Projection-based Inference for High-dimensional Linear Models

Sangyoon Yi and Xianyang Zhang

Texas A&M University

Supplementary Material

We empirically investigate the sensitiveness of our method to the choice of tuning parameters in Section S1. Section S2 provides every figure and table for Sections 2 and 6 in the main paper. Technical details and additional numerical results are gathered in Sections S3 and S4, respectively.

S1 Empirical analysis of tuning parameters

We empirically investigate the sensitiveness of our method to the choice of tuning parameters. Throughout this subsection, we suppose the rows of $\mathbf{X} \in \mathbb{R}^{100\times 500}$ are i.i.d realizations from $N(0, \mathbf{\Sigma})$ with $\Sigma_{j,k} = 0.9^{|j-k|}$ (Toeplitz) or $\Sigma_{j,k} = 0.8$ (Equicorrelation) for $j \neq k$ and $\Sigma_{jj} = 1$. Regression coefficients β_j's are generated by either Case 1 with $s_0 = 10$ or Case 2 with $s_0 = 4$ as described in Section 6. The errors are independently generated from the standard normal distribution. The nominal level is 95% and results are based on 100 independent simulation runs.

We first explore the effect of C_0 on the estimation of the surrogate set and the impact of C_1 and C_2 on the coverage rate and interval width of the BRP-based confidence interval. The results for β_j generated from Case 2 with $s_0 = 4$ and Toeplitz covariance $\boldsymbol{\Sigma}$ are summarized in Figure 1. As seen from Panel A, the surrogate set $\mathcal{A}(\tau)$ with $\tau = 2$ correctly identifies the large coefficients when $C_0 \geq 2$. Panels B-D provide the average coverage rate, bias and length of the BRP-based confidence intervals for the active set over a prespecified set of grid points for (C_1, C_2). The coverage probability and interval width both tend to increase with the values of C_1 and C_2. These results appear to suggest that fixing one parameter at a reasonably large value while choosing the other parameter to balance the coverage probability and interval width would generally deliver similar results as simultaneously selecting the two parameters.

To confirm this intuition, we set $C_0 = 2$, $C_1 = 8$ and use the procedure in Section 5 to select C_2 over the following prespecified grid points

$$\left\{c_{2,j,(k)}\right\}_{k=1}^{K} = \{0.3, 0.6, \cdots, 14.7, 15.0\}. \tag{S1.1}$$

We denote the corresponding procedures by "Fix-BRP" and "Fix-MBRP" and compare their performance with the procedures that select all tuning parameters automatically using the method in Section 5. Notice that fixing

C_0 and C_1 would significantly ease the computational burden. Figure 2 presents the empirical coverage probabilities and lengths of the 95% confidence intervals and the normalized overall bias as in (6.4). Fix-BRP and Fix-MBRP perform equally well in terms of the coverage accuracy and bias as compared to BRP and MBRP but with a much lower computational cost. Indeed similar results are observed for the other simulation setups in Section 6.1. For the rest of the paper, we shall adopt the above procedure by fixing C_0 and C_1 to implement the proposed method.

Finally, we study the impact of B and τ. Figure 3 summarizes the performance of the BRP and MBRP-based confidence intervals with different values of B and τ. The results are not sensitive to the bootstrap sample size B. We also observe that a larger τ tends to deliver higher coverage for MBRP in the equicorrelation case. Unreported numerical studies show that similar phenomenon can be observed for the other simulation setups. In Section 6 below, we shall fix $B = 200$ and $\tau = 2$.

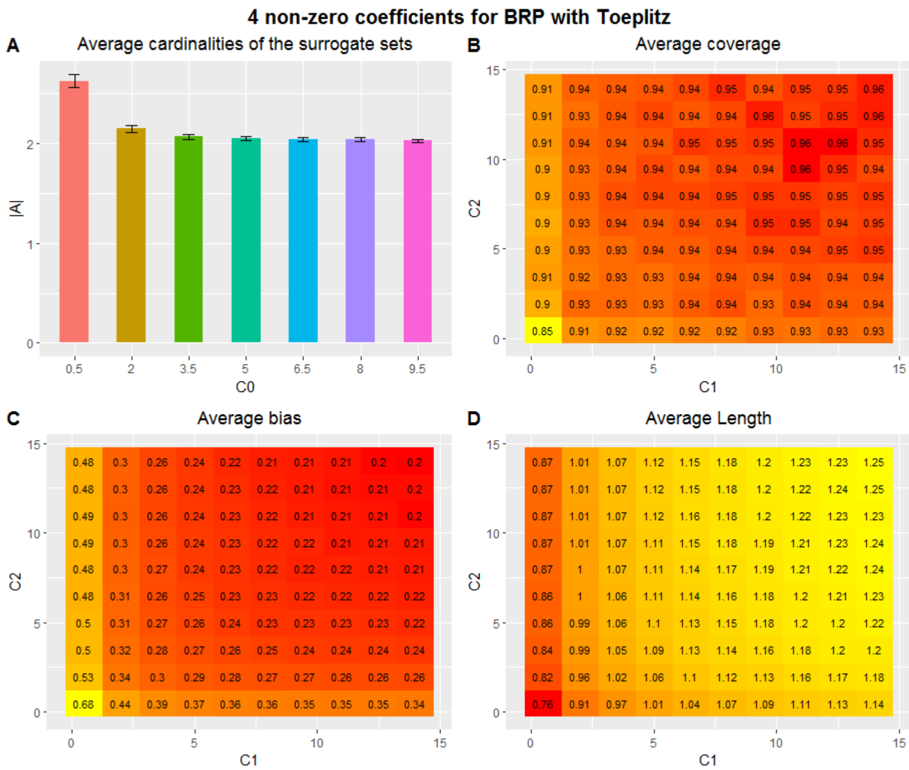

Figure 1: Panel A shows the barplots of the average cardinality of $\mathcal{A}(\tau)$ against C_0. Error bars in the barplots represent the interval within one standard error of the average value. Panel B (C or D) shows the heatmap of the average coverage rates (bias or length) by the BRP estimator over a prespecified grid points for (C_1, C_2). The number represents the average coverage probability (bias or length) of the 95% confidence intervals for the active set.

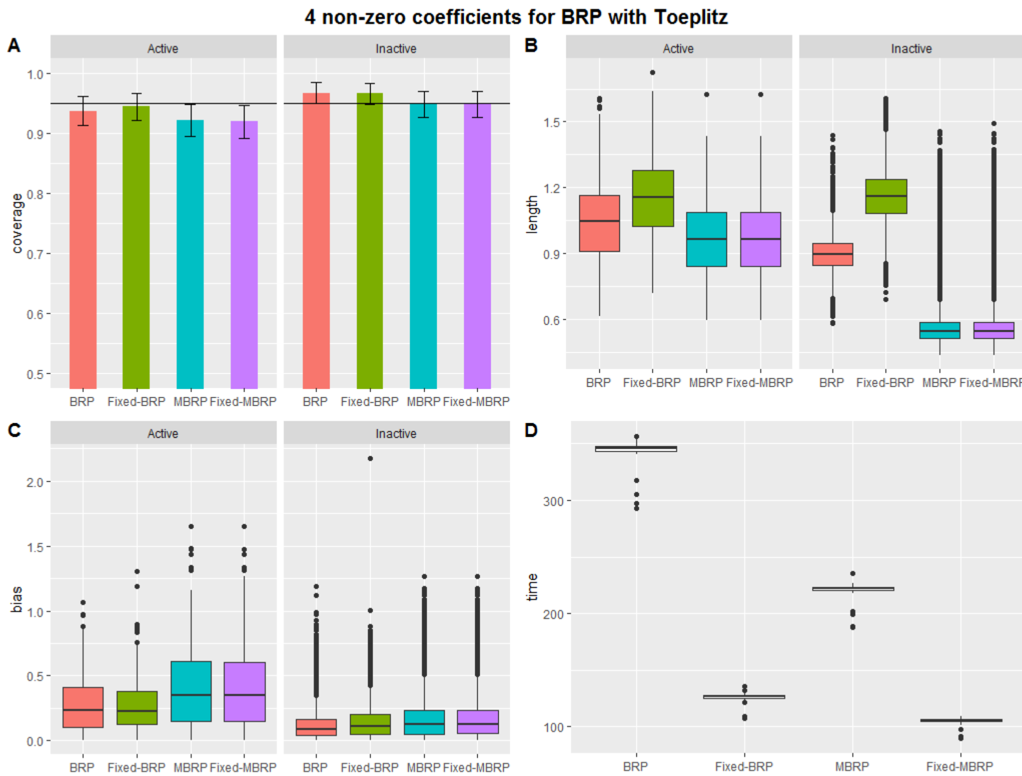

Figure 2: Panel A shows the barplots of the empirical coverage and Panels B-C display the boxplots for the length and bias of the 95% confidence intervals of each method. In Panel A, the horizontal line indicates the nominal level and error bars represent the interval within one standard deviation of the empirical coverage. Panel D shows the boxplots of the computation time (in seconds) for each method.

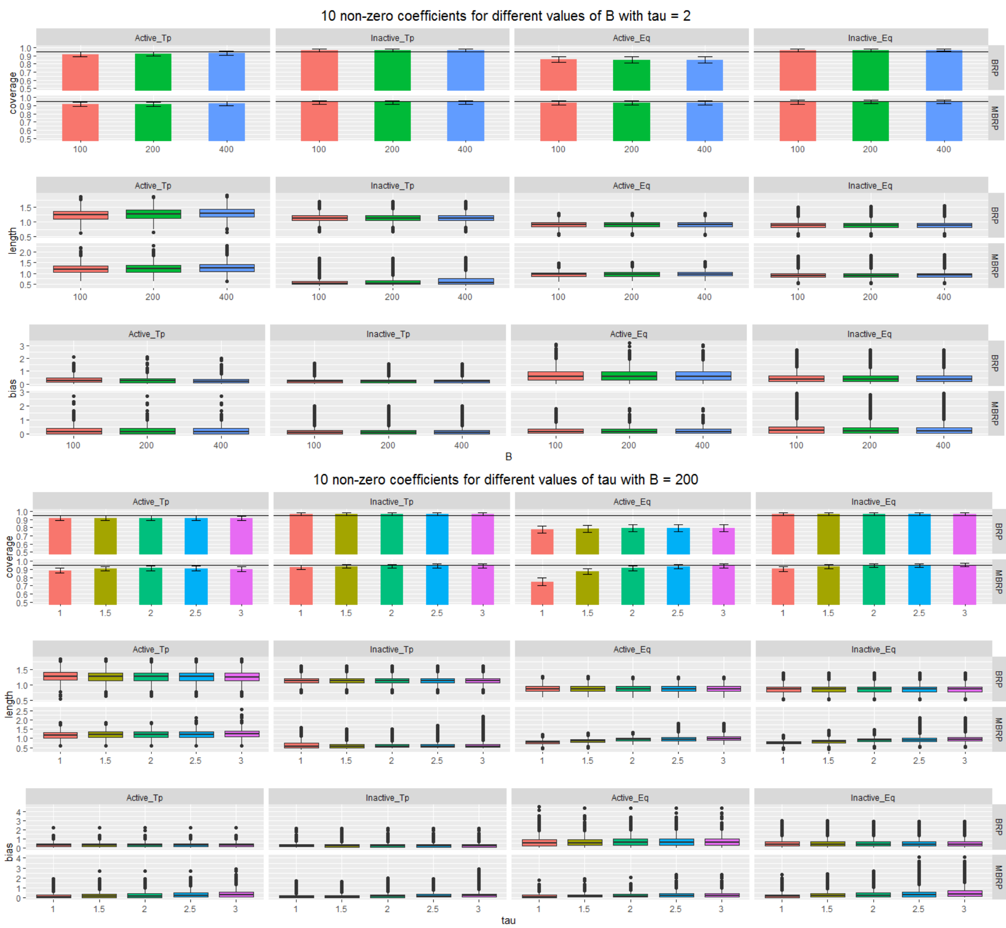

Figure 3: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for both the active and inactive sets with different values of B and τ. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage. Data sets are independently generated from Case 1 with $s_0 = 10$ and standard normal error as in Section 6.

