SUPPLEMENT TO "GAUSSIAN APPROXIMATIONS AND MULTIPLIER BOOTSTRAP FOR MAXIMA OF SUMS OF HIGH-DIMENSIONAL RANDOM VECTORS"*

By Victor Chernozhukov[†], Denis Chetverikov[‡] and Kengo Kato[§]

MIT[†], UCLA[‡], and University of Tokyo[§]

Supplementary Material I

Deferred Proofs for Results from Main Text

APPENDIX D: DEFERRED PROOFS FOR SECTION 2

D.1. Proof of Lemma A.7. Claim (a). Define $I_{ij} = 1\{|x_{ij}| \leqslant u(\bar{\mathrm{E}}[x_{ij}^2])^{1/2}\}$, and observe that

$$
\begin{aligned}
(\bar{\mathrm{E}}[|\tilde{x}_{ij}|^q])^{1/q} &\leqslant (\bar{\mathrm{E}}[|x_{ij}I_{ij}|^q])^{1/q} + (\mathbb{E}_n[|\mathrm{E}[x_{ij}I_{ij}]|^q])^{1/q} \\
&\leqslant (\bar{\mathrm{E}}[|x_{ij}I_{ij}|^q])^{1/q} + (\bar{\mathrm{E}}[|x_{ij}I_{ij}|^q])^{1/q} \leqslant 2(\bar{\mathrm{E}}[|x_{ij}|^q])^{1/q}.
\end{aligned}
$$

Claim (b). Observe that

$$
\begin{aligned}
\bar{\mathrm{E}}[|\tilde{x}_{ij}\tilde{x}_{ik} - x_{ij}x_{ik}|] &\leqslant \bar{\mathrm{E}}[|(\tilde{x}_{ij} - x_{ij})\tilde{x}_{ik}|] + \bar{\mathrm{E}}[|x_{ij}(\tilde{x}_{ik} - x_{ik})|] \\
&\leqslant \sqrt{\bar{\mathrm{E}}[(\tilde{x}_{ij} - x_{ij})^2]}\sqrt{\bar{\mathrm{E}}[\tilde{x}_{ik}^2]} + \sqrt{\bar{\mathrm{E}}[(\tilde{x}_{ik} - x_{ik})^2]}\sqrt{\bar{\mathrm{E}}[x_{ij}^2]} \\
&\leqslant 2\varphi(u)\sqrt{\bar{\mathrm{E}}[x_{ij}^2]}\sqrt{\bar{\mathrm{E}}[x_{ik}^2]} + \varphi(u)\sqrt{\bar{\mathrm{E}}[x_{ik}^2]}\sqrt{\bar{\mathrm{E}}[x_{ij}^2]} \\
&\leqslant (3/2)\varphi(u)(\bar{\mathrm{E}}[x_{ij}^2] + \bar{\mathrm{E}}[x_{ik}^2]),
\end{aligned}
$$

where the first inequality follows from the triangle inequality, the second from the Cauchy-Schwarz inequality, the third from the definition of $\varphi(u)$ together with claim (a), and the last from inequality $|ab| \leqslant (a^2 + b^2)/2$.

Claim (c). This follows from the Cauchy-Schwarz inequality and the definition of $\varphi(u)$.

Claim (d). We shall use the following lemma.

*V. Chernozhukov and D. Chetverikov are supported by a National Science Foundation grant. K. Kato is supported by the Grant-in-Aid for Young Scientists (B) (25780152), the Japan Society for the Promotion of Science.

LEMMA D.1 (Tail Bounds for Self-Normalized Sums). *Let* $\xi_1, \dots, \xi_n$ *be independent real-valued random variables such that* $\mathrm{E}[\xi_i] = 0$ *and* $\mathrm{E}[\xi_i^2] < \infty$ *for all* $1 \leqslant i \leqslant n$. *Let* $S_n = \sum_{i=1}^n \xi_i$. *Then for every* $x > 0$,

$$\mathrm{P}(|S_n| > x(4B_n + V_n)) \leqslant 4\exp(-x^2/2),$$

where $B_n^2 = \sum_{i=1}^n \mathrm{E}[\xi_i^2]$ *and* $V_n^2 = \sum_{i=1}^n \xi_i^2$.

PROOF OF LEMMA D.1. See [12], Theorem 2.16. ∎

Define

$$\Lambda_j := 4\sqrt{\bar{\mathrm{E}}[(x_{ij} - \tilde{x}_{ij})^2]} + \sqrt{\mathbb{E}_n[(x_{ij} - \tilde{x}_{ij})^2]}.$$

Then by Lemma D.1 and the union bound, with probability at least $1 - 4\gamma$,

$$|X_j - \tilde{X}_j| \leqslant \Lambda_j \sqrt{2\log(p/\gamma)}, \text{ for all } 1 \leqslant j \leqslant p.$$

By claim (c), for $u \geqslant u(\gamma)$, with probability at least $1 - \gamma$, for all $1 \leqslant j \leqslant p$,

$$\begin{aligned}\Lambda_j = 4\sqrt{\bar{\mathrm{E}}[(x_{ij} - \tilde{x}_{ij})^2]} + \sqrt{\mathbb{E}_n[(\mathrm{E}[x_{ij}1\{|x_{ij}| > u(\bar{\mathrm{E}}[x_{ij}^2])^{1/2}\}])^2]}& \\ \leqslant 5\sqrt{\bar{\mathrm{E}}[x_{ij}^2]}\varphi(u).&\end{aligned}$$

The last two assertions imply claim (d). ∎

D.2. Proof of Theorem 2.2. Since M_2 is bounded from below and above by positive constants, we may normalize $M_2 = 1$, without loss of generality. In this proof, let $C > 0$ denote a generic constant depending only on c_1 and C_1, and its value may change from place to place.

For given $\gamma \in (0,1)$, denote $\ell_n := \log(pn/\gamma) \geqslant 1$ and let

$$u_1 := n^{3/8}\ell_n^{-5/8}M_3^{3/4} \text{ and } u_2 := n^{3/8}\ell_n^{-5/8}M_4^{1/2}.$$

Define $u := u(\gamma) \vee u_1 \vee u_2$ and $\beta := \sqrt{n}/(2\sqrt{2}u)$. Then $u \geqslant u(\gamma)$ and the choice of β trivially obeys $2\sqrt{2}u\beta \leqslant \sqrt{n}$. So, by Theorem 2.1 and using the argument as that in the proof of Corollary I.1, for every $\psi > 0$ and any $\bar{\varphi}(u) \geqslant \varphi(u)$, we have

$$\begin{aligned}\rho \leqslant C\big[&n^{-1/2}(\psi^3 + \psi^2\beta + \psi\beta^2)M_3^3 + (\psi^2 + \psi\beta)\bar{\varphi}(u) \\ &+\psi\bar{\varphi}(u)\sqrt{\log(p/\gamma)} + (\beta^{-1}\log p + \psi^{-1})\sqrt{1 \vee \log(p\psi)} + \gamma\big].\end{aligned} \tag{1}$$

Step 1. We claim that we can take $\bar{\varphi}(u) := CM_4^2/u$ for all $u > 0$. Since $\bar{\mathrm{E}}[x_{ij}^2] \geqslant c_1$, we have $1\{|x_{ij}| > u(\bar{\mathrm{E}}[x_{ij}^2])^{1/2}\} \leqslant 1\{|x_{ij}| > c_1^{1/2}u\}$. Hence

$$\begin{aligned}&\bar{\mathrm{E}}[x_{ij}^2 1\{|x_{ij}| > u(\bar{\mathrm{E}}[x_{ij}^2])^{1/2}\}] \leqslant \bar{\mathrm{E}}[x_{ij}^2 1\{|x_{ij}| > c_1^{1/2}u\}] \\ &\quad\leqslant \bar{\mathrm{E}}[x_{ij}^4 1\{|x_{ij}| > c_1^{1/2}u\}]/(c_1u^2) \leqslant \bar{\mathrm{E}}[x_{ij}^4]/(c_1u^2) \leqslant M_4^4/(c_1u^2).\end{aligned}$$

This implies $\varphi_x(u) \leqslant CM_4^2/u$. For $\varphi_y(u)$, note that

$$\bar{\mathrm{E}}[y_{ij}^4] = \mathbb{E}_n[\mathrm{E}[y_{ij}^4]] = 3\mathbb{E}_n[(\mathrm{E}[y_{ij}^2])^2] = 3\mathbb{E}_n[(\mathrm{E}[x_{ij}^2])^2] \leqslant 3\mathbb{E}_n[\mathrm{E}[x_{ij}^4]] = 3\bar{\mathrm{E}}[x_{ij}^4],$$

and hence $\varphi_y(u) \leqslant CM_4^2/u$ as well. This implies the claim of this step.

Step 2. We shall bound the right side of (1) by suitably choosing ψ depending on the range of u. In order to set up this choice we define $u^\star$ by the following equation:

$$\bar{\varphi}(u^\star) n^{3/8}/(M_3^3 \ell_n^{5/6})^{3/4} = 1.$$

We then take

$$\psi = \psi(u) := \begin{cases} n^{1/8}\ell_n^{-3/8}M_3^{-3/4} & \text{if } u \geqslant u^\star, \\ \ell_n^{-1/6}(\bar{\varphi}(u))^{-1/3} & \text{if } u < u^\star. \end{cases} \tag{2}$$

We note that for $u < u^\star$,

$$\psi(u) \leqslant \psi(u^\star) = n^{1/8}\ell_n^{-3/8}M_3^{-3/4}.$$

That is, when $u < u^\star$ the smoothing parameter ψ is smaller than when $u \geqslant u^\star$.

Using these choices of parameters β and ψ and elementary calculations (which will be done in Step 3 below), we conclude from (1) that whether $u < u^\star$ or $u \geqslant u^\star$,

$$\rho \leqslant C(n^{-1/2}u\ell_n^{3/2} + \gamma).$$

The bound in the theorem follows from this inequality.

Step 3. (Computation of the bound on ρ). Note that since $\rho \leqslant 1$, we only had to consider the case where $n^{-1/2}u\ell_n^{3/2} \leqslant 1$ since otherwise the inequality is trivial by taking, say, $C = 1$. Since $u_1 = n^{3/8}M_3^{3/4}/\ell_n^{5/8}$ and $u_2 = n^{3/8}M_4^{1/2}/\ell_n^{5/8}$, we have

$$\begin{aligned} (\bar{\varphi}(u^\star))^{4/3} &= n^{-1/2}\ell_n^{5/6}M_3^3, \\ \bar{\varphi}(u_1) &\leqslant Cn^{-3/8}\ell_n^{5/8}M_4^2/M_3^{3/4}, \\ \bar{\varphi}(u_2) &\leqslant Cn^{-3/8}\ell_n^{5/8}M_4^{3/2}. \end{aligned}$$

Also note that $\psi \leqslant n^{1/8}$, and so $1 \vee \log(p\psi) \lesssim \log(pn) \leqslant \ell_n$. Therefore,

$$\beta^{-1}\log p\sqrt{1 \vee \log(p\psi)} \lesssim \beta^{-1}\ell_n^{3/2} \lesssim n^{-1/2}u\ell_n^{3/2}.$$

In addition, note that $\beta \lesssim \sqrt{n}/u \leqslant \sqrt{n}/u_1 = n^{1/8}\ell_n^{5/8}M_3^{-3/4} =: \bar{\beta}$ and $\psi \leqslant \bar{\beta}$ under either case. This implies that $(\psi^3 + \psi^2\beta + \psi\beta^2) \lesssim \psi\bar{\beta}^2$ and

$(\psi^2+\psi\beta) \leqslant \psi\bar{\beta}$.
Using these inequalities, we can compute the bounds claimed above.

(a). Bounding ρ when $u \geqslant u^\star$. Then

$$n^{-1/2}(\psi^3+\psi^2\beta+\psi\beta^2)M_3^3 \lesssim n^{-1/2}\psi\bar{\beta}^2 M_3^3 \leqslant n^{-1/8}\ell_n^{7/8}M_3^{3/4} \leqslant n^{-1/2}u\ell_n^{3/2};$$
$$(\psi^2+\psi\beta)\bar{\varphi}(u) \lesssim \psi\bar{\beta}\bar{\varphi}(u) \leqslant \psi\bar{\beta}\bar{\varphi}(u^\star) \leqslant n^{-1/8}\ell_n^{7/8}M_3^{3/4} \leqslant n^{-1/2}u\ell_n^{3/2};$$
$$\psi\bar{\varphi}(u)\sqrt{\log(p/\gamma)} \leqslant \psi\bar{\beta}\bar{\varphi}(u)\sqrt{\ell_n}/\bar{\beta} \leqslant \psi\bar{\beta}\bar{\varphi}(u^\star) \leqslant n^{-1/2}u\ell_n^{3/2};\text{ and}$$
$$\psi^{-1}\sqrt{\ell_n} \leqslant n^{-1/8}\ell_n^{7/8}M_3^{3/4} \leqslant n^{-1/2}u\ell_n^{3/2};$$

where we have used Step 1 and the fact that

$$\sqrt{\ell_n}/\bar{\beta} = \ell_n^{-1/2}\psi^{-1} \leqslant n^{-1/8}\ell_n^{-1/8}M_3^{3/4} \leqslant n^{-1/2}u\ell_n^{3/2} \leqslant 1.$$

The claimed bound on ρ now follows.

