Supplement to "Simultaneous Inference for High-dimensional Linear Models"

Xianyang Zhang* and Guang Cheng†

Texas A&M University and Purdue University

February 22, 2016

This supplementary material provides proofs of the main results in the paper as well as some additional numerical results.

1 Technical details

We first present two lemmas that will be used in the rest proofs. Define $\xi_{ij} = \Theta_j^T \widetilde{X}_i \epsilon_i$. Denote by c, c', C, C', C_i be some generic constants which can be different from line to line.

Lemma **1.1**. *Under Assumptions 2.1-2.3, we have for any $G \subseteq \{1, 2, \ldots, p\}$,*

$$\sup_{x \in \mathbb{R}} \left| P\left(\max_{j \in G} \sum_{i=1}^{n} \xi_{ij}/\sqrt{n} \le x \right) - P\left(\max_{j \in G} \sum_{i=1}^{n} z_{ij}/\sqrt{n} \le x \right) \right| \lesssim n^{-c'}, \quad c' > 0,$$

where $\{z_i = (z_{i1}, \ldots, z_{ip})'\}$ is a sequence of mean zero independent Gaussian vector with $\mathbb{E} z_i z_i' = \Theta_j^T \Sigma \Theta_j \sigma_\epsilon^2$.

*Assistant Professor, Department of Statistics, Texas A&M University, College Station, TX 77843. E-mail: zhangxiany@stat.tamu.edu.

†Associate Professor, Department of Statistics, Purdue University, West Lafayette, IN 47906. E-mail: chengg@purdue.edu. Tel: +1 (765) 496-9549. Fax: +1 (765) 494-0558. Research Sponsored by NSF CAREER Award DMS-1151692, DMS-1418042, Simons Fellowship in Mathematics, Office of Naval Research (ONR N00014-15-1-2331) and a grant from Indiana Clinical and Translational Sciences Institute. Guang Cheng was on sabbatical at Princeton while part of this work was carried out; he would like to thank the Princeton ORFE department for its hospitality and support.

Proof of Lemma 1.1. We apply Corollary 2.1 of Chernozhukov et al. (2013) to the sequence $\{\xi_{ij}\}$ by verifying its Condition (E.1). For the sake of clarity, we state the condition below, i.e.

$$c_0 \le \mathbb{E}\xi_{ij}^2 \le C_0, \quad \max_{k=1,2} \mathbb{E}|\xi_{ij}|^{2+k}/B^k + \mathbb{E}\exp(|\xi_{ij}|/B) \le 4, \tag{1}$$

uniformly over j, where $c_0, C_0 > 0$, and B is some large enough constant. In what follows, we consider two cases for $\mathbf{X}$: (i) $\mathbf{X}$ has i.i.d. sub-Gaussian rows; (ii) $\mathbf{X}$ is strongly bounded.

(i) By Assumption 2.2, $\mathbb{E}(\Theta_j^T \widetilde{X}_i)^2 = \Theta_j^T \Sigma \Theta_j = \theta_{jj} := 1/\tau_j^2$, and $1/c < \Lambda_{\min}^2 \le \tau_j^2 \le \Sigma_{j,j} = C$, for some constants $c, C > 0$. Recall that $\Lambda_{\min}^2$ is the minimal eigenvalue of Σ. Thus we have $c_1\sigma_\epsilon^2 \le \mathbb{E}\xi_{ij}^2 \le C_1\sigma_\epsilon^2$. By the independence between $\{\widetilde{X}_i\}$ and $\{\epsilon_i\}$, we have for large enough C and uniformly for all j,

$$\begin{aligned}\mathbb{E}\exp(|\xi_{ij}|/C) =& 1 + \sum_{k=1}^{+\infty} \frac{\mathbb{E}|\xi_{ij}|^k}{C^k k!} = 1 + \sum_{k=1}^{+\infty} \frac{\mathbb{E}|\Theta_j^T \widetilde{X}_i|^k \mathbb{E}|\epsilon_i|^k}{C^k k!} \\ \le& 1 + \sum_{k=1}^{+\infty} \frac{k^k}{(C')^k k!} \le 1 + \sum_{k=1}^{+\infty} (e/C')^k < \infty,\end{aligned}$$

where we have used the fact that $k! \ge (k/e)^k$, $||\Theta_j||_2 \lesssim \Lambda_{\min}^{-1} = O(1)$ (because $||\Theta_j||_2^2 \Lambda_{\min}^2 \le c$) and $\mathbb{E}|X|^k \le (C'')^k k^{k/2}$ with C'' being some positive constant for sub-Gaussian variable X. Thus we have $\max_{k=1,2} \mathbb{E}|\xi_{ij}|^{2+k}/B^k + \mathbb{E}\exp(|\xi_{ij}|/B) \le 4$ uniformly for some large enough constant B.

(ii) In the strongly bounded case, using the fact that $||\Theta_j||_2^2 \lesssim \Lambda_{\min}^{-2} = O(1)$ and $||\Theta_j||_1 \le \sqrt{s_j}||\Theta_j||_2$, we have $|\Theta_j^T \widetilde{X}_i| \le ||\Theta_j||_1 ||\widetilde{X}_i||_\infty \le K_n \sqrt{s_j} ||\Theta_j||_2$. It is straightforward to verify that $\max_{k=1,2} \mathbb{E}|\xi_{ij}|^{2+k}/B_n^k + \mathbb{E}\exp(|\xi_{ij}|/B_n) \le 4$ uniformly with some $B_n \asymp K_n \max_j \sqrt{s_j}$ and $B_n^2(\log(pn))^7/n \le C_2 n^{-c_2}$ under part (ii) of Assumption 2.3. $\diamondsuit$

Remark **1.1**. The conclusion in Lemma 1.1 still holds if we assume that (i) $\max_{i,j}|X_{ij}| \le K_n$ with $\max_{1\le j\le p} s_j^2 K_n^4 (\log(pn))^7/n \le C_1 n^{-c_1}$ for some constants $c_1, C_1 > 0$; and (ii) $\{\epsilon_i\}$ are i.i.d with with $\mathbb{E}|\epsilon_i|^4 < \infty$ and $c' < \sigma_\epsilon^2$ for $c' > 0$.