S2 Appendices for Sections 2 and 6

S2.1 For Section 2

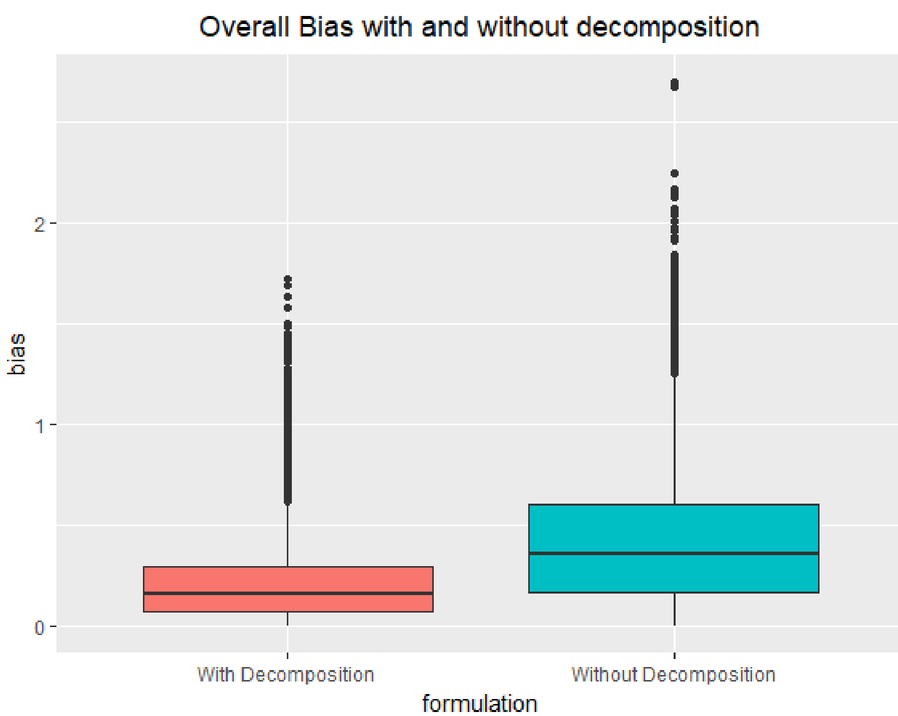

Figure 4: Boxplots of the absolute values of the normalized bias terms defined in (6.4) by "With Decomposition" and "Without Decomposition." The non-zero β_j's are independently generated from $U(0,4)$ with $s_0 = 10$. All the simulation settings are the same as the case with the Toeplitz covariance structure and standard normal error in Section 6. The results are based on 100 simulation runs.

S2.2 For Section 6

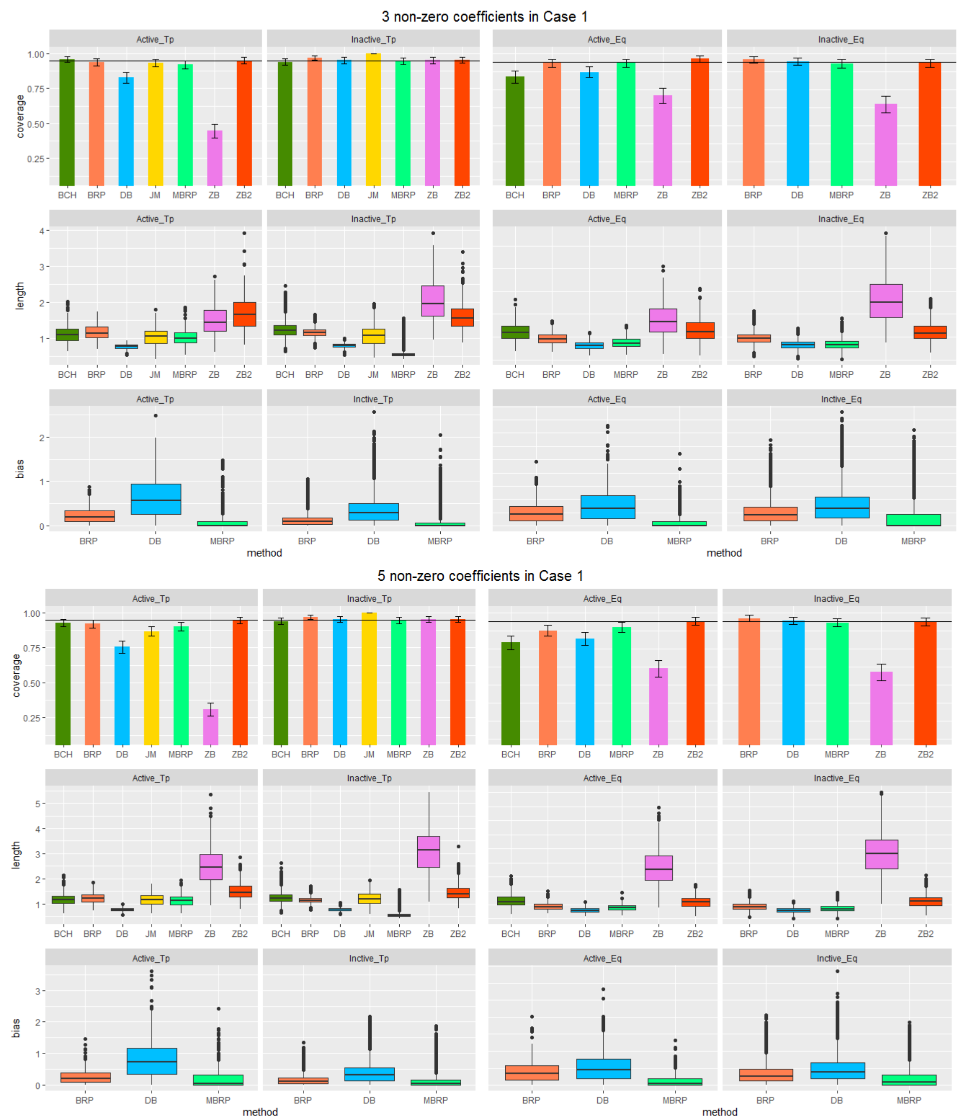

Figure 5: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 3, 5$ and standard normal error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

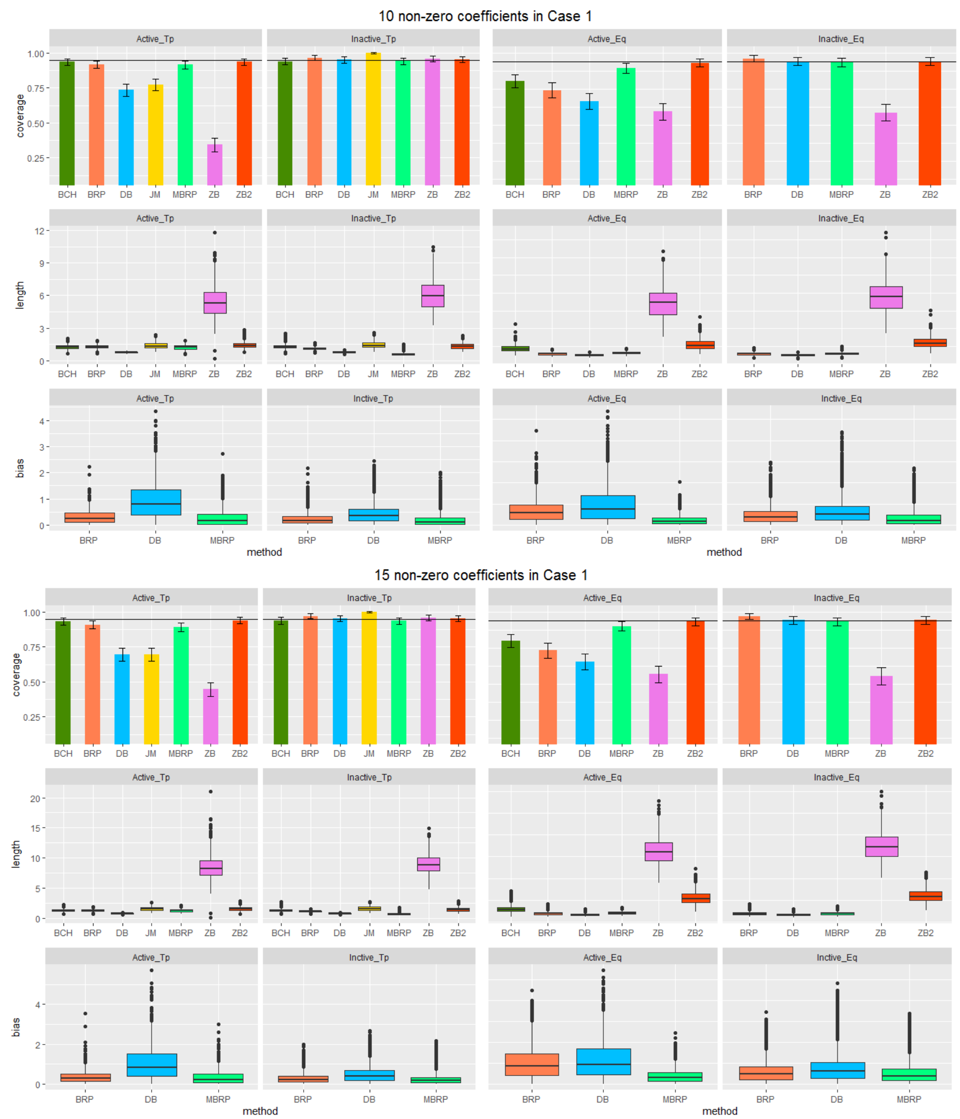

Figure 6: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 10, 15$ and standard normal error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

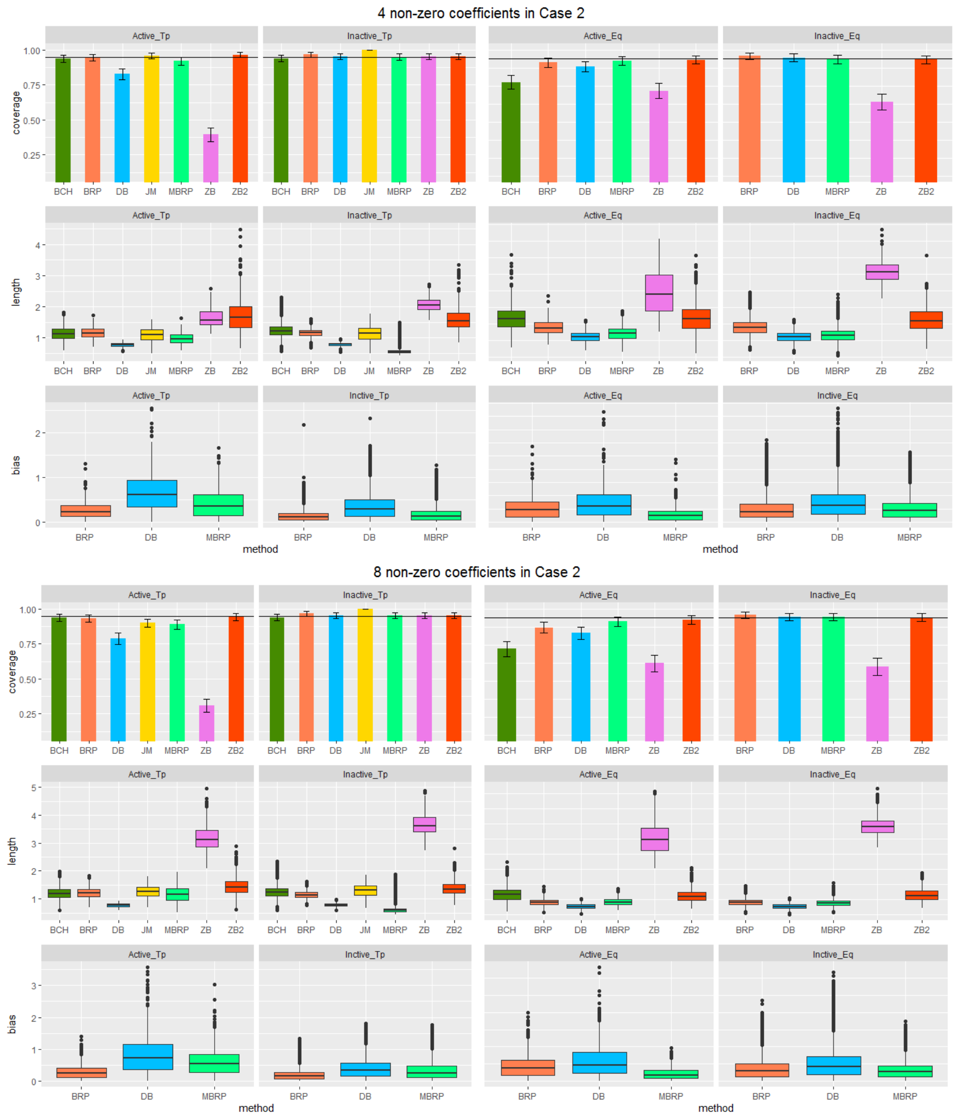

Figure 7: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 4, 8$ and standard normal error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