(b). Bounding ρ when $u < u^\star$. Since ψ is smaller than in case (a), by the calculations in Step (a)

$$n^{-1/2}(\psi^3+\psi^2\beta+\psi\beta^2)M_3^3/\sqrt{n} \lesssim n^{-1/2}u\ell_n^{3/2}.$$

Moreover, using definition of ψ, $u > u_2$, definition of u_2, we have

$$\psi\beta\bar{\varphi}(u) \leqslant \beta\bar{\varphi}(u)^{2/3}\ell_n^{-1/6} \leqslant \beta\bar{\varphi}(u_2)^{2/3}\ell_n^{-1/6} \leqslant n^{-1}\beta u_2^2\ell_n^{5/3-1/6} \lesssim n^{-1/2}u\ell_n^{3/2};$$
$$\psi^2\bar{\varphi}(u) \leqslant \bar{\varphi}(u)^{1/3}\ell_n^{-1/3} \leqslant \bar{\varphi}(u_2)^{1/3}\ell_n^{-1/3} \leqslant n^{-1/2}u_2\sqrt{\ell_n} \leqslant n^{-1/2}u\ell_n^{3/2}.$$

Analogously and using $n^{-1/2}u\ell_n^{3/2} \leqslant 1$, we have

$$\psi\bar{\varphi}(u)\sqrt{\log(p/\gamma)} \leqslant \bar{\varphi}(u)^{2/3}\ell_n^{1/3} \leqslant \bar{\varphi}(u_2)^{2/3}\ell_n^{1/3} \leqslant n^{-1}u_2^2\ell_n^2 \leqslant n^{-1/2}u\ell_n^{3/2}.$$
$$\psi^{-1}\sqrt{\ell_n} = \bar{\varphi}(u)^{1/3}\ell_n^{2/3} \leqslant n^{-1/2}u\ell_n^{3/2}.$$

This completes the proof. ■

APPENDIX E: DEFERRED PROOFS FOR SECTION 4

E.1. Proof of Theorem 4.1. The proof proceeds in three steps. In the proof $(\widehat{\beta},\lambda)$ denotes $(\widehat{\beta}^{(k)},\lambda^{(k)})$ with k either 0 or 1.

Step 1. Here we show that there exist some constants $c>0$ and $C>0$ (depending only c_1, C_1 and σ^2) such that

$$\mathrm{P}(T_0 \leqslant \lambda) \geqslant 1-\alpha-\nu_n, \tag{3}$$

with $\nu_n = Cn^{-c}$. We first note that $T_0 = \sqrt{n}\max_{1\leqslant k\leqslant 2p}\mathbb{E}_n[\tilde{z}_{ik}\varepsilon_i]$, where $\tilde{z}_i = (z_i', -z_i')'$. Application of Corollary 2.1-(ii) gives

$$|\mathrm{P}(T_0 \leqslant \lambda) - \mathrm{P}(Z_0 \leqslant \lambda)| \leqslant Cn^{-c},$$

where $c > 0$ and $C > 0$ are constants depending only on c_1, C_1 and σ^2. The claim follows since $\lambda \geqslant c_{Z_0}(1-\alpha)$, which holds because $\lambda^{(1)} = c_{Z_0}(1-\alpha)$, and $\lambda^{(1)} \leqslant \lambda^{(0)} = c_0(1-\alpha) := \sigma\Phi^{-1}(1-\alpha/(2p))$ (by the union bound $\mathrm{P}(Z_0 \geqslant c_0(1-\alpha)) \leqslant 2p\mathrm{P}(\sigma N(0,1) \geqslant c_0(1-\alpha)) = \alpha$).

Step 2. We claim that with probability $\geqslant 1-\alpha-\nu_n$, $\widehat{\delta} = \widehat{\beta} - \beta$ obeys:

$$\sqrt{n} \max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[z_{ij}(z_i'\widehat{\delta})]| \leqslant 2\lambda.$$

Indeed, by definition of $\widehat{\beta}$, $\sqrt{n}\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}(y_i - z_i'\widehat{\beta})]| \leqslant \lambda$, which by the triangle inequality implies $\sqrt{n}\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}(z_i'\widehat{\delta})]| \leqslant T_0 + \lambda$. The claim follows from Step 1.

Step 3. By Step 1, with probability $\geqslant 1-\alpha-\nu_n$, the true value β obeys the constraint in optimization problem (16) in the main text, in which case by definition of $\widehat{\beta}$, $\|\widehat{\beta}\|_{\ell_1} \leqslant \|\beta\|_{\ell_1}$. Therefore, with the same probability, $\widehat{\delta} \in \mathcal{R}(\beta) = \{\delta \in \mathbb{R}^d : \|\beta+\delta\|_{\ell_1} \leqslant \|\beta\|_{\ell_1}\}$. By definition of $\kappa_I(\beta)$ we have that with the same probability,

$$\kappa_I(\beta)\|\widehat{\delta}\|_I \leqslant \max_{1\leqslant j\leqslant p} |\mathbb{E}_n[z_{ij}(z_i'\widehat{\delta})]|.$$

Combining this inequality with Step 2 gives the claim of the theorem. ∎

E.2. Proof of Theorem 4.2. The proof has four steps. In the proof, we let $\varrho_n = Cn^{-c}$ for sufficiently small $c > 0$ and sufficiently large $C > 0$ depending only on $c_1, C_1, \underline{\sigma}^2, \sigma^2$, where c and C (and hence ϱ_n) may change from place to place.

Step 0. The same argument as in the previous proof applies to $\widehat{\beta}^{(0)}$ with $\lambda = \lambda^{(0)} := c_0(1-1/n)$, where now σ^2 is the upper bound on $\mathrm{E}[\varepsilon_i^2]$. Thus, we conclude that with probability at least $1-\varrho_n$,

$$\|\widehat{\beta}^{(0)} - \beta\|_{\mathrm{pr}} \leqslant \frac{2c_0(1-1/n)}{\sqrt{n}\kappa_{\mathrm{pr}}(\beta)}.$$

Step 1. We claim that with probability at least $1-\varrho_n$,

$$\max_{1\leqslant j\leqslant p} \left(\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i - \varepsilon_i)^2]\right)^{1/2} \leqslant B_n \frac{2c_0(1-1/n)}{\sqrt{n}\kappa_{\mathrm{pr}}(\beta)} =: \iota_n.$$

Application of Hölder's inequality and identity $\varepsilon_i - \widehat{\varepsilon}_i = z_i'(\widehat{\beta}^{(0)} - \beta)$ gives

$$\max_{1\leqslant j\leqslant p} \left(\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i - \varepsilon_i)^2]\right)^{1/2} \leqslant B_n(\mathbb{E}_n[z_i'(\widehat{\beta}^{(0)} - \beta)]^2)^{1/2} \leqslant B_n\|\widehat{\beta}^{(0)} - \beta\|_{\mathrm{pr}}.$$

The claim follows from Step 0.

Step 2. In this step, we apply Corollary 3.1-(ii) to

$$T = T_0 = \sqrt{n} \max_{1 \leqslant j \leqslant 2p} \mathbb{E}_n[\tilde{z}_{ij}\varepsilon_i], \; W = \sqrt{n} \max_{1 \leqslant j \leqslant 2p} \mathbb{E}_n[\tilde{z}_{ij}\widehat{\varepsilon}_i e_i], \text{ and}$$
$$W_0 = \sqrt{n} \max_{1 \leqslant j \leqslant 2p} \mathbb{E}_n[\tilde{z}_{ij}\varepsilon_i e_i],$$

where $\tilde{z}_i = (z_i', -z_i')'$, to conclude that uniformly in $\alpha \in (0,1)$

$$\mathrm{P}(T_0 \leqslant c_W(1-\alpha)) \geqslant 1 - \alpha - \varrho_n. \tag{4}$$

To show applicability of Corollary 3.1-(ii), we note that for any $\zeta_1 > 0$,

$$\mathrm{P}_e(|W - W_0| > \zeta_1) \leqslant \mathrm{E}_e[|W - W_0|]/\zeta_1 \leqslant \sqrt{n}\mathrm{E}_e\left[\max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[z_{ij}(\widehat{\varepsilon}_i - \varepsilon_i)e_i]|\right]/\zeta_1$$
$$\lesssim \sqrt{\log p} \max_{1 \leqslant j \leqslant p} (\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i - \varepsilon_i)^2])^{1/2}/\zeta_1,$$

where the third inequality is due to Pisier's inequality. The last quantity is bounded by $(\iota_n^2 \log p)^{1/2}/\zeta_1$ with probability $\geqslant 1 - \varrho_n$ by Step 1.

Since $\iota_n \log p \leqslant C_1 n^{-c_1}$ by assumption (vi) of the theorem, we can take ζ_1 in such a way that $\zeta_1(\log p)^{1/2} \leqslant \varrho_n$ and $(\iota_n^2 \log p)^{1/2}/\zeta_1 \leqslant \varrho_n$. Then all the conditions of Corollary 3.1-(ii) with so defined ζ_1 and $\zeta_2 = \varrho_n \vee ((\iota_n^2 \log p)^{1/2}/\zeta_1)$ are satisfied, and hence application of the corollary gives that uniformly in $\alpha \in (0,1)$,

$$|\mathrm{P}(T_0 \leqslant c_W(1-\alpha)) - 1 - \alpha| \leqslant \varrho_n, \tag{5}$$

which implies the claim of this step.

Step 3. In this step we claim that with probability at least $1 - \varrho_n$,

$$c_W(1-\alpha) \leqslant c_{Z_0}(1 - \alpha + 2\varrho_n).$$

Combining Step 2 and Lemma 3.3 gives that with probability at least $1-\zeta_2$, $c_W(1-\alpha) \leqslant c_{W_0}(1-\alpha+\zeta_2)+\zeta_1$, where ζ_1 and ζ_2 are chosen as in Step 2. In addition, Lemma 3.2 shows that $c_{W_0}(1-\alpha+\zeta_2) \leqslant c_{Z_0}(1-\alpha+\varrho_n)$. Finally, Lemma 2.1 yields $c_{Z_0}(1-\alpha+\varrho_n)+\zeta_1 \leqslant c_{Z_0}(1-\alpha+2\varrho_n)$. Combining these bounds gives the claim of this step.

Step 4. Given (4), the rest of the proof is identical to Steps 2-3 in the proof of Theorem 4.1 with $\lambda = c_W(1-\alpha)$. The result follows for $\nu_n = 2\varrho_n$.■

APPENDIX F: DEFERRED PROOFS FOR SECTION 5

F.1. Proof of Theorem 5.1. The multiplier bootstrap critical values $c_{1-\alpha,w}$ clearly satisfy $c_{1-\alpha,w} \leqslant c_{1-\alpha,w'}$ whenever $w \subset w'$, so inequality (22) in the main text is satisfied. Therefore, it suffices to prove (23) in the main text.

Let w denote the set of true null hypotheses. Then for any $j \in w$,

$$t_j = \sqrt{n}(\widehat{\beta}_j - \beta_{0j}) \leqslant \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{ij} + r_{nj}.$$

Therefore, for all $j \in w$, we can and will assume that $\beta_j = \beta_{0j}$.

For $w \subset \mathcal{W} = \{1, \dots, p\}$, define

$$T := T(w) := \max_{j \in w} \sqrt{n}(\widehat{\beta}_j - \beta_{0j}),\ W := W(w) := \max_{j \in w} \frac{1}{\sqrt{n}} \sum_{i=1}^n \widehat{x}_{ij} e_i.$$

In addition, define

$$T_0 := T_0(w) := \max_{j \in w} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{ij},\ W_0 := W_0(w) := \max_{j \in w} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{ij} e_i.$$

To prove (23), we will apply Corollary 3.1. By assumption, either (i) or (ii) of Corollary 3.1 holds. Therefore, it remains to verify conditions in equations (14) and (15) in the main text with $\zeta_1 \sqrt{\log p} + \zeta_2 \leqslant Cn^{-c}$ for some $c > 0$ and $C > 0$ uniformly over all $w \subset \mathcal{W}$.

Set $\zeta_1 = (C/2)n^{-c}/\sqrt{\log p}$ and $\zeta_2 = (C/2)n^{-c}$ for sufficiently small $c > 0$ and large $C > 0$ depending on c_1, C_1, c_2, and C_2 only. Note that $\zeta_1 \sqrt{\log p} + \zeta_2 \leqslant Cn^{-c}$. Also note that

$$|T - T_0| \leqslant \max_{1 \leqslant j \leqslant p} |r_{nj}| = \Delta_1$$

for all $w \subset \mathcal{W}$. Therefore, it follows from assumption (i) that $\mathrm{P}(|T - T_0| > \zeta_1) < \zeta_2$ for all $w \subset \mathcal{W}$, i.e. condition in equation (14) holds uniformly over all $w \subset \mathcal{W}$. Further, note that $\sum_{i=1}^n (\widehat{x}_{ij} - x_{ij}) e_i / \sqrt{n}$ conditional on $(x_i)_{i=1}^n$ and $(\widehat{x}_i)_{i=1}^n$ is distributed as $N(0, \mathbb{E}_n[(\widehat{x}_{ij} - x_{ij})^2])$ random variable and

$$|W - W_0| \leqslant \max_{1 \leqslant j \leqslant p} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\widehat{x}_{ij} - x_{ij}) e_i \right|$$

for all $w \subset \mathcal{W}$. Therefore, $\mathrm{E}_e[|W - W_0|] \leqslant (C/2)\sqrt{\Delta_2 \log p}$, and so it follows from Borell inequality and assumption (ii) that $\mathrm{P}(\mathrm{P}_e(|W - W_0| > \zeta_1) > \zeta_2) < \zeta_2$ for all $w \subset \mathcal{W}$, i.e. condition in equation (15) holds uniformly over all $w \subset \mathcal{W}$. This completes the proof by applying Corollary 3.1. ∎

APPENDIX G: MONTE CARLO EXPERIMENTS IN SUPPORT OF SECTION 4

In this section, we present results of Monte Carlo simulations that illustrate our theoretical results on Dantzig selector given in Section 4. We consider Gaussian and Non-Gaussian noise with homoscedasticity and heteroscedasticity. We study 3 types of Dantzig selector depending on the choice of the penalty level: canonical, ideal (based on gaussian approximation, GAR), and multiplier bootstrap (MB).