Next we quantify the effect by replacing ξ_i with $\widehat{\xi}_i$.

LEMMA **1.2**. *Suppose Assumptions 2.1-2.3 hold. Assume* $\max_j K_0^2 s_j^2(\log(pn))^3(\log(n))^2/n = o(1)$. *Recall that* $K_0 = 1$ *in the sub-Gaussian case and* $K_0 = K_n$ *in the strongly bounded case. Then with* $\lambda_j \asymp K_0\sqrt{\log(p)/n}$ *uniformly for* j, *there exist* $\zeta_1, \zeta_2 > 0$ *such that*

$$P\left(\max_{1\le j\le p}\left|\sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n} - \sum_{i=1}^n \xi_{ij}/\sqrt{n}\right| \ge \zeta_1\right) < \zeta_2,$$

where $\zeta_1\sqrt{1\vee\log(p/\zeta_1)} = o(1)$ *and* $\zeta_2 = o(1)$.

Proof of Lemma 1.2. Let $\widetilde{K}_0 = \log(np)\log(n)$ in the sub-Gaussian case and $\widetilde{K}_0 = K_n\log(n)$ in the strongly bounded case. Using Lemma A.1 in Chernozhukov et al. (2013), we deduce that

$$\begin{aligned}
\mathbb{E}\left\{\max_{1\le j\le p}\left|\sum_{i=1}^n X_{ij}\epsilon_i/n\right|\right\} \lesssim & \sigma_\epsilon\sqrt{\max_j \Sigma_{j,j}}\sqrt{\log(p)/n} + \sqrt{\mathbb{E}\max_{i,j}|X_{ij}\epsilon_i|^2}\log(p)/n \\
\lesssim & \sqrt{\log(p)/n} + \sqrt{\mathbb{E}\max_{i,j}X_{ij}^2}\sqrt{\mathbb{E}\max_i \epsilon_i^2}\log(p)/n \\
\lesssim & \sqrt{\log(p)/n} + \widetilde{K}_0\log(p)/n,
\end{aligned}$$

where we have used the fact that $\sqrt{\mathbb{E}\max_i \epsilon_i^2} \lesssim \log(n)\max_{1\le i\le n}||\epsilon_i||_{\psi_1} \lesssim \log n$ with $\psi_1(x) = \exp(x) - 1$ and $||\cdot||_{\psi_1}$ being the corresponding Orlicz norm, and similar result for $\sqrt{\mathbb{E}\max_{i,j}X_{i,j}^2}$ (see Lemma 2.2.2 in van der Vaart and Wellner 1996). Because $||\widehat{\Theta}_j - \Theta_j||_1 = O_P(K_0 s_j\sqrt{\log(p)/n})$ uniformly for j, we obtain,

$$\begin{aligned}
\left|\sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n} - \sum_{i=1}^n \xi_{ij}/\sqrt{n}\right| = & \left|(\widehat{\Theta}_j^T - \Theta_j^T)\sum_{i=1}^n \widetilde{X}_i\epsilon_i/\sqrt{n}\right| \le ||\widehat{\Theta}_j - \Theta_j||_1\left|\left|\sum_{i=1}^n \widetilde{X}_i\epsilon_i/\sqrt{n}\right|\right|_\infty \\
= & O_P\left(K_0 s_j\sqrt{\log(p)/n}\left|\left|\sum_{i=1}^n \widetilde{X}_i\epsilon_i/\sqrt{n}\right|\right|_\infty\right) \\
= & O_P\left(K_0 s_j\log(p)/\sqrt{n} + \sqrt{n}K_0\widetilde{K}_0 s_j(\log(p)/n)^{3/2}\right) \\
\le & O_P\left(\max_j s_j K_0\log(p)/\sqrt{n}\right),
\end{aligned}$$

uniformly for all j. Choosing ζ_1 such that $\max_j K_0 s_j\log(p)/(\sqrt{n}\zeta_1) = o(1)$ and $\zeta_1\sqrt{1\vee\log(p/\zeta_1)} =$

$o(1)$ (e.g. $\zeta_1^2 = O(\max_j K_0 s_j \sqrt{\log(p)/n})$), we deduce that

$$P\left(\max_{1\le j\le p}\left|\sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n} - \sum_{i=1}^n \xi_{ij}/\sqrt{n}\right| \ge \zeta_1\right) < \zeta_2, \quad \zeta_2 = o(1).$$

$\diamondsuit$

Remark **1.2**. With a more delicate analysis, one can specify the order of ζ_2 in Lemma 1.2; see e.g., Theorem 6.1 and Lemma 6.2 of Bühlmann and van de Geer (2011).

Proof of Theorem 2.2. Without loss of generality, we set $G = \{1, 2, \ldots, p\}$. Define

$$T_G = \max_{j\in G} \sqrt{n}(\breve{\beta}_j - \beta_j^0), \quad T_{0,G} = \max_{j\in G} \sum_{i=1}^n \xi_{ij}/\sqrt{n}.$$

Let $\pi(v) = C_2 v^{1/3}(1 \vee \log(p/v))^{2/3}$ with $C_2 > 0$, and

$$\Gamma = \max_{1\le j,k\le p} |\widehat{\sigma}_\epsilon^2 \widehat{\Theta}_j^T \widehat{\Sigma} \widehat{\Theta}_k - \sigma_\epsilon^2 \Theta_j^T \Sigma \Theta_k|, \quad \widehat{\Sigma} = \mathbf{X}^T\mathbf{X}/n.$$

Notice that

$$|T_G - T_{0,G}| \le \max_{1\le j\le p}\left|\sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n} - \sum_{i=1}^n \xi_{ij}/\sqrt{n}\right| + ||\Delta||_\infty.$$

By similar arguments in the proof of Theorem 2.4 of van de Geer et al. (2014) and the results in Theorem 2.1, we have

$$||\Delta||_\infty \le ||\widehat{\beta} - \beta^0||_1 \max_j \sqrt{n}\lambda_j/\widehat{\tau}_j^2 = O_P(K_0\sqrt{\log(p)}||\widehat{\beta} - \beta^0||_1) = O_P(K_0^2 s_0 \log(p)/\sqrt{n}),$$

where we use the fact that $\max_j \lambda_j/\widehat{\tau}_j^2 = O_P(K_0\sqrt{\log(p)/n})$ and $||\widehat{\beta} - \beta^0||_1 = O_P(s_0\lambda)$ with $\lambda = O(K_0\sqrt{\log(p)/n})$. Thus by Lemma 1.2 and the assumption that $K_0^4 s_0^2(\log(p))^3/n = o(1)$, we have

$$P(|T_G - T_{0,G}| > \zeta_1) < \zeta_2,$$

for $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$.