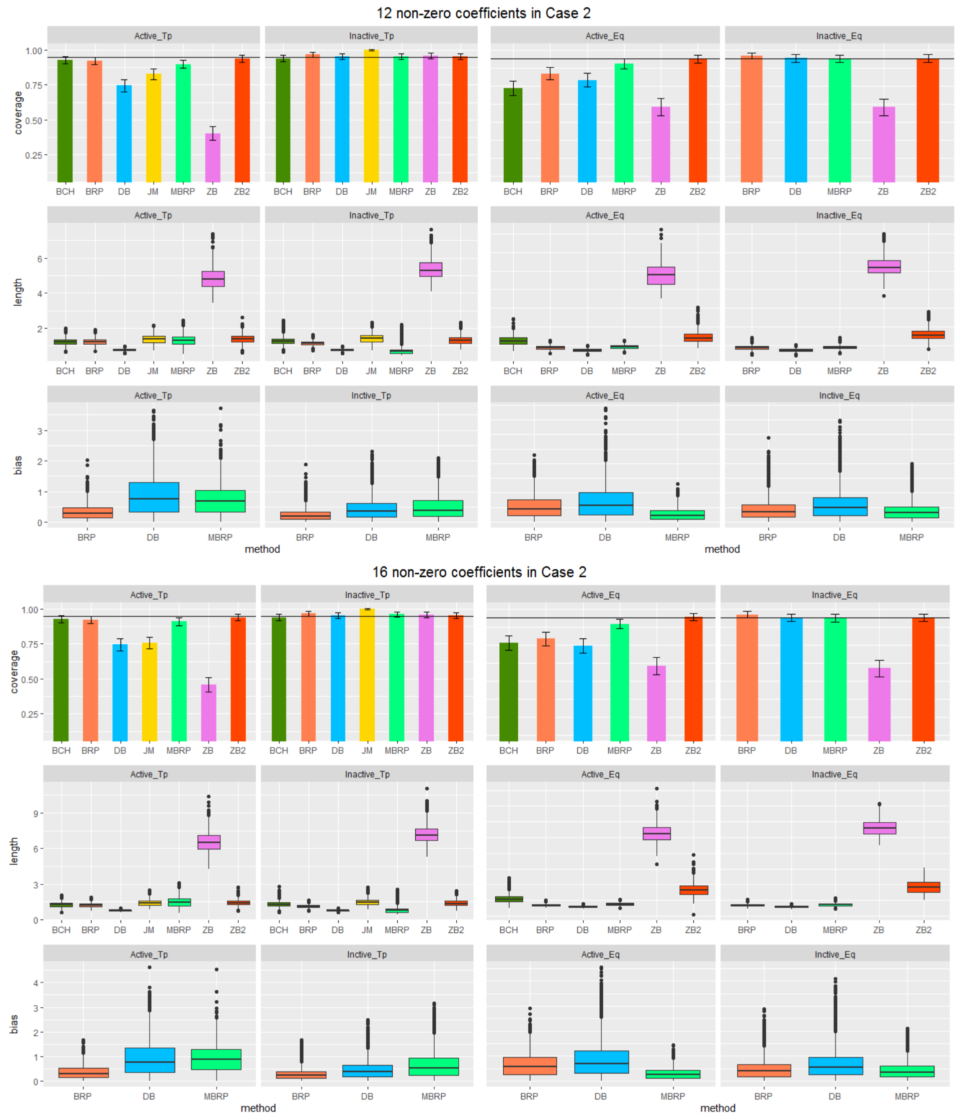

Figure 8: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 12, 16$ and standard normal error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

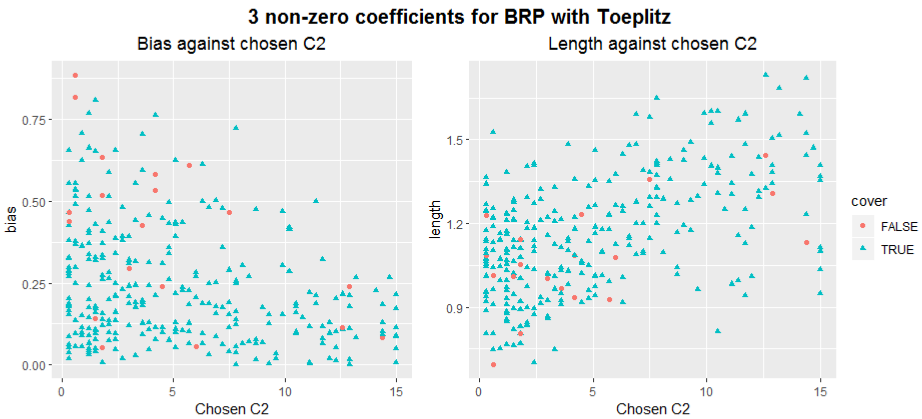

Figure 9: Scatterplots of the bias and length of the BRP-based confidence interval for the active set with $s_0 = 3$ and Toeplitz covariance structure for $\mathbf{X}$ against the selected C_2. The point shapes and colors indicate whether the constructed confidence intervals include the true parameter or not.

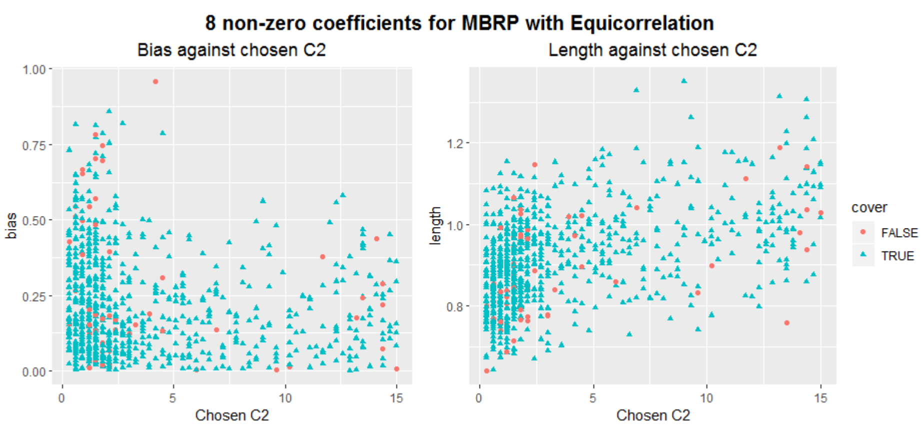

Figure 10: Scatterplots of the bias and length of the MBRP-based confidence interval for the active set with $s_0 = 8$ and equicorrelation covariance structure for $\mathbf{X}$ against the selected C_2. The point shapes and colors indicate whether the constructed confidence intervals include the true parameter or not.

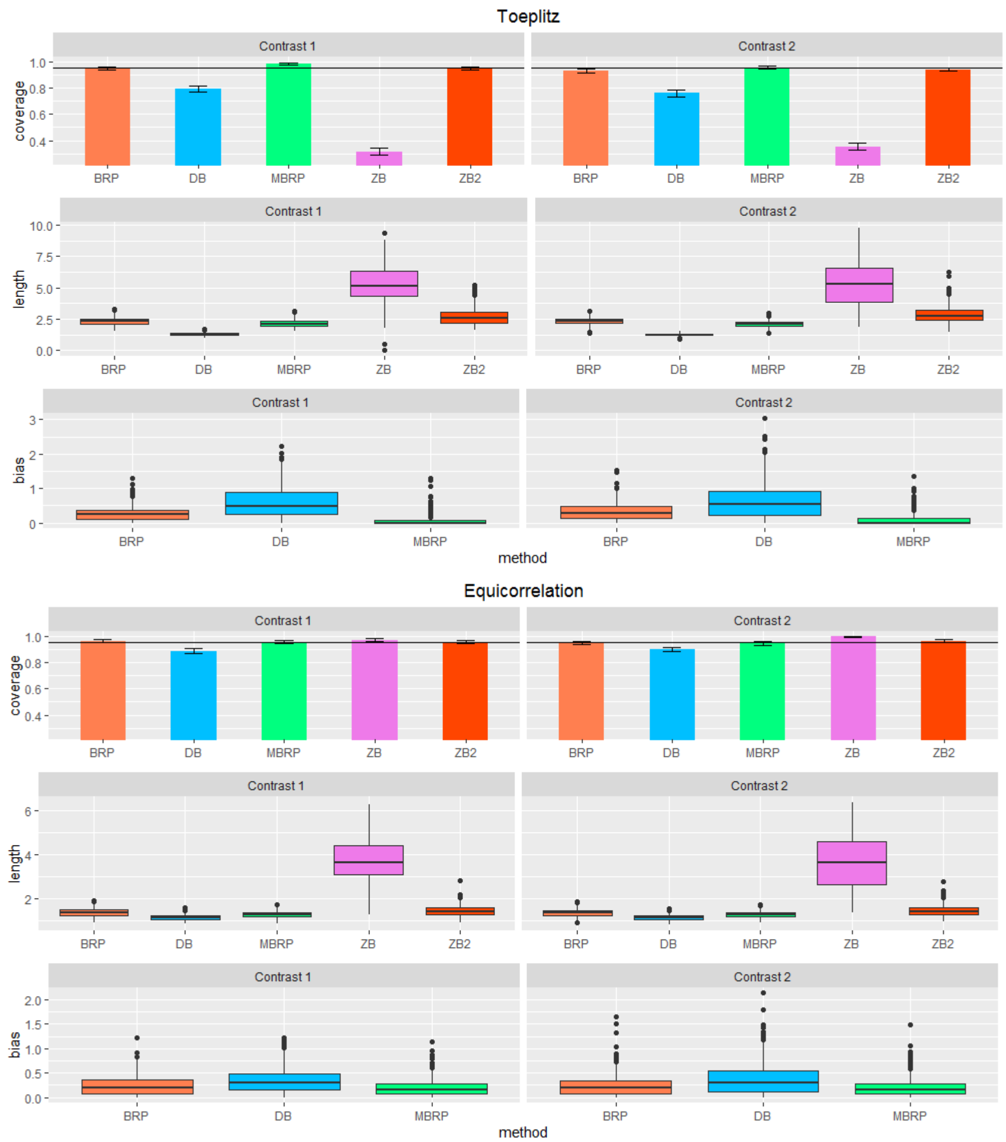

Figure 11: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for different contrasts with standard normal error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

	Min	Q1	Median	Q3	Max
BRP	107.60	125.30	126.70	127.80	135.20
MBRP	89.65	104.34	105.21	106.35	109.03
DB	26.29	33.45	34.41	35.66	38.20
ZB	457.90	471.30	476.50	483.20	499.50

Table 1: Computation time (in seconds) of each method for constructing 500 confidence intervals calculated by the R package `microbenchmark`. The five number summaries are obtained based on 100 independent simulation runs.

S3 Technical Details

S3.1 Concentration Inequalities

We first define several quantities which will appear throughout the supplementary material. Let $\theta_j = X_j - \mathbf{X}_{-j} b_{-j}$ and

$$b_{-j} = \underset{\tilde{b}\in\mathbb{R}^{p-1}}{\operatorname{argmin}} E||X_j - \mathbf{X}_{-j}\tilde{b}||_2^2 = \mathbf{\Sigma}_{-j,-j}^{-1}\mathbf{\Sigma}_{-j,j}.$$

Define $\kappa_1 = 2\kappa^2, \kappa_{2j} = 2\kappa^2\sqrt{\Lambda_{\min}^{-1}\mathbf{\Sigma}_{j,j}}$ and $\kappa_{3j} = 2\kappa^2\Lambda_{\min}^{-1}\mathbf{\Sigma}_{j,j}$.

The following lemmas shows the concentration inequalities for sub-exponential and sub-gaussian random variables which are motivated by Lemmas 5.5, 5.15 and Propositions 5.10, 5.16 in Vershynin (2010).

