n such that the following bound holds when $s_{1,n} =$*

$c_1 k_n^{-\beta(1-\tau)/2}(2 \vee \log(n))$,

$$\mathbf{P}\Big(B(s_{1,n})\Big) \lesssim \tfrac{1}{n}.$$

Proof. Let $q = 2 \vee \log(n)$ and observe that

$$\Big\|\max_{j\in J(k_n)^c} \tfrac{\sum_{i=1}^n \hat{Y}_{ij}}{\hat{\sigma}_j^\tau \sqrt{n}}\Big\|_{L^q|X'}^q \leq \sum_{j\in J(k_n)^c} \left(2^q \Big\|\tfrac{\sum_{i=1}^n \hat{Y}_{ij} - \mathbf{E}(\hat{Y}_{ij}|X')}{\hat{\sigma}_j^\tau \sqrt{n}}\Big\|_{L^q|X'}^q + 2^q \Big|\tfrac{\sum_{i=1}^n \mathbf{E}(\hat{Y}_{ij}|X')}{\hat{\sigma}_j^\tau \sqrt{n}}\Big|^q\right)$$

holds almost surely. By Rosenthal's inequality (Lemma 16), the following event holds almost surely,

$$\begin{aligned}\Big\|\frac{\sum_{i=1}^n \hat{Y}_{ij} - \mathbf{E}(\hat{Y}_{ij}|X')}{\hat{\sigma}_j^\tau \sqrt{n}}\Big\|_{L^q|X'} &\leq \frac{cq}{\hat{\sigma}_j^\tau} \max\Big\{\sqrt{\mathrm{var}(\hat{Y}_{1j}|X')}\ , n^{-1/2+1/q}\|\hat{Y}_{1j}\|_{L^q|X'}\Big\}\\ &\leq \frac{cq}{\hat{\sigma}_j^\tau} \max\Big\{\sigma_j\ , n^{1/q}\hat{\sigma}_j\Big\},\end{aligned}$$

where the second step follows from $\mathrm{var}(\hat{Y}_{1j}|X') \leq \mathbf{E}(X_{1j}^2) = \sigma_j^2$ and $|\hat{Y}_{1j}| \leq n^{1/2}\hat{\sigma}_j$. Lemmas 13(iv) and 15 imply that there is an constant $c > 0$ not depending on n such that both of the bounds

$$\max_{j\in J(k_n)^c} \frac{\hat{\sigma}_j^{1-\tau}}{\sigma_j^{(1-\tau)/2}} \leq c \quad \text{and} \quad \max_{j\in J(k_n)^c} \frac{\sigma_j}{\hat{\sigma}_j^\tau \sigma_j^{(1-\tau)/2}} \leq c$$

hold simultaneously with probability at least $1 - cn^{-(2+\delta)}$. Consequently, the bound

$$\begin{aligned}\Big\|\frac{\sum_{i=1}^n \hat{Y}_{ij} - \mathbf{E}(\hat{Y}_{ij}|X')}{\hat{\sigma}_j^\tau \sqrt{n}}\Big\|_{L^q|X'} &\leq cq \max\Big\{\sigma_j^{(1-\tau)/2},\ \sigma_j^{(1-\tau)/2} n^{1/q}\Big\}\\ &\leq cq\sigma_j^{(1-\tau)/2}\end{aligned}$$

holds simultaneously over all $j \in J(k_n)^c$ with probability at least $1 - cn^{-(2+\delta)}$. Using Lemma 10 and similar reasoning, it can also be shown that

$$\left|\frac{\mathbf{E}(\hat{Y}_{ij}|X')}{\hat{\sigma}_j^{\tau}\sqrt{n}}\right| \leq c\sigma_j^{(1-\tau)/2}n^{-2}$$

holds simultaneously over all $j \in J(k_n)^c$ with probability at least $1 - cn^{-(2+\delta)}$. Combining the last several steps and Assumption 1(iii), we conclude that the bound

$$\begin{aligned}\left\|\max_{j\in J(k_n)^c}\frac{\sum_{i=1}^n \hat{Y}_{ij}}{\hat{\sigma}_j^{\tau}\sqrt{n}}\right\|_{L^q|X'}^q &\leq (cq)^q\sum_{j\geq k_n} j^{-q\beta(1-\tau)/2}\\ &\leq c^q\frac{q^q}{(1-\tau)q\beta/2-1}k_n^{1-(1-\tau)q\beta/2}\end{aligned}$$

holds with probability at least $1 - cn^{-(2+\delta)}$, where in the second step we have used the fact that $q\beta(1-\tau) > 2$ when n is sufficiently large. Also, since $q \asymp \log(n)$, we have

$$\Big(\frac{1}{(1-\tau)q\beta/2-1}\Big)^{1/q} \lesssim 1.$$

Thus, there is a sufficiently large choice of $c_1 > 0$, such that if $s_{1,n} = c_1 q k_n^{-\beta(1-\tau)/2}$, then the bound

$$\mathbf{P}(B(s_{1,n})|X') \leq \frac{1}{s_{1,n}^q}\left\|\max_{j\in J(k_n)^c}\frac{\sum_{i=1}^n \hat{Y}_{ij}}{\hat{\sigma}_j^{\tau}\sqrt{n}}\right\|_{L^q|X'}^q \leq e^{-q} \lesssim \tfrac{1}{n}$$

holds with probability at least $1 - cn^{-(2+\delta)}$. This implies the stated result.

$\square$

Proposition 4. *If the conditions of Theorem 1 hold, then*

$$d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}(\hat{Y})), \mathcal{L}(M_{k_n}(Y))\Big) \lesssim n^{-\frac{1}{2}+\epsilon}.$$

Proof. Using the decomposition (S1.3) for the Kolmogorov metric, followed by the bound for anti-concentration probabilities in (S1.4), we have

$$\begin{aligned} d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(\hat{Y})), \mathcal{L}(M_{k_n}(Y))) \leq\ & \sup_{t\in\mathbb{R}} \mathbf{P}\big(|\tilde{M}_{k_n}(X) - t| \leq n^{-\frac{1}{2}+\frac{3\epsilon}{4}} \log(n)\big) \\ & +\ 2d_{\mathrm{K}}(\mathcal{L}(\tilde{M}_{k_n}(X)), \mathcal{L}(M_{k_n}(Y))) \\ & +\ \mathbf{P}\Big(\big|M_{k_n}(\hat{Y}) - M_{k_n}(Y)\big| \geq n^{-\frac{1}{2}+\frac{3\epsilon}{4}} \log(n)\Big). \end{aligned}$$

For the first term on the right side, Nazarov's inequality (Lemma 20) and Assumption 1(iii) imply

$$\begin{aligned} \sup_{t\in\mathbb{R}} \mathbf{P}\Big(|\tilde{M}_{k_n}(X) - t| \leq n^{-\frac{1}{2}+\frac{3\epsilon}{4}} \log(n)\Big) & \lesssim\ n^{-\frac{1}{2}+\frac{3\epsilon}{4}} \log(n) k_n^{\beta(1-\tau)} \sqrt{\log(k_n)} \\ & \lesssim\ n^{-\frac{1}{2}+\epsilon}. \end{aligned} \tag{S3.18}$$

Next, it follows from Proposition 5 that

$$d_{\mathrm{K}}\Big(\mathcal{L}(\tilde{M}_{k_n}(X)), \mathcal{L}(M_{k_n}(Y))\Big)\ \lesssim\ n^{-\frac{1}{2}+\epsilon}.$$

Finally, Lemma 3 implies

$$\mathbf{P}\Big(\big|M_{k_n}(\hat{Y}) - M_{k_n}(Y)\big| \geq n^{-\frac{1}{2}+\frac{\epsilon}{2}} \log(n)\Big) \lesssim\ n^{-\frac{1}{2}+\epsilon},$$

which completes the proof. ☐

Lemma 3. *Lemma: MOM variance If the conditions of Theorem 1 hold, then*

$$\mathbf{P}\Big(\big|M_{k_n}(\hat{Y}) - M_{k_n}(Y)\big| \geq n^{-1/2+\epsilon/2} \log(n)\Big) \lesssim n^{-1/2+\epsilon}.$$

Proof. First observe that $\left|M_{k_n}(\hat{Y}) - M_{k_n}(Y)\right|$ can be bounded by

$$\max_{j\in J(k_n)}\left|\frac{\sum_{i=1}^n \hat{Y}_{ij} - Y_{ij}}{\hat{\sigma}_j^\tau\sqrt{n}}\right| + \max_{j\in J(k_n)}\left|\frac{\sum_{i=1}^n Y_{ij}}{\hat{\sigma}_j^\tau\sqrt{n}} - \frac{\sum_{i=1}^n Y_{ij}}{\sigma_j^\tau\sqrt{n}}\right| + \max_{j\in J(k_n)}\frac{\sqrt{n}|\mathbf{E}(Y_{1j})|}{\sigma_j^\tau}. \tag{S3.19}$$

The first term in the bound (S3.19) is 0 with probability at least $1-\frac{ck_n}{n}$ by Lemma 4, and the (deterministic) third term is $\mathcal{O}(n^{-1})$ by Lemma 10.