Let $c_{z,G}(\alpha) = \inf\{t \in \mathbb{R} : P(\max_{j\in G} \sum_{i=1}^n z_{ij}/\sqrt{n} \le t) \ge 1 - \alpha\}$, where the sequence $\{z_{ij}\}$ is defined in Lemma 1.1. Following the arguments in the proof of Lemma 3.2 in Chernozhukov et al. (2013), we have

$$P(c_G(\alpha) \le c_{z,G}(\alpha + \pi(v))) \ge 1 - P(\Gamma > v), \tag{2}$$

$$P(c_{z,G}(\alpha) \le c_G(\alpha + \pi(v))) \ge 1 - P(\Gamma > v). \tag{3}$$

By Lemma 1.1, (2) and (3), we have for every $v > 0$,

$$\begin{aligned}\sup_{\alpha \in (0,1)} \left| P\left(T_{0,G} > c_G(\alpha)\right) - \alpha \right| \lesssim & \sup_{\alpha \in (0,1)} \left| P\left(\max_{j \in G} \sum_{i=1}^{n} z_{ij}/\sqrt{n} > c_G(\alpha) \right) - \alpha \right| + n^{-c'} \\ \lesssim & \pi(v) + P(\Gamma > v) + n^{-c'}.\end{aligned}$$

Moreover, by the arguments in the proof of Theorem 3.2 in Chernozhukov et al. (2013), we have

$$\sup_{\alpha \in (0,1)} \left| P\left(T_G > c_G(\alpha)\right) - \alpha \right| \lesssim \pi(v) + P(\Gamma > v) + n^{-c'} + \zeta_1 \sqrt{1 \vee \log(p/\zeta_1)} + \zeta_2.$$

By Lemma 5.3 and Lemma 5.4 of van de Geer et al. (2014), we have

$$\max_{1 \le j,k \le p} |\widehat{\Theta}_j^T \widehat{\Sigma} \widehat{\Theta}_k - \Theta_j^T \Sigma \Theta_k| = O_P(\max_j \lambda_j \sqrt{s_j}).$$

Since $|\Theta_j^T \Sigma \Theta_k| \le 1/(\tau_j \tau_k) = O(1)$ uniformly for $1 \le j, k \le p$, we have

$$\Gamma = O_P\left(|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| + \max_j \lambda_j \sqrt{s_j} \right).$$

Under Assumption 2.4, choosing $v = 1/(\alpha_n (\log(p))^2)$, we deduce that

$$\sup_{\alpha \in (0,1)} \left| P(\max_{1 \le j \le p} \sqrt{n}(\breve{\beta}_j - \beta_j^0) > c_G(\alpha)) - \alpha \right| = o(1),$$

which completes the proof. $\diamondsuit$

Proof of Theorem 2.3. From the arguments in the proof of Theorem 2.2, we have

$$\Gamma = \max_{1\le j,k\le p} |\widehat{\sigma}_\epsilon^2 \widehat{\Theta}_j^T \widehat{\Sigma} \widehat{\Theta}_k - \sigma_\epsilon^2 \Theta_j^T \Sigma \Theta_k| = O_P\left(|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| + \max_j \lambda_j \sqrt{s_j}\right),$$

which implies that $\max_{1\le j\le p} |\widehat{\omega}_{jj} - \omega_{jj}| = O_P\left(|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| + \max_j \lambda_j \sqrt{s_j}\right)$ with $\omega_{jj} = \sigma_\epsilon^2 \theta_{jj}$. We then have

$$P(\omega_{jj}/2 < \widehat{\omega}_{jj} < 2\omega_{jj} \text{ for all } 1 \le j \le p) \to 1. \tag{4}$$

The fact that $1/c < \Lambda_{\min}^2 \le \tau_j^2 = 1/\theta_{jj} \le \Sigma_{j,j} = C$ implies that ω_{jj} is uniformly bounded away from zero and infinity.

Define $\bar{T}_G = \max_{j\in G} \sqrt{n}(\breve{\beta}_j - \beta_j^0)/\sqrt{\widehat{\omega}_{jj}}$ and $\bar{T}_{0,G} = \max_{j\in G} \sum_{i=1}^n \xi_{ij}/\sqrt{n\omega_{jj}}$. Denote by $\Delta = (\Delta_1, \ldots, \Delta_p)^T$ and $\bar{\Delta} = (\bar{\Delta}_1, \ldots, \bar{\Delta}_p)^T$ with $\bar{\Delta}_j = \Delta_j/\sqrt{\widehat{\omega}_{jj}}$. Then we have

$$\begin{aligned}
&|\bar{T}_G - \bar{T}_{0,G}| \\
\le& \max_{1\le j\le p} \left| \sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n\widehat{\omega}_{jj}} - \sum_{i=1}^n \xi_{ij}/\sqrt{n\omega_{jj}} \right| + ||\bar{\Delta}||_\infty \\
\le& \max_{1\le j\le p} \left| \sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n\widehat{\omega}_{jj}} - \sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n\omega_{jj}} \right| + \max_{1\le j\le p} \left| \sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n\omega_{jj}} - \sum_{i=1}^n \xi_{ij}/\sqrt{n\omega_{jj}} \right| + ||\bar{\Delta}||_\infty \\
\le& C' \max_{1\le j\le p} \left| \sum_{i=1}^n \widehat{\xi}_{ij}/\sqrt{n} \right| \max_{1\le j\le p} \left| \sqrt{\omega_{jj}/\widehat{\omega}_{jj}} - 1 \right| + C'' \max_{1\le j\le p} \left| \sum_{i=1}^n (\widehat{\xi}_{ij} - \xi_{ij})/\sqrt{n} \right| + ||\bar{\Delta}||_\infty, \\
=& I_1 + I_2 + I_3,
\end{aligned}$$

where $C', C'' > 0$.