We consider the following regression model:

$$y_i = z_i'\beta + \varepsilon_i,$$

where observations are independent across i, y_i is a scalar dependent variable, z_i is a p-dimensional vector of covariates, and ε_i is noise. The first component of z_i equals 1 in all experiments (an intercept). Other $p-1$ components are simulated as follows: first, we simulate a vector $w_i \in \mathbb{R}^{p-1}$ from the Gaussian distribution with zero mean so that $\mathrm{E}[w_{ij}^2] = 1$ for all $1 \leqslant j \leqslant p-1$ and $\mathrm{E}[w_{ij}w_{ik}] = \rho$ for all $1 \leqslant j, k \leqslant p-1$ with $j \neq k$; second, we set $z_{ij+1} = w_{ij}/(\mathbb{E}_n[w_{ij}^2])^{1/2}$ (equicorrelated design). Depending on the experiment, we set $\rho = 0$, 0.5, 0.9, or 0.99. We simulate $\epsilon_i = \sigma_0\sigma(z_i)e_i$ where depending on the experiment, $\sigma_0 = 0.5$ or 1.0 and e_i is taken either from $N(0,1)$ distribution (Gaussian noise) or from t-distribution with 5 degrees of freedom normalized to have variance 1 (Non-Gaussian noise). To investigate the effect of heteroscedasticity on the properties of different estimators, we set

$$\sigma(z_i) = \frac{2\exp(\gamma z_{i2})}{1+\exp(\gamma z_{i2})}$$

where γ is either 0 (homoscedastic case) or 1 (heteroscedastic case).

Tables 1 and 2 present results on prediction error of Dantzig selector for the case of Non-Gaussian and Gaussian noise, respectively. Prediction error is defined as

$$\|\widehat{\beta} - \beta\|_{pr} = \sqrt{\mathbb{E}_n[z_i'(\widehat{\beta} - \beta)]}$$

where $\widehat{\beta}$ is the Dantzig selector; see Section 4 for the definition of the Dantzig selector. Recall that implementing the Dantzig selector requires selecting the penalty level λ. Both tables show results for 3 different choices of the penalty level. Canonical penalty is $\lambda = \bar{\sigma}\Phi^{-1}(1-\alpha/(2p))$ where $\bar{\sigma} = \sigma_0(1 + I\{|\gamma| > 0\})$, the upper bound on the variance of ε_i's. Ideal (based on gaussian approximation, GAR) penalty is $\lambda = c_{Z_0}(1-\alpha)$, the conditional $(1-\alpha)$ quantile of Z_0 given $(z_i)_{i=1}^n$ where

$$Z_0 = \sqrt{n}\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}\sigma_0\sigma(z_i)e_i]|$$

where $e_i \sim N(0,1)$ independently across i. Finally, multiplier bootstrap (MB) penalty is defined as follows. First, we calculate the Dantzig selector with the canonical choice of the penalty level, $\widehat{\beta}$, and select regressors corresponding to non-zero components of $\widehat{\beta}$. Second, we run the OLS regression

Table 1
Results of Monte Carlo experiments for prediction error. Non-Gaussian noise.

Distribution ε	σ_0	ρ	Method		
			Canonical	GAR	MB
Homoscedastic	0.5	0.00	0.224	0.207	0.208
		0.50	0.390	0.353	0.379
		0.90	0.352	0.317	0.340
		0.99	0.107	0.057	0.058
	1.0	0.00	0.648	0.604	0.674
		0.50	0.695	0.643	0.644
		0.90	0.538	0.406	0.412
		0.99	0.348	0.139	0.137
Heteroscedastic	0.5	0.00	0.656	0.252	0.393
		0.50	0.660	0.407	0.469
		0.90	0.516	0.326	0.342
		0.99	0.339	0.066	0.064
	1.0	0.00	1.588	0.661	0.909
		0.50	1.590	0.722	0.755
		0.90	1.336	0.445	0.454
		0.99	1.219	0.153	0.156

Table 2
Results of Monte Carlo experiments for prediction error. Gaussian noise.

Distribution ε	σ_0	ρ	Method		
			Canonical	GAR	MB
Homoscedastic	0.5	0.00	0.229	0.211	0.210
		0.50	0.421	0.386	0.417
		0.90	0.365	0.315	0.350
		0.99	0.109	0.059	0.059
	1.0	0.00	0.663	0.618	0.671
		0.50	0.679	0.627	0.624
		0.90	0.554	0.424	0.429
		0.99	0.326	0.127	0.120
Heteroscedastic	0.5	0.00	0.674	0.257	0.395
		0.50	0.643	0.412	0.451
		0.90	0.503	0.310	0.324
		0.99	0.319	0.060	0.059
	1.0	0.00	1.690	0.708	0.976
		0.50	1.537	0.665	0.679
		0.90	1.334	0.439	0.452
		0.99	1.189	0.155	0.148

of y_i on the set of selected regressors, and take residuals from this regression, $(\widehat{\epsilon}_i)_{i=1}^n$. Then the multiplier bootstrap penalty level is $\lambda = c_W(1-\alpha)$, the conditional $(1-\alpha)$ quantile of W given $(z_i, \widehat{\varepsilon}_i)_{i=1}^n$ where

$$W = \sqrt{n} \max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[z_{ij}\widehat{\varepsilon}_i e_i]|$$

where $e_i \sim N(0,1)$ independently across i.

The results show that the GAR penalty always yields smaller prediction error than that of the canonical penalty. Moreover, as predicted by the theory, GAR penalty works especially good in comparison with the canonical penalty in heteroscedastic case and/or in the case with high correlation between regressors (high ρ). In addition, in most cases, the results for the MB penalty are similar to those for the GAR penalty. In particular, the MB penalty in most cases yields smaller prediction error than that of the canonical penalty. Finally, the GAR penalty in most cases is slightly better than the MB penalty. Note, however, that when heteroscedasticity function $\sigma(z_i)$ is unknown, the GAR penalty becomes infeasible but the MB penalty is feasible given that the upper bound on the variance of ε_i's exists.

Supplementary Material II

Additional Results and Discussions

APPENDIX H: A NOTE ON SLEPIAN-STEIN TYPE METHODS FOR NORMAL APPROXIMATIONS

To keep the notation simple, consider a random vector X in $\mathbb{R}^p$ and a standard normal vector Y in $\mathbb{R}^p$. We are interested in bounding

$$\mathrm{E}[g(X)] - \mathrm{E}[g(Y)],$$

over some collection of test functions $g \in \mathcal{G}$. Without loss of generality, suppose that Y and X are independent.

Consider Stein's partial differential equation:

$$g(x) - \mathrm{E}[g(Y)] = \triangle h(x) - x'\nabla h(x)$$

where $\triangle h(X)$ and $\nabla h(X)$ refer to the Laplacian and the gradient of $h(X)$. It is well known, e.g. [14] and [8], that an explicit solution for h in this equation is given by

$$h(x) := -\int_0^1 \frac{1}{2t}\left[\mathrm{E}[g(\sqrt{t}x + \sqrt{1-t}Y)] - \mathrm{E}[g(Y)]\right] dt,$$

so that

$$\mathrm{E}[g(X)] - \mathrm{E}[g(Y)] = \mathrm{E}[\triangle h(X) - X'\nabla h(X)].$$

The Stein type method for normal approximation bounds the right side for $g \in \mathcal{G}$.

Next, let us consider the Slepian smart path interpolation:

$$Z(t) = \sqrt{t}X + \sqrt{1-t}Y.$$

Then we have

$$\mathrm{E}[g(X)] - \mathrm{E}[g(Y)] = \mathrm{E}\left[\int_0^1 \frac{1}{2}\nabla g(Z(t))'\left(\frac{X}{\sqrt{t}} - \frac{Y}{\sqrt{1-t}}\right)\right] dt.$$

The Slepian type method, as used in our paper, bounds the right side for $g \in \mathcal{G}$. We also refer the reader to [27] for a related discussion and interesting results (see in particular Lemma 2.1 in [27]).

Elementary calculations and integration by parts yield the following observation.

LEMMA H.1. *Suppose that $g : \mathbb{R}^p \to \mathbb{R}$ is a C^2-function with uniformly bounded derivatives up to order two. Then*

$$I := \mathrm{E}\left[\int_0^1 \frac{1}{2}\nabla g(Z(t))'\left(\frac{X}{\sqrt{t}}\right)\right] = -\mathrm{E}[X'\nabla h(X)]$$

and

$$II := \mathrm{E}\left[\int_0^1 \frac{1}{2}\nabla g(Z(t))'\left(\frac{Y}{\sqrt{1-t}}\right)\right] = -\mathrm{E}[\triangle h(X)].$$

Hence the Slepian and Stein methods both show that difference between I and II is small or approaches zero under suitable conditions on X; therefore, they are very similar in spirit, if not identical. The details of treating terms may be different from application to application; see more on this in [27].

PROOF OF LEMMA H.1. By definition of h, we have

$$-\mathrm{E}[X'\nabla h(X)] = \mathrm{E}\left[X'\int_0^1 \frac{1}{2t}\nabla g(Z(t))\sqrt{t}dt\right] = \mathrm{E}\left[\int_0^1 \nabla g(Z(t))'\frac{X}{2\sqrt{t}}dt\right].$$

On the other hand, by definition of h and Stein's identity (Lemma H.2),

$$-\mathrm{E}[\triangle h(X)] = \mathrm{E}\left[\frac{1}{2}\int_0^1 \triangle g(Z(t))dt\right] = \mathrm{E}\left[\frac{1}{2}\int_0^1 \nabla g(Z(t))'\left(\frac{Y}{\sqrt{1-t}}\right)dt\right].$$

This completes the proof. ∎

LEMMA H.2 (Stein's identity). *Let $W = (W_1, \dots, W_p)^T$ be a centered Gaussian random vector in $\mathbb{R}^p$. Let $f : \mathbb{R}^p \to \mathbb{R}$ be a C^1-function such that $\mathrm{E}[|\partial_j f(W)|] < \infty$ for all $1 \leqslant j \leqslant p$. Then for every $1 \leqslant j \leqslant p$,*

$$\mathrm{E}[W_j f(W)] = \sum_{k=1}^p \mathrm{E}[W_j W_k]\mathrm{E}[\partial_k f(W)].$$

PROOF OF LEMMA H.2. See Section A.6 of [29], and also [28]. ∎

APPENDIX I: A SIMPLE GAUSSIAN APPROXIMATION RESULT

This section can be helpful to the reader wishing to see how Slepian-Stein methods can be used to prove a simple Gaussian approximation (whose applicability is limited however.) We start with the following elementary lemma.

LEMMA I.1 (A Simple Comparison of Gaussian to Non-Gaussian Maxima). *For every $g \in C_b^3(\mathbb{R})$ and $\beta > 0$,*

$$|\mathrm{E}[g(F_\beta(X)) - g(F_\beta(Y))]| \lesssim n^{-1/2}(G_3 + G_2\beta + G_1\beta^2)\bar{\mathrm{E}}[S_i^3],$$

and hence

$$|\mathrm{E}[g(T_0) - g(Z_0)]| \lesssim n^{-1/2}(G_3 + G_2\beta + G_1\beta^2)\bar{\mathrm{E}}[S_i^3] + \beta^{-1}G_1 \log p.$$

The optimal value of the last bound is given by taking the minimum over β. We postpone choices of β to the proof of the subsequent corollary, leaving ourselves more flexibility in optimizing bounds in the corollary.

COMMENT I.1. The bound above per se seems new, though it is merely a simple extension of results in [6], who obtained the bound for the case with X having a special structure like in our example (E.4), related to spin glasses, using classical Lindeberg's method. We give a proof using a variant of Slepian-Stein method, since this is the tool we end up using to prove our main results, as the Lindeberg's method, in its pure form, did not yield the same sharp results. Our proof is related but rather different in details from the more abstract/general arguments based on Stein triplets given in [27] (Lemma 2.1), but given for the special case of data $(x_i)_{i=1}^n$ with coordinates x_i's that $\mathbb{R}$-valued, in contrast to the $\mathbb{R}^p$-valued case treated here. [27] reanalyzed [6]'s setup under local dependence and gave a number of other interesting applications. ■

The next result states a bound on the Kolmogorov distance between distributions of T_0 and Z_0. The result follows from Lemma I.1 and the anti-concentration inequality for maxima of Gaussian random variables stated in Lemma 2.1. Note that this result was not included in either [6] or [27] for the cases that they have analyzed.

COROLLARY I.1 (**A Simple Gaussian Approximation**). *Suppose that there are some constants $c_1 > 0$ and $C_1 > 0$ such that $c_1 \leqslant \bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ for all $1 \leqslant j \leqslant p$. Then there exists a constant $C > 0$ depending only on c_1 and C_1 such that*

$$\rho := \sup_{t \in \mathbb{R}} |\mathrm{P}(T_0 \leqslant t) - \mathrm{P}(Z_0 \leqslant t)| \leqslant C(n^{-1}(\log(pn))^7)^{1/8}(\bar{\mathrm{E}}[S_i^3])^{1/4}.$$

Theorem I.1 and Corollary I.1 imply that the error of approximating the maximum coordinate in the sum of independent random vectors by its Gaussian analogue depends on p (possibly) only through $\log p$. This is the main qualitative feature of all the results in this paper. Both Lemma I.1 and Corollary I.1 and all the results in this paper do not limit the dependence among the coordinates in x_i.