It remains to handle the middle term in the bound (S3.19), which satisfies

$$\max_{j\in J(k_n)}\left|\frac{\sum_{i=1}^n Y_{ij}}{\hat{\sigma}_j^\tau\sqrt{n}} - \frac{\sum_{i=1}^n Y_{ij}}{\sigma_j^\tau\sqrt{n}}\right| \le \max_{j\in J(k_n)}\left|\left(\frac{\sigma_j}{\hat{\sigma}_j}\right)^\tau - 1\right| \cdot \max_{j\in J(k_n)}\left|\frac{\sum_{i=1}^n Y_{ij}}{\sqrt{n}\sigma_j^\tau}\right|. \tag{S3.20}$$

Since $|a^\tau - 1| \le |a^2-1|$ for any $a\ge 0$ and $\tau\in[0,1)$, the first factor on the right is of order $n^{-1/2+\epsilon/2}$ with probability at least $1-cn^{-(2+\delta)}$ by Lemma 13(iii).

Next we will show the second factor in the bound (S3.20) is of order $\log(n)$ with probability at least $1-\frac{c}{n}$. Let $q = 2\vee\log(n)$ and observe that under Assumption 1(iii), Lemma 10 implies

$$\left\|\max_{j\in J(k_n)}\left|\frac{\sum_{i=1}^n Y_{ij}}{\sqrt{n}\sigma_j^\tau}\right|\right\|_{L^q} \lesssim \left\|\max_{j\in J(k_n)}\left|\sum_{i=1}^n\frac{Y_{ij}-\mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j^\tau}\right|\right\|_{L^q} + \frac{1}{n}.$$

Furthermore, Lemma 11 gives

$$\begin{aligned}
\Big\| \max_{j\in J(k_n)} \Big| \sum_{i=1}^{n} \frac{Y_{ij}-\mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j^{\tau}} \Big| \Big\|_{L^q}^{q} &\leq \sum_{j\in J(k_n)} \Big\| \sum_{i=1}^{n} \frac{Y_{ij}-\mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j^{\tau}} \Big\|_{L^q}^{q} \\
&\leq c^q q^q \sum_{j\in J(k_n)} \sigma_j^{(1-\tau)q} \\
&\leq c^q q^q.
\end{aligned}$$

Therefore, Chebyshev's inequality implies that the second factor in the bound (S3.20) is of order $\log(n)$ with probability $1-\frac{c}{n}$. □

Lemma 4. *If the conditions of Theorem 1 hold, then*

$$\mathbf{P}\Big(\max_{j\in J(k_n)} \max_{1\leq i\leq n} |Y_{ij}-\hat{Y}_{ij}| > 0 \Big) \lesssim \frac{k_n}{n}.$$

Proof. First observe that for each i and j, the definitions of Y_{ij} and $\hat{Y}_{ij}$ give

$$\mathbf{P}\big(|Y_{ij}-\hat{Y}_{ij}|>0\big) \;\leq\; \mathbf{P}\Big(|X_{ij}| > \min\{t_j, \hat{t}_j\}\Big).$$

Recalling that $\hat{t}_j = \hat{\sigma}_j\sqrt{n}$ and $t_j = \sigma_j\sqrt{n}$, the events $\{\hat{t}_j > t_j/2\}$ for $j \in J(k_n)$ occur simultaneously with probability at least $1-cn^{-(2+\delta)}$ by Lemma 13(iii). Combining this with a union bound over $i = 1, \ldots, n$, we have

$$\begin{aligned}
\mathbf{P}\Big(\max_{1\leq i\leq n} |Y_{ij}-\hat{Y}_{ij}| > 0 \Big) &\lesssim \sum_{i=1}^{n} \Big(\mathbf{P}\big(|X_{1j}| > \tfrac{t_j}{2}\big) \;+\; n^{-(2+\delta)} \Big) \\
&\lesssim \frac{n\,\mathbf{E}|X_{1j}|^4}{t_j^4} + \; n^{-(1+\delta)} \qquad \text{(S3.21)} \\
&\lesssim \frac{1}{n},
\end{aligned}$$

where the last step uses Assumption 1(i). Finally, taking a union bound over $j \in J(k_n)$,

$$\mathbf{P}\Big(\max_{j\in J(k_n)}\max_{1\leq i\leq n}|Y_{ij}-\hat{Y}_{ij}|>0\Big)\lesssim\frac{k_n}{n}$$

as needed. □

Lemma 5. *If the conditions of Theorem 1 hold, then*

$$\mathbf{P}\Big(|M_{k_n}(Y)-M_{k_n}(X)|\geq n^{-\frac{1}{2}}\Big)\ \lesssim\ n^{-\frac{1}{2}+\epsilon},$$

Proof. For each $i=1,\dots,n$ and $j\in J(k_n)$, let

$$\Delta_{ij}=\frac{1}{\sigma_j^{\tau}\sqrt{n}}\Big(Y_{ij}-\mathbf{E}(Y_{ij})-X_{ij}\Big),$$

so that

$$|M_{k_n}(Y)-M_{k_n}(X)|\ \leq\ \max_{j\in J(k_n)}\Big|\sum_{i=1}^{n}\Delta_{ij}\Big|.$$

Noting that Y_{ij} and X_{ij} only differ when $|X_{ij}|>t_j$, we have

$$\begin{aligned}\mathbf{E}\big(|\Delta_{ij}|\big)\ &\lesssim\ \mathbf{E}\Big(\tfrac{1}{\sqrt{n}\sigma_j^{\tau}}|X_{ij}|1\{|X_{ij}|>t_j\}\Big)+\frac{1}{\sqrt{n}\sigma_j^{\tau}}\big|\mathbf{E}(Y_{ij})\big|\\&\lesssim\ \frac{1}{\sqrt{n}\sigma_j^{\tau}}\|X_{ij}\|_{L^{\frac{4}{1+\epsilon}}}\|1\{|X_{ij}|>t_j\}\|_{L^{\frac{4}{3-\epsilon}}}+n^{-2}\\&\lesssim\ n^{-2+\frac{\epsilon}{2}}.\end{aligned}$$

where we have used Hölder's inequality and Lemma 10 in the first step, followed by Chebyshev's inequality in bounding $\|1\{|X_{ij}|>t_j\}\|_{L^{\frac{4}{3-\epsilon}}}$. Therefore,

$$\begin{aligned}
\mathbf{P}\Big(\max_{j\in J(k_n)} \Big|\sum_{i=1}^{n} \Delta_{ij}\Big| \geq n^{-\frac{1}{2}}\Big) &\leq \sum_{j\in J(k_n)} \mathbf{P}\Big(\Big|\sum_{i=1}^{n} \Delta_{ij}\Big| \geq n^{-\frac{1}{2}}\Big) \\
&\lesssim k_n n^{-\frac{1}{2}+\frac{\epsilon}{2}} \\
&\lesssim n^{-\frac{1}{2}+\epsilon},
\end{aligned}$$

as needed. □

Proposition 5. *If the conditions of Theorem 1 hold, then*

$$d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(Y)), \mathcal{L}(\tilde{M}_{k_n}(X))) \lesssim n^{-\frac{1}{2}+\epsilon}.$$

Proof. First observe that

$$\begin{aligned}
d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(Y)), \mathcal{L}(\tilde{M}_{k_n}(X))) &\leq d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(Y)), \mathcal{L}(M_{k_n}(X))) \\
&\quad + d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(X)), \mathcal{L}(\tilde{M}_{k_n}(X))).
\end{aligned}$$

The second term on the right side is of $n^{-1/2+\epsilon}$ by Lemma 6. Using the decomposition (S1.3) for the Kolmogorov metric, the first term on the right side can be bounded by

$$\begin{aligned}
d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(Y)), \mathcal{L}(M_{k_n}(X))) \leq &\sup_{t\in\mathbb{R}} \mathbf{P}(|M_{k_n}(X) - t| \leq n^{-\frac{1}{2}}) \\
&+ \mathbf{P}(|M_{k_n}(X) - M_{k_n}(Y)| > n^{-\frac{1}{2}}).
\end{aligned}$$

Lemma 5 shows that

$$\mathbf{P}(|M_{k_n}(X) - M_{k_n}(Y)| > n^{-\frac{1}{2}}) \lesssim n^{-\frac{1}{2}+\epsilon}.$$

Using the bound for anti-concentration probabilities in (S1.4), we have

$$\sup_{t\in\mathbb{R}} \mathbf{P}(|M_{k_n}(X) - t| \leq n^{-\frac{1}{2}}) \leq \sup_{t\in\mathbb{R}} \mathbf{P}(|\tilde{M}_{k_n}(X) - t| \leq n^{-\frac{1}{2}})$$
$$+ 2d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(X)), \mathcal{L}(\tilde{M}_{k_n}(X))).$$

Nazarov's inequality (Lemma 20) and Assumption 1(iii) imply

$$\sup_{t\in\mathbb{R}} \mathbf{P}(|\tilde{M}_{k_n}(X) - t| \leq n^{-\frac{1}{2}})) \lesssim \; n^{-\frac{1}{2}+\epsilon}, \tag{S3.22}$$

which completes the proof.