On the event $\omega_{jj}/2 < \widehat{\omega}_{jj} < 2\omega_{jj}$ for all $1 \leq j \leq p$,

$$
\begin{aligned}
\max_{1\leq j\leq p}\left|\sqrt{\omega_{jj}/\widehat{\omega}_{jj}}-1\right| \leq & \max_{1\leq j\leq p}|\sqrt{\omega_{jj}}-\sqrt{\widehat{\omega}_{jj}}|\max_{1\leq j\leq p}\sqrt{2/\omega_{jj}} \\
\leq & \max_{1\leq j\leq p}\left|\frac{\omega_{jj}-\widehat{\omega}_{jj}}{\sqrt{\omega_{jj}}+\sqrt{\widehat{\omega}_{jj}}}\right|\max_{1\leq j\leq p}\sqrt{2/\omega_{jj}} \\
\leq & \max_{1\leq j\leq p}|\omega_{jj}-\widehat{\omega}_{jj}|\max_{1\leq j\leq p}1/\omega_{jj} \\
= & O_P\left(|\widehat{\sigma}_\epsilon^2-\sigma_\epsilon^2|+\max_j\lambda_j\sqrt{s_j}\right).
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\max_{1\leq j\leq p}\left|\sum_{i=1}^n\widehat{\xi}_{ij}/\sqrt{n}\right| \leq & \max_{1\leq j\leq p}\left|\sum_{i=1}^n(\widehat{\xi}_{ij}-\xi_{ij})/\sqrt{n}\right|+\max_{1\leq j\leq p}\left|\sum_{i=1}^n\xi_{ij}/\sqrt{n}\right| \\
= & O_P(\sqrt{\log(p)}+\max_j\sqrt{s_j}\widetilde{K}_0\log(p)/\sqrt{n})=O_P(\sqrt{\log p}),
\end{aligned}
$$

where $\widetilde{K}_0=\log(np)\log(n)$ in the sub-Gaussian case and $\widetilde{K}_0=K_n\log(n)$ in the strongly bounded case. Therefore, on the above event, $I_1 \leq O_P\left(\sqrt{\log(p)}|\widehat{\sigma}_\epsilon^2-\sigma_\epsilon^2|+\sqrt{\log(p)}\max_j\lambda_j\sqrt{s_j}\right)$. Under Assumption 2.4, we can find ζ_1' such that $P(I_1>\zeta_1')=o(1)$ and $\zeta_1'\sqrt{1\vee\log(p/\zeta_1')}=o(1)$. Using the fact that $||\Delta||_\infty \leq O_P(K_0^2 s_0\log(p)/\sqrt{n})$, we can prove the same result for $||\bar{\Delta}||_\infty$ conditional on the event $\{\omega_{jj}/2<\widehat{\omega}_{jj}<2\omega_{jj}$ for all $1\leq j\leq p\}$. Thus by Lemma 1.2 and (4), we have

$$
P(|\bar{T}_G-\bar{T}_{0,G}|>\zeta_1)\leq P(I_1+I_2+I_3>\zeta_1)<\zeta_2,
$$

for $\zeta_1\sqrt{1\vee\log(p/\zeta_1)}=o(1)$ and $\zeta_2=o(1)$.

Let $\bar{\Gamma}=\max_{1\leq j,k\leq p}|\widehat{\sigma}_\epsilon^2\widehat{\Theta}_j^T\widehat{\Sigma}\widehat{\Theta}_k/\sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}-\sigma_\epsilon^2\Theta_j^T\Sigma\Theta_k/\sqrt{\omega_{jj}\omega_{kk}}|$. Note that

$$
|\sqrt{\omega_{jj}\omega_{kk}}-\sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}|=\frac{|\omega_{jj}\omega_{kk}-\widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}}+\sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}}.
$$

On the event $\omega_{jj}/2 < \widehat{\omega}_{jj} < 2\omega_{jj}$ for all $1 \le j \le p$, we have

$$\frac{|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}} + \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}} \le \frac{|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}|}{\sqrt{\omega_{jj}\omega_{kk}} + \sqrt{\omega_{jj}\omega_{kk}/4}} \le (2/3)|\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}| \max_{1\le j\le p} 1/\omega_{jj},$$

which implies that

$$\begin{aligned}\max_{1\le j,k\le p} |\sqrt{\omega_{jj}\omega_{kk}/\widehat{\omega}_{jj}\widehat{\omega}_{kk}} - 1| \le &\max_{1\le j,k\le p} |\sqrt{\omega_{jj}\omega_{kk}} - \sqrt{\widehat{\omega}_{jj}\widehat{\omega}_{kk}}| \max_{1\le j\le p} 2/\omega_{jj} \\ \le &(4/3) \max_{1\le j,k\le p} |\omega_{jj}\omega_{kk} - \widehat{\omega}_{jj}\widehat{\omega}_{kk}| \max_{1\le j\le p} 1/\omega_{jj}^2 \\ = &O_P\left(|\widehat{\sigma}_\epsilon^2 - \sigma_\epsilon^2| + \max_j \lambda_j \sqrt{s_j}\right).\end{aligned}$$

Using similar arguments above, we can show that $P(\bar{\Gamma} > v) = o(1)$ for $v = 1/(\alpha_n(\log(p))^2)$. The rest of the proofs is similar to those in the proof of Theorem 2.2. We skip the details $\diamondsuit$