While Lemma I.1 and Corollary I.1 convey an important qualitative aspect of the problem and admit easy-to-grasp proofs, an important disadvantage of these results is that the bounds depend on $\bar{\mathrm{E}}[S_i^3]$. When $\bar{\mathrm{E}}[S_i^3]$ increases with n, for example when $|x_{ij}| \leqslant B_n$ for all i and j and B_n grows with n, the simple bound above may be too poor, and can be improved considerably using several inputs. We derive in Theorem 2.1 in the main text a bound that can be much better in the latter scenario. The improvement there comes at a cost of more involved statements and proofs.

PROOF OF LEMMA I.1. Without loss of generality, we are assuming that sequences $(x_i)_{i=1}^n$ and $(y_i)_{i=1}^n$ are independent. For $t \in [0,1]$, we consider the Slepian interpolation between Y and X:

$$Z(t) := \sqrt{t}X + \sqrt{1-t}Y = \sum_{i=1}^n Z_i(t),\ Z_i(t) := \frac{1}{\sqrt{n}}(\sqrt{t}x_i + \sqrt{1-t}y_i).$$

We shall also employ Stein's leave-one-out expansions:

$$Z^{(i)}(t) := Z(t) - Z_i(t).$$

Let $\Psi(t) = \mathrm{E}[m(Z(t))]$ for $m := g \circ F_\beta$. Then by Taylor's theorem,

$$\begin{aligned}\mathrm{E}[m(X) - m(Y)] = \Psi(1) - \Psi(0) &= \int_0^1 \Psi'(t)dt \\ &= \frac{1}{2}\sum_{j=1}^p \sum_{i=1}^n \int_0^1 \mathrm{E}[\partial_j m(Z(t))\dot{Z}_{ij}(t)]dt = \frac{1}{2}(I + II + III),\end{aligned}$$

where

$$\begin{aligned}\dot{Z}_{ij}(t) &= \frac{d}{dt}Z_{ij}(t) = \frac{1}{\sqrt{n}}\left(\frac{1}{\sqrt{t}}x_{ij} - \frac{1}{\sqrt{1-t}}y_{ij}\right), \text{ and} \\ I &= \sum_{j=1}^p \sum_{i=1}^n \int_0^1 \mathrm{E}[\partial_j m(Z^{(i)}(t))\dot{Z}_{ij}(t)]dt, \\ II &= \sum_{j,k=1}^p \sum_{i=1}^n \int_0^1 \mathrm{E}[\partial_j\partial_k m(Z^{(i)}(t))\dot{Z}_{ij}(t)Z_{ik}(t)]dt, \\ III &= \sum_{j,k,l=1}^p \sum_{i=1}^n \int_0^1 \int_0^1 (1-\tau)\mathrm{E}[\partial_j\partial_k\partial_l m(Z^{(i)}(t) + \tau Z_i(t))\dot{Z}_{ij}(t)Z_{ik}(t)Z_{il}(t)]d\tau dt.\end{aligned}$$

Note that random vector $Z^{(i)}(t)$ is independent of $(\dot{Z}_{ij}(t), Z_{ij}(t))$, and $\mathrm{E}[\dot{Z}_{ij}(t)] = 0$. Hence we have $I = 0$; moreover, since $\mathrm{E}[\dot{Z}_{ij}(t)Z_{ik}(t)] = n^{-1}\mathrm{E}[x_{ij}x_{ik} - y_{ij}y_{ik}] = 0$ by construction of $(y_i)_{i=1}^n$, we also have $II = 0$. Consider the third term III. We have that

$$
\begin{aligned}
|III| &\lesssim_{(1)} (G_3 + G_2\beta + G_1\beta^2)n \int \bar{\mathrm{E}}\left[\max_{1\leqslant j,k,l\leqslant p} |\dot{Z}_{ij}(t)Z_{ik}(t)Z_{il}(t)|\right] dt, \\
&\lesssim_{(2)} n^{-1/2}(G_3 + G_2\beta + G_1\beta^2)\bar{\mathrm{E}}\left[\max_{1\leqslant j\leqslant p} (|x_{ij}| + |y_{ij}|)^3\right],
\end{aligned}
$$

where (1) follows from $|\partial_j\partial_k\partial_l m(Z^{(i)}(t) + \tau Z_i(t))| \leqslant U_{jkl}(Z^{(i)}(t) + \tau Z_i(t)) \lesssim (G_3 + G_2\beta + G_1\beta^2)$ holding by Lemma A.5, and (2) is shown below. The first claim of the theorem now follows. The second claim follows directly from property (8) in the main text of the smooth max function.

It remains to show (2). Define $\omega(t) = 1/(\sqrt{t} \wedge \sqrt{1-t})$ and note,

$$
\begin{aligned}
&\int_0^1 n\bar{\mathrm{E}}\left[\max_{1\leqslant j,k,l\leqslant p} |\dot{Z}_{ij}(t)Z_{ik}(t)Z_{il}(t)|\right] dt \\
&= \int_0^1 \omega(t)n\bar{\mathrm{E}}\left[\max_{1\leqslant j,k,l\leqslant p} |\dot{Z}_{ij}(t)/\omega(t))Z_{ik}(t)Z_{il}(t)|\right] dt \\
&\leqslant n\int_0^1 \omega(t)\left(\bar{\mathrm{E}}[\max_{1\leqslant j\leqslant p} |\dot{Z}_{ij}(t)/\omega(t)|^3]\bar{\mathrm{E}}[\max_{1\leqslant j\leqslant p} |Z_{ij}(t)|^3]\bar{\mathrm{E}}[\max_{1\leqslant j\leqslant p} |Z_{ij}(t)|^3]\right)^{1/3} dt \\
&\leqslant n^{-1/2}\left\{\int_0^1 \omega(t)dt\right\}\bar{\mathrm{E}}\left[\max_{1\leqslant j\leqslant p} (|x_{ij}| + |y_{ij}|)^3\right]
\end{aligned}
$$

where the first inequality follows from Hölder's inequality, and the second from the fact that $|\dot{Z}_{ij}(t)/\omega(t)| \leqslant (|x_{ij}| + |y_{ij}|)/\sqrt{n}$, $|Z_{ij}(t)| \leqslant (|x_{ij}| + |y_{ij}|)/\sqrt{n}$. Finally we note that $\int_0^1 \omega(t)dt \lesssim 1$, so inequality (2) follows. This completes the overall proof. ∎

Proof of Corollary I.1. In this proof, let $C > 0$ denote a generic constant depending only on c_1 and C_1, and its value may change from place to place. For $\beta > 0$, define $e_\beta := \beta^{-1}\log p$. Recall that $S_i := \max_{1\leqslant j\leqslant p}(|x_{ij}| + |y_{ij}|)$. Consider and fix a $C_b^3(\mathbb{R})$-function $g_0 : \mathbb{R} \to [0,1]$ such that $g_0(s) = 1$ for $s \leqslant 0$ and $g_0(s) = 0$ for $s \geqslant 1$. Fix any $t \in \mathbb{R}$, and define $g(s) = g_0(\psi(s - t - e_\beta))$. For this function g, $G_0 = 1$, $G_1 \lesssim \psi$, $G_2 \lesssim \psi^2$ and $G_3 \lesssim \psi^3$.

Observe now that

$$
\begin{aligned}
\mathrm{P}(T_0 \leqslant t) &\leqslant \mathrm{P}(F_\beta(X) \leqslant t + e_\beta) \leqslant \mathrm{E}[g(F_\beta(X))] \\
&\leqslant \mathrm{E}[g(F_\beta(Y))] + C(\psi^3 + \beta\psi^2 + \beta^2\psi)(n^{-1/2}\bar{\mathrm{E}}[S_i^3]) \\
&\leqslant \mathrm{P}(F_\beta(Y) \leqslant t + e_\beta + \psi^{-1}) + C(\psi^3 + \beta\psi^2 + \beta^2\psi)(n^{-1/2}\bar{\mathrm{E}}[S_i^3]) \\
&\leqslant \mathrm{P}(Z_0 \leqslant t + e_\beta + \psi^{-1}) + C(\psi^3 + \beta\psi^2 + \beta^2\psi)(n^{-1/2}\bar{\mathrm{E}}[S_i^3]).
\end{aligned}
$$

where the first inequality follows from (8), the second from construction of g, the third from Theorem I.1, and the fourth from construction of g, and the last from (8). The remaining step is to compare $\mathrm{P}(Z_0 \leqslant t + e_\beta + \psi^{-1})$ with $\mathrm{P}(Z_0 \leqslant t)$ and this is where Lemma 2.1 plays its role. By Lemma 2.1,

$$\mathrm{P}(Z_0 \leqslant t + e_\beta + \psi^{-1}) - \mathrm{P}(Z_0 \leqslant t) \leqslant C(e_\beta + \psi^{-1})\sqrt{1 \vee \log(p\psi)}.$$

by which we have

$$\mathrm{P}(T_0 \leqslant t) - \mathrm{P}(Z_0 \leqslant t) \leqslant C[(\psi^3 + \beta\psi^2 + \beta^2\psi)(n^{-1/2}\bar{\mathrm{E}}[S_i^3]) + (e_\beta + \psi^{-1})\sqrt{1 \vee \log(p\psi)}].$$

We have to minimize the right side with respect to β and ψ. It is reasonable to choose β in such a way that e_β and ψ^{-1} are balanced, i.e., $\beta = \psi \log p$. With this β, the bracket on the right side is bounded from above by

$$C[\psi^3 (\log p)^2 (n^{-1/2}\bar{\mathrm{E}}[S_i^3]) + \psi^{-1}\sqrt{1 \vee \log(p\psi)}],$$

which is approximately minimized by $\psi = (\log p)^{-3/8}(n^{-1/2}\bar{\mathrm{E}}[S_i^3])^{-1/4}$. With this ψ, $\psi \leqslant (n^{-1/2}\bar{\mathrm{E}}[S_i^3])^{-1/4} \leqslant Cn^{1/8}$ (recall that $p \geqslant 3$), and hence $\log(p\psi) \leqslant C\log(pn)$. Therefore,

$$\mathrm{P}(T_0 \leqslant t) - \mathrm{P}(Z_0 \leqslant t) \leqslant C(n^{-1/2}\bar{\mathrm{E}}[S_i^3])^{1/4}(\log(pn))^{7/8}.$$

This gives one half of the claim. The other half follows similarly. ∎

APPENDIX J: GAUSSIAN APPROXIMATION AND MULTIPLIER BOOTSRAP RESULTS, ALLOWING FOR LOW VARIANCES

The purpose of this section is to provide results without an assumption that $\bar{\mathrm{E}}[x_{ij}^2] > c$ for all $1 \leqslant j \leqslant p$ and some constant $c > 0$.

J.1. Gaussian Approximation Results. In this subsection, we use the same setup and notation as those in Section 2. In particular, $x_1, \dots, x_n$ is a sequence of independent centered random vectors in $\mathbb{R}^p$, $y_1, \dots, y_n$ is a sequence of independent centered Gaussian random vectors such that $\mathrm{E}[y_i y_i'] = \mathrm{E}[x_i x_i']$, $T_0 = \max_{1 \leqslant j \leqslant p} X_j$ where $X = \sum_{i=1}^n x_i/\sqrt{n}$, $Z_0 = \max_{1 \leqslant j \leqslant p} Y_j$ where $Y = \sum_{i=1}^n y_i/\sqrt{n}$, and

$$\rho = \sup_{t \in \mathbb{R}} |\mathrm{P}(T_0 \leqslant t) - \mathrm{P}(Z_0 \leqslant t)|.$$

In addition, denote

$$M_{k,2} := \max_{1 \leqslant j \leqslant p} \frac{\bar{\mathrm{E}}[|x_{ij}|^k]^{1/k}}{\bar{\mathrm{E}}[x_{ij}^2]^{1/2}} \text{ and } \ell_n := \log(pn/\gamma).$$

We will impose the following condition:

(SM) There exists $J \subset \{1, \dots, p\}$ such that $|J| \geqslant \nu p$ and for all $(j,k) \in J \times J$ with $j \neq k$, $\bar{\mathrm{E}}[x_{ij}^2] \geqslant c_1$ and $|\bar{\mathrm{E}}[x_{ij}x_{ik}]| \leqslant (1-\nu')(\bar{\mathrm{E}}[x_{ij}^2]\bar{\mathrm{E}}[x_{ik}^2])^{1/2}$ for some strictly positive constants ν, ν', and c_1 independent of n.

THEOREM J.1. *Suppose that condition (SM) holds. In addition, suppose that there is some constant $C_1 > 0$ such that $\bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ for $1 \leqslant j \leqslant p$. Then for every $\gamma \in (0,1)$,*

$$\rho \leqslant C \left\{ n^{-1/8}(M_3^{3/4} \vee M_{4,2}^{1/2})\ell_n^{7/8} + n^{-1/2}\ell_n^{3/2}u(\gamma) + p^{-c}\ell_n^{1/2} + \gamma \right\},$$

where $c, C > 0$ are constants that depend only on ν, ν', c_1 and C_1.