Lemma 6. *If the conditions of Theorem 1 hold, then*

$$d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(X)), \mathcal{L}(\tilde{M}_{k_n}(X))) \lesssim n^{-1/2+\epsilon}.$$

Proof. For each $i = 1, \ldots, n$, let $X_i(k_n)$ denote the vector in $\mathbb{R}^{k_n}$ corresponding to the coordinates of X_i indexed by $J(k_n)$, and let $\mathcal{R}_t = \prod_{j\in J(k_n)}(-\infty, t\sigma_j^\tau]$, so that

$$\mathbf{P}(M_{k_n}(X) \leq t) = \mathbf{P}\Big(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n X_i(k_n) \in \mathcal{R}_t\Big).$$

If the rank of the covariance matrix of $X_1(k_n)$ is denoted by r, let $\Pi_r \in \mathbb{R}^{k_n\times r}$ be the matrix whose columns correspond to the leading r eigenvectors of the covariance matrix of $X_1(k_n)$. In particular, we have $X_1(k_n) = \Pi_r\Pi_r^\top X_1(k_n)$ almost surely, and it follows that

$$\mathbf{P}(M_{k_n}(X) \leq t) = \mathbf{P}\Big(\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n \Pi_r^\top X_i(k_n) \in \Pi_r^{-1}(\mathcal{R}_t)\Big),$$

where $\Pi_r^{-1}(\mathcal{R}_t)$ refers to the pre-image. Next, the definition of Π_r ensures that the covariance matrix of $\Pi_r^\top X_1(k_n)$, denoted by $\mathfrak{S}_r$, is invertible. So, if we define the random vector

$$V_i = \mathfrak{S}_r^{-1/2}\Pi_r^\top X_i(k_n)$$

for each $i = 1, \dots, n$ and the set $\mathcal{C}_t = \mathfrak{S}_r^{-1/2}\Pi_r^{-1}(\mathcal{R}_t)$ for each $t \in \mathbb{R}$, then

$$\mathbf{P}(M_{k_n}(X) \leq t) = \mathbf{P}\Big(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n V_i \in \mathcal{C}_t\Big).$$

It can also be shown by similar reasoning that

$$\mathbf{P}(\tilde{M}_{k_n}(X) \leq t) = \gamma_r(\mathcal{C}_t),$$

where γ_r denotes the standard Gaussian distribution on $\mathbb{R}^r$. Due to the fact that the i.i.d. random vectors $V_1, \dots, V_n$ are centered and isotropic, Bentkus' multivariate Berry-Esseen theorem (Lemma 19) ensures there is an absolute constant $c > 0$ such that

$$\begin{aligned} d_{\mathrm{K}}(\mathcal{L}(M_{k_n}(X)), \mathcal{L}(\tilde{M}_{k_n}(X))) \leq \ & \sup_{\mathcal{C}\in\mathscr{C}} \Big|\mathbf{P}\Big(\tfrac{1}{\sqrt{n}}\textstyle\sum_{i=1}^n V_i \in \mathcal{C}\Big) - \gamma_r(\mathcal{C})\Big| \\ \leq \ & \frac{cr^{1/4}\mathbf{E}\|V_1\|_2^3}{\sqrt{n}}, \end{aligned} \tag{S3.23}$$

where $\mathscr{C}$ denotes the collection of all Borel convex subsets of $\mathbb{R}^r$. To bound the third moment on the right side, observe that

$$\mathbf{E}\|V_1\|_2^3 \ = \ \Big\|\textstyle\sum_{j=1}^r V_{1j}^2\Big\|_{L^{3/2}}^{3/2} \ \leq \ \Big(\textstyle\sum_{j=1}^r \|V_{1j}\|_{L^3}^2\Big)^{3/2}.$$

Since V_{1j} can be expressed as $\langle v, X_1\rangle$ for some vector $v \in \mathbb{R}^p$, Assumption 1(i) implies $\|V_{1j}\|_{L^3}^2 \lesssim \operatorname{var}(V_{1j}) = 1$ for each $j = 1, \ldots, r$. Thus $\mathbf{E}\|V_1\|_2^3 \lesssim r^{3/2} \leq k_n^{3/2}$, and combining with (S3.23) completes the proof. $\square$

S4 Gaussian comparison

Recall that $M_{k_n}^*(X) = \max_{j\in J(k_n)} \frac{1}{\sqrt{n}\sigma_j^\tau} \sum_{i=1}^n \xi_i(X_{ij} - \bar{X}_j)$ from (S1.2).

Proposition 6. *Suppose the conditions of Theorem 1 hold. Then, there is a constant $c > 0$ not depending on n such that the event*

$$d_{\mathrm{K}}\Big(\mathcal{L}(\tilde{M}_{k_n}(X))\,,\, \mathcal{L}(M_{k_n}^*(X)|X)\Big) \leq cn^{-\frac{1}{2}+\epsilon}$$

holds with probability at least $1 - cn^{-\delta/4}$.

Proof. Let $V_1, \ldots, V_n \in \mathbb{R}^r$ be as in the proof of Lemma 6 and put $\bar{V} = \frac{1}{n}\sum_{i=1}^n V_i$. The reasoning used in the proof of Lemma 6 shows that for any $t \in \mathbb{R}$, there is a convex Borel set $\mathcal{C}_t \subset \mathbb{R}^r$ such that

$$\mathbf{P}\big(M_{k_n}^*(X) \leq t|X\big) \;=\; \mathbf{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i(V_i - \bar{V}) \in \mathcal{C}_t\Big|X\Big),$$

and also, that

$$\mathbf{P}(\tilde{M}_{k_n}(X) \leq t) = \gamma_r(\mathcal{C}_t),$$

where γ_r is the standard Gaussian distribution on $\mathbb{R}^r$. Next, define the sample covariance matrix

$$W_r = \frac{1}{n}\sum_{i=1}^{n}(V_i - \bar{V})(V_i - \bar{V})^\top$$

and observe that Lemma 18 gives the following almost-sure bound,

$$\sup_{t\in\mathbb{R}} \Big| \mathbf{P}\big(M^*_{k_n}(X) \le t \big| X\big) \ - \ \mathbf{P}\big(\tilde{M}_{k_n}(X) \le t\big)\Big| \ \le \ 2\big\|W_r - I_r\big\|_F,$$

where I_r denotes the identity matrix of size $r \times r$. To handle the Frobenius norm, Assumption 1(i) implies

$$\max_{1\le j,j'\le r} \mathbf{E}\Big|e_j^\top(V_1V_1^\top - I_r)e_{j'}\Big|^{\frac{4+\delta}{2}} \lesssim 1,$$

$$\max_{1\le j\le r} \mathbf{E}\big|e_j^\top V_1\big|^{4+\delta} \lesssim 1.$$

Consequently, the Fuk-Nagaev inequality (Lemma 21) with the choices $q = (4+\delta)/2$ and $q = 4+\delta$ in the notation of that result ensures that for each $j, j' = 1, \ldots, r$,

$$\mathbf{P}\bigg(\Big|\frac{1}{n}\sum_{i=1}^{n} e_j^\top(V_iV_i^\top - I_r)e_{j'}\Big| \ge n^{-1/2+\epsilon/2}\bigg) \ \lesssim \ n^{-(\delta/4+\epsilon)},$$

$$\mathbf{P}\Big(\big|e_j^\top \bar{V}\big| \ge n^{-1/4+\epsilon/4}\Big) \le n^{-(2+3\delta/4+\epsilon)}.$$

So, using the identity

$$W_r - I_r = \Big(\frac{1}{n}\sum_{i=1}^{n} V_iV_i^\top - I_r\Big) - \bar{V}\bar{V}^\top$$

it is straightforward to check that the event

$$\big\|W_r - I_r\big\|_F \ge 2k_n n^{-1/2+\epsilon/2}$$

holds with probability at most of order $k_n^2 n^{-(\delta/4+\epsilon)} \lesssim n^{-\delta/4}$, which leads to the stated result. □

S5 Bootstrap approximation

Proposition 7. *Suppose the conditions of Theorem 1 hold. Then, there is a constant* $c > 0$ *not depending on* n *such that the event*

$$d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}^*(X)|X), \mathcal{L}(M_p^*(\hat{Z})|X))\Big) \le cn^{-\frac{1}{2}+\epsilon}$$

holds with probability at least $1 - cn^{-\delta/4}$.

Proof. Consider the inequality

$$d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}^*(X)|X), \mathcal{L}(M_p^*(\hat{Z})|X))\Big) \le \mathrm{I}_n' + \mathrm{II}_n',$$

where we define

$$\mathrm{I}_n' = d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}^*(X)|X), \mathcal{L}(M_{k_n}^*(\hat{Z})|X)\Big),$$

$$\mathrm{II}_n' = d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}^*(\hat{Z})|X), \mathcal{L}(M_p^*(\hat{Z})|X)\Big).$$

Both I_n' and II_n' are of order at most $n^{-1/2+\epsilon}$ with probability at least $1 - cn^{-\delta/4}$, as shown in Lemma 7 and Proposition 8. □

Lemma 7. *Suppose the conditions of Theorem 1 hold. Then, there is a constant* $c > 0$ *not depending on* n *such that the event*

$$d_{\mathrm{K}}\Big(\mathcal{L}(M_{k_n}^*(X)|X), \mathcal{L}(M_{k_n}^*(\hat{Z})|X)\Big) \le cn^{-1/2+\epsilon}$$

holds with probability at least $1 - cn^{-\delta/4}$.