Proof of Theorem 2.4. Define $\widetilde{T}_G = \max_{j\in G} |\sqrt{n}(\breve{\beta}_j - \beta_j^0)/\sqrt{\widehat{\omega}_{jj}}|$ and $\widetilde{T}_{0,G} = \max_{j\in G} \sum_{i=1}^n |\xi_{ij}/\sqrt{n\omega_{jj}}|$. Under the assumptions in Theorem 2.3, we can show that $P(|\widetilde{T}_G - \widetilde{T}_{0,G}| > \zeta_1) < \zeta_2$ for $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$. In another word, the distribution of $\max_{j\in G} \sqrt{n}|\breve{\beta}_j - \beta_0|/\sqrt{\widehat{\omega}_{jj}}$ can be approximated by $\max_{j\in G} |Z_j|$ with $Z = (Z_1, \ldots, Z_p) \sim^d N(0, \widetilde{\Theta})$. Under Assumption 2.5, by Lemma 6 of Cai et al. (2014), we have for any $x \in \mathbb{R}$ and as $|G| \to +\infty$,

$$P\left(\max_{j\in G} |Z_i|^2 - 2\log(|G|) + \log\log(|G|) \le x\right) \to F(x) := \exp\left\{-\frac{1}{\sqrt{\pi}} \exp\left(-\frac{x}{2}\right)\right\}.$$

It implies that

$$P\left(\max_{j\in G} n|\breve{\beta}_j - \beta_j^0|^2/\widehat{\omega}_{jj} \le 2\log(|G|) - \log\log(|G|)/2\right) \to 1. \tag{5}$$

The bootstrap consistency result implies that

$$|(\bar{c}_G^*(\alpha))^2 - 2\log(|G|) + \log\log(|G|) - q_\alpha| = o_P(1), \tag{6}$$

where q_α is the $100(1-\alpha)$th quantile of $F(x)$. Consider any $j^* \in G$ such that $|\widetilde{\beta}_{j^*} - \beta_{j^*}^0|/\sqrt{\omega_{j^*j^*}} >$

$(\sqrt{2}+\varepsilon_0)\sqrt{(\log|G|)/n}$. Using the inequality $2a_1a_2 \le \delta^{-1}a_1^2 + \delta a_2^2$ for any $\delta > 0$, we have

$$n|\widetilde{\beta}_{j^*} - \beta^0_{j^*}|^2/\widehat{\omega}_{j^*j^*} \le (1+\delta^{-1})n|\breve{\beta}_{j^*} - \beta^0_{j^*}|^2/\widehat{\omega}_{j^*j^*} + (1+\delta)n|\breve{\beta}_{j^*} - \widetilde{\beta}_{j^*}|^2/\widehat{\omega}_{j^*j^*}, \tag{7}$$

where $n|\breve{\beta}_{j^*} - \beta^0_{j^*}|^2/\widehat{\omega}_{j^*j^*} = o_p(\log|G|)$ as j^* is fixed and $|G|$ grows. From the proof of Theorem 2.3, we know the difference between $n|\widetilde{\beta}_{j^*} - \beta^0_{j^*}|^2/\widehat{\omega}_{j^*j^*}$ and $n|\widetilde{\beta}_{j^*} - \beta^0_{j^*}|^2/\omega_{j^*j^*}$ is asymptotically negligible. Thus by (7) and the fact that $\beta^0 \in \mathcal{U}_G(\sqrt{2}+\varepsilon_0)$, we have,

$$\max_{j\in G} n|\breve{\beta}_j - \widetilde{\beta}_j|^2/\widehat{\omega}_{jj} \ge \frac{1}{1+\delta}\left\{(\sqrt{2}+\varepsilon_0)^2(\log|G|) - o_p(\log|G|)\right\}. \tag{8}$$

The conclusion thus follows from (8) and (6) provided that δ is small enough. ◇

Proof of Proposition 3.1. Similar to the proof of Theorem 2.4, the distribution of $\max_{1\le j\le p}\sqrt{n}|\breve{\beta}_j - \beta_0|/\sqrt{\widehat{\omega}_{jj}}$ can be approximated by $\max_{1\le j\le p}|Z_j|$ with $Z=(Z_1,\ldots,Z_p)\overset{d}{\sim}N(0,\widetilde{\Theta})$. Under Assumption 2.5, by Lemma 6 of Cai et al. (2014), we have for any $x\in\mathbb{R}$ and as $p\to+\infty$,

$$P\left(\max_{1\le i\le p}|Z_i|^2 - 2\log(p) + \log\log(p) \le x\right) \to \exp\left\{-\frac{1}{\sqrt{\pi}}\exp\left(-\frac{x}{2}\right)\right\}.$$

It implies that

$$P\left(\max_{j\in\mathcal{S}_0^c} n|\breve{\beta}_j|^2/\widehat{\omega}_{jj} \le 2\log(p) - \log\log(p)/2\right) \to 1. \tag{9}$$

On the other hand, we note that

$$\min_{j\in\mathcal{S}_0} n|\beta^0_j|^2/\widehat{\omega}_{jj} \le 2\max_{j\in\mathcal{S}_0} n|\breve{\beta}_j - \beta^0_j|^2/\widehat{\omega}_{jj} + 2\min_{j\in\mathcal{S}_0} n|\breve{\beta}_j|^2/\widehat{\omega}_{jj}$$

Because the difference between $\min_{j\in\mathcal{S}_0} n|\beta^0_j|^2/\widehat{\omega}_{jj}$ and $\min_{j\in\mathcal{S}_0} n|\beta^0_j|^2/\omega_{jj}$ is asymptotically negligible, and $P(2\max_{j\in\mathcal{S}_0} n|\breve{\beta}_j - \beta^0_j|^2/\widehat{\omega}_{jj} \le 4\log(p) - \log\log(p)) \to 1$, we obtain

$$\begin{aligned} &P\left(\min_{j\in\mathcal{S}_0} n|\breve{\beta}_j|^2/\widehat{\omega}_{jj} > 2\log p\right) \\ \ge &P\left(2\min_{j\in\mathcal{S}_0} n|\breve{\beta}_j|^2/\widehat{\omega}_{jj} + 4\log(p) - \log\log(p) > 8\log(p)\right) \to 1. \end{aligned} \tag{10}$$

Hence, (16) follows from (9) and (10).