Lemma 1. *Let $X_1, \cdots, X_N$ be i.i.d. mean-zero sub-exponential random variables with $\|X_i\|_{\psi_1} = K_1$. Then, for every $a = (a_1, \cdots, a_N)^\top \in \mathbb{R}^{N\times 1}$*

and any $t \geq 0$, we have

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} a_i X_i\right| \geq t\right) \leq 2\exp\left\{-\min\left(\frac{t^2}{8e^2\|a\|^2K_1^2}, \frac{t}{4eK_1\|a\|_\infty}\right)\right\}.$$

Proof of Lemma 1. We first derive an upper bound of the moment generating function of X_i. By expanding the exponential function in the Taylor series, we have

$$\begin{aligned} E[\exp(\lambda X_i)] = E\left[1 + \lambda X_i + \sum_{p=2}^{\infty} \frac{(\lambda X_i)^p}{p!}\right] &= 1 + \sum_{p=2}^{\infty} \frac{\lambda^p E[X_i^p]}{p!} \\ \leq 1 + \sum_{p=2}^{\infty} \frac{\lambda^p (K_1 p)^p}{(p/e)^p} = 1 + \sum_{p=2}^{\infty} (e\lambda K_1)^p &= 1 + \frac{(e\lambda K_1)^2}{1-(e\lambda K_1)} \end{aligned}$$

provided that $|e\lambda K_1| < 1$. The inequality follows by the definition of sub-exponential norm

$$E[X_i^p] \leq (K_1 p)^p$$

and Stirling's approximation $p! \geq (p/e)^p$. In addition, if $|e\lambda K_1| < 0.5$, the quantity on the right hand side can be bounded above by

$$1 + 2(e\lambda K_1)^2 \leq \exp(2(e\lambda K_1)^2).$$

Thus, combining all of the above implies

$$E[\exp(\lambda X_i)] \leq \exp(2(e\lambda K_1)^2) \quad \text{for} \quad |\lambda| < \frac{1}{2eK_1}. \tag{S3.1}$$

Next, for $\lambda > 0$, we have

$$\mathbb{P}\left(\sum_{i=1}^{N} a_i X_i \geq t\right) = \mathbb{P}\left(\exp\left(\lambda \sum_{i=1}^{N} a_i X_i\right) \geq \exp(\lambda t)\right)$$
$$\leq \exp(-\lambda t) E\left[\exp\left(\lambda \sum_{i=1}^{N} a_i X_i\right)\right] = \exp(-\lambda t) \prod_{i=1}^{N} E[\exp(\lambda a_i X_i)]$$

by the exponential Markov inequality for $\sum_{i=1}^{N} a_i X_i$. If λ is small enough so that $|\lambda| < (2eK_1\|a\|_\infty)^{-1}$, (S3.1) gives

$$\mathbb{P}\left(\sum_{i=1}^{N} a_i X_i \geq t\right) \leq \exp(-\lambda t) \prod_{i=1}^{N} \exp(2(e\lambda a_i K_1)^2) = \exp(-\lambda t + 2e^2\lambda^2\|a\|^2 K_1^2).$$

By choosing $\lambda = \min\left(t(4e^2\|a\|^2 K_1^2)^{-1}, (2eK_1\|a\|_\infty)^{-1}\right)$, we obtain

$$\mathbb{P}\left(\sum_{i=1}^{N} a_i X_i \geq t\right) \leq \exp\left\{-\min\left(\frac{t^2}{8e^2\|a\|^2 K_1^2}, \frac{t}{4eK_1\|a\|_\infty}\right)\right\}.$$

The second term in min can be obtained as follows. When $\lambda = (2eK_1\|a\|_\infty)^{-1}$, we have

$$-\lambda t + 2e^2\lambda^2\|a\|^2 K_1^2 = -\frac{t}{2eK_1\|a\|_\infty} + \frac{\|a\|^2}{2\|a\|_\infty^2} \leq -\frac{t}{4eK_1\|a\|_\infty}$$

where the last inequality follows as

$$\lambda = \frac{1}{2eK_1\|a\|_\infty} \leq \frac{t}{(4e^2\|a\|^2 K_1^2)}$$

which implies

$$\frac{\|a\|^2}{\|a\|_\infty} \leq \frac{t}{2eK_1}.$$

By repeating the same argument for $-X_i$, we get the same bound for $\mathbb{P}(-\sum_{i=1}^{N} a_i X_i \geq t)$, which completes the proof. $\square$

Lemma 2. *Let* $X_1, \cdots, X_N$ *be i.i.d. mean-zero sub-gaussian random variables with* $\|X_i\|_{\psi_2} = K_2$. *Then, we have the following results.*

1. *For any* $|\omega_1| \leq 1$,

$$E\left[\exp\left(\omega_1^2 \frac{X_i^2}{4eK_2^2}\right)\right] \leq \exp(\omega_1^2). \tag{S3.2}$$

2. *For* $\omega_2 \in \mathbb{R}$,

$$E[\exp(\omega_2 X_i)] \leq \exp(8eK_2^2\omega_2^2). \tag{S3.3}$$

3. *For every* $a = (a_1, \cdots, a_N) \in \mathbb{R}^N$ *and any* $t \geq 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^N a_i X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{32eK_2^2\|a\|^2}\right). \tag{S3.4}$$

Proof of Lemma 2. Let $Y_i = X_i/(2\sqrt{e}K_2)$. We note that, for $|\omega_1^2/2| < 1$,

$$\begin{aligned} E[\exp(\omega_1^2 Y_i^2)] &= 1 + \sum_{k=1}^\infty \frac{\omega_1^{2k} E[Y_i^{2k}]}{k!} \\ &\leq 1 + \sum_{k=1}^\infty \frac{1}{(4e)^k} \frac{(2\omega_1^2 k)^k}{(k/e)^k} = \sum_{k=0}^\infty \left(\frac{\omega_1^2}{2}\right)^k = \left(1 - \frac{\omega_1^2}{2}\right)^{-1} \end{aligned}$$

by the Taylor series expansion of the exponential function and Stirling's approximation. We can further bound

$$E[\exp(\omega_1^2 Y_i^2)] \leq \exp(\omega_1^2) \quad \text{for} \quad |\omega_1| \leq 1$$

by using the inequality $(1-x)^{-1} \leq \exp(2x)$ for $0 \leq x \leq 0.5$, which completes (S3.2).

For (S3.3), we notice that

$$E[\exp(\omega Y_i)] \leq E[\omega Y_i + \exp(\omega^2 Y_i^2)] \leq \exp(\omega^2) \tag{S3.5}$$

for $|\omega| \leq 1$ where the first inequality follows by $e^x \leq x + e^{x^2}$ for any $x \in \mathbb{R}$ and the second one does by (S3.2). If $|\omega| \geq 1$, we have

$$E[\exp(\omega Y_i)] \leq \exp(\omega^2) E[\exp(Y_i^2)] \leq \exp(\omega^2 + 1) \leq \exp(2\omega^2) \tag{S3.6}$$

due to $\omega Y_i \leq \omega^2 + Y_i^2$ for any ω, Y_i and (S3.2). Thus, combining (S3.5) with (S3.6) gives

$$E[\exp(\omega Y_i)] \leq \exp(2\omega^2).$$

for any $\omega \in \mathbb{R}$. Letting $\omega_2 = \omega/(2\sqrt{e}K_2)$ completes (S3.3).

For (S3.4), notice that

$$\begin{aligned} E\left[\exp(\omega_2 \sum_{i=1}^N a_i X_i)\right] &= \prod_{i=1}^N E\left[\exp(\omega_2 a_i X_i)\right] \\ &\leq \prod_{i=1}^N \exp(8eK_2^2 \omega_2^2 a_i^2) = \exp(8eK_2^2 \omega_2^2 \|a\|^2). \end{aligned}$$

For $\omega_2 \geq 0$, we have

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^N a_i X_i \geq t\right) &= \mathbb{P}\left(\exp\left(\omega_2 \sum_{i=1}^N a_i X_i\right) \geq \exp(\omega_2 t)\right) \\ &\leq \exp(-\omega_2 t) E\left[\exp\left(\omega_2 \sum_{i=1}^N a_i X_i\right)\right] \\ &\leq \exp(-\omega_2 t + 8e\omega_2^2 K_2^2 \|a\|^2) \\ &\leq \exp\left(-\frac{t^2}{32eK_2^2 \|a\|^2}\right) \end{aligned}$$

and the same bound can be obtained for $\mathbb{P}\left(-\sum_{i=1}^{N} a_i X_i \geq t\right)$. Thus, combining those bounds gives (S3.4). $\square$

S3.2 Technical details in Section 3

Lemma 3. *Under Assumption 5,*

$$\mathbb{P}\left(n^{-1}\|\theta_j^\top \mathbf{X}_{-j}\|_\infty \geq \varepsilon_{0j}\sqrt{\frac{\log p}{n}}\right) \leq 2\exp\left\{\left(1-\frac{1}{8e^2}\frac{\varepsilon_{0j}^2}{(\kappa_{0j})^2}\right)\log p\right\}$$

for $0 < \varepsilon_{0j} \leq \kappa_{0j}\sqrt{n(\log p)^{-1}}$.

Proof of Lemma 3. Let $Z = (Z_1 \cdots Z_{p-1}) = n^{-1}(X_j^\top \mathbf{X}_{-j} - b_{-j}^\top \mathbf{X}_{-j}{}^\top \mathbf{X}_{-j})$.

Then we have

$$Z = \frac{1}{n}\sum_{i=1}^{n}(X_{i,j} - b_{-j}^\top X_{i,-j})X_{i,-j}^\top$$

where $X_{i,j}$ is the value of the jth predictor of the ith observation and

$$X_{i,-j}^\top = (X_{i,1} \cdots X_{i,j-1}, X_{i,j+1} \cdots X_{i,p}).$$

Fix some $k \in \{1, 2, \ldots, p\} \setminus \{j\}$ and let $Z_{i,j}^{(k)} = (X_{i,j} - b_{-j}^\top X_{i,-j})X_{i,-j}^{(k)}$, where $X_{i,-j}^{(k)}$ denotes the kth element of $X_{i,-j}$. Then $Z_k = n^{-1}\sum_{i=1}^{n} Z_{ij}^{(k)}$, where $E[Z_{i,j}^{(k)}] = 0$ and $Z_{i,j}^{(k)}$'s are independent across $1 \leq i \leq n$.

We derive an upper bound for $\|Z_{i,j}^{(k)}\|_{\psi_1}$. Notice that

$$\begin{aligned}\|Z_{i,j}^{(k)}\|_{\psi_1} = \|(X_{i,j} - b_{-j}^\top X_{i,-j})X_{i,-j}^{(k)}\|_{\psi_1} &\leq 2\|X_{i,j} - b_{-j}^\top X_{i,-j}\|_{\psi_2}\|X_{i,-j}^{(k)}\|_{\psi_2} \\ &= 2\|X_{i,\cdot}^\top \gamma_{-j}\|_{\psi_2}\|X_{i,-j}^{(k)}\|_{\psi_2} \\ &\leq 2\kappa^2\|\gamma_{-j}\|_2 \\ &\leq 2(1 + \sqrt{\Lambda_{\min}^{-1}\mathbf{\Sigma}_{j,j}})\kappa^2,\end{aligned}$$

where $X_{i,\cdot}^\top = (X_{i,j}, X_{i,-j}^\top)$ and $\gamma_{-j}^\top = (1, -b_{-j}^\top)$. Here, the first inequality holds from the fact that $\|XY\|_{\psi_1} \leq 2\|X\|_{\psi_2}\|Y\|_{\psi_2}$ for any two random variables X, Y; the second inequality comes from

$$q^{-1/2}(E|X_{i,\cdot}^\top \gamma_{-j}|^q)^{1/q} = \|\gamma_{-j}\|q^{-1/2}\{E|X_{i,\cdot}^\top(\gamma_{-j}/\|\gamma_{-j}\|)|^q\}^{1/q} \leq \|\gamma_{-j}\|\kappa$$

and the third inequality follows from

$$\|\gamma_{-j}\|_2 = \sqrt{1 + \|b_{-j}\|^2} \leq 1 + \|b_{-j}\| \leq 1 + \sqrt{\lambda_{\max}(\mathbf{\Sigma}_{-j,-j}^{-1})\mathbf{\Sigma}_{j,j}}.$$

By Lemma 1, for any $\varepsilon > 0$, we have

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n Z_{i,j}^{(k)}\right| \geq \varepsilon\right) \leq 2\exp\left\{-n\min\left(\frac{1}{8e^2}\left(\frac{\varepsilon}{\kappa_{0j}}\right)^2, \frac{1}{4e}\frac{\varepsilon}{\kappa_{0j}}\right)\right\}.$$

Choosing $\varepsilon = \varepsilon_{0j}\sqrt{n^{-1}\log p}$ and assuming that $n \geq \varepsilon_{0j}^2(\kappa_{0j})^{-2}\log p$, then

$$\mathbb{P}\left(\frac{1}{n}\left|\sum_{i=1}^n Z_{i,j}^{(k)}\right| \geq \varepsilon_{0j}\sqrt{\frac{\log p}{n}}\right) \leq 2\exp\left\{-\frac{1}{8e^2}\frac{\varepsilon_{0j}^2}{(\kappa_{0j})^2}\log p\right\}.$$

The result follows from the union bound over $k \in \{1, 2, \dots, p-1\}$. $\square$

An implication of Lemma 3 is that

$$n^{-1}\|\theta_j^\top \mathbf{X}_{-j}\|_\infty \leq \varepsilon_{0j}\sqrt{\frac{\log p}{n}} \tag{S3.1}$$

with probability tending to 1 for a fixed ε_{0j} such that $\varepsilon_0^2 > (\kappa_{0j})^2 8e^2$. We introduce an additional result below for a later use.