Proof. The coupling and anti-concentration decomposition in (S1.3) shows that for any $\eta > 0$, we have

$$d_{\mathrm{K}}\Big(\mathcal{L}\big(M^*_{k_n}(X)|X\big), \mathcal{L}\big(M^*_{k_n}(\hat{Z})|X\big)\Big) \;\leq\; \sup_{t\in\mathbb{R}} \mathbf{P}\Big(|M^*_{k_n}(X) - t| \leq \eta\Big|X\Big) + \mathbf{P}\Big(\big|M^*_{k_n}(X) - M^*_{k_n}(\hat{Z})\big| \geq \eta\Big|X\Big). \tag{S5.24}$$

We will take $\eta = c\log(n)n^{-\frac{1}{2}+\frac{3\epsilon}{4}}$ for some constant $c > 0$ not depending on n. Using the generic bound for anti-concentration probabilities in (S1.4), the first term on the right side of (S5.24) is upper bounded by

$$\sup_{t\in\mathbb{R}} \mathbf{P}\big(|\tilde{M}_{k_n}(X) - t| \leq \eta\big) + \; 2d_{\mathrm{K}}\Big(\mathcal{L}\big(\tilde{M}_{k_n}(X)\big), \mathcal{L}\big(M^*_{k_n}(X)|X\big)\Big),$$

which is at most of order $n^{-1/2+\epsilon}$ with probability at least $1 - cn^{-\delta/4}$, due to (S3.18) and Proposition 6.

To address the coupling term in (S5.24), observe that for any $q \geq 2$, the basic inequalities (S1.5), (S1.7) and (S1.8) ensure there is a constant $c > 0$ not depending on n such that

$$\big\|M^*_{k_n}(X) - M^*_{k_n}(\hat{Z})\big\|_{L^q|X} \;\leq\; c\sqrt{q}k_n^{1/q} \max_{j\in J(k_n)} \bigg(\frac{1}{n}\sum_{i=1}^n \Big(\tfrac{X_{ij}-\bar{X}_j}{\sigma_j^\tau} - \tfrac{\hat{Z}_{ij}-\hat{Z}_j}{\hat{\sigma}_j^\tau}\Big)^2\bigg)^{1/2}.$$

To decompose this bound, define the random variables

$$T_1 = \max_{j\in J(k_n)} \left(\frac{1}{n}\sum_{i=1}^n \left(\tfrac{X_{ij}-\hat{Z}_{ij}}{\sigma_j^\tau}\right)^2\right)^{1/2},$$

$$T_2 = \max_{j\in J(k_n)} \left|\frac{1}{\sigma_j^\tau} - \frac{1}{\hat{\sigma}_j^\tau}\right| \left(\frac{1}{n}\sum_{i=1}^n \hat{Z}_{ij}^2\right)^{1/2}.$$

Using Jensen's inequality for sample averages, it follows that

$$\left\|M_{k_n}^*(X) - M_{k_n}^*(\hat{Z})\right\|_{L^q|X} \ \le \ c\sqrt{q}k_n^{1/q}(T_1+T_2).$$

With regard to T_1, note that

$$|X_{ij}-\hat{Z}_{ij}| \ \le \ |\tilde{X}_j| 1\{|X_{ij}-\tilde{X}_j| \le \hat{t}_j\} \ + \ (|X_{ij}|+\hat{t}_j)1\big\{|X_{ij}-\tilde{X}_j| > \hat{t}_j\big\}.$$

Also, by Lemma 13(iii), the events

$$\max_{j\in J(k_n)} \frac{\hat{t}_j}{t_j} \ \le \ 2 \qquad \text{and} \qquad \min_{j\in J(k_n)} \frac{\hat{t}_j}{t_j} \ \ge \ \tfrac{1}{2}$$

hold simultaneously with probability at least $1-cn^{-(2+\delta)}$ for some constant $c>0$ not depending on n. Based on this and $\min_{j\in J(k_n)} \sigma_j^\tau \gtrsim k_n^{-\beta\tau}$ under Assumption 1(iii), the bound

$$\left(\tfrac{X_{ij}-\hat{Z}_{ij}}{\sigma_j^\tau}\right)^2 \ \le \ ck_n^{2\beta\tau}\tilde{X}_j^2 \ + \ ck_n^{2\beta\tau}(X_{ij}^2+t_j^2)\Big(1\big\{|X_{ij}| > \tfrac{t_j}{4}\big\} + 1\big\{|\tilde{X}_j| > \tfrac{t_j}{4}\big\}\Big)$$

holds simultaneously for all $i\in\{1,\dots,n\}$ and $j\in J(k_n)$ with probability at least $1-cn^{-(2+\delta)}$. Therefore, the bound

$$T_1 \le ck_n^{\beta\tau} \max_{j\in J(k_n)} \left(|\tilde{X}_j| + \left(\frac{1}{n}\sum_{i=1}^n \left(X_{ij}^2+t_j^2\right)1\{|X_{ij}| > \tfrac{t_j}{4}\}\right)^{\frac{1}{2}} + 1\{|\tilde{X}_j| > \tfrac{t_j}{4}\}\left(\frac{1}{n}\sum_{i=1}^n X_{ij}^2+t_j^2\right)^{\frac{1}{2}}\right)$$

holds with the same probability. The terms on the right side are handled as follows. First, $\max_{j\in J(k_n)}|\tilde{X}_j|$ is of order $n^{-1/2+\epsilon/2}$ with probability at least $1-cn^{-(2+\delta)}$ by Assumption 1(iii) and Lemma 12(iii). The indicators $1\{|X_{ij}|\geq \frac{t_j}{4}\}$ and $1\{|\tilde{X}_j|\geq \frac{t_j}{4}\}$ are 0 with probability at least $1-cn^{-2}$ due to the argument associated with the bounds in (S3.21), as well as Lemma 12(iii) and the fact that $\sigma_j n^{-\frac{1}{2}+\frac{\epsilon}{2}}\lesssim \frac{t_j}{4}$. After taking a union bound over $i\in\{1,\dots,n\}$ and $j\in J(k_n)$, the event

$$T_1\leq ck_n^{\beta\tau}n^{-1/2+\epsilon/2}\leq cn^{-1/2+3\epsilon/4}$$

holds with probability at least $1-ck_n/n$.

Turning our attention to T_2, Lemma 13(iii) implies that the bound

$$\max_{j\in J(k_n)}\left|\frac{1}{\sigma_j^\tau}-\frac{1}{\hat{\sigma}_j^\tau}\right|\leq ck_n^{\beta\tau}n^{-1/2+\epsilon/2}\leq cn^{-1/2+3\epsilon/4},$$

holds with probability at least $1-cn^{-(2+\delta)}$. Also, it will be shown in Equation (S5.26) that the bound

$$\max_{j\in J(k_n)}\Big(\frac{1}{n}\sum_{i=1}^n\hat{Z}_{ij}^2\Big)^{1/2}\leq c\log(n)$$

holds with probability at least $1-ck_n/n$. Combining the last two steps shows that T_2 is of order $\log(n)n^{-1/2+3\epsilon/4}$ with the same probability. $\square$

Proposition 8. *Suppose the conditions of Theorem 1 hold. Then, there is a constant $c>0$ not depending on n such that the event*

$$d_{\mathrm{K}}(\mathcal{L}(M_{k_n}^*(\hat{Z})|X),\mathcal{L}(M_p^*(\hat{Z})|X))\leq cn^{-\frac{1}{2}+\epsilon},$$

holds with probability at least $1 - cn^{-\delta/4}$.

Proof. We may assume without loss of generality that $k_n < p$, for otherwise the quantity $d_{\mathrm{K}}(\mathcal{L}(M_p(\hat{Y})), \mathcal{L}(M_{k_n}(\hat{Y})))$ is zero. For any $t \in \mathbb{R}$, define the events

$$A'(t) = \Big\{ \max_{j \in J(k_n)} \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}\hat{\sigma}_j^\tau} \le t \Big\} \quad \text{and}$$
$$B'(t) = \Big\{ \max_{j \in J(k_n)^c} \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}\hat{\sigma}_j^\tau} > t \Big\}.$$

Using the argument in the proof of Proposition 3, it can be shown that the following bound holds almost surely for any real numbers $s'_{1,n} \le s'_{2,n}$,

$$d_{\mathrm{K}}\Big(\mathcal{L}(M^*_{k_n}(\hat{Z})|X), \mathcal{L}(M^*_p(\hat{Z})|X)\Big) \ \le\ \mathbf{P}\big(A'(s'_{2,n})|X\big) \ +\ \mathbf{P}\big(B'(s'_{1,n})|X\big).$$

If we choose

$$s'_{1,n} = c'_1 k_n^{-\beta(1-\tau)/4} \log(n)^{3/2}, \quad s'_{2,n} = c'_2 l_n^{-\beta(1-\tau)} \sqrt{\log(d_n)},$$

where $c'_1, c'_2 > 0$ are constants not depending on n, then $s'_{1,n} \le s'_{2,n}$ holds when n is sufficiently large (regardless of the particular values of c'_1 and c'_2). Recall also that d_n is defined in (S3.15). Lemma 8 shows that there is a choice of c'_1 such that random variable $\mathbf{P}\big(B'(s'_{1,n})|X\big)$ is at most n^{-1} with probability at least $1 - cn^{-(1+\delta)}$. To deal with and $\mathbf{P}\big(A'(s'_{2,n})|X\big)$, notice that

$$\mathbf{P}\big(A'(s'_{2,n})|X\big) \le \mathbf{P}(\tilde{M}_{k_n}(X) \le s'_{2,n}) \ +\ d_{\mathrm{K}}(\mathcal{L}(\tilde{M}_{k_n}(X), \mathcal{L}(M^*_{k_n}(\hat{Z})|X)).$$