Theorem J.1 has the following applications. Let $C_1 > 0$ be some constant that is independent of n, and let $B_n \geqslant 1$ be a sequence of constants. We allow for the case where $B_n \to \infty$ as $n \to \infty$. We will assume that one of the following conditions is satisfied *uniformly in* $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant p$:

(E.5) $\bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ and $\max_{k=1,2} M_{k+2,2}^{k+2}/B_n^k + \mathrm{E}[\exp(|x_{ij}|/B_n)] \leqslant 2$;

(E.6) $\bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ and $\max_{k=1,2} M_{k+2,2}^{k+2}/B_n^k + \mathrm{E}[(\max_{1 \leqslant j \leqslant p} |x_{ij}|/B_n)^4] \leqslant 2$.

COROLLARY J.1. *Suppose that there exist constants $c_2 > 0$ and $C_2 > 0$ such that one of the following conditions is satisfied: (i) (E.5) holds and $B_n^2(\log(pn))^7/n \leqslant C_2 n^{-c_2}$ or (ii) (E.6) holds and $B_n^4(\log(pn))^7/n \leqslant C_2 n^{-c_2}$. In addition, suppose that condition (SM) holds and $p \geqslant C_3 n^{c_3}$ for some constants $c_3 > 0$ and $C_3 > 0$. Then there exist constants $c > 0$ and $C > 0$ depending only on $\nu, \nu', c_1, C_1, c_2, C_2, c_3$, and C_3 such that*

$$\rho \leqslant C n^{-c}.$$

J.2. Multiplier Bootstrap Results. In this subsection, we use the same setup and notation as those in Section 3. In particular, in addition to the notation used above, we assume that random variables T and W satisfy conditions (14) and (15) in the main text, respectively, where $\zeta_1 \geqslant 0$ and $\zeta_2 \geqslant 0$ depend on n and where W_0 appearing in condition (15) is defined in equation (13) in the main text. Recall that $\Delta = \max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[x_{ij}] - \bar{\mathrm{E}}[x_{ij}]|$.

THEOREM J.2. *Suppose that there is some constant $C_1 > 0$ such that $\bar{\sigma} := \max_{1 \leqslant j \leqslant p} \bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ for all $1 \leqslant j \leqslant p$. In addition, suppose that condition (SM) holds. Moreover, suppose that conditions (14) and (15) are satisfied. Then for every $\vartheta > 0$,*

$$\begin{aligned}\rho_{\ominus} &:= \sup_{\alpha \in (0,1)} \mathrm{P}(\{T \leqslant c_W(\alpha)\} \ominus \{T_0 \leqslant c_{Z_0}(\alpha)\}) \\ &\leqslant 2(\rho + \pi(\vartheta) + \mathrm{P}(\Delta > \vartheta)) + C(\zeta_1 \vee p^{-c})\sqrt{1 \vee \log(p/\zeta_1)} + 5\zeta_2,\end{aligned}$$

where

$$\pi(\vartheta) := C\vartheta^{1/3}(1 \vee \log(p/\vartheta))^{2/3} + Cp^{-c}\sqrt{1 \vee \log(p/\vartheta)}$$

and $c, C > 0$ *depend only on* ν, ν', c_1 *and* C_1. *In addition,*

$$\sup_{\alpha \in (0,1)} |\mathrm{P}(T \leqslant c_W(\alpha)) - \alpha| \leqslant \rho_\ominus + \rho.$$

COROLLARY J.2. *Suppose that there exist constants* $c_2, C_2 > 0$ *such that conditions (14) and (15) hold with* $\zeta_1\sqrt{\log p} + \zeta_2 \leqslant C_2 n^{-c_2}$. *Moreover, suppose that one of the following conditions is satisfied: (i) (E.5) holds and* $B_n^2(\log(pn))^7/n \leqslant C_2 n^{-c_2}$ *or (ii) (E.6) holds and* $B_n^4(\log(pn))^7/n \leqslant C_2 n^{-c_2}$. *Finally, suppose that condition (SM) holds and* $p \geqslant C_3 n^{c_3}$ *for some constants* $c_3 > 0$ *and* $C_3 > 0$. *Then there exist constants* $c > 0$ *and* $C > 0$ *depending only on* $\nu, \nu', c_1, C_1, c_2, C_2, c_3,$ *and* C_3 *such that*

$$\rho_\ominus = \sup_{\alpha \in (0,1)} \mathrm{P}(\{T \leqslant c_W(\alpha)\} \ominus \{T_0 \leqslant c_{Z_0}(\alpha)\}) \leqslant Cn^{-c}.$$

In addition, $\sup_{\alpha \in (0,1)} |\mathrm{P}(T \leqslant c_W(\alpha)) - \alpha| \leqslant \rho_\ominus + \rho \leqslant Cn^{-c}$.

The proofs rely on the following auxiliary lemmas, whose proofs will be given below.

LEMMA J.1. *(a) Let* $Y_1, \dots, Y_p$ *be jointly Gaussian random variables with* $\mathrm{E}[Y_j] = 0$ *and* $\sigma_j^2 := \mathrm{E}[Y_j^2]$ *for all* $1 \leqslant j \leqslant p$. *Let* $b_p := \mathrm{E}[\max_{1 \leqslant j \leqslant p} Y_j]$ *and* $\bar{\sigma} = \max_{1 \leqslant j \leqslant p} \sigma_j > 0$. *Assume that* $b_p \geqslant c_1\sqrt{\log p}$ *for some* $c_1 > 0$. *Then for every* $\varsigma > 0$,

$$\sup_{z \in \mathbb{R}} \mathrm{P}\left(|\max_{1 \leqslant j \leqslant p} Y_j - z| \leqslant \varsigma\right) \leqslant C(\varsigma \vee p^{-c})\left(b_p + \sqrt{1 \vee \log(\bar{\sigma}/\varsigma)}\right) \tag{6}$$

where $c, C > 0$ *are some constants depending only on* c_1 *and* $\bar{\sigma}$. *(b) Furthermore, the worst case bound is obtained by bounding* b_p *by* $\bar{\sigma}\sqrt{2\log p}$.

LEMMA J.2. *Let* V *and* Y *be centered Gaussian random vectors in* $\mathbb{R}^p$ *with covariance matrices* Σ^V *and* Σ^Y, *respectively. Let* $\Delta_0 := \max_{1 \leqslant j,k \leqslant p} |\Sigma_{jk}^V - \Sigma_{jk}^Y|$. *Suppose that there are some constants* $0 < c_1 < C_1$ *such that* $\bar{\sigma} := \max_{1 \leqslant j \leqslant p} \mathrm{E}[Y_j^2] \leqslant C_1$ *for all* $1 \leqslant j \leqslant p$ *and* $b_p := \mathrm{E}[\max_{1 \leqslant j \leqslant p} Y_j] \geqslant c_1\sqrt{\log p}$. *Then there exist constants* $c > 0$ *and* $C > 0$ *depending only on* c_1 *and* C_1 *such that*

$$\begin{aligned}\sup_{t \in \mathbb{R}} \left|\mathrm{P}\left(\max_{1 \leqslant j \leqslant p} V_j \leqslant t\right) - \mathrm{P}\left(\max_{1 \leqslant j \leqslant p} Y_j \leqslant t\right)\right| \leqslant & C\Delta_0^{1/3}(1 \vee \log(p/\Delta_0))^{2/3} \\ & + Cp^{-c}\sqrt{1 \vee \log(p/\Delta_0)}.\end{aligned}$$

Proof of Theorem J.1. It follows from Theorem 2.3.16 in [13] that condition (SM) implies that $\mathrm{E}[Z_0] \geqslant c\sqrt{\log p}$ for some $c > 0$ that depends only on ν, ν', and c_1. Therefore, using the argument like that in the proof of Theorem 2.2 with an application of Lemma J.1 instead of Lemma 2.1, we obtain

$$\begin{aligned} \rho \leqslant C\big[&n^{-1/2}(\psi^3 + \psi^2\beta + \psi\beta^2)M_3^3 + (\psi^2 + \psi\beta)\bar{\varphi}(u) \\ &+\psi\bar{\varphi}(u)\sqrt{\log(p/\gamma)} + (\beta^{-1}\log p + \psi^{-1} + p^{-c})\sqrt{1 \vee \log(p\psi)}\big] \end{aligned} \tag{7}$$

where all notation is taken from the proof of Theorem 2.2. Recall that $\bar{\varphi}(\cdot)$ is any function satisfying $\bar{\varphi}(u) \geqslant \varphi(u)$ for all $u > 0$ and $\varphi(u) = \varphi_x(u) \vee \varphi_y(u)$. To bound $\varphi_x(u)$, we have

$$\begin{aligned} \bar{\mathrm{E}}[x_{ij}^2 1\{|x_{ij}| > u(\bar{\mathrm{E}}[x_{ij}^2])^{1/2}\}] &\leqslant \bar{\mathrm{E}}[x_{ij}^4]/(u^2\bar{\mathrm{E}}[x_{ij}^2]) \\ &= \bar{\mathrm{E}}[x_{ij}^4]/(u^2\bar{\mathrm{E}}[x_{ij}^2]^2)\bar{\mathrm{E}}[x_{ij}^2] \leqslant (M_{4,2}^4/u^2)\bar{\mathrm{E}}[x_{ij}^2]. \end{aligned}$$

This implies that $\varphi_x(u) \leqslant M_{4,2}^2/u$. To bound $\varphi_y(u)$, note that $\bar{\mathrm{E}}[y_{ij}^4] \leqslant 3\bar{\mathrm{E}}[x_{ij}^4]$, which was shown in the proof of Theorem 2.2. In addition, $\bar{\mathrm{E}}[y_{ij}^2] = \bar{\mathrm{E}}[x_{ij}^2]$. Therefore, $\varphi_y(u) \leqslant C\varphi_x(u)$. Hence, we can set $\bar{\varphi}(u) := CM_{4,2}^2/u$ for all $u > 0$. The rest of the proof is the same as that for Theorem 2.2 with M_4 replaced by $M_{4,2}$. ∎

Proof of Corollary J.1. Note that in both cases, $M_{4,2}^2 \leqslant CB_n$ and

$$M_3^3 = \max_{1\leqslant j\leqslant p} \bar{\mathrm{E}}[|x_{ij}|^3] \leqslant M_{3,2}^3 \max_{1\leqslant j\leqslant p} \bar{\mathrm{E}}[x_{ij}^2]^{3/2} \leqslant CM_{3,2}^3 \leqslant CB_n.$$

Therefore, the claim of the corollary follows from Theorem J.1 by the same argument as that leading to Corollary 2.1 from Theorem 2.2. ∎

Proof of Theorem J.2. The proof is the same as that for Theorem 3.2 with Lemmas J.1 and J.2 replacing Lemmas 2.1 and 3.1. ∎

Proof of Corollary J.2. Since $B_n \geqslant 1$, both under (E.5) and under (E.6) we have $(\log(pn))^7/n \leqslant C_2 n^{-c_2}$. Let $\tilde{\zeta}_1 := \zeta_1 \vee n^{-1}$. Then conditions (14) and (15) hold with $(\tilde{\zeta}_1, \zeta_2)$ replacing (ζ_1, ζ_2) and $\tilde{\zeta}_1\sqrt{\log p} + \zeta_2 \leqslant Cn^{-c}$. Further, since $p \geqslant C_3 n^{c_3}$, we have $p^{-c}(1 \vee \log(p/\tilde{\zeta}_1))^{1/2} \leqslant Cn^{-c}$.

Let $\vartheta = \vartheta_n := ((\mathrm{E}[\Delta])^{1/2}/\log p) \vee n^{-1}$. Then $p^{-c}(1 \vee \log(p/\vartheta))^{1/2} \leqslant Cn^{-c}$. In addition, if $\vartheta = n^{-1}$, then $\vartheta^{1/3}(\log(p/\vartheta))^{2/3} \leqslant Cn^{-c}$. Finally,

$$M_4^2 = \max_{1\leqslant j\leqslant p} \bar{\mathrm{E}}[x_{ij}^4]^{1/2} \leqslant M_{4,2}^2 \max_{1\leqslant j\leqslant p} \bar{\mathrm{E}}[x_{ij}^2] \leqslant CM_{4,2}^2 \leqslant CB_n.$$

The rest of the proof is similar to that for Corollary 3.1. ∎

Proof of Lemma J.1. In the proof, several constants will be introduced. All of these constants are implicitly assumed to depend only on c_1 and $\bar{\sigma}$.

We choose $c > 0$ such that $4\bar{\sigma}\sqrt{c} = c_1$. Fix $\varsigma > 0$. It suffices to consider the case $\varsigma \geqslant \bar{\sigma}p^{-c}$. Let $\underline{\sigma} := c_2 b_p/\sqrt{\log p}$ for sufficiently small $c_2 > 0$ to be chosen below. Note that $\underline{\sigma} \geqslant c_1 c_2$. So, if $\varsigma > \underline{\sigma}$, (6) holds trivially by selecting sufficiently large C.

Consider the case $\varsigma \leqslant \underline{\sigma}$. Assume that $z > b_p + \varsigma + \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}$. Then

$$\begin{aligned} \mathrm{P}(|\max_{1\leqslant j\leqslant p} Y_j - z| \leqslant \varsigma) &\leqslant \mathrm{P}(\max_{1\leqslant j\leqslant p} Y_j \geqslant z - \varsigma) \\ &\leqslant \mathrm{P}(\max_{1\leqslant j\leqslant p} Y_j \geqslant b_p + \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}) \leqslant \varsigma/\underline{\sigma} \end{aligned}$$

where the last inequality follows from Borell's inequality. So, (6) holds by selecting sufficiently large C.