Lemma 4. *Under Assumption 5, we have*

$$\mathbb{P}\left(\left|\frac{n}{\theta_j^\top X_j} - \frac{1}{\mathbf{\Sigma}_{j\setminus -j}}\right| \le \varepsilon_{1j}\right) \ge 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\setminus -j}^2\varepsilon_{1j}}{4\kappa_1}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\setminus -j}^2\varepsilon_{1j}}{4\kappa_{2j}}\right)^2 n\right\}$$

for $0 < \varepsilon_{1j} \le \min\{(\mathbf{\Sigma}_{j\setminus -j})^{-1}, 4\min(\kappa_1, \kappa_{2j})(\mathbf{\Sigma}_{j\setminus -j})^{-2}\}$ *and*

$$\mathbb{P}\left(\left|\frac{\|\theta_j\|^2}{n} - \mathbf{\Sigma}_{j\setminus -j}\right| \le \varepsilon_{2j}\right) \ge 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{3\kappa_1}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{6\kappa_{2j}}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{3\kappa_{3j}}\right)^2 n\right\}$$

for $0 < \varepsilon_{2j} \le 3\min(\kappa_1, 2\kappa_{2j}, \kappa_{3j})$.

Proof of Lemma 4. We notice that

$$\theta_j^\top X_j = X_j^\top X_j - \sum_{i=1}^n \sum_{k=1}^{p-1} b_{-j,k} X_{i,-j}^{(k)} X_{i,j},$$

where $b_{-j,k}$ is the kth element of b_{-j} and $X_{i,-j}^{(k)}$ is the kth element of $\mathbf{X}_{-j}$.

Then, we see that

$$\frac{\theta_j^\top X_j}{n} - \mathbf{\Sigma}_{j\setminus -j} = \frac{1}{n}\sum_{i=1}^n (X_{i,j}^2 - \mathbf{\Sigma}_{j,j}) - \frac{1}{n}\sum_{i=1}^n\left(\sum_{k=1}^{p-1} b_{-j,k}X_{i,-j}^{(k)}X_{i,j} - \mathbf{\Sigma}_{j,-j}\mathbf{\Sigma}_{-j,-j}^{-1}\mathbf{\Sigma}_{-j,j}\right).$$

By Lemma 1,

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n (X_{i,j}^2 - \mathbf{\Sigma}_{j,j})\right| \le \delta_j\right) \ge 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\delta_j}{\kappa_1}\right)^2 n\right\}$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\left(\sum_{k=1}^{p-1} b_{-j,k}X_{i,-j}^{(k)}X_{i,j} - \mathbf{\Sigma}_{j,-j}\mathbf{\Sigma}_{-j,-j}^{-1}\mathbf{\Sigma}_{-j,j}\right)\right| \le \delta_j\right) \ge 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\delta_j}{\kappa_{2j}}\right)^2 n\right\}$$

for $0 < \delta_j \le \min(\kappa_1, \kappa_{2j})$. Also, for $\varepsilon_{1j} \le (\boldsymbol{\Sigma}_{j\backslash -j})^{-1}$, we have

$$\begin{aligned}
&\left\{ \left| \frac{n}{\theta_j^\top X_j} - \frac{1}{\boldsymbol{\Sigma}_{j\backslash -j}} \right| \ge \varepsilon_{1j} \right\} \\
\subset &\left[\left\{ \left| \frac{1}{n} \sum_{i=1}^n (X_{i,j}^2 - \boldsymbol{\Sigma}_{j,j}) \right| \ge \frac{\boldsymbol{\Sigma}_{j\backslash -j}^2}{4} \varepsilon_{1j} \right\} \right. \\
&\left. \bigcup \left\{ \left| \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^{p-1} b_{-j,k} X_{i,-j}^{(k)} X_{i,j} - \boldsymbol{\Sigma}_{j,-j} \boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\Sigma}_{-j,j} \right) \right| \ge \frac{\boldsymbol{\Sigma}_{j\backslash -j}^2}{4} \varepsilon_{1j} \right\} \right].
\end{aligned}$$

Thus, for $\varepsilon_{1j} \le \min\{(\boldsymbol{\Sigma}_{j\backslash -j})^{-1}, 4\min(\kappa_1, \kappa_{2j})(\boldsymbol{\Sigma}_{j\backslash -j})^{-2}\}$, we have

$$\begin{aligned}
\mathbb{P}\left(\left| \frac{n}{\theta_j^\top X_j} - \frac{1}{\boldsymbol{\Sigma}_{j\backslash -j}} \right| \le \varepsilon_{1j} \right) \ge 1 - 2\exp\left\{ -\frac{1}{8e^2} \left(\frac{\boldsymbol{\Sigma}_{j\backslash -j}^2 \varepsilon_{1j}}{4\kappa_1} \right)^2 n \right\} \\
- 2\exp\left\{ -\frac{1}{8e^2} \left(\frac{\boldsymbol{\Sigma}_{j\backslash -j}^2 \varepsilon_{1j}}{4\kappa_{2j}} \right)^2 n \right\}
\end{aligned}$$

which proves the first inequality. Next, we note that

$$\begin{aligned}
\frac{\|\theta_j\|^2}{n} - \boldsymbol{\Sigma}_{j\backslash -j} = &\underbrace{\left(\frac{X_j^\top X_j}{n} - \boldsymbol{\Sigma}_{j,j} \right)}_{(*)} \underbrace{-2 \left(\frac{X_j^\top \mathbf{X}_{-j}}{n} - \boldsymbol{\Sigma}_{j,-j} \right) \boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\Sigma}_{-j,j}}_{(**)} \\
&+ \underbrace{\boldsymbol{\Sigma}_{j,-j} \boldsymbol{\Sigma}_{-j,-j}^{-1} \left(\frac{{\mathbf{X}_{-j}}^\top \mathbf{X}_{-j}}{n} - \boldsymbol{\Sigma}_{-j,-j} \right) \boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\Sigma}_{-j,j}}_{(***)}.
\end{aligned}$$

The concentration inequalities for $(*)$ and $(**)$ are given respectively as

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n (X_{i,j}^2 - \boldsymbol{\Sigma}_{j,j}) \right| \le \frac{\varepsilon_{2j}}{3} \right) \ge 1 - 2\exp\left\{ -\frac{1}{8e^2} \left(\frac{\varepsilon_{2j}}{3\kappa_1} \right)^2 n \right\}, \quad \text{(S3.2)}$$

and

$$\begin{aligned}
&\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^n \left(\sum_{k=1}^{p-1} b_{-j,k} X_{i,-j}^{(k)} X_{i,j} - \boldsymbol{\Sigma}_{j,-j} \boldsymbol{\Sigma}_{-j,-j}^{-1} \boldsymbol{\Sigma}_{-j,j} \right) \right| \le \frac{\varepsilon_{2j}}{6} \right) \\
\ge &1 - 2\exp\left\{ -\frac{1}{8e^2} \left(\frac{\varepsilon_{2j}}{6\kappa_{2j}} \right)^2 n \right\}, \qquad \text{(S3.3)}
\end{aligned}$$

for $0 < \varepsilon_{2j} \leq \min(3\kappa_1, 6\kappa_{2j})$. Also, we notice that

$$(***) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{p-1} X_{i,-j}^{(k)} b_{-j,k}\right)^2 - \mathbf{\Sigma}_{j,-j}\mathbf{\Sigma}_{-j,-j}^{-1}\mathbf{\Sigma}_{-j,j}.$$

Lemma 1 gives us

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\left(\sum_{k=1}^{p-1} X_{i,-j}^{(k)} b_{-j,k}\right)^2 - \mathbf{\Sigma}_{j,-j}\mathbf{\Sigma}_{-j,-j}^{-1}\mathbf{\Sigma}_{-j,j}\right)\right| \leq \frac{\varepsilon_{2j}}{3}\right) \\
\geq& 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{3\kappa_{3j}}\right)^2 n\right\} \qquad \text{(S3.4)}
\end{aligned}$$

for $0 < \varepsilon_{2j} \leq 3\kappa_{3j}$. Combining (S3.2), (S3.3) and (S3.4) finishes the proof.

□

The following result directly follows by Lemmas 3 and 4.

Corollary 1. *Let $\breve{v}_j = n\theta_j/(\theta_j^\top X_j)$ and $\breve{u}_j = n^{-1}\|\breve{v}_j^\top \mathbf{X}_{-j}\|_\infty$. Under Assumption 5, $\breve{v}_j$ satisfies $\breve{v}_j^\top X_j = n$,*

$$\begin{aligned}
&\mathbb{P}\left(n^{-1}\|\breve{v}_j\|^2 \leq \left(\frac{1}{\mathbf{\Sigma}_{j\backslash -j}} + \varepsilon_{1j}\right)^2 (\mathbf{\Sigma}_{j\backslash -j} + \varepsilon_{2j})\right) \\
\geq& 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{3\kappa_{1j}}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{6\kappa_{2j}}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\varepsilon_{2j}}{3\kappa_{3j}}\right)^2 n\right\} \\
&- 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\backslash -j}^2\varepsilon_{1j}}{4\kappa_{1j}}\right)^2 n\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\backslash -j}^2\varepsilon_{1j}}{4\kappa_{2j}}\right)^2 n\right\},
\end{aligned}$$

and

$$\mathbb{P}\left(\breve{u}_j \le \varepsilon_{0j}\sqrt{\frac{\log p}{n}}\left(\frac{1}{\mathbf{\Sigma}_{j\setminus -j}} + \varepsilon_{1j}\right)\right)$$
$$\ge 1 - 2\exp\left\{\left(1 - \frac{1}{8e^2}\frac{\varepsilon_{0j}^2}{(\kappa_{0j})^2}\right)\log p\right\} - 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\setminus -j}^2\varepsilon_{1j}}{4\kappa_1}\right)^2 n\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\frac{\mathbf{\Sigma}_{j\setminus -j}^2\varepsilon_{1j}}{4\kappa_{2j}}\right)^2 n\right\},$$

for $\varepsilon_{0j}, \varepsilon_{1j}, \varepsilon_{2j}$ given in Lemmas 3 and 4.

Lemma 5. *Let $R_l = n^{-1}\hat{v}_l^\top \mathbf{X}_{-l}(\beta_{-l} - \hat{\beta}_{-l})$ where $\hat{v}_l$ and $\hat{u}_l$ denote the solution to (3.2). Then,*

$$\max_l R_l = o_p(1).$$

Proof of Lemma 5. According to the definition of $\hat{v}_l$,

$$C_2\frac{n}{\log p}\hat{u}_l^2 + n^{-1}\|\hat{v}_l\|^2 \le C_2\frac{n}{\log p}\breve{u}_l^2 + n^{-1}\|\breve{v}_l\|^2$$

where $\hat{u}_l = n^{-1}\|\hat{v}_l^\top \mathbf{X}_{-l}\|_\infty$, $\breve{v}_l = n\theta_l/\theta_l^\top X_l$ and $\breve{u}_l = n^{-1}\|\breve{v}_l \mathbf{X}_{-l}\|_\infty$. Then, we have

$$\sqrt{C_2\frac{n}{\log p}}\hat{u}_l \le C_2\frac{n}{\log p}\breve{u}_l^2 + n^{-1}\|\breve{v}_l\|^2$$

which implies

$$\sqrt{C_2\frac{n}{\log p}}\max_l n^{-1}\|\hat{v}_l^\top \mathbf{X}_{-l}\|_\infty \le \left(\max_l (\mathbf{\Sigma}_{l\setminus -l})^{-1} + \varepsilon_1'\right)^2\left(C_2(\varepsilon_0')^2 + \max_l \mathbf{\Sigma}_{l\setminus -l} + \varepsilon_2'\right)$$

with probability tending to 1 by (S3.6). Therefore,

$$\max_l R_l \le \max_l n^{-1}\|\hat{v}_l^\top \mathbf{X}_{-l}\|_\infty\sqrt{n}\|\hat{\beta} - \beta\|_1 = o_p(1)$$

by Assumptions 3 and 6. $\square$

The following inequalities are direct consequences of Lemmas 3-4 and the definition of $\hat{v}_l$.