Lemma 1 shows there is a choice of c_2' such that the first term on the right hand side is of order $n^{-1/2}$. Finally, the second term is of order $n^{-1/2+\epsilon}$ with probability at least $1 - cn^{-\delta/4}$ by Proposition 6 and Lemma 7. □

Lemma 8. *If the conditions of Theorem 1 hold, then there are constants $c, c_1' > 0$ not depending on n such that the following event holds with probability at least $1 - cn^{-(1+\delta)}$ when $s_{1,n}' = c_1' k_n^{-\beta(1-\tau)/4} \log(n)^{3/2}$,*

$$\mathbf{P}\big(B'(s_{1,n}')|X\big) \leq \frac{1}{n}.$$

Proof. Notice that

$$\max_{j \in J(k_n)^c} \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}\hat{\sigma}_j^\tau} \leq \max_{j \in J(k_n)^c} \frac{\sigma_j^{(\tau+1)/2}}{\hat{\sigma}_j^\tau} \cdot \max_{j \in J(k_n)^c} \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}\sigma_j^{(\tau+1)/2}}.$$

The first factor on the right side is of order 1 with probability at least $1 - cn^{-(2+\delta)}$ by Lemma 15. To handle the second factor, let $q = \max\{2, (1+\delta)\log(n)\}$. The idea of the rest of the proof is to construct a number b_n such that the following event holds with high probability for every realization of the data,

$$\Big\| \max_{j \in J(k_n)^c} \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}\sigma_j^{(\tau+1)/2}} \Big\|_{L^q|X} \leq b_n.$$

This will lead to the statement of the lemma via Chebyshev's inequality, because it will turn out that $s'_{1,n} \asymp b_n$. To construct b_n, first observe that

$$\begin{aligned}\Big\|\max_{j\in J(k_n)^c}\frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij}-\hat{Z}_j)}{\sqrt{n}\sigma_j^{(\tau+1)/2}}\Big\|_{L^q|X}^q &\leq \sum_{j\in J(k_n)^c}\sigma_j^{-q(\tau+1)/2}\Big\|\frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij}-\hat{Z}_j)}{\sqrt{n}}\Big\|_{L^q|X}^q\\ &\leq c^q q^{3q/2}\sum_{j\in J(k_n)^c}\sigma_j^{(1-\tau)q/4}\\ &\leq c^q\frac{q^{3q/2}}{(1-\tau)q\beta/4-1}k_n^{1-(1-\tau)q\beta/4}\end{aligned}$$

where the second inequality holds with probability at least $1-cn^{-(1+\delta)}$ by Lemma 9, and the third inequality follows from the fact that $q\beta(1-\tau)/4>1$ when n is sufficiently large. Since $q\asymp\log(n)$, we have

$$\Big(\frac{1}{(1-\tau)q\beta/4-1}k_n\Big)^{1/q}\lesssim 1,$$

and so the event

$$\Big\|\max_{j\in J(k_n)^c}\frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij}-\hat{Z}_j)}{\sqrt{n}\sigma_j^{(\tau+1)/2}}\Big\|_{L^q|X}\leq cq^{3/2}k_n^{-(1-\tau)\beta/4}$$

holds with probability at least $1-cn^{-(1+\delta)}$. Thus, we may take b_n to be of the form $b_n=cq^{3/2}k_n^{-(1-\tau)\beta/4}$, and there is a choice of c_1' such that the stated result holds. □

Lemma 9. *Let $q=\max\{2,(1+\delta)\log(n)\}$ and suppose the conditions of Theorem 1 hold. Then, there is a constant $c>0$ not depending on n such that the bound*

$$\Big\|\frac{1}{\sqrt{n}}\sum_{i=1}^n\xi_i(\hat{Z}_{ij}-\hat{Z}_j)\Big\|_{L^q|X}\leq cq^{3/2}\sigma_j^{\frac{\tau+3}{4}}$$

holds simultaneously over all $j \in J(k_n)^c$ *with probability at least* $1-cn^{-(1+\delta)}$.

Proof. The L^q-norm bound for centered Gaussian random variables in (S1.5) gives

$$\bigg\| \frac{\sum_{i=1}^n \xi_i(\hat{Z}_{ij} - \hat{Z}_j)}{\sqrt{n}} \bigg\|_{L^q|X} \;\leq\; c\sqrt{q}\bigg(\frac{1}{n}\sum_{i=1}^n \hat{Z}_{ij}^2\bigg)^{1/2}. \tag{S5.25}$$

To develop a high-probability bound for the right side, note that we have $|\hat{Z}_{ij}| \leq |\hat{Y}_{ij}| + |\tilde{X}_j|$ for any fixed j, and so

$$\frac{1}{n}\sum_{i=1}^n \hat{Z}_{ij}^2 \;\leq\; \frac{2}{n}\sum_{i=1}^n \hat{Y}_{ij}^2 \;+\; 2|\tilde{X}_j|^2.$$

Here, we apply Lemma 12(iv) with $\theta = (1-\tau)/4$ in the notation used there. This implies there is a constant $c > 0$ not depending on n such that the bound

$$|\tilde{X}_j|^2 \leq c\sigma_j^{\frac{\tau+3}{2}}$$

holds simultaneously over all $j \in J(k_n)^c$ with probability at least $1 - cn^{-(1+\delta)}$. Next, Rosenthal's inequality for non-negative random variables (Lemma 16) gives

$$\begin{aligned}
\bigg\|\frac{1}{n}\sum_{i=1}^n \hat{Y}_{ij}^2\bigg\|_{L^q|X'} \;&\leq\; cq\max\Big\{\|\hat{Y}_{1j}^2\|_{L^1|X'}\,,\; n^{-1+1/q}\|\hat{Y}_{1j}^2\|_{L^q|X'}\Big\} \\
&\leq\; cq\max\Big\{\|X_{1j}^2\|_{L^1}\,,\; n^{-1+1/q}\hat{t}_j^2\Big\} \\
&\leq\; cq(\sigma_j^2 + \hat{\sigma}_j^2),
\end{aligned}$$

where the second step uses $|\hat{Y}_{1j}| \leq |X_{1j}| \wedge \hat{t}_j$. Applying Chebyshev's inequality conditionally on X' shows that the event

$$\mathbf{P}\Big(\frac{1}{n}\sum_{i=1}^n \hat{Y}_{ij}^2 \;\geq\; ceq(\sigma_j^2+\hat{\sigma}_j^2)\sigma_j^{\frac{\tau-1}{4}}\Big|X'\Big) \;\leq\; e^{-q}\sigma_j^{\frac{(1-\tau)q}{4}} \;\leq cn^{-(1+\delta)}\sigma_j^{\frac{(1-\tau)q}{4}}$$

holds with probability 1, and thus the unconditional version of the left hand side is also at most $cn^{-(1+\delta)}\sigma_j^{\frac{(1-\tau)q}{4}}$. Consequently, the event

$$\frac{1}{n}\sum_{i=1}^n \hat{Y}_{ij}^2 \;\leq\; ceq(\sigma_j^2+\hat{\sigma}_j^2)\sigma_j^{\frac{\tau-1}{4}}$$

holds simultaneously over all $j \in J(k_n)^c$ with probability at least $1-cn^{-(1+\delta)}$ since

$$\sum_{j\in J(k_n)^c} j^{-\frac{(1-\tau)q\beta}{4}} \lesssim k_n^{-\frac{(1-\tau)q\beta}{4}+1} \lesssim 1.$$

Finally, Lemma 13(iv) implies there is a constant $c > 0$ not depending on n that that the bound

$$\hat{\sigma}_j^2 \leq c\sigma_j^{\frac{\tau+7}{4}}$$

holds simultaneously over all $j \in J(k_n)^c$ with probability at least $1 - cn^{-(2+\delta)}$. Combining results above, we have that the bound

$$\frac{1}{n}\sum_{i=1}^n \hat{Z}_{ij}^2 \;\leq\; cq\sigma_j^{\frac{\tau+3}{2}} \tag{S5.26}$$

holds simultaneously over $j \in J(k_n)^c$ with probability at least $1 - cn^{-(1+\delta)}$, which completes the proof. $\square$

Lemma 10. *B orderIf the conditions of Theorem 1 hold, then there is a constant $c > 0$ not depending on n such that the following bounds hold for all $j, k \in \{1, \dots, p\}$,*

$$\begin{aligned} |\mathbf{E}(Y_{1j})| &\leq c\sigma_j n^{-\frac{3}{2}} \\ |\mathbf{E}(\hat{Y}_{1j}|X')| &\leq c\sigma_j^4 \hat{\sigma}_j^{-3} n^{-\frac{3}{2}} \quad \text{(almost surely)} \\ |\text{cov}(Y_{1j}, Y_{1k}) - \text{cov}(X_{1j}, X_{1k})| &\leq c\sigma_j \sigma_k n^{-1}. \end{aligned}$$