We next prove the optimality of $\tau^* = 2$, i.e., (17). For large enough p, we can choose a set G^* such that $\beta_j = 0$ for $j \in G^*$, and $|G^*| = \lfloor p^{\tau_2} \rfloor$ with $\tau/2 < \tau_2 < 1$. Following the above arguments, we know that the distribution of $\max_{j \in G^*} \sqrt{n}|\breve{\beta}_j - \beta_j^0|/\sqrt{\widehat{\omega}_{jj}}$ can be approximated by $\max_{j \in G^*} |Z_j|$ with $Z = (Z_1, \ldots, Z_p) \sim^d N(0, \widetilde{\Theta})$. Then we have

$$P\left(\max_{j \in G^*} n|\breve{\beta}_j|^2/\widehat{\omega}_{jj} \geq c\log(p)\right) \to 1,$$

where $\tau < c < 2\tau_2 < 2$. The conclusion thus follows immediately. $\diamondsuit$

Proof of Theorem 4.1. For simplicity, we only prove the result for the one-sided case (the arguments below can be easily modified for the two-sided case). Define $T_G = \max_{j \in G} \sqrt{n}(\breve{\beta}_j - \beta_j^0)$ and $T_{0,G} = \max_{j \in G} \sum_{i=1}^n \xi_{ij}/\sqrt{n}$. Let $\widetilde{c}_G(\alpha)$ be the bootstrap critical value for the one-sided test at level α. We first show that there exist $\zeta_1, \zeta_2 > 0$ such that

$$P\left(|T_G - T_{0,G}| \geq \zeta_1\right) < \zeta_2, \tag{11}$$

where $\zeta_1\sqrt{1 \vee \log(p/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$. Notice that

$$|T_G - T_{0,G}| \leq \max_{j \in G} \sqrt{n}|(\Theta_j^T - \widehat{\Theta}_j^T)\mathbb{E}_n \dot{L}_{\beta_0}| + ||\Delta||_\infty + \sqrt{n}||\widehat{\Theta}\mathcal{R}||.$$

Under the Lipschitz continuity in Assumption 4.1, we have

$$\widehat{\Theta}_j^T \mathbb{E}_n \dot{L}_{\widehat{\beta}} = \widehat{\Theta}_j^T \mathbb{E}_n \dot{L}_{\beta_0} + \widehat{\Theta}_j^T \mathbb{E}_n \ddot{L}_{\widehat{\beta}}(\widehat{\beta} - \beta^0) + \mathcal{R}_j,$$

where $\mathcal{R}_j = \widehat{\Theta}_j^T \mathcal{R} \leq \max_i |\widehat{\Theta}_j^T x_i| \cdot ||\mathbf{X}(\widehat{\beta} - \beta^0)||_2^2/n = O_P(K_n s_0 \lambda^2)$ (see the proof of Theorem 3.1 in van de Geer et al. 2014). It thus implies that $\sqrt{n}||\widehat{\Theta}\mathcal{R}||_\infty = O_P(\sqrt{n}K_n s_0 \lambda^2)$. By Assumptions 4.3-4.4, we have

$$||\Delta||_\infty = ||\sqrt{n}(\widehat{\Theta}\widehat{\Sigma} - I)(\widehat{\beta} - \beta_0)||_\infty \leq ||\widehat{\Theta}\widehat{\Sigma} - I||_\infty \sqrt{n}||\widehat{\beta} - \beta_0||_1 = O_P(\sqrt{n}\lambda\lambda_* s_0)$$

Following the arguments in the proof of Lemma 1.2, it can be shown that under Assumption 4.5

$$\max_{j\in G}\sqrt{n}|(\Theta_j^T-\widehat{\Theta}_j^T)\mathbb{E}_n\dot{L}_{\beta_0}| = \max_{j\in G}\left|\sum_{i=1}^n\widehat{\xi}_{ij}/\sqrt{n}-\sum_{i=1}^n\xi_{ij}/\sqrt{n}\right|$$
$$=O_P(K_n\max_j s_j\log(p)/\sqrt{n})+O_P\left(K_n^2 s_0\left(\lambda^2\sqrt{n}\vee\lambda\sqrt{\log(p)}\right)\right)$$

Thus (11) follows from a proper choice of ζ_1.

By Lemma 1.1, we have

$$\sup_{x\in\mathbb{R}}\left|P\left(\max_{j\in G}\sum_{i=1}^n\xi_{ij}/\sqrt{n}\le x\right)-P\left(\max_{j\in G}\sum_{i=1}^n z_{ij}/\sqrt{n}\le x\right)\right|\lesssim n^{-c'},\quad c'>0,$$

where $\{z_i=(z_{i1},\dots,z_{ip})'\}$ is a sequence of mean zero independent Gaussian vector with $\mathbb{E}z_iz_i'=\Theta_j^T\Sigma_{\beta_0}\Theta_j$. By the arguments in the proof of Theorem 3.2 in Chernozhukov et al. (2013), we have

$$\sup_{\alpha\in(0,1)}|P(T_G>\widetilde{c}_G^*(\alpha))-\alpha|\lesssim\pi(v)+P(\widetilde{\Gamma}>v)+n^{-c'}+\zeta_1\sqrt{1\vee\log(p/\zeta_1)}+\zeta_2, \tag{12}$$

where $\pi(v)=C_2v^{1/3}(1\vee\log(p/v))^{2/3}$. The conclusion follows by choosing $v=1/(\alpha_n(\log(p))^2)$ in (12). ♢

2 Additional numerical results

We consider the linear models where the rows of $\mathbf{X}$ are fixed i.i.d realizations from $N_p(0,\Sigma)$ with $\Sigma=(\Sigma_{i,j})_{i,j=1}^p$ under two scenarios: (i) Toeplitz: $\Sigma_{i,j}=0.9^{|i-j|}$; (ii) Exchangeable/Compound symmetric: $\Sigma_{i,i}=1$ and $\Sigma_{i,j}=0.8$ for $i\neq j$. The active set is $\mathcal{S}_0=\{1,2,\dots,s_0\}$ with $s_0=3$ or 15. To obtain the main Lasso estimator, we implemented the scaled Lasso with the tuning parameter $\lambda_0=\sqrt{2}\tilde{L}_n(k_0/p)$ with $\tilde{L}_n(t)=n^{-1/2}\Phi^{-1}(1-t)$, where Φ is the cumulative distribution function for $N(0,1)$, and k_0 is the solution to $k=\tilde{L}_1^4(k/p)+2\tilde{L}_1^2(k/p)$. We estimate the noise level σ^2 using the modified variance estimator.