Now assume that $z < b_p - \varsigma - \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}$. Then

$$\begin{aligned} \mathrm{P}(|\max_{1\leqslant j\leqslant p} Y_j - z| \leqslant \varsigma) &\leqslant \mathrm{P}(\max_{1\leqslant j\leqslant p} Y_j \leqslant z + \varsigma) \\ &\leqslant \mathrm{P}(\max_{1\leqslant j\leqslant p} Y_j \leqslant b_p - \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}) \leqslant \varsigma/\underline{\sigma} \end{aligned}$$

where the last inequality follows from Borell's inequality. So, (6) holds by selecting sufficiently large C.

Finally, assume that

$$b_p - \varsigma - \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)} \leqslant z \leqslant b_p + \varsigma + \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}.$$

Then

$$\mathrm{P}(|\max_{1\leqslant j\leqslant p} Y_j - z| \leqslant \varsigma) \leqslant \mathrm{P}(|\max_{j\in\tilde{J}} Y_j - z| \leqslant \varsigma) + \mathrm{P}(|\max_{j\in J\setminus\tilde{J}} Y_j - z| \leqslant \varsigma) =: I + II$$

where $J := \{1,\dots,p\}$ and $\tilde{J} := \{j \in J : \sigma_j \leqslant \underline{\sigma}\}$. Consider I. We have

$$b_p \geqslant_{(1)} c_1\sqrt{\log p} =_{(2)} 4\bar{\sigma}\sqrt{c\log p} = 4\bar{\sigma}\sqrt{\log(p^c)} \geqslant_{(3)} 4\bar{\sigma}\sqrt{\log(\bar{\sigma}/\varsigma)}$$

where (1) holds by assumption, (2) follows from the definition of c, and (3) holds because $\varsigma \geqslant \bar{\sigma}p^{-c}$. In addition, there exists $C_2 > 0$ such that $\mathrm{E}[\max_{j\in\tilde{J}} Y_j] \leqslant c_2 C_2 b_p$, and there exist $C_3 > 0$ such that $b_p \leqslant C_3\bar{\sigma}\sqrt{\log p}$, so that $\underline{\sigma} \leqslant c_2 C_3\bar{\sigma}$. We choose c_2 so that $c_2 C_2 \leqslant 1/4$, $c_2/\sqrt{\log p} \leqslant 1/8$, and $c_2 C_3 \leqslant 1$. Then $\underline{\sigma} \leqslant \bar{\sigma}$, $\underline{\sigma} \leqslant b_p/8$, and $\mathrm{E}[\max_{j\in\tilde{J}} Y_j] \leqslant b_p/4$. Also recall that $\varsigma \leqslant \underline{\sigma}$. Therefore,

$$\begin{aligned} b_p - 2\varsigma - \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)} - \mathrm{E}[\max_{j\in\tilde{J}} Y_j] &\geqslant b_p/2 - \bar{\sigma}\sqrt{2\log(\bar{\sigma}/\varsigma)} \\ &\geqslant \bar{\sigma}\sqrt{2\log(\bar{\sigma}/\varsigma)} \geqslant \underline{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}. \end{aligned}$$

So, Borell's inequality yields

$$I \leqslant \mathrm{P}(\max_{j\in\tilde{J}} Y_j \geqslant z - \varsigma) \leqslant \mathrm{P}(\max_{j\in\tilde{J}} Y_j \geqslant b_p - 2\varsigma - \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)}) \leqslant \varsigma/\underline{\sigma}$$

because $\sigma_j \leqslant \underline{\sigma}$ for all $j \in \tilde{J}$.

Consider II. It is proved in [11] that

$$II \leqslant 4\varsigma\{(1/\underline{\sigma} - 1/\bar{\sigma})|z| + a_p + 1\}/\underline{\sigma} \tag{8}$$

where $a_p := \mathrm{E}[\max_{j\in J\setminus\tilde{J}} Y_j/\sigma_j]$. See, in particular, equation (16) in that paper. Note that $a_p \leqslant b_p/\underline{\sigma}$. Therefore, (8) combined with our restriction on z yields

$$\begin{aligned} II &\leqslant 4\varsigma\left(2b_p + \varsigma + \bar{\sigma}\sqrt{2\log(\underline{\sigma}/\varsigma)} + \underline{\sigma}\right)/\underline{\sigma}^2 \\ &\leqslant 4\varsigma\left(2b_p + 2\underline{\sigma} + \bar{\sigma}\sqrt{2\log(\bar{\sigma}/\varsigma)}\right)/\underline{\sigma}^2 \end{aligned}$$

where in the second line we used the facts that $\varsigma \leqslant \underline{\sigma}$ and $\underline{\sigma} \leqslant \bar{\sigma}$ by assumption and by construction, respectively. Now (6) holds by selecting sufficiently large $C > 0$, and using the fact that $\underline{\sigma} \geqslant c_1c_2 > 0$. This completes the proof. ■

Proof of Lemma J.2. The proof is the same as that for Theorem 2 in [11] with Lemma J.1 replacing Lemma 2.1. ■

APPENDIX K: VALIDITY OF EFRON'S EMPIRICAL BOOTSTRAP

In this section, we study the validity of the empirical (or Efron's) bootstrap in approximating the distribution of T_0 in the simple case where x_{ij}'s are uniformly bounded (the bound can increase with n). Moreover, we consider here the asymptotics where $n \to \infty$ and possibly $p = p_n \to \infty$. Recall the setup in Section 2: let $x_1, \dots, x_n$ be independent centered random vectors in $\mathbb{R}^p$ and define

$$T_0 = \max_{1\leqslant j\leqslant p} \frac{1}{\sqrt{n}} \sum_{i=1}^n x_{ij}.$$

The empirical bootstrap procedure is described as follows. Let $x_1^*, \dots, x_n^*$ be i.i.d. draws from the empirical distribution of $x_1, \dots, x_n$. Conditional on $(x_i)_{i=1}^n$, $x_1^*, \dots, x_n^*$ are i.i.d. with mean $\mathbb{E}_n[x_i]$ and covariance matrix $\mathbb{E}_n[(x_i - \mathbb{E}_n[x_i])(x_i - \mathbb{E}_n[x_i])']$. Define

$$T_0^* = \max_{1\leqslant j\leqslant p} \frac{1}{\sqrt{n}} \sum_{i=1}^n (x_{ij}^* - \mathbb{E}_n[x_{ij}]).$$

The empirical bootstrap approximates the distribution of T_0 by the conditional distribution T_0^* given $(x_i)_{i=1}^n$.

Recall

$$W_0 = \max_{1 \leqslant j \leqslant p} \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i x_{ij},$$

where $e_1, \dots, e_n$ are i.i.d. $N(0,1)$ random variables independent of $(x_i)_{i=1}^n$. We shall here compare the conditional distribution of T_0^* to that of W_0.

THEOREM K.1. *Suppose that there exists constants $C_1 > c_1 > 0$ and a sequence $B_n \geqslant 1$ of constants such that $c_1 \leqslant \bar{\mathrm{E}}[x_{ij}^2] \leqslant C_1$ for all $1 \leqslant j \leqslant p$ and $|x_{ij}| \leqslant B_n$ for all $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant p$. Then provided that $B_n^2(\log(pn))^7 = o(n)$, with probability $1 - o(1)$,*

$$\sup_{t \in \mathbb{R}} |\mathrm{P}\{T_0^* \leqslant t \mid (x_i)_{i=1}^n\} - \mathrm{P}\{W_0 \leqslant t \mid (x_i)_{i=1}^n\}| = o(1).$$

This theorem shows the asymptotic equivalence of the empirical and Gaussian multiplier bootstraps. The validity of the empirical bootstrap (in the form similar to that in Theorem 3.1) follows relatively directly from the validity of the Gaussian multiplier bootstrap.

PROOF OF THEOREM K.1. The proof consists of three steps.

Step 1. We first show that with probability $1 - o(1)$, $c_1/2 \leqslant \mathbb{E}_n[(x_{ij} - \mathbb{E}_n[x_{ij}])^2] \leqslant 2C_1$ for all $1 \leqslant j \leqslant p$. By Lemma A.1,

$$\mathrm{E}\left[\max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[x_{ij}]|\right] \lesssim \sqrt{C_1(\log p)/n} + B_n(\log p)/n = o((\log p)^{-1/2}),$$

$$\mathrm{E}\left[\max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[x_{ij}^2] - \bar{\mathrm{E}}[x_{ij}^2]|\right] \lesssim \sqrt{C_1 B_n^2(\log p)/n} + B_n^2(\log p)/n = o(1),$$

so that uniformly in $1 \leqslant j \leqslant p$, $|\mathbb{E}_n[(x_{ij} - \mathbb{E}_n[x_{ij}])^2] - \bar{\mathrm{E}}[x_{ij}^2]| = o_{\mathrm{P}}(1)$, which implies the desired assertion.

Step 2. Define

$$\check{W}_0 = \max_{1 \leqslant j \leqslant p} \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i(x_{ij} - \mathbb{E}_n[x_{ij}]).$$

We show that with probability $1 - o(1)$,

$$\sup_{t \in \mathbb{R}} |\mathrm{P}\{T_0^* \leqslant t \mid (x_i)_{i=1}^n\} - \mathrm{P}\{\check{W}_0 \leqslant t \mid (x_i)_{i=1}^n\}| = o(1). \tag{9}$$

Conditional on $(x_i)_{i=1}^n$, $x_i^* - \mathbb{E}_n[x_i]$ are independent centered random vector in $\mathbb{R}^p$ with covariance matrix $\mathbb{E}_n[(x_i - \mathbb{E}_n[x_i])(x_i - \mathbb{E}_n[x_i])']$. Hence conditional on $(x_i)_{i=1}^n$, we can apply Corollary 2.1 to T_0^* to deduce (9).

Step 3. We show that with probability $1 - o(1)$,

$$\sup_{t \in \mathbb{R}} |\mathrm{P}\{\check{W}_0 \leqslant t \mid (x_i)_{i=1}^n\} - \mathrm{P}\{W_0 \leqslant t \mid (x_i)_{i=1}^n\}| = o(1).$$

By definition, we have

$$|\check{W}_0 - W_0| \leqslant \max_{1 \leqslant j \leqslant p} |\mathbb{E}_n[x_{ij}]| \times \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n e_i \right| = o_{\mathrm{P}}((\log p)^{-1/2}).$$

Hence using the anti-concentration inequality together with Step 1, we deduce the desired assertion.

The conclusion of Theorem K.1 follows from combining Steps 1-3. ∎

APPENDIX L: COMPARISON OF OUR GAUSSIAN APPROXIMATION RESULTS TO OTHER ONES

We first point out that our Gaussian approximation result (4) can be viewed as a version of multivariate central limit theorem, which is concerned with conditions under which

$$|\mathrm{P}\,(X \in A) - \mathrm{P}\,(Y \in A)| \to 0, \tag{10}$$

uniformly in a collection of sets A, typically *all* convex sets. Recall that

$$X = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i,$$

where $x_1, \dots, x_n$ are independent centered random vectors in $\mathbb{R}^p$ with possibly $p = p_n \to \infty$, and

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n y_i,$$

$y_1, \dots, y_n$ independent random vectors with $y_i \sim N(0, \mathrm{E}[x_i x_i'])$. In fact, the result (4) in the main text can be rewritten as

$$\sup_{t \in \mathbb{R}} |\mathrm{P}\,\{X \in A_{\max}(t)\} - \mathrm{P}\,\{Y \in A_{\max}(t)\}| \to 0, \tag{11}$$

where $A_{\max}(t) = \{a \in \mathbb{R}^p : \max_{1 \leqslant j \leqslant p} a_j \leqslant t\}$.

Hence, our paper contributes to the literature on multivariate central limit theorems with growing number of dimensions (see, among others, [21, 24, 1, 15, 4]). These papers are concerned with results of the form (10), but either explicitly or implicitly require the condition that $p^c/n \to 0$ for some $c > 0$ (when specialized to a setting like our setup). Results in these papers rely on the anti-concentration results for Gaussian random vectors on the δ-expansions of boundaries of arbitrary convex sets A (see [2]). We restrict our attention to the class of sets of the form $A_{\max}(t)$ in (11). These sets have a special structure that allows us to deal with the case where $p \gg n$: in particular, concentration of measure on the δ-expansion of boundary of $A_{\max}(t)$ is at most of order

$$\delta \mathrm{E}[\max_{j \leqslant p} Y_j]$$

for Gaussian random vectors with unit variance (and separable Gaussian processes more generally), as shown in [11] (see also Lemma 2.1).

There is large literature on bounding the difference:

$$|\mathrm{E}[H(X)] - \mathrm{E}[H(Y)]|, \tag{12}$$

for various smooth functions $H(\cdot) : \mathbb{R}^p \to \mathbb{R}$, in particular the recent work includes [14, 8, 25, 9]. Any such bounds lead to Gaussian approximations, though the structure of H's plays an important role in limiting the scope of this approximation. Two methods in the literature that turned out most fruitful for deriving gaussian approximation results in high dimensional settings ($p \to \infty$ as $n \to \infty$ in our context) are those of Lindeberg and Stein. The history of the Lindeberg method dates back to Lindeberg's original proof of the central limit theorem ([20]), which has been revived in the recent literature. We refer to the introduction of [7] for a brief history on the Lindeberg's method; see also [6]. The recent development on Stein's method when x_i's are multivariate can be found in [14, 8, 25, 9]. See also [5] for a comprehensive overview of different methods. In contrast to these papers, our paper analyzes a rather particular, yet important case $H(\cdot) = g(F_\beta(\cdot))$, with $g : \mathbb{R} \to \mathbb{R}$ (progressively less) smooth function and $0 < \beta \to \infty$, and in our case, self-normalized truncation, some fine properties of the smooth potential F_β, maximal fourth order moments of variables and of their envelops, play a critical role (as we comment further below), and so our main results can not be (immediately) deduced from these prior results (nor do we attempt to follow this route).