Proof. We may assume without loss of generality that $\mathbf{E}(X_{1j}) = 0$ for all $j \in \{1, \dots, p\}$. Observe that Assumption 1(i) gives

$$\begin{aligned} |\mathbf{E}(Y_{1j})| &\leq \mathbf{E}(|X_{1j}| 1\{|X_{1j}| \geq t_j\}) \\ &\leq \|X_{1j}\|_{L^4} \|1\{|X_{1j}| \geq t_j\}\|_{L^{4/3}} \\ &\leq c\sigma_j \Big(\tfrac{\|X_{1j}\|_{L^4}^4}{t_j^4}\Big)^{3/4} \\ &\leq c\sigma_j n^{-\frac{3}{2}}. \end{aligned}$$

Second, the stated bound on $|\mathbf{E}(\hat{Y}_{1j}|X')|$ can also be obtained from essentially the same argument. Third, to bound the difference between the covariances, we use the fact that $|Y_{1j}Y_{1k} - X_{1j}X_{1k}|$ vanishes on the intersection of the events $\{|X_{1j}| \leq t_j\}$ and $\{|X_{1k}| \leq t_k\}$, and otherwise it is at most $2|X_{1j}X_{1k}|$. Therefore, we have

$$\begin{aligned}\big|\mathbf{E}(Y_{1j}Y_{1k}) - \mathbf{E}(X_{1j}X_{1k})\big| &\leq 2\mathbf{E}\big(|X_{1j}X_{1k}|1\{|X_{1j}| \geq t_j\}\big)\\ &\quad + 2\mathbf{E}\big(|X_{1j}X_{1k}|1\{|X_{1k}| \geq t_k\}\big).\end{aligned}$$

The two terms on the right hand side can be handled via

$$\begin{aligned}\mathbf{E}\big(|X_{1j}X_{1k}|1\{|X_{1j}| \geq t_j\}\big) &\leq \|X_{1j}\|_{L^4}\|X_{1k}\|_{L^4}\big\|1\{|X_{1j}| \geq t_j\}\big\|_{L^2}\\ &\leq c\sigma_j\sigma_k\Big(\tfrac{\|X_{1j}\|_{L^4}^4}{t_j^4}\Big)^{1/2}\\ &\leq \frac{c\sigma_j\sigma_k}{n},\end{aligned}$$

which yields the stated result. □

Lemma 11. *If the conditions of Theorem 1 hold and* $q = \max\{\log(n), 2\}$, *then*

$$\Bigg\|\sum_{i=1}^n \frac{Y_{ij} - \mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j}\Bigg\|_{L^q} \lesssim q.$$

Proof. Due to Rosenthal's inequality (Lemma 16), we have

$$\begin{aligned}\Bigg\|\sum_{i=1}^n \frac{Y_{ij} - \mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j}\Bigg\|_{L^q} &\lesssim q\max\Bigg\{\Bigg\|\sum_{i=1}^n \frac{Y_{ij} - \mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j}\Bigg\|_{L^2}, \Bigg(\sum_{i=1}^n\Bigg\|\frac{Y_{ij} - \mathbf{E}(Y_{ij})}{\sqrt{n}\sigma_j}\Bigg\|_{L^q}^q\Bigg)^{1/q}\Bigg\}\\ &\lesssim q\max\{1, n^{1/q}\}\\ &\lesssim q,\end{aligned}$$

where the second step uses the almost-sure bound $|Y_{ij}| \leq \sqrt{n}\sigma_j$, as well as Lemma 10 to relate the variance of Y_{ij} with σ_j. □

S6 Results on median-of-means estimators

The results in this section will continue to follow the convention that $\max_{1\leq j\leq p}\sigma_j^2 = 1$, as discussed on p.6. However, to make the results easier to interpret, we will state them so that they explicitly account for the coordinate-wise means $\mu_j = \mathbf{E}(X_{1j})$, $j = 1,\dots,p$ (even though these parameters may be assumed to be zero without loss of generality in proving Theorem 1).

Lemma 12. *lemma: XZ kn-kn Fix any constant $\theta \in (0,1)$ and suppose that the conditions of Theorem 1 hold. Then, there is a constant $c \geq 1$ not depending on n, such that for any $j \in \{1,\dots,p\}$, the median-of-means estimator $\tilde{X}_j$ with $b_n \asymp \log(n)$ blocks satisfies*

$$\mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq \sigma_j n^{-1/2+\epsilon/2}\Big) \lesssim \Big(\frac{c}{n}\Big)^{\frac{\epsilon}{c}b_n} \tag{i}$$

$$\mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq \sigma_j^{1-\theta} n^{-1/2+\epsilon/2}\Big) \lesssim (c\sigma_j)^{\frac{\theta}{c}b_n}. \tag{ii}$$

Furthermore, we have

$$\sum_{j\in J(k_n)} \mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq \sigma_j n^{-1/2+\epsilon/2}\Big) \lesssim n^{-(2+\delta)} \tag{iii}$$

$$\sum_{j=1}^{p} \mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq C\sigma_j^{1-\theta} n^{-1/2+\epsilon/2}\Big) \lesssim k_n^{-\log(n)/c}. \tag{iv}$$

Proof. Recall the notation $\bar{X}_j(l) = \frac{1}{\ell_n}\sum_{i\in\mathcal{B}_l} X_{ij}$ where $l = 1,\dots,b_n$. Fix $t > 0$ and let $\xi_{jl} = 1\big\{|\bar{X}_j(l) - \mu_j| \geq t\big\}$. Since the event $\{|\tilde{X}_j(l) - \mu_j| \geq t\}$ can only occur if at least half of the random variables $\xi_{j1},\dots,\xi_{jb_n}$ are 1, we must have

$$\mathbf{P}\big(|\tilde{X}_j - \mu_j| \geq t\big) \;\leq\; \mathbf{P}\bigg(\frac{1}{b_n}\sum_{l=1}^{b_n}\xi_{jl} \;\geq \frac{1}{2}\bigg).$$

Next, applying Kiefer's inequality (Lemma 22) to the right side gives

$$\mathbf{P}(|\tilde{X}_j - \mu_j| \geq t) \;\lesssim\; \big(e\mathbf{E}(\xi_{1j})\big)^{b_n(\frac{1}{2}-\mathbf{E}(\xi_{j1}))^2}. \tag{S6.27}$$

Furthermore, by Chebyshev's inequality, $\mathbf{E}(\xi_{jl}) \lesssim \frac{\sigma_j^2}{\ell_n t^2}$, and so if we take $t = \sigma_j n^{-1/2+\epsilon/2}$, then

$$\mathbf{E}(\xi_{jl}) \;\lesssim\; \frac{n^{1-\epsilon}}{\ell_n} \;\lesssim\; n^{-\epsilon/2},$$

where the last step uses $n/\ell_n \asymp m_n/\ell_n = b_n \asymp \log(n)$. Thus, combining this bound on $\mathbf{E}(\xi_{jl})$ with (S6.27) establishes the first claim (i). Similarly, choosing $t = \sigma_j^{1-\theta} n^{-1/2+\epsilon/2}$ in the previous argument leads to the second claim (ii).

For the fourth claim (iv), we decompose the sum over $j = 1, \ldots, p$ along the indices in $J(k_n)$ and $J(k_n)^c$. To bound the sum over $J(k_n)^c$, we may use (ii) to obtain

$$\begin{aligned}
\sum_{j\in J(k_n)^c} \mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq C\sigma_j^{1-\theta} n^{-1/2+\epsilon/2}\Big) &\lesssim \sum_{j\in J(k_n)^c} (c\sigma_j)^{\frac{\theta}{c}b_n} \\
&\lesssim \sum_{j\geq k_n} (Cj^{-\beta})^{\frac{\theta}{c}b_n} \\
&\lesssim k_n^{-\frac{\beta\theta}{c}b_n+1} \\
&\lesssim k_n^{-\log(n)/c}.
\end{aligned}$$

Regarding the sum over $j \in J(k_n)$, note that $C\sigma_j^{1-\theta} \geq \sigma_j$ holds for all $j = 1, \ldots, p$. Therefore the bound (i) gives

$$\begin{aligned}
\sum_{j\in J(k_n)} \mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq C\sigma_j^{1-\theta} n^{-1/2+\epsilon/2}\Big) &\leq \sum_{j\in J(k_n)} \mathbf{P}\Big(|\tilde{X}_j - \mu_j| \geq C\sigma_j n^{-1/2+\epsilon/2}\Big) \\
&\lesssim k_n \Big(\frac{c}{n}\Big)^{\frac{\epsilon}{c}b_n} \\
&\lesssim n^{-(2+\delta)}.
\end{aligned}$$

This leads to the third claim (iii) and completes the proof. □

Lemma 13. *Fix any constant $\theta \in (0,1)$, and suppose that the conditions of Theorem 1 hold. Then, there is a constant $c \geq 1$ not depending on n, such that the following bounds hold for any $j \in \{1, \ldots, p\}$,*

$$\mathbf{P}\Big(|\hat{\sigma}_j^2 - \sigma_j^2| > \sigma_j^2 n^{-1/2+\epsilon/2}\Big) \lesssim \Big(\frac{c}{n}\Big)^{\frac{\epsilon}{c}b_n} \tag{i}$$

$$\mathbf{P}\Big(|\hat{\sigma}_j^2 - \sigma_j^2| \geq \sigma_j^{2-2\theta} n^{-1/2+\epsilon/2}\Big) \lesssim (c\sigma_j^2)^{\frac{\theta}{c}b_n}. \tag{ii}$$

Furthermore, we have

$$\sum_{j\in J(k_n)} \mathbf{P}\Big(|\hat\sigma_j^2-\sigma_j^2| > C^2\sigma_j^2 n^{-1/2+\epsilon/2}\Big) \;\lesssim\; n^{-(2+\delta)}. \tag{iii}$$

$$\sum_{j=1}^{p} \mathbf{P}\Big(|\hat\sigma_j^2-\sigma_j^2| > C^2\sigma_j^{2-2\theta} n^{-1/2+\epsilon/2}\Big) \;\lesssim\; k_n^{-\log(n)/c}. \tag{iv}$$

Proof. The proof of Lemma 12 can be repeated with the i.i.d. random variables $\frac{1}{2}(X_{ij}-X_{i'j})^2$, playing the role that X_{ij} previously did. Also note that because $\operatorname{var}\big(\frac{1}{2}(X_{ij}-X_{i'j})^2\big) \lesssim \sigma_j^4$ holds under Assumption 1(i), the parameter σ_j^4 plays the role that σ_j^2 did in the context of Lemma 12. □

For the next lemma, recall that for each $l \in \{1,\dots,b_n\}$, the lth blockwise variance estimate $\bar\sigma_j^2(l)$ for σ_j^2 is defined in (2.5) of the main text.