Corollary 2. *Let $\hat{v}_l$ be the solution to (3.2). Then, we have*

$$\mathbb{P}\left(\max_l n^{-1}\|\theta_l^\top \mathbf{X}_{-l}\|_\infty \geq \varepsilon_0' \sqrt{\frac{\log p}{n}}\right) \leq 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{\kappa_{0l}^2}\right)(\varepsilon_0')^2 \log p + 2\log p\right\},$$

$$\mathbb{P}\left(\max_l \frac{n}{|\theta_l^\top X_l|} \leq \max_l (\mathbf{\Sigma}_{l\backslash -l})^{-1} + \varepsilon_1'\right) \geq 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{(\mathbf{\Sigma}_{l\backslash -l})^4}{(4\kappa_1)^2}\right)(\varepsilon_1')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{(\mathbf{\Sigma}_{l\backslash -l})^4}{(4\kappa_{2l})^2}\right)(\varepsilon_1')^2 n + \log p\right\},$$

$$\mathbb{P}\left(\max_l \frac{\|\theta_l\|^2}{n} \leq \max_l \mathbf{\Sigma}_{l\backslash -l} + \varepsilon_2'\right) \geq 1 - 2\exp\left\{-\frac{1}{8e^2(3\kappa_1)^2}(\varepsilon_2')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{(6\kappa_{2l})^2}\right)(\varepsilon_2')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{(3\kappa_{3l})^2}\right)(\varepsilon_2')^2 n + \log p\right\},$$

and

$$\mathbb{P}\left(\max_l n^{-1}\|\hat{v}_l\|^2 \leq M'\right) \geq 1 - 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{(\kappa_{0l})^2}\right)(\varepsilon_0')^2 \log p + 2\log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{(\mathbf{\Sigma}_{l\backslash -l})^4}{(4\kappa_1)^2}\right)(\varepsilon_1')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{(\mathbf{\Sigma}_{l\backslash -l})^4}{(4\kappa_{2l})^2}\right)(\varepsilon_1')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2(3\kappa_1)^2}(\varepsilon_2')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{(6\kappa_{2l})^2}\right)(\varepsilon_2')^2 n + \log p\right\}$$
$$- 2\exp\left\{-\frac{1}{8e^2}\left(\min_l \frac{1}{(3\kappa_{3l})^2}\right)(\varepsilon_2')^2 n + \log p\right\},$$

where

$$M' = (\max_l (\mathbf{\Sigma}_{l\backslash -l})^{-1} + \varepsilon_1')^2 (C_2 \varepsilon_0'^2 + \max_l \mathbf{\Sigma}_{l\backslash -l} + \varepsilon_2') \tag{S3.5}$$

for $0 < \varepsilon_0' \leq (\min_l \kappa_{0l})\sqrt{n(\log p)^{-1}}, 0 < \varepsilon_1' \leq \min_l \left\{\min((\mathbf{\Sigma}_{l\backslash -l})^{-1}, 4\min(\kappa_1, \kappa_{2l})(\mathbf{\Sigma}_{l\backslash -l})^{-2})\right\}$ *and* $0 < \varepsilon_2' \leq \min_l(\min(3\kappa_1, 6\kappa_{2l}, 3\kappa_{3l}))$.

Under Assumption 6, Corollary 2 implies that

$$\begin{aligned}
&\max_l n^{-1}\|\theta_l^\top \mathbf{X}_{-l}\|_\infty \leq \varepsilon_0' \sqrt{\frac{\log p}{n}},\\
&\max_l \frac{n}{|\theta_l^\top X_l|} \leq \max_l (\mathbf{\Sigma}_{l\backslash -l})^{-1} + \varepsilon_1',\\
&\max_l n^{-1}\|\theta_l\|^2 \leq \max_l \mathbf{\Sigma}_{l\backslash -l} + \varepsilon_2',\\
&\max_l n^{-1}\|\hat{v}_l\|^2 \leq M',
\end{aligned} \tag{S3.6}$$

with probability tending to 1 for a fixed $(\varepsilon_0')^2 \min_l(\kappa_{0l})^{-2} > 16e^2$ and fixed $\varepsilon_1', \varepsilon_2'$ as in Corollary 2.

Proof of Proposition 1. Noting that $(n^{-1/2}\|\hat{v}_l\|)^{-1} \leq \|X_l\|/\sqrt{n}$, we have

$$\begin{aligned}
|T_l| &= \frac{\sigma}{\hat{\sigma}} \left| \frac{\sqrt{n}(\tilde{\beta}_l(\hat{v}_l) - \beta_l)}{\sigma n^{-1/2}\|\hat{v}_l\|} + \frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|} \right| \\
&= \frac{\sigma}{\hat{\sigma}} \left| \frac{1}{\sigma n^{-1/2}\|\hat{v}_l\|} \left(\frac{1}{\sqrt{n}} \hat{v}_l^\top \epsilon + \sqrt{n} R_l \right) + \frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|} \right| \\
&\leq \frac{\sigma}{\hat{\sigma}} \left(|Z_l| + \left| \frac{\sqrt{n} R_l}{\sigma n^{-1/2}\|\hat{v}_l\|} \right| + \left| \frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|} \right| \right) \\
&\leq \frac{\sigma}{\hat{\sigma}} \left(|Z_l| + \frac{\|X_l\|}{\sqrt{n}} \left| \frac{\sqrt{n} R_l}{\sigma} \right| + \frac{\|X_l\|}{\sqrt{n}} \left| \frac{\sqrt{n}\beta_l}{\sigma} \right| \right),
\end{aligned}$$

and

$$
\begin{aligned}
|T_l| &\geq \frac{\sigma}{\hat{\sigma}}\left(\left|\frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right| - |Z_l| - \left|\frac{\sqrt{n}R_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right|\right) \\
&\geq \frac{\sigma}{\hat{\sigma}}\left(\left|\frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right| - |Z_l| - \frac{\|X_l\|}{\sqrt{n}}\left|\frac{\sqrt{n}R_l}{\sigma}\right|\right),
\end{aligned}
$$

where $R_l = n^{-1}\hat{v}_l^\top \mathbf{X}_{-l}(\beta_{-l} - \hat{\beta}_{-l})$. We also observe that

$$
\begin{aligned}
&\left[\left\{\left|\frac{\hat{\sigma}}{\sigma} - 1\right| \leq \varepsilon\right\} \bigcap \left\{\max_{l \in \mathcal{B}_j^{(2)}} |Z_l| + D' \max_{l \in \mathcal{B}_j^{(2)}} \left|\frac{\sqrt{n}R_l}{\sigma}\right| + D' \max_{l \in \mathcal{B}_j^{(2)}} \left|\frac{\sqrt{n}\beta_l}{\sigma}\right| \leq \sqrt{\tau \log p}\right\}\right. \\
&\qquad \left. \bigcap \left\{\max_l \frac{\|X_l\|}{\sqrt{n}} \leq D'\right\}\right] \subset \left\{\max_{l \in \mathcal{B}_j^{(2)}} |T_l| \leq \sqrt{\tau \log p}\right\},
\end{aligned}
$$

and

$$
\begin{aligned}
&\left[\left\{\min_{l \in \mathcal{B}_j^{(1)}} \left|\frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right| - \max_{l \in \mathcal{B}_j^{(1)}} |Z_l| - D' \max_{l \in \mathcal{B}_j^{(1)}} \left|\frac{\sqrt{n}R_l}{\sigma}\right| > \sqrt{\tau \log p}\right\}\right. \\
&\left. \bigcap \left\{\left|\frac{\hat{\sigma}}{\sigma} - 1\right| \leq \varepsilon\right\} \bigcap \left\{\max_l \frac{\|X_l\|}{\sqrt{n}} \leq D'\right\}\right] \subset \left\{\min_{l \in \mathcal{B}_j^{(1)}} |T_l| > \sqrt{\tau \log p}\right\}
\end{aligned}
$$

and where $D' = \sqrt{\max_l \mathbf{\Sigma}_{l,l} + \varepsilon'}$ for $0 < \varepsilon' \leq 2\kappa^2$. Note that

$$
\mathbb{P}\left\{\max_l \frac{\|X_l\|}{\sqrt{n}} \leq D'\right\} \geq 1 - 2\exp\left\{-\frac{1}{8e^2}\frac{(\varepsilon')^2}{4\kappa^4}n + \log p\right\}. \tag{S3.7}
$$

We prove Proposition 1 in the following two steps.

1. Under Assumption 1, it suffices to show that

$$
\mathbb{P}\left(\max_{l \in \mathcal{B}_j^{(2)}} |Z_l| + \frac{D'}{\sigma} \max_{l \in \mathcal{B}_j^{(2)}} |\sqrt{n}R_l| \leq c_1\sqrt{\log p}\right) \to 1,
$$

where $c_1 = \sqrt{\tau} - D'\sqrt{d_0}$. We have, for $\varepsilon'' > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(2)}}|Z_l| + \frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(2)}}|\sqrt{n}R_l| \le c_1\sqrt{\log p}\right)\\
\ge&\mathbb{P}\left(\left\{\max_{l\in\mathcal{B}_j^{(2)}}|Z_l| \le c_1\sqrt{\log p} - \varepsilon''\right\}\bigcap\left\{\frac{D'}{\sigma}\max_{k\in\mathcal{B}_j^{(2)}}|\sqrt{n}R_l| \le \varepsilon''\right\}\right)\\
\ge&\mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(2)}}|Z_l| \le c_1\sqrt{\log p} - \varepsilon''\right) + \mathbb{P}\left(\frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(2)}}|\sqrt{n}R_l| \le \varepsilon''\right) - 1\\
\ge&\mathbb{P}\left(\frac{D'}{\sigma}\max_{k\in\mathcal{B}_j^{(2)}}|\sqrt{n}R_k| \le \varepsilon''\right) - 2p\exp\left\{-\frac{\sigma^2(c_1\sqrt{\log p} - \varepsilon'')^2}{32e\kappa_\epsilon^2}\right\}.
\end{aligned}$$

Here the last inequality follows by Lemma 2 under Assumption 2, i.e.,

$$\begin{aligned}
&\mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(2)}}|Z_l| \ge c_1\sqrt{\log p} - \varepsilon''\right)\\
\le&\mathbb{P}\left(\bigcup_{l\in\mathcal{B}_j^{(2)}}\left\{|Z_l| \ge c_1\sqrt{\log p} - \varepsilon''\right\}\right)\\
\le&|\mathcal{B}_j^{(2)}| \times \mathbb{P}\left(\left|\frac{1}{\sigma\|\hat{v}_l\|}\sum_{i=1}^n \hat{v}_{li}\epsilon_i\right| \ge c_1\sqrt{\log p} - \varepsilon''\right)\\
\le&|\mathcal{B}_j^{(2)}| \times 2 \times \exp\left\{-\frac{\sigma^2(c_1\sqrt{\log p} - \varepsilon'')^2}{32e\kappa_\epsilon^2}\right\}
\end{aligned} \tag{S3.8}$$

conditional on $\hat{v}_l$. By the assumption

$$\frac{\sigma^2}{32e\kappa_\epsilon^2}(\sqrt{\tau} - \sqrt{d_0\max_l \mathbf{\Sigma}_{l,l}})^2 > 1,$$

we have

$$\frac{\sigma^2}{32e\kappa_\epsilon^2}c_1^2 > 1$$

for small enough ε'. Together with (S3.7), Lemma 5 and Assumption

4, we obtain

$$\mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(2)}} |T_l| \leq \sqrt{\tau \log p}\right) \to 1.$$

2. We define $c_2 = \sqrt{d_1/M''} - \sqrt{\tau}$, where

$$M'' = \left(\min_l \mathbf{\Sigma}_{l\backslash -l} + \varepsilon_1'\right)^2 \left(\frac{2C_2}{8e^2}\left(\min_l \frac{1}{(\kappa_{0l})^2}\right)^{-1} + \max_l \mathbf{\Sigma}_{l\backslash -l} + 2\varepsilon_2'\right)$$

by letting $(\varepsilon_0')^2 = 2\left((8e^2)^{-1}\min_l(\kappa_{0l})^{-2}\right)^{-1} + \varepsilon_2'$ in (S3.5). We have, for $\varepsilon'' > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\min_{l\in\mathcal{B}_j^{(1)}}\left|\frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right| - \max_{l\in\mathcal{B}_j^{(1)}}|Z_l| - D'\max_{l\in\mathcal{B}_j^{(1)}}\left|\frac{\sqrt{n}R_l}{\sigma}\right| > \sqrt{\tau\log p}\right) \\
\geq&\mathbb{P}\left(\left\{\min_{l\in\mathcal{B}_j^{(1)}}\left|\frac{\sqrt{n}\beta_l}{\sigma n^{-1/2}\|\hat{v}_l\|}\right| - \max_{l\in\mathcal{B}_j^{(1)}}|Z_l| > \sqrt{\tau\log p} + \varepsilon''\right\}\bigcap\left\{\frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(1)}}\left|\sqrt{n}R_l\right| \leq \varepsilon''\right\}\right) \\
\geq&\mathbb{P}\left(\left\{\min_{l\in\mathcal{B}_j^{(1)}}\left|\frac{\sqrt{n}\beta_l}{\sigma\sqrt{M''}}\right| - \max_{l\in\mathcal{B}_j^{(1)}}|Z_l| > \sqrt{\tau\log p} + \varepsilon''\right\}\bigcap\left\{\min_{l\in\mathcal{B}_j^{(1)}}\frac{1}{n^{-1/2}\|\hat{v}_l\|} \geq \frac{1}{\sqrt{M''}}\right\}\right) \\
&+\mathbb{P}\left(\frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(1)}}|\sqrt{n}R_l| \leq \varepsilon''\right) - 1 \\
\geq&\mathbb{P}\left(\min_{l\in\mathcal{B}_j^{(1)}}\left|\frac{\sqrt{n}\beta_l}{\sigma\sqrt{M''}}\right| - \max_{l\in\mathcal{B}_j^{(1)}}|Z_l| > \sqrt{\tau\log p} + \varepsilon''\right) + \mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(1)}} n^{-1}\|\hat{v}_l\|^2 \leq M''\right) \\
&+\mathbb{P}\left(\frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(1)}}|\sqrt{n}R_l| \leq \varepsilon''\right) - 2 \\
\geq&\mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(1)}}|Z_l| < c_2\sqrt{\log p} - \varepsilon''\right) + \mathbb{P}\left(\max_{l\in\mathcal{B}_j^{(1)}} n^{-1}\|\hat{v}_l\|^2 \leq M''\right) + \mathbb{P}\left(\frac{D'}{\sigma}\max_{l\in\mathcal{B}_j^{(1)}}|\sqrt{n}R_l| \leq \varepsilon''\right) - 2 \\
\geq&1 - 2\left|\mathcal{B}_j^{(1)}\right|\exp\left\{-\frac{\sigma^2}{32e\kappa_\epsilon^2}(b\sqrt{\log p} - \varepsilon'')^2\right\} + \mathbb{P}\left(\max_{k\in\mathcal{B}_j^{(1)}} n^{-1}\|\check{v}_k\|^2 \leq M''\right) \\
&+\mathbb{P}\left(\frac{D'}{\sigma}\max_{k\in\mathcal{B}_j^{(1)}}|\sqrt{n}R_k| \leq \varepsilon''\right) - 2
\end{aligned}$$

where the last inequality follows from (S3.8). By the assumption $\sqrt{d_1/M} - \sqrt{\tau} > 0$, we have $c_2 = \sqrt{d_1/M''} - \sqrt{\tau} > 0$ for small enough $\varepsilon_1', \varepsilon_2'$. Since $\left|\mathcal{B}_j^{(1)}\right| \leq s_0 \ll p$, by (S3.6), (S3.7), Lemma 5 and Assumption 4, we get $\mathbb{P}\left(\min_{l \in \mathcal{B}_j^{(1)}} |T_l| > \sqrt{\tau \log p}\right) \to 1$.

□

Proof of Theorem 1. The argument below is conditional on the event $\{\mathcal{A}_j^{(k)}(\tau) = \mathcal{B}_j^{(k)}$ for $k = 1, 2\}$ which occurs almost surely by Proposition 1. Let $\breve{u}_{j1} = \max_{k \in \mathcal{A}_j^{(1)}(\tau)} n^{-1}|\breve{v}_j^\top X_k|$ and $\breve{u}_{j1} = \max_{k \in \mathcal{A}_j^{(2)}(\tau)} n^{-1}|\breve{v}_j^\top X_k|$ where $\breve{v}_j$ is as in Corollary 1. Then, $(\breve{u}_{j1}, \breve{u}_{j1}, \breve{v}_j)$ is a feasible point to problem (2.8). By the definition of $\tilde{v}_j$,

$$C_1 \frac{n}{\log p} \tilde{u}_{j1}^2 + C_2 \frac{n}{\log p} \tilde{u}_{j2}^2 + n^{-1}\|\tilde{v}_j\|^2 \leq C_1 \frac{n}{\log p} \breve{u}_{j1}^2 + C_2 \frac{n}{\log p} \breve{u}_{j2}^2 + n^{-1}\|\breve{v}_j\|^2,$$

where $\tilde{u}_{j1} = \max_{k \in \mathcal{A}_j^{(1)}} n^{-1}|\tilde{v}_j^\top X_k|$ and $\tilde{u}_{j2} = \max_{k \in \mathcal{A}_j^{(2)}} n^{-1}|\tilde{v}_j^\top X_k|$. Then, for $i = 1, 2$, we must have

$$\sqrt{C_i \frac{n}{\log p} \tilde{u}_{ji}} \leq \max\{C_1, C_2\}\varepsilon_{0j}^2 \left(\frac{1}{\mathbf{\Sigma}_{j\backslash -j}} + \varepsilon_{1j}\right)^2 + \left(\frac{1}{\mathbf{\Sigma}_{j\backslash -j}} + \varepsilon_{1j}\right)^2 (\mathbf{\Sigma}_{j\backslash -j} + \varepsilon_{2j}),$$

with probability tending to 1 by Corollary 1. Then, by Assumptions 3 and 6,

$$|\sqrt{n} R(\tilde{v}_j, \beta_{-j})| = n^{-1/2}|\tilde{v}_j^\top \mathbf{X}_{-j}(\beta_{-j} - \hat{\beta}_{-j})| \leq n^{-1} \max_{k \neq j} |\tilde{v}_j^\top X_k| \sqrt{n} \|\hat{\beta}_{-j} - \beta_{-j}\|_1 = o_p(1).$$

Hence, we obtain

$$\sqrt{n}(\tilde{\beta}_j(\tilde{v}_j) - \beta_j) = \frac{1}{\sqrt{n}}\tilde{v}_j^\top \epsilon + o_p(1). \tag{S3.9}$$

Note that

$$\frac{\sum_{i=1}^n E[(\tilde{v}_{j,i}\epsilon_i)^{2+\delta}|\tilde{v}_j]}{\sigma^{2+\delta}\|\tilde{v}_j\|^{2+\delta}} = \frac{E\epsilon_1^{2+\delta}}{\sigma^{2+\delta}}\frac{\|\tilde{v}_j\|_{2+\delta}^{2+\delta}}{\|\tilde{v}_j\|^{2+\delta}} = o_{a.s.}(1).$$

Conditional on the event that $\{\|\tilde{v}_j\|_{2+\delta}/\|\tilde{v}_j\| \to 0\}$, the Lyapunov condition is satisfied and thus $\tilde{v}_j^\top\epsilon/\{\sigma\|\tilde{v}_j\|\}$ converges to $N(0,1)$. If $\epsilon \sim N(0,\sigma^2\mathbf{I})$, $\tilde{v}_j^\top\epsilon/\{\sigma\|\tilde{v}_j\|\} \sim N(0,1)$ conditional on $\tilde{v}_j$. The conclusion thus follows from (S3.9) and Assumption 4 by the Slutsky's theorem. $\square$

Proof of Proposition 2. All the arguments below are conditional on the event $\{\mathcal{A}_j^{(2)} = \mathcal{B}_j^{(2)}\}$ which occurs almost surely by Proposition 1. With the projection direction $\bar{v}_j$ from (3.8) and the refitted least square estimator $\check{\beta}$, the bias (2.6) reduces to

$$\begin{aligned}
\sqrt{n}R(\bar{v}_j, \beta_{-j}) &= \frac{1}{\sqrt{n}}\sum_{k\neq j}\bar{v}_j^\top X_k(\beta_k - \check{\beta}_k) \\
&= \frac{1}{\sqrt{n}}\sum_{k\in\mathcal{B}_j^{(1)}}\bar{v}_j^\top X_k(\beta_k - \check{\beta}_k) + \frac{1}{\sqrt{n}}\sum_{k\in\mathcal{B}_j^{(2)}}\bar{v}_j^\top X_k(\beta_k - \check{\beta}_k) \\
&= \frac{1}{\sqrt{n}}\sum_{k\in\mathcal{A}_j^{(1)}}\bar{v}_j^\top X_k(\beta_k - \check{\beta}_k) + \frac{1}{\sqrt{n}}\sum_{k\in\mathcal{A}_j^{(2)}}\bar{v}_j^\top X_k(\beta_k - \check{\beta}_k) \\
&= \frac{1}{\sqrt{n}}\sum_{k\in\mathcal{A}_j^{(2)}}\bar{v}_j^\top X_k\beta_k,
\end{aligned}$$

where we have used the fact that $\bar{v}_j^\top X_k = 0$ for $k \in \mathcal{A}_j^{(1)}$ from (3.8) and $\check{\beta}_{\mathcal{A}_j^{(2)}} = 0$ by (3.9). Thus, we have

$$\begin{aligned}
|\sqrt{n}R(\bar{v}_j, \beta_{-j})| &\leq \|n^{-1}\bar{v}_j^\top \mathbf{X}_{-j}\|_\infty \sqrt{n}\|\beta_{\mathcal{A}_j^{(2)}}\|_1 \\
&\leq \|n^{-1}\bar{v}_j^\top \mathbf{X}_{-j}\|_\infty \sigma\sqrt{d_0 \log p}\|\beta_{\mathcal{B}_j^{(2)}}\|_0 \\
&\leq O_p\left(\sqrt{\frac{\log p}{n}}\right)\sigma\sqrt{d_0 \log p}\|\beta_{\mathcal{B}_j^{(2)}}\|_0
\end{aligned}$$

where the second inequality holds by Assumption 1 under the event $\{\mathcal{A}_j^{(2)} = \mathcal{B}_j^{(2)}\}$. The last inequality follows from the fact that $\|n^{-1}\bar{v}_j^\top \mathbf{X}_{-j}\|_\infty = O_p(\sqrt{\log p/n})$, which can be verified by using similar arguments as in the proof of Corollary 1 together with the definition of $\bar{v}_j$ under Assumption 5. The last statement follows immediately from condition (3.12). $\square$

S3.3 Technical details in Section 4

We first state the following results which are parallel to Lemma 3 and the first inequality in Corollary 2. As the proof is similar to the one in Lemma 3, we omit the details.

Corollary 3. *Let $\theta_l = X_l - \mathbf{X}_{-S}b_l$ with $b_l = \text{argmin}_{\tilde{b}} E\|X_l - \mathbf{X}_{-S}\tilde{b}\|^2$ for $l \in S$. Under Assumption 5,*

$$\mathbb{P}\left(n^{-1}\|\theta_l^\top \mathbf{X}_{-S}\|_\infty \geq \xi_{0l}\sqrt{\frac{\log p}{sn}}\right) \leq 2\exp\left\{\left(1 - c_{l,S}\frac{\delta_{0l}^2}{s(\xi_{0l})^2}\right)\log p\right\}$$

for $0 < \xi_{0l} \leq \kappa_{0l}\sqrt{sn(\log p)^{-1}}$ *where* $c_{l,S} > 0$ *is an absolute constant and* $\kappa_{0l} = 2\left(1 + \sqrt{\Lambda_{\min}^{-1}\mathbf{\Sigma}_{l,l}}\right)\kappa^2$. *As a consequence, we have*

$$\mathbb{P}\left(\max_{l\in S} n^{-1}\|\theta_l^\top \mathbf{X}_{-S}\|_\infty \geq \xi_0'\sqrt{\frac{\log p}{sn}}\right) \leq 2\exp\left\{-\left(\min_l \frac{c_{l,S}}{s(\kappa_{0l})^2}\right)(\xi_0')^2\log p + 2\log p\right\}.$$

for $0 < \xi_0' \leq \min_l \kappa_{0l}\sqrt{sn(\log p)^{-1}}$.