Using the Lindeberg's method and the smoothing technique of Bentkus [3], [22] derived in their Theorem 5 a Gaussian approximation result that is of similar form to our *simple* (non-main) Gaussian approximation result presented in Section I of the SM. However, in [22], the anti-concentration property is *assumed* (see equation (1.4) in their paper); in our notation, their assumption on anti-concentration says that there exists a constant C independent of n and p such that

$$\mathrm{P}(r < \max_{1 \leqslant j \leqslant p} |Y_j| \leqslant r + \epsilon) \leqslant C\epsilon(1+r)^{-3}.$$

This assumption is useful for the analysis of the Donsker case, but does not apply in our (non-Donsker) cases. In fact, it rules out the simple case where $Y_1, \dots, Y_p$ are independent (i.e., the coordinates in x_i are uncorrelated) or $Y_1, \dots, Y_p$ are weakly dependent (i.e., the coordinates in x_i are weakly correlated). In addition, it is worth pointing out that the use of Bentkus's [3] smoothing[1] instead of the smoothing by potentials from spin glasses used here, does not lead to optimal results in our case, since very subtle properties (stability property noted in Lemma A.6) of potentials play an important role in our proofs, and in particular, is crucial for getting a reasonable exponent in the dependence on $\log p$.

Chatterjee [6], who also used the Lindeberg method, analyzed a spin-glass example like our example (E.4) and also derived a result similar to our *simple* (non-main) Gaussian approximation result presented in Section I of the SM, where $x_i = z_{ij}\epsilon_i$, where $(z_{ij})_{i=1}^p$ are fixed and $(\epsilon_i)_{i=1}^n$ are i.i.d $\mathbb{R}$-valued and centered with bounded third moment. We note [6] only provided a result for smooth functionals, but the extension to non-smooth cases follows from our Lemma 2.1 along with standard kernel smoothing. In fact, all of our paper is inspired by Chatterjee's work, and an early version employed (combinatorial) Lindeberg's method. We discuss the limitations of Lindeberg's approach (in the canonical form) in the main text, where we motivate the use of a combination of Slepian-Stein method in conjunction with self-normalized truncation and subtle properties of the potential function that approximates the maximum function. Generalization of results of Chatterjee to the case where $\mathbb{R}$-valued $(\epsilon_i)_{i=1}^n$ are locally dependent are given in [27], who uses Slepian-Stein methods for proofs and gave a result similar to our *simple* (non-main) Gaussian approximation result presented in Section I of the SM. Like [27], we also use a version of Slepian-Stein methods, but the proof details (as well as results and applications) are quite different, since we instead analyze the case where the data $(x_i)_{i=1}^n$ are $\mathbb{R}^p$-valued (instead of $\mathbb{R}$-valued) independent vectors, and since we have to perform truncation (to get good dependencies on the size of the envelopes, $\max_{1 \leqslant j \leqslant p} |x_{ij}|$) and use subtle properties of the potential function to get our main results (to get good dependencies on $\log p$).

Using an interesting modification of the Lindeberg method, [17] obtained an invariance principle for a sequence of sub-Gaussian $\mathbb{R}$-valued random variables $(\epsilon_i)_{i=1}^n$ (instead of $\mathbb{R}^p$-valued case consider here). Specifically, they looked at the large-sample probability of $(\epsilon_i)_{i=1}^n$ hitting a polytope formed

[1] Bentkus gave a proof of existence of a smooth function that approximates a supremum norm, with certain properties. The use of these properties alone does not lead to sharp results in our case. We rely instead on smoothing by potentials from spin glasses, and their detailed properties, most importantly the stability property, noted in Lemma A.6, which is readily established for this smoothing method.

by p half-spaces, which is on the whole a different problem than studied here (though tools are insightful, e.g. the novel use of results developed by Nazarov [23]). These results have no intersection with our results, except for a special case of "sub-exponential regression/spin-glass example" (E.3), if we further require in that example, that $\varepsilon_{ij} = \varepsilon_i$ for all $1 \leqslant j \leqslant p$, that ε_i's are sub-Gaussian, that $\mathrm{E}[\varepsilon_i^3] = 0$, and that $B_n^2 (\log p)^{16}/n \to 0$. All of these conditions and especially the last one are substantively more restrictive than what is obtained for the example (E.3) in our Corollary 2.1.

Finally, we note that when x_i's are identically distributed in addition to being independent, the theory of strong approximations and, in particular, Hungarian coupling can also be used to obtain results like that in (4) in the main text under conditions permitting $p \gg n$; see, for example, Theorem 3.1 in Koltchinskii [19] and Rio [26]. However, in order for this theory to work, x_i have to be well approximable in a Haar basis when considered as functions on the underlying probability space – e.g., $x_{ij} = f_{j,n}(u_i)$, where $f_{j,n}$ should have a total variation norm with respect to $u_i \sim U(0,1)^d$ (where d is fixed) that does not grow too quickly to enable the expansion in the Haar basis. This technique has been proven fruitful in many applications, but this requires a radically different structure than what our leading applications impose; instead our results, based upon Slepian-Stein methods, are more readily applicable in these settings (instead of controlling total variation bounds, they rely on control of maxima moments and moments of envelopes of $\{x_{ij}, j \leqslant p\}$). For further theoretical comparisons of the two methods in the context of strong approximations of suprema of non-Donsker emprical processes by those of Gaussian processes, in the classical kernel and series smoothing examples, we refer to our companion work [10] (there is no winner in terms of guaranteed rates of approximation, though side conditions seem to be weaker for the Slepian-Stein type methods; in particular Hungarian couplings often impose the boundedness conditions, e.g. $\|x_i\|_\infty \leqslant B_n$).

Supplementary Material III

Additional Application

APPENDIX M: ADAPTIVE SPECIFICATION TESTING

In this section, we study the problem of adaptive specification testing. Let $(v_i, y_i)_{i=1}^n$ be a sample of independent random pairs where y_i is a scalar dependent random variable, and $v_i \in \mathbb{R}^d$ is a vector of non-stochastic covariates. The null hypothesis, H_0, is that there exists $\beta \in \mathbb{R}^d$ such that

$$\mathrm{E}[y_i] = v_i'\beta;\ i = 1, \dots, n. \tag{13}$$

The alternative hypothesis, H_a, is that there is no β satisfying (13). We allow for triangular array asymptotics so that everything in the model may depend on n. For brevity, however, we omit index n.

Let $\varepsilon_i = y_i - \mathrm{E}[y_i]$, $i = 1, \dots, n$. Then $\mathrm{E}[\varepsilon_i] = 0$, and under H_0, $y_i = v_i'\beta + \varepsilon_i$. To test H_0, consider a set of test functions $P_j(v_i)$, $j = 1, \dots, p$. Let $z_{ij} = P_j(v_i)$. We choose test functions so that $\mathbb{E}_n[z_{ij}v_i] = 0$ and $\mathbb{E}_n[z_{ij}^2] = 1$ for all $j = 1, \dots, p$. In our analysis, p may be higher or even much higher than n. Let $\widehat{\beta} = (\mathbb{E}_n[v_i v_i'])^{-1}(\mathbb{E}_n[v_i y_i])$ be an OLS estimator of β, and let $\widehat{\varepsilon}_i = y_i - z_i'\widehat{\beta}$; $i = 1, \dots, n$ be corresponding residuals. Our test statistic is

$$T := \max_{1 \leqslant j \leqslant p} \frac{|\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2 \widehat{\varepsilon}_i^2]}}.$$

The test rejects H_0 if T is significantly large.

Note that since $\mathbb{E}_n[z_{ij}v_i] = 0$, we have

$$\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i/\sqrt{n} = \sum_{i=1}^n z_{ij}(\varepsilon_i + v_i'(\beta - \widehat{\beta}))/\sqrt{n} = \sum_{i=1}^n z_{ij}\varepsilon_i/\sqrt{n}.$$

Therefore, under H_0,

$$T = \max_{1 \leqslant j \leqslant p} \frac{|\sum_{i=1}^n z_{ij}\varepsilon_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2 \widehat{\varepsilon}_i^2]}}.$$

This suggests that we can use the multiplier bootstrap to obtain a critical value for the test. More precisely, let $(e_i)_{i=1}^n$ be a sequence of independent $N(0, 1)$ random variables that are independent of the data, and let

$$W := \max_{1 \leqslant j \leqslant p} \frac{|\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2 \widehat{\varepsilon}_i^2]}}.$$

The multiplier bootstrap critical value $c_W(1-\alpha)$ is the conditional $(1-\alpha)$-quantile of W given the data. To prove the validity of multiplier bootstrap, we will impose the following condition:

(S) There are some constants $c_1 > 0, C_1 > 0, \bar{\sigma}^2 > 0, \underline{\sigma}^2 > 0$, and a sequence $B_n \geqslant 1$ of constants such that for all $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$, $1 \leqslant k \leqslant d$: (i) $|z_{ij}| \leqslant B_n$; (ii) $\mathbb{E}_n[z_{ij}^2] = 1$; (iii) $\underline{\sigma}^2 \leqslant \mathrm{E}[\varepsilon_i^2] \leqslant \bar{\sigma}^2$; (iv) $|v_{ik}| \leqslant C_1$; (v) $d \leqslant C_1$; and (vi) the minimum eigenvalue of $\mathbb{E}_n[v_i v_i']$ is bounded from below by c_1.

THEOREM M.1 (Size Control of Adaptive Specification Test). *Let $c_2 > 0$ be some constant. Suppose that condition (S) is satisfied. Moreover, suppose that either*

(a) $\mathrm{E}[\varepsilon_i^4] \leqslant C_1$ *for all* $1 \leqslant i \leqslant n$ *and* $B_n^4(\log(pn))^7/n \leqslant C_1 n^{-c_2}$*; or*
(b) $\mathrm{E}[\exp(|\varepsilon_i|/C_1)] \leqslant 2$ *for all* $1 \leqslant i \leqslant n$ *and* $B_n^2(\log(pn))^7/n \leqslant C_1 n^{-c_2}$.

Then there exist constants $c > 0$ and $C > 0$, depending only on $c_1, c_2, C_1, \underline{\sigma}^2$ and $\bar{\sigma}^2$, such that under H_0, $|\mathrm{P}(T \leqslant c_W(1-\alpha)) - (1-\alpha)| \leqslant Cn^{-c}$.

COMMENT M.1. The literature on specification testing is large. In particular, [18] and [16] developed adaptive tests that are suitable for inference in L_2-norm. In contrast, our test is most suitable for inference in sup-norm. An advantage of our procedure is that selecting a wide class of test functions leads to a test that can effectively adapt to a wide range of alternatives, including those that can not be well-approximated by Hölder-continuous functions. ∎

PROOF OF THEOREM M.1. We only consider case (a). The proof for case (b) is similar and hence omitted. In this proof, let c, c', C, C' denote generic positive constants depending only on $c_1, c_2, C_1, \underline{\sigma}^2, \bar{\sigma}^2$ and their values may change from place to place. Let

$$T_0 := \max_{1 \leqslant j \leqslant p} \frac{|\sum_{i=1}^n z_{ij}\varepsilon_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2\sigma_i^2]}} \text{ and } W_0 := \max_{1 \leqslant j \leqslant p} \frac{|\sum_{i=1}^n z_{ij}\varepsilon_i e_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2\sigma_i^2]}}.$$

We make use of Corollary 3.1-(ii). To this end, we shall verify conditions (14) and (15) in Section 3 of the main text, which will be separately done in Steps 1 and 2, respectively.

Step 1. We show that $\mathrm{P}(|T - T_0| > \zeta_1) < \zeta_2$ for some ζ_1 and ζ_2 satisfying $\zeta_1\sqrt{\log p} + \zeta_2 \leqslant Cn^{-c}$.