Lemma 14. *Fix any constant $\theta \in (0,1)$, and suppose that the conditions of Theorem 1 hold. Then, there is a constant $c \geq 1$ not depending on n, such that the following bound holds for any $j \in \{1,\dots,p\}$ and any $l \in \{1,\dots,b_n\}$,*

$$\mathbf{P}\big(\bar\sigma_j^2(l) \leq \sigma_j^{2+2\theta}\big) \;\lesssim\; (c\ell_n\sigma_j^{2\theta})^{\ell_n/4}.$$

Proof. Because the $\ell_n/2$ terms in the definition of $\bar\sigma_j^2(l)$ are i.i.d., it follows that

$$\begin{aligned}\mathbf{P}\big(\bar\sigma_j^2(l) \;\leq\; \sigma_j^{2+2\theta}\big) \;&\leq\; \mathbf{P}\Bigg(\max_{\substack{i,i'\in\mathcal{B}_l\\ i'-i=\ell_n/2}} \tfrac{1}{\ell_n}(X_{ij}-X_{i'j})^2 \leq \sigma_j^{2+2\theta}\Bigg)\\ &=\; \mathbf{P}\Bigg(\tfrac{1}{2\sigma_j^2}(X_{1j}-X_{(\ell_n/2+1)j})^2 \leq \tfrac{1}{2}\ell_n\sigma_j^{2\theta}\Bigg)^{\ell_n/2}.\end{aligned} \tag{S6.28}$$

Since the independent random variables X_{1j}/σ_j and $X_{(1+\ell_n/2)j}/\sigma_j$ have densities whose L^∞ norms are $\mathcal{O}(1)$ under Assumption 1(ii), it follows from Young's convolution inequality (Stein and Weiss, 1971, p.178) that the random variable $\frac{1}{\sqrt{2}\sigma_j}(X_{1j} - X_{(\ell_n/2+1)j})$ also has a density whose L^∞ norm is $\mathcal{O}(1)$, and so

$$\mathbf{P}\left(\frac{1}{2\sigma_j^2}\big(X_{1j} - X_{(\ell_n/2+1)j}\big)^2 \leq \frac{1}{2}\ell_n\sigma_j^{2\theta}\right) \lesssim \big(\ell_n\sigma_j^{2\theta}\big)^{1/2}.$$

Combining this with (S6.28) completes the proof. □

Lemma 15. *Fix any constant $\theta \in (0,1)$, and suppose that the conditions of Theorem 1 hold. Then, there is a constant $c \geq 1$ not depending on n, such that the event*

$$\max_{1\leq j\leq p} \frac{\sigma_j^{2+2\theta}}{\hat{\sigma}_j^2} \leq c$$

holds with probability at least $1 - cn^{-(2+\delta)}$.

Proof. Let $r_n = \lceil n^{\epsilon/(\theta\beta)} \wedge p\rceil$. It follows from Lemma 13(i) that there is a constant $c > 0$ not depending on n such that the bound

$$\max_{j\in J(r_n)} \frac{\sigma_j^2}{\hat{\sigma}_j^2} \leq c \tag{S6.29}$$

holds with probability at least $1 - cn^{-(2+\delta)}$. Therefore, a bound of the same form must also hold for $\max_{j\in J(r_n)} \frac{\sigma_j^{2+2\theta}}{\hat{\sigma}_j^2}$, since $\max_{1\leq j\leq p}\sigma_j^{2\theta} \lesssim 1$.

To complete the proof, it remains to handle the maximum of $\sigma_j^{2+2\theta}/\hat{\sigma}_j^2$ over indices j in the complementary set $J(r_n)^c$. Letting $\xi_{jl} = 1\{\bar{\sigma}_j^2(l) \leq$

$\sigma_j^{2+2\theta}\}$ for $l = 1, \ldots, b_n$, Kiefer's inequality (Lemma 22) implies that the following bound holds for any $j \in J(r_n)$,

$$\mathbf{P}\big(\hat{\sigma}_j^2 \le \sigma_j^{2+2\theta}\big) \;\le\; \mathbf{P}\bigg(\frac{1}{b_n}\sum_{l=1}^{b_n} \xi_{jl} \ge \frac{1}{2}\bigg) \;\lesssim\; (e\mathbf{E}(\xi_{j1}))^{b_n(\frac{1}{2}-\mathbf{E}(\xi_{j1}))^2}. \quad \text{(S6.30)}$$

Also, Lemma 14 gives

$$\mathbf{E}(\xi_{j1}) \;\lesssim\; (c\ell_n j^{-2\theta\beta})^{\ell_n/4}.$$

Therefore, combining with with (S6.30), we conclude that

$$\begin{aligned}
\sum_{j\in J(r_n)^c} \mathbf{P}(\hat{\sigma}_j^2 \le \sigma_j^{2+\theta}) \;&\lesssim\; (c\ell_n)^{\ell_n/4} \sum_{j \ge n^{\frac{\epsilon}{\theta\beta}}} (j^{-\theta\beta\ell_n/2})^{b_n/c} \\
&\lesssim\; (c\ell_n)^{\ell_n/4} n^{-\frac{n}{2c}+1} \\
&\lesssim n^{-(2+\delta)},
\end{aligned}$$

where the last step uses $\ell_n \asymp n/\log(n)$. Note also that the final bound $n^{-(2+\delta)}$ can be replaced with any fixed positive power of n^{-1}, but the current form is all that is needed. □

S7 Proof of Proposition 1

To ease notation, we let $q = 4 + \delta$ throughout the proof.

Elliptical case. Suppose X_1 is a centered elliptical random vector of the form $X_1 = \eta_1 \Sigma^{1/2} Z_1 / \|Z_1\|_2$, where Z_1 is a standard Gaussian p-dimensional

Gaussian vector, and η_1 is independent of Z_1 with $\mathbf{E}(\eta_1^2) = p$. We first check the L^q-L^2 moment equivalence condition (i) with $q = 4 + \delta$. Letting $w = \Sigma^{1/2} v$ for a generic vector $v \in \mathbb{R}^p$, a direct calculation gives

$$\begin{aligned} \|\langle v, X_1 \rangle\|_{L^2}^2 &= \|\eta_1 \langle w, Z_1/\|Z_1\|_2 \rangle\|_{L^2} \\ &= \|\eta_1\|_{L^2}^2 \|w\|_2^2/p \\ &= \|w\|_2^2. \end{aligned}$$

Because the distribution of $U_1 = Z_1/\|Z_1\|_2$ is invariant to orthogonal transformations, it follows that the random variables $\langle w, U_1 \rangle$ and $\|w\|_2 \langle e_1, U_1 \rangle$ are equal in distribution, where e_1 is the first standard basis vector. Therefore,

$$\|\langle v, X_1 \rangle\|_{L^q} = \|w\|_2 \|\eta_1\|_{L^q} \|U_{11}\|_{L^q}.$$

The quantity $\|U_{11}\|_{L^q}$ is at most of order $1/\sqrt{p}$, which can be shown as follows. Due to the independence of U_{11} and $\|Z_1\|_2$, we have $\|Z_{11}\|_{L^q} = \|\|Z_1\|_2\|_{L^q} \|U_{11}\|_{L^q}$. Furthermore, Lyapunov's inequality gives $\|\|Z_1\|_2\|_{L^q} = \|\|Z_1\|_2^2\|_{L^{q/2}}^{1/2} \geq \|\|Z_1\|_2^2\|_{L^1}^{1/2} = \sqrt{p}$. Combining the last several steps and the assumption that $\|\eta_1\|_{L^q} \lesssim \sqrt{p}$, we conclude that

$$\|\langle v, X_1 \rangle\|_{L^q} \lesssim \|w\|_2 = \|\langle v, X_1 \rangle\|_{L^2}, \tag{S7.31}$$

which verifies condition (i).

Regarding the density condition (ii), note that if e_j is the jth standard basis vector, then the vector $w = \Sigma^{1/2}e_j/\sigma_j$ satisfies $\|w\|_2 = 1$. So, the discussion above shows that all the random variables $X_{11}/\sigma_1, \ldots, X_{1p}/\sigma_p$ have the same distribution, which is that of $\eta_1\langle e_1, U_1\rangle$. In particular, if the random variable X_{11}/σ_1 has a Lebesgue density f_1 such that $\|f_1\|_{L^\infty} \lesssim 1$, then condition (ii) holds, which completes the proof in the elliptical case.