2.1 Modified variance estimator

Figure S.1 provides boxplots of $\widehat{\sigma}/\sigma$ for the variance estimator delivered by the scaled Lasso (denoted by "SLasso") and for the modified variance estimator in (24) of the paper (denoted by"SLasso*"). Clearly, the modified variance estimator corrects the noise underestimation issue and thus is preferable.

2.2 Impact of the remainder term

We discuss the impact of the (normalized) remainder term Δ on the coverage accuracy. Recall the linear expansion $\sqrt{n}(\breve{\beta}-\beta^0)=\widehat{\Theta}\mathbf{X}^T\epsilon/\sqrt{n}+\Delta$, where $\Delta=(\Delta_1,\ldots,\Delta_p)^T=-\sqrt{n}(\widehat{\Theta}\widehat{\Sigma}-I)(\widehat{\beta}-\beta^0)$ with $\widehat{\Sigma}$ being the Gram matrix and $\widehat{\beta}$ being the Lasso estimator. The studentized maximum type test statistic can be written as

$$\max_{1\leq j\leq p}\frac{\sqrt{n}|\breve{\beta}_j-\beta_j^0|}{\sqrt{\widehat{\omega}_{jj}}}=\max_{1\leq j\leq p}\left|\frac{\sum_{i=1}^n\widehat{\xi}_{ij}}{\sqrt{n\widehat{\omega}_{jj}}}+\frac{\Delta_j}{\sqrt{\widehat{\omega}_{jj}}}\right|. \tag{13}$$

Thus the coverage accuracy can be greatly affected by the term $\Delta_j^*:=\frac{\Delta_j}{\sqrt{\widehat{\omega}_{jj}}}$. Note that this (normalized) remainder term is determined by $\widehat{\Theta}$. We now consider three different methods in estimating Θ : (i) nodewise Lasso with λ_js chosen by 10-fold cross validation; (ii) nodewise Lasso with $\lambda_j=0.01$; (iii) the method in Javanmard and Montanari (2014) with the tuning parameters chosen automatically by their algorithm. To empirically evaluate $\Delta^*:=(\Delta_1^*,\ldots,\Delta_p^*)^T$, we consider the linear models with $t(4)/\sqrt{2}$ errors, $n=100$ and $p=500$. Define $\Delta_{\text{ac}}^*=(\Delta_j^*)_{j\in\mathcal{S}_0}$ and $\Delta_{\text{in}}^*=(\Delta_j^*)_{j\in\mathcal{S}_0^c}$. Figure S.2 presents the boxplots for $||\Delta_{\text{ac}}^*||_\infty$ and $||\Delta_{\text{in}}^*||_\infty$. The nodewise Lasso clearly outperforms the method in Javanmard and Montanari (2014), and the choice of $\lambda_j=0.01$ yields the smallest $||\Delta_{\text{ac}}^*||_\infty$ in all cases. In addition, $||\Delta_{\text{ac}}^*||_\infty$ is relatively large when Σ is exchangeable, $s_0=15$ and $p=500$, which explains the lack of performance/undercoverage in this case. We observe that the maximum norms of Δ_{ac}^* and Δ_{in}^* generally increase with s_0. Overall, the above discussions support our observations in Tables 1-2 of the paper in the sense that the lower the (normalized) remainder term is, the more accurate the coverage is.

References

[1] Javanmard, A. and Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. arXiv:1306.3171.