The following results are introduced for the proof of Theorem 2 which follows from a direct application of Proposition 2.1 in Vershynin (2012).

Lemma 6. *For every* $\delta > 0$, *we have*

$$\mathbb{P}\left(\|n^{-1}\mathbf{X}_S^\top\mathbf{X}_S - \mathbf{\Sigma}_{S,S}\| \leq \sqrt{\frac{4}{C_\kappa}\frac{s}{n}\log\frac{2}{\delta}}\right) \geq 1-\delta,$$

where $C_\kappa > 0$ *is an absolute constant which only depends on* δ *and* κ.

We next introduce the following lemma which provides an upper bound for the operator norm of a matrix.

Lemma 7. *Let* $\mathbf{B}$ *be a* $m\times m$ *matrix and* $\mathcal{N}_\varepsilon$ *be an* ε*-net of the unit sphere* $\mathcal{S}^{m-1}$ *for some* $\varepsilon \in (0, 1/2)$. *Then*

$$\|\mathbf{B}\| \leq (1-2\varepsilon)^{-1}\sup_{c,d\in\mathcal{N}_\varepsilon}\left|c^\top\mathbf{B}d\right|.$$

Proof of Lemma 7. For any $c, d \in \mathcal{S}^{m-1}$, we can choose $c_\mathcal{N}, d_\mathcal{N} \in \mathcal{N}_\varepsilon$ such that $\max\{\|c - c_\mathcal{N}\|, \|d - d_\mathcal{N}\|\} \leq \varepsilon$. Some algebra gives us

$$c^\top\mathbf{B}d = c_\mathcal{N}^\top\mathbf{B}d_\mathcal{N} + (c - c_\mathcal{N})^\top\mathbf{B}d + c_\mathcal{N}^\top\mathbf{B}(d - d_\mathcal{N}),$$

which implies that

$$\left|c^\top \mathbf{B} d\right| \le 2\varepsilon \|\mathbf{B}\| + \sup_{c_\mathcal{N}, d_\mathcal{N} \in \mathcal{N}_\varepsilon} \left|c_\mathcal{N}^\top \mathbf{B} d_\mathcal{N}\right|.$$

Taking supremum over all $c, d \in \mathcal{S}^{m-1}$ and rearranging terms give us the desired result. $\square$

Lemma 8. *For every $\delta > 0$, we have*

$$\mathbb{P}\left(\|(n^{-1}\mathbf{X}_S^\top \mathbf{X}_{-S} - \mathbf{\Sigma}_{S,-S})\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\| \le 3\sqrt{\frac{8}{C_{\kappa'}} \log \frac{2}{\delta} \frac{s}{n}}\right) \ge 1 - \delta$$

where $C_{\kappa'} > 0$ denotes an absolute constant which only depends on $\kappa' = 2\kappa^2\sqrt{\Lambda_{\min}^{-1} D^2}$.

Proof of Lemma 8. We prove the result in several steps. First, we bound the operator norm by using the so-called ε-net argument. Then we apply the concentration inequality for sub-exponential random variables and finally use the union bound to finish the proof. For two vectors $a, b \in \mathbb{R}^{q\times 1}$, write $\langle a, b\rangle = a^\top b$. By Lemma 7 and Lemma 5.2 in Vershynin (2012), we have

$$\begin{aligned}
&\|(n^{-1}\mathbf{X}_S^\top \mathbf{X}_{-S} - \mathbf{\Sigma}_{S,-S})\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\| \\
=& \sup_{c,d\in\mathcal{S}^{s-1}} \left|\frac{1}{n}\sum_{i=1}^n \left(\langle X_{i,S}, c\rangle \langle \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1} X_{i,-S}, d\rangle - c^\top \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S} d\right)\right| \\
\le& 3 \sup_{c,d\in\mathcal{N}_{1/3}} \left|\frac{1}{n}\sum_{i=1}^n \left(\langle X_{i,S}, c\rangle \langle \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1} X_{i,-S}, d\rangle - c^\top \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S} d\right)\right|
\end{aligned}$$

where $\mathcal{N}_{1/3}$ denotes a 1/3-net of $\mathcal{S}^{s-1}$ with the covering number $|\mathcal{N}_{1/3}| \le 7^s$.

Let us fix $c, d \in \mathcal{N}_{1/3}$. Because each row of $\mathbf{X}_S$ and $\mathbf{X}_{-S}$ is independent sub-gaussian random vector, we can apply the concentration inequality in Corollary 5.17 of Vershynin (2010). Specifically, we have

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left(\langle X_{i,S}, c\rangle\langle \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}X_{i,-S}, d\rangle - c^{\top}\mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}d\right)\right| \geq \epsilon\right)$$
$$\leq 2\exp\left(-cn\frac{\epsilon^2}{(\kappa')^2}\right)$$

provided that $\epsilon^2 \leq (\kappa')^2$, where $\|\langle X_{i,S}, c\rangle\langle \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}X_{i,-S}, d\rangle\|_{\psi_1} \leq \kappa'$ and $c > 0$ is an absolute constant. Applying the union bound over $c, d \in \mathcal{N}_{1/3}$, we have

$$\sup_{c,d\in\mathcal{N}_{1/3}}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\langle X_{i,S}, c\rangle\langle \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}X_{i,-S}, d\rangle - c^{\top}(\mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S})d\right)\right| \geq \epsilon$$

with probabiliity at most $2|\mathcal{N}_{1/3}|^2\exp\left[-cn\epsilon^2/(\kappa')^2\right]$, which implies

$$\begin{aligned}\mathbb{P}\left(\|(n^{-1}\mathbf{X}_S^{\top}\mathbf{X}_{-S} - \mathbf{\Sigma}_{S,-S})\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\| < 3\epsilon\right) &\geq 1 - 2|\mathcal{N}_{1/3}|^2\exp\left[-cn\left(\frac{\epsilon}{\kappa'}\right)^2\right]\\ &\geq 1 - 2\exp\left[4s - n\epsilon^2 C_{\kappa'}\right]\end{aligned}$$

where $C_{\kappa'} = c/(\kappa')^2$. Then by letting $\epsilon^2 = (8/C_{\kappa'})\log(2/\delta)(s/n)$, we have

$$\mathbb{P}\left(\|(n^{-1}\mathbf{X}_S^{\top}\mathbf{X}_{-S} - \mathbf{\Sigma}_{S,-S})\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\| \leq 3\sqrt{\frac{8}{C_{\kappa'}}\log\frac{2}{\delta}\frac{s}{n}}\right) \geq 1 - \delta$$

which completes the proof. □

Lemma 9. *Let* $\hat{\mathbf{A}} = n^{-1}\mathbf{X}_S^{\top}\mathbf{\Theta}$ *and* $\mathbf{A} = \mathbf{\Sigma}_{S,S} - \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}$. *Under the assumption that* $s/n = o(1)$ *and* $\|\mathbf{A}^{-1}\| \leq B$ *for some constant* $B > 0$, *we have* $\|w\| = O_p(\|a_S\|)$.

Proof of Lemma 9. Note that

$$\|w\| = \|(n^{-1}\mathbf{X}_S^\top \mathbf{\Theta})^{-1} a_S\| \leq \|\hat{\mathbf{A}}^{-1}\| \|a_S\|.$$

We want to bound $\|\hat{\mathbf{A}}^{-1}\|$. Using the properties of operator norm, we have

$$\|\hat{\mathbf{A}}^{-1}\| \leq \|\hat{\mathbf{A}}^{-1} - \mathbf{A}^{-1}\| + \|\mathbf{A}^{-1}\| \leq \|\hat{\mathbf{A}}^{-1}\| \|\mathbf{A}^{-1}\| \|\mathbf{A} - \hat{\mathbf{A}}\| + \|\mathbf{A}^{-1}\|.$$

Rearranging the terms, we obtain

$$\|\hat{\mathbf{A}}^{-1}\| (1 - \|\mathbf{A}^{-1}\| \|\mathbf{A} - \hat{\mathbf{A}}\|) \leq \|\mathbf{A}^{-1}\|.$$

With the assumption $\|\mathbf{A}^{-1}\| \leq B$, we have

$$1 - \|\mathbf{A}^{-1}\| \|\mathbf{A} - \hat{\mathbf{A}}\| \geq 1 - B\|\mathbf{A} - \hat{\mathbf{A}}\|.$$

Under the assumption $s/n = o(1)$, by Lemmas 6 and 8, we have $\|\mathbf{A} - \hat{\mathbf{A}}\| = o_p(1)$. Thus $1 - \|\mathbf{A}^{-1}\| \|\mathbf{A} - \hat{\mathbf{A}}\|$ is bounded from below by a positive constant with probability tending to one. Thus

$$\|\hat{\mathbf{A}}^{-1}\| \leq (1 - \|\mathbf{A}^{-1}\| \|\mathbf{A} - \hat{\mathbf{A}}\|)^{-1} \|\mathbf{A}^{-1}\| \leq (1 - B\|\mathbf{A} - \hat{\mathbf{A}}\|)^{-1} \|\mathbf{A}^{-1}\|$$

which implies that $\|\hat{\mathbf{A}}^{-1}\| = O_p(1)$. The conclusion follows directly. □

Lemma 10. *Let $\mathbf{\Theta} \in \mathbb{R}^{n \times s}$ where the l-th column vector is θ_l for $l \in S$ as in Corollary 3 and $\breve{v}_a = \mathbf{\Theta} w$ where $w = (n^{-1}\mathbf{X}_S^\top \mathbf{\Theta})^{-1} a_S$. Then, under Assumption 5 and $\|a_S\| = O(1)$, we have $n^{-1}\|\breve{v}_a\|^2 = O_p(1)$.*

Proof of Lemma 10. We note that

$$\|\mathbf{\Theta} w\|^2 \leq \|\mathbf{X}_S - \mathbf{X}_{-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\|^2\|w\|_2^2 \leq 2\underbrace{\left\{\|\mathbf{X}_S\|^2 + \|\mathbf{X}_{-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\|^2\right\}}_{I}\|w\|_2^2.$$

We shall control I below. Lemma 5.3 in Vershynin (2012) gives us

$$\|\mathbf{X}_S\|^2 \leq 4 \max_{c\in\mathcal{N}_{1/2}} c^\top \mathbf{X}_S'\mathbf{X}_S c,$$

$$\|\mathbf{X}_{-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\|^2 \leq 4 \max_{d\in\mathcal{N}_{1/2}} d^\top \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{X}_{-S}'\mathbf{X}_{-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}d.$$

Let $\mathbf{Q} = \mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\left(n^{-1}\mathbf{X}_{-S}^\top\mathbf{X}_{-S} - \mathbf{\Sigma}_{-S,-S}\right)\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}$. Since the elements of the terms inside the maximization can be expressed as a sum of independent sub-exponential random variables, we can use similar arguments as in the proof of Lemma 3 to show that for every $\delta > 0$,

$$\mathbb{P}\left(\|\mathbf{Q}\| \leq \sqrt{\frac{4}{C_{\kappa'}}\frac{s}{n}\log\frac{2}{\delta}}\right) \leq 1-\delta$$

where $C_{\kappa'} > 0$ is an absolute constant which only depends on $\kappa' = 2\kappa^2\sqrt{\Lambda_{\min}^{-1}D^2}$. Together with Lemma 6, we have

$$\begin{aligned} n^{-1}\{\|\mathbf{X}_S\|^2 + \|\mathbf{X}_{-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S}\|^2\} \leq& C_0\left\{o_p(1) + \lambda_{\max}(\mathbf{\Sigma}_{S,S}) + \lambda_{\max}(\mathbf{\Sigma}_{S,-S}\mathbf{\Sigma}_{-S,-S}^{-1}\mathbf{\Sigma}_{-S,S})\right\} \\