Separable case. Suppose that X_1 is has a centered separable distribution so that $X_1 = \Sigma^{1/2}\zeta_1$, where $\zeta_1 = (\zeta_{11}, \ldots, \zeta_{1p})$ has i.i.d. entries with $\mathbf{E}(\zeta_{11}) = 0$, and $\mathrm{var}(\zeta_{11}) = 1$. To check the L^q-L^2 moment equivalence condition (i), it suffices to show that $\|\langle w, \zeta_1\rangle\|_{L^q} \lesssim \|\langle w, \zeta_1\rangle\|_{L^2}$ for any $w \in \mathbb{R}^p$. Using Rosenthal's inequality (Lemma 16), and the assumption that $\max_{1\leq j\leq p}\|\zeta_{1j}\|_{L^q} \lesssim 1$, we have

$$
\begin{aligned}
\|\langle w, \zeta_1\rangle\|_{L^q} &\lesssim \max\left\{\big\|\langle w, \zeta_1\rangle\big\|_{L^2}, \Big(\sum_{j=1}^{p}\big\|w_j\zeta_{1j}\big\|_{L^q}^q\Big)^{1/q}\right\} \\
&\lesssim \max\big\{\|w\|_2, \|w\|_q\big\} \\
&= \|\langle w, \zeta_1\rangle\|_{L^2}
\end{aligned}
$$

as needed. Finally, the density condition (ii) is a direct consequence of Theorem 1.2 in the paper (Rudelson and Vershynin, 2015) and the assumption that $\max_{1\leq j\leq p}\|g_j\|_{L^\infty} \lesssim 1$, where g_j is the Lebesgue density of ζ_{1j}. $\square$

S8 Background results

Lemma 16 (Rosenthal inequalities (Johnson et al., 1985)). *Fix $q \geq 1$, and let $\xi_1, \ldots, \xi_n$ be independent random variables. Then, there is an absolute constant $c > 0$ such that the following two statements are true.*

(i). If $\xi_1, \ldots, \xi_n$ are non-negative, then

$$\Big\| \sum_{i=1}^{n} \xi_i \Big\|_{L^q} \leq c \cdot q \cdot \max\left\{ \Big\| \sum_{i=1}^{n} \xi_i \Big\|_{L^1}, \Big(\sum_{i=1}^{n} \|\xi_i\|_{L^q}^q \Big)^{1/q} \right\}.$$

(ii). If $q \geq 2$, and $\xi_1, \ldots, \xi_n$ are centered, then

$$\Big\| \sum_{i=1}^{n} \xi_i \Big\|_{L^q} \leq c \cdot q \cdot \max\left\{ \Big\| \sum_{i=1}^{n} \xi_i \Big\|_{L^2}, \Big(\sum_{i=1}^{n} \|\xi_i\|_{L^q}^q \Big)^{1/q} \right\}.$$

For the next lemma, recall that we denote the stable rank of a non-zero positive semidefinite matrix A as $\mathtt{r}(A) = \mathrm{tr}(A)^2/\|A\|_F^2$.

Lemma 17 (Lower-tail bound for Gaussian maxima (Lopes and Yao, 2022)). *Let $\xi \sim \mathcal{N}(0, R)$ be a Gaussian random vector in $\mathbb{R}^d$ for some correlation matrix R, and fix two constants $a, b \in (0,1)$ with respect to d. Then, there is a constant $c > 0$ depending only on (a, b) such that the following inequality holds for any integer k satisfying $2 \leq k \leq \frac{b^2}{4}\mathtt{r}(R)$,*

$$\mathbf{P}\Big(\max_{1 \leq j \leq d} \xi_j \leq a\sqrt{2(1-b)\log(k)} \Big) \leq c\, k^{\frac{-(1-b)(1-a)^2}{b}} \big(\log(k) \big)^{\frac{1-b(2-a)-a}{2b}}$$

The next result is a variant of Lemma A.7 in Spokoiny and Zhilova (2015) that can be proven in essentially the same way.

Lemma 18 (Gaussian comparison inequality)**.** *Let $\zeta \sim N(0, I_d)$ and $\xi \sim N(0, A)$ for some positive semidefinite matrix $A \in \mathbb{R}^{d\times d}$. Then,*

$$\sup_{s\in\mathbb{R}} \left|\mathbf{P}\Big(\max_{1\leq j\leq d} \xi_j \leq s\Big) - \mathbf{P}\Big(\max_{1\leq j\leq d} \zeta_j \leq s\Big)\right| \ \leq\ 2\|A - I_d\|_F.$$

Lemma 19 (Bentkus' Berry-Esseen Theorem (Bentkus, 2003))**.** *Let $V_1, \dots, V_n$ be i.i.d. random vectors in $\mathbb{R}^d$ with zero mean and identity covariance matrix. Furthermore, let γ_d denote the standard Gaussian distribution on $\mathbb{R}^d$, and let $\mathscr{A}$ denote the collection of all Borel convex subsets of $\mathbb{R}^d$. Then, there is an absolute constant $c > 0$ such that*

$$\sup_{\mathcal{A}\in\mathscr{A}} \left|\mathbf{P}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i \in \mathcal{A}\Big) - \gamma_d(\mathcal{A})\right| \leq \frac{c\cdot d^{1/4}\cdot \mathbf{E}\,\|V_1\|_2^3}{n^{1/2}}.$$

Lemma 20 (Nazarov's inequality)**.** *Let $(\zeta_1, \dots, \zeta_d)$ be a Gaussian random vector, and suppose the parameter $\underline{\sigma}^2 := \min_{1\leq j\leq d} \operatorname{var}(\zeta_j)$ is positive. Then, for any fixed $\epsilon > 0$,*

$$\sup_{s\in\mathbb{R}} \mathbf{P}\Big(\big|\max_{1\leq j\leq d} \zeta_j - s\big| \leq \epsilon\Big) \ \leq\ \frac{2\epsilon}{\underline{\sigma}}(\sqrt{2\log(d)} + 2).$$

This version of Nazarov's inequality appears in Lemma 4.3 of (Chernozhukov et al., 2016) and originates from Nazarov (2003).

Lemma 21 (Fuk-Nagaev inequality)**.** *Fix $q \geq 1$, and let $\xi_1, \dots, \xi_n$ be centered independent random variables. Then, for any fixed $s > 0$,*

$$\mathbf{P}\Big(\Big|\sum_{i=1}^n \xi_i\Big| \geq s\Big) \ \leq\ 2\Big(\frac{q+2}{qs}\Big)^q \sum_{i=1}^n \mathbf{E}|\xi_i|^q + 2\exp\left(\frac{-2s^2}{(q+2)^2 e^q \sum_{i=1}^n \mathbf{E}(\xi_i^2)}\right).$$

This statement of the Fuk-Nagaev inequality is based on (Rio, 2017, eqn. 1.7).

Lemma 22 (Kiefer's inequality). *If $\xi_1, \ldots, \xi_n$ are i.i.d. Bernoulli random variables, then*

$$\mathbf{P}\Big(\frac{1}{n}\sum_{i=1}^{n}\xi_i \geq \frac{1}{2}\Big) \;\leq\; 2\big(e\mathbf{E}(\xi_1)\big)^{n(1/2-\mathbf{E}(\xi_1))^2}.$$

The result above is the modification of Corollary A.6.3 in (van der Vaart and Wellner, 1996). In that reference, the success probability $\mathbf{E}(\xi_1)$ of the Bernoulli random variables is assumed to be less than $1/e$, but in the formulation above, the result holds for all values of $\mathbf{E}(\xi_1)$.

Bibliography

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

Chernozhukov, V., D. Chetverikov, and K. Kato (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and their Applications 126*(12), 3632–3651.

Chernozhukov, V., D. Chetverikov, K. Kato, and Y. Koike (2023). High-dimensional data bootstrap. *Annual Review of Statistics and Its Application 10*(1), 427–449.

Johnson, W. B., G. Schechtman, and J. Zinn (1985). Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. *The Annals of Probability*, 234–253.

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

Lopes, M. E. and J. Yao (2022). A sharp lower-tail bound for Gaussian maxima with application to bootstrap methods in high dimensions. *Electronic Journal of Statistics 16*(1), 58–83.

Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to a Gaussian measure. In *Geometric Aspects of Functional Analysis: Israel Seminar 2001-2002*, pp. 169–187.

Rio, E. (2017). About the constants in the Fuk-Nagaev inequalities. *Electronic Communications in Probability 22*, 1–12.

Rudelson, M. and R. Vershynin (2015). Small ball probabilities for linear images of high-dimensional distributions. *International Mathematics Research Notices 2015*(19), 9594–9617.

Spokoiny, V. and M. Zhilova (2015). Bootstrap confidence sets under model misspecification. *The Annals of Statistics 43*(6), 2653–2675.

Stein, E. M. and G. Weiss (1971). *Introduction to Fourier Analysis on Euclidean Spaces*. Princeton University Press.

van der Vaart, A. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. Springer.

Vershynin, R. (2012). Introduction to the non-asymptotic analysis of random matrices. *Compressed Sensing*, 210–268.