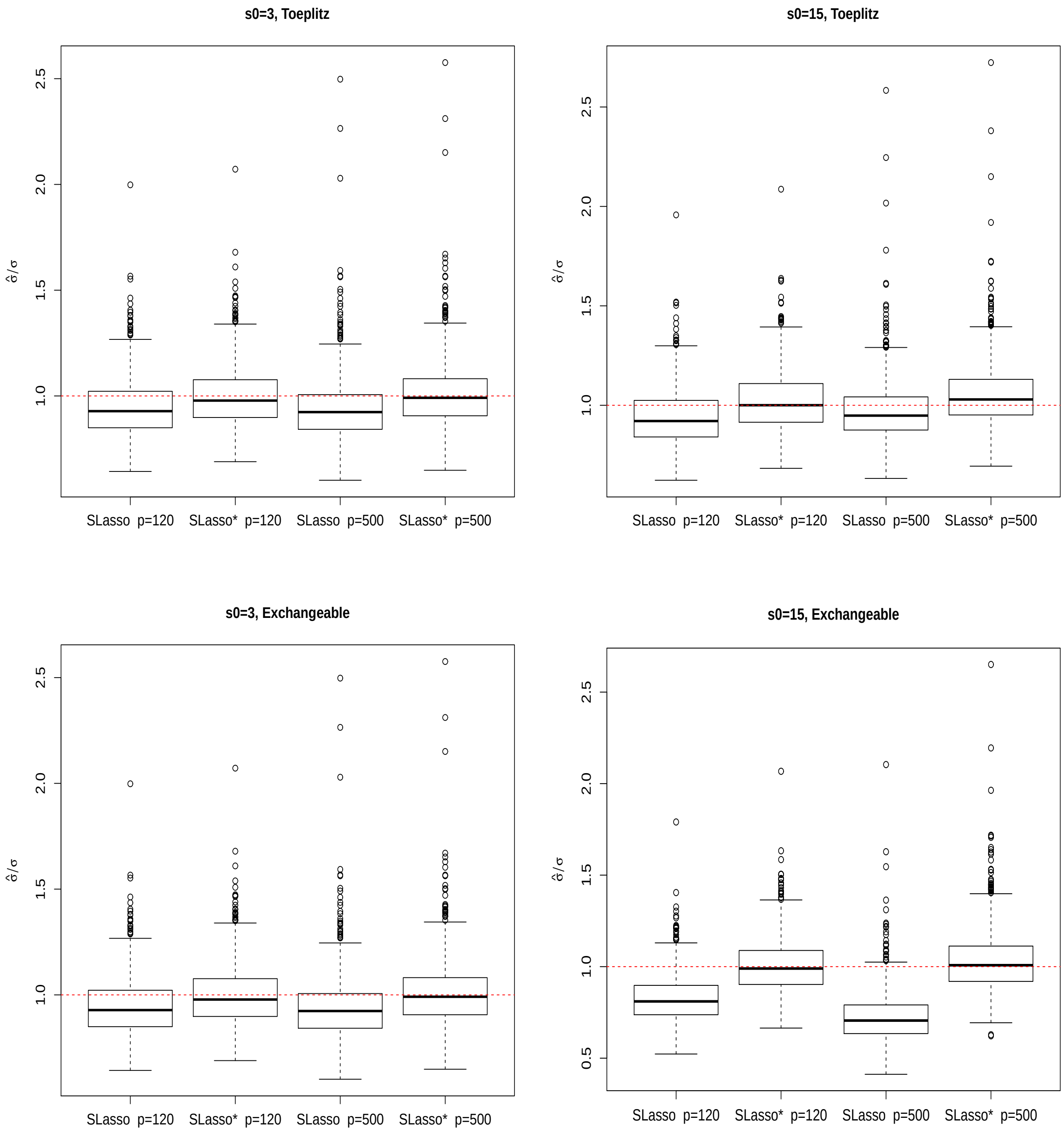

Figure S.1: Boxplots for $\widehat{\sigma}/\sigma$, where $s_0 = 3$ or 15, Σ is Toeplitz or exchangeable, and the errors are generated from the studentized $t(4)$ distribution. Here "SLasso" corresponds to the variance estimator delivered by the scaled Lasso and 'SLasso*" corresponds to the modified variance estimator.

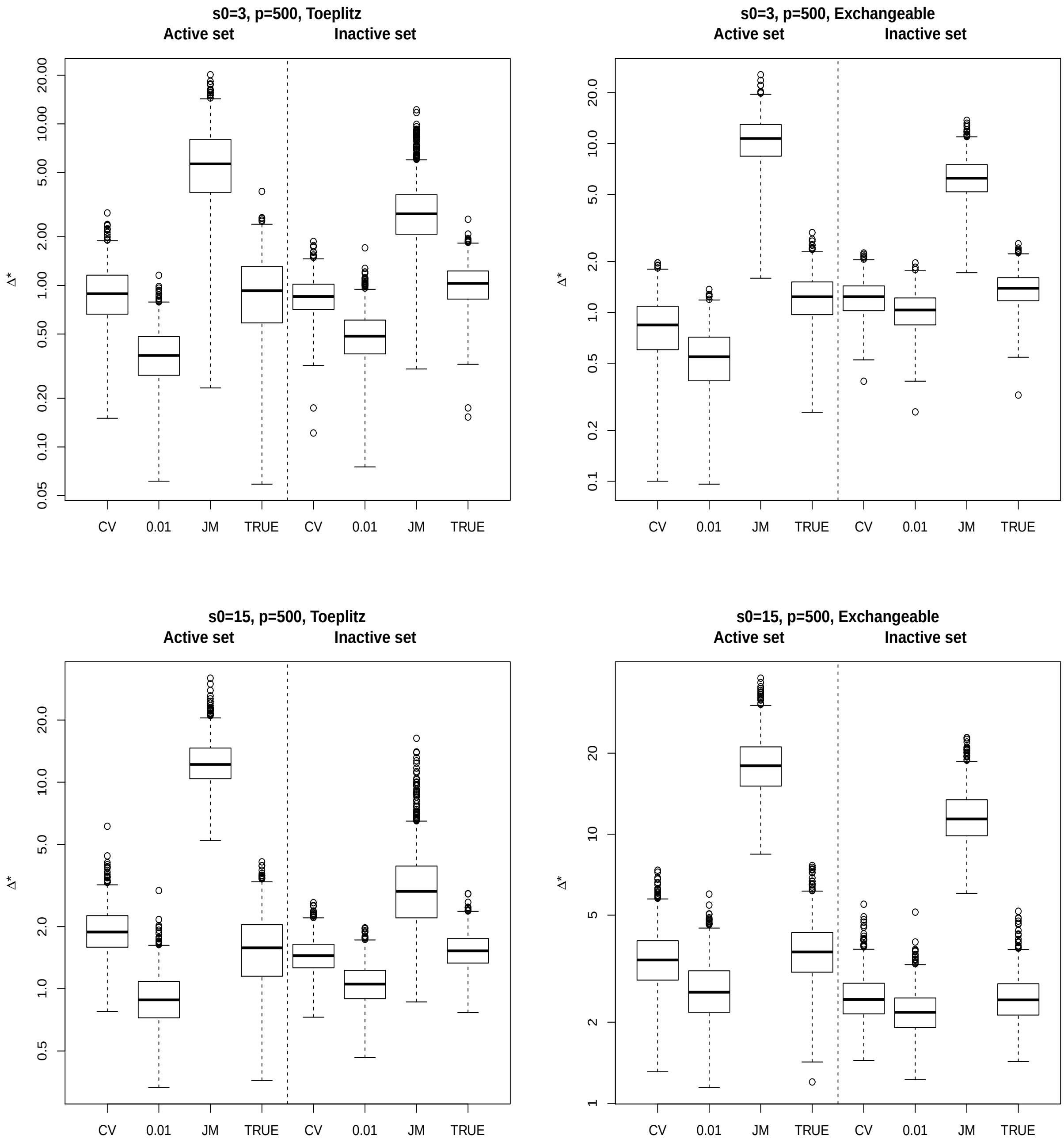

Figure S.2: Boxplots for $||\Delta^*_{\text{ac}}||_\infty$ and $||\Delta^*_{\text{in}}||_\infty$, where $s_0 = 3$ or 15, $p = 500$, Σ is Toeplitz or exchangeable, and the errors are $t(4)/\sqrt{2}$. Here "CV", "0.01", "JM" and "TRUE" denote the nodewise Lasso with λ_js chosen by 10-fold cross validation and $\lambda_j = 0.01$, the method in Javanmard and Montanari (2014) and the method with the true Θ respectively. Note that the y-axis is plotted on a log scale.