By Corollary 2.1-(ii), we have

$$
\begin{aligned}
&\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\textstyle\sum_{i=1}^n z_{ij}\varepsilon_i/\sqrt{n}| > t\right)\\
&\quad\leqslant \mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\textstyle\sum_{i=1}^n z_{ij}\sigma_i e_i/\sqrt{n}| > t\right) + Cn^{-c},
\end{aligned}
$$

uniformly in $t\in\mathbb{R}$. By the Gaussian concentration inequality [Proposition A.2.1 30], for every $t>0$, we have

$$
\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\textstyle\sum_{i=1}^n z_{ij}\sigma_i e_i/\sqrt{n}| > \mathrm{E}[\displaystyle\max_{1\leqslant j\leqslant p}|\textstyle\sum_{i=1}^n z_{ij}\sigma_i e_i/\sqrt{n}|] + Ct\right) \leqslant e^{-t^2}.
$$

Since $\mathrm{E}[\max_{1\leqslant j\leqslant p}|\sum_{i=1}^n z_{ij}\sigma_i e_i/\sqrt{n}|]\leqslant C\sqrt{\log p}$, we conclude that

$$
\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\textstyle\sum_{i=1}^n z_{ij}\varepsilon_i/\sqrt{n}| > C\sqrt{\log(pn)}\right) \leqslant C'n^{-c}. \tag{14}
$$

Moreover,

$$
\begin{aligned}
\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2-\sigma_i^2)] &= \mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i-\varepsilon_i)^2] + \mathbb{E}_n[z_{ij}^2(\varepsilon_i^2-\sigma_i^2)] + 2\mathbb{E}_n[z_{ij}^2\varepsilon_i(\widehat{\varepsilon}_i-\varepsilon_i)]\\
&=: I_j + II_j + III_j.
\end{aligned}
$$

Consider I_j. We have

$$
I_j \leqslant_{(1)} \max_{1\leqslant i\leqslant n}(\widehat{\varepsilon}_i-\varepsilon_i)^2 \leqslant_{(2)} C\|\widehat{\beta}-\beta\|^2 \leqslant_{(3)} C'\|\mathbb{E}_n[v_i\varepsilon_i]\|^2,
$$

where (1) follows from assumption S-(ii), (2) from S-(iv) and S-(v), and (3) from S-(vi). Since $\mathrm{E}[\|\mathbb{E}_n[v_i\varepsilon_i]\|^2]\leqslant C/n$, by Markov's inequality, for every $t>0$,

$$
\mathrm{P}\left(\max_{1\leqslant j\leqslant p}\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i-\varepsilon_i)^2] > t\right) \leqslant C/(nt). \tag{15}
$$

Consider II_j. By Lemma A.1 and Markov's inequality, we have

$$
\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2(\varepsilon_i^2-\sigma_i^2)]| > t\right) \leqslant CB_n^2(\log p)/(\sqrt{n}t). \tag{16}
$$

Consider III_j. We have $|III_j|\leqslant 2|\mathbb{E}_n[z_{ij}^2 v_i'(\beta-\widehat{\beta})\varepsilon_i]|\leqslant 2\|\mathbb{E}_n[z_{ij}^2\varepsilon_i v_i]\|\|\widehat{\beta}-\beta\|$. Hence

$$
\begin{aligned}
&\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2\varepsilon_i(\widehat{\varepsilon}_i-\varepsilon_i)]| > t\right)\\
&\quad\leqslant \mathrm{P}\left(\max_{1\leqslant j\leqslant p}\|\mathbb{E}_n[z_{ij}^2\varepsilon_i v_i]\| > t\right) + \mathrm{P}(\|\widehat{\beta}-\beta\|>1)\\
&\quad\leqslant C[B_n^2(\log p)/(\sqrt{n}t) + 1/n]. \qquad (17)
\end{aligned}
$$

By (15)-(17), we have

$$\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2-\sigma_i^2)]|>t\right)\leqslant C[B_n^2(\log p)/(\sqrt{n}t)+1/(nt)+1/n]. \tag{18}$$

In particular,

$$\mathrm{P}\left(\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2-\sigma_i^2)]|>\underline{\sigma}^2/2\right)\leqslant Cn^{-c}.$$

Since $\mathbb{E}_n[z_{ij}^2\sigma_i^2]\geqslant\underline{\sigma}^2>0$ (which is guaranteed by S-(iii) and S-(ii)), on the event $\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2-\sigma_i^2)]|\leqslant\underline{\sigma}^2/2$, we have

$$\min_{1\leqslant j\leqslant p}\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2]\geqslant\min_{1\leqslant j\leqslant p}\mathbb{E}_n[z_{ij}^2\sigma_i^2]-\underline{\sigma}^2/2\geqslant\underline{\sigma}^2/2,$$

and hence

$$\begin{aligned}|T-T_0|&=\max_{1\leqslant j\leqslant p}\left|\frac{\sqrt{\mathbb{E}_n[z_{ij}^2\sigma_i^2]}-\sqrt{\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2]}}{\sqrt{\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2]}}\right|\times T_0\\
&\leqslant C\max_{1\leqslant j\leqslant p}\left|\sqrt{\mathbb{E}_n[z_{ij}^2\sigma_i^2]}-\sqrt{\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2]}\right|\times T_0\\
&\leqslant C\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2\sigma_i^2]-\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2]|\times T_0,\end{aligned}$$

where the last step uses the simple fact that

$$|\sqrt{a}-\sqrt{b}|=\frac{|a-b|}{\sqrt{a}+\sqrt{b}}\leqslant\frac{|a-b|}{\sqrt{a}}.$$

By (14) and (18), for every $t>0$,

$$\mathrm{P}\left(|T-T_0|>Ct\sqrt{\log(pn)}\right)\leqslant C'[n^{-c}+B_n^2(\log p)/(\sqrt{n}t)+1/(nt)].$$

By choosing $t=(\log(pn))^{-1}n^{-c'}$ with sufficiently small $c'>0$, we obtain the claim of this step.

Step 2. We show that $\mathrm{P}(\mathrm{P}_e(|W-W_0|>\zeta_1)>\zeta_2)<\zeta_2$ for some ζ_1 and ζ_2 satisfying $\zeta_1\sqrt{\log p}+\zeta_2\leqslant Cn^{-c}$.

For $0<t\leqslant\underline{\sigma}^2/2$, consider the event

$$\mathcal{E}=\left\{(\varepsilon_i)_{i=1}^n:\max_{1\leqslant j\leqslant p}|\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2-\sigma_i^2)]|\leqslant t,\max_{1\leqslant i\leqslant p}(\widehat{\varepsilon}_i-\varepsilon_i)^2\leqslant t^2\right\}.$$

By calculations in Step 1, $\mathrm{P}(\mathcal{E}) \geqslant 1 - C[B_n^2(\log p)/(\sqrt{n}t) + 1/(nt^2) + 1/n]$. We shall show that, on this event,

$$\mathrm{P}_e\left(\max_{1\leqslant j\leqslant p} |\textstyle\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}| > C\sqrt{\log(pn)}\right) \leqslant n^{-1}, \tag{19}$$

$$\mathrm{P}_e\left(\max_{1\leqslant j\leqslant p} |\textstyle\sum_{i=1}^n z_{ij}(\widehat{\varepsilon}_i - \varepsilon_i)e_i/\sqrt{n}| > Ct\sqrt{\log(pn)}\right) \leqslant n^{-1}. \tag{20}$$

For (19), by the Gaussian concentration inequality, for every $s > 0$,

$$\mathrm{P}_e\left(\max_{1\leqslant j\leqslant p} |\textstyle\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}| > \mathrm{E}_e[\max_{1\leqslant j\leqslant p} |\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}|] + Cs\right) \leqslant e^{-s^2}.$$

where we have used the fact $\mathbb{E}_n[z_{ij}^2\widehat{\varepsilon}_i^2] = \mathbb{E}_n[z_{ij}^2\sigma_i^2] + \mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2 - \sigma_i^2)] \leqslant \bar{\sigma}^2 + t \leqslant \bar{\sigma}^2 + \underline{\sigma}^2/2$ on the event $\mathcal{E}$. Here $\mathrm{E}_e[\cdot]$ means the expectation with respect to $(e_i)_{i=1}^n$ conditional on $(\varepsilon_i)_{i=1}^n$. Moreover, on the event $\mathcal{E}$,

$$\mathrm{E}_e[\max_{1\leqslant j\leqslant p} |\textstyle\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}|] \leqslant C\sqrt{\log p}.$$

Hence by choosing $s = \sqrt{\log n}$, we obtain (19). Inequality (20) follows similarly, by noting that $(\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i - \varepsilon_i)^2])^{1/2} \leqslant \max_{1\leqslant i\leqslant n} |\widehat{\varepsilon}_i - \varepsilon_i| \leqslant t$ on the event $\mathcal{E}$.

Define

$$W_1 := \max_{1\leqslant j\leqslant p} \frac{|\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}|}{\sqrt{\mathbb{E}_n[z_{ij}^2\sigma_i^2]}}.$$

Note that $\mathbb{E}_n[z_{ij}^2\sigma_i^2] \geqslant \underline{\sigma}^2$. Since on the event $\mathcal{E}$, $\max_{1\leqslant j\leqslant p} |\mathbb{E}_n[z_{ij}^2(\widehat{\varepsilon}_i^2 - \sigma_i^2)]| \leqslant t \leqslant \underline{\sigma}^2/2$, in view of Step 1, on this event, we have

$$\begin{aligned}
|W - W_0| &\leqslant |W - W_1| + |W_1 - W_0| \\
&\leqslant CtW_1 + |W_1 - W_0| \\
&\leqslant Ct \max_{1\leqslant j\leqslant p} |\textstyle\sum_{i=1}^n z_{ij}\widehat{\varepsilon}_i e_i/\sqrt{n}| + C \max_{1\leqslant j\leqslant p} |\sum_{i=1}^n z_{ij}(\widehat{\varepsilon}_i - \varepsilon_i)e_i/\sqrt{n}|.
\end{aligned}$$

Therefore, by (19) and (20), on the event $\mathcal{E}$, we have

$$\mathrm{P}_e\left(|W - W_0| > Ct\sqrt{\log(pn)}\right) \leqslant 2n^{-1}.$$

By choosing $t = (\log(pn))^{-1}n^{-c}$ with sufficiently small $c > 0$, we obtain the claim of this step.

Step 3. Steps 1 and 2 verified conditions (14) and (15) in Section 3 of the main text. Theorem M.1 case (a) follows from Corollary 3.1-(ii). ∎

REFERENCES

[1] Asriev, A.V. and Rotar, V.I. (1989). On the convergence rate in the infinite-dimensional central limit theorem for probabilities of hitting parallelepipeds. *Theory of Probability and Its Applications* **30** 691-701.

[2] Ball, K. (1993). The reverse isoperimetric problem for Gaussian measure. *Discrete Comput. Geom.* **10** 411-420.

[3] Bentkus, V. (1990). Smooth approximations of the norm and differentiable functions with bounded support in Banach space l^k_∞. *Lith. Math. Journal* **30** 223-230.

[4] Bentkus, V. (2003). On the dependence of the Berry-Esseen bound on dimension. *J. Statist. Plann. Infer.* **113** 385-402.

[5] Bentkus, V., Götze, F., Paulauskas, V., and Rackauskus, L. (2000). The accuracy of Gaussian approximation in Banach spaces. In: *Limit Theorems of Probability Theory* (Eds. Y.V. Prokholov and V. Statulevičius), Springer, pp. 25-111.

[6] Chatterjee, S. (2005a). A simple invariance theorem. arXiv:math/0508213.

[7] Chatterjee, S. (2006). A generalization of Lindeberg's principle. *Ann. Probab.* **34** 2061-2076.

[8] Chatterjee, S. and Meckes, E. (2008). Multivariate normal approximation using exchangeable pairs. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** 257-283.

[9] Chen, L., and Fang, X. (2011). Multivariate normal approximation by Stein's method: the concentration inequality approach. arXiv:1111.4073.

[10] Chernozhukov, V., Chetverikov, D. and Kato, K. (2012a). Gaussian approximation of suprema of empirical processes. arXiv:1212.6906.

[11] Chernozhukov, V., Chetverikov, D. and Kato, K. (2012b). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. arXiv:1301.4807. Submitted to *Probab. Theory Related Fields*.

[12] de la Peña, V., Lai, T. and Shao, Q.-M. (2009). *Self-Normalized Processes: Limit Theory and Statistical Applications*. Springer.

[13] Dudley, R.M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press.

[14] Goldstein, L., and Rinott, Y. (1996) Multivariate normal approximations by Stein's method and size bias couplings. *J. Appl. Probab.* **33** 1-17.

[15] Götze, F. (1991). On the rate of convergence in the multivariate CLT. *Ann. Probab.* **19** 724-739.

[16] Guerre, E. and Lavergne, P. (2005). Data-driven rate-optimal specification testing in regression models. *Ann. Statist.* **33** 840-870.

[17] Harsha, P., Klivans, A., and Meka, R. (2012). An invariance principle for polytopes. *Journal of the ACM* **59**.

[18] Horowitz, J. L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599-631.

[19] Koltchinskii, V.I. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoret. Probab.* **7** 73-118.

[20] Lindeberg, J. W. (1922). Eine neue herleitung des exponentialgesetzes in der wahrscheinlichkeitsrechnung. *Math. Z.* **15** 211-225.

[21] Nagaev, S. (1976). An estimate of the remainder term in the multidimensional central limit theorem. *Proc. Third Japan-USSR Symp. Probab. Theory.* Lecture Notes in Math. pp. 419-438.

[22] Norvaisa, R. and Paulauskas, V. (1991). Rate of convergence in the Central Limit Theorem for empirical processes. *Journal of Theoretical Probability.* **4** 511-534.

[23] Nazarov, F. (2003). On the maximal perimeter of a convex set in $\mathbb{R}^n$ with respect to

a Gaussian measure. In: *Geometric Aspects of Functional Analysis*, Lecture Notes in Mathematics Vol. 1807/2003, Springer, pp. 169-187.

[24] Portnoy, S. (1986). On the central limit theorem in $\mathbb{R}^p$ when $p \to \infty$. *Probab. Theory Related Fields* **73** 571-583.

[25] Reinert, G. and Röllin, A. (2009). Multivariate normal approximation with Stein's method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37** 2150-2173.

[26] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probab. Theory Related Fields* **98** 21-45.

[27] Röllin, A. (2011). Stein's method in high dimensions with applications, *Ann. Inst. H. Poincaré Probab. Statist.* **49**, 529-549.

[28] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135-1151.

[29] Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians*. Springer.

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

Department of Economics and
Operations Research Center, MIT
50 Memorial Drive
Cambridge, MA 02142, USA.
E-mail: vchern@mit.edu

Department of Economics, UCLA
Bunche Hall, 8283
315 Portola Plaza
Los Angeles, CA 90095, USA.
E-mail: chetverikov@econ.ucla.edu

Graduate School of Economics
University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 113-0033, Japan.
E-mail: kkato@e.u-tokyo.ac.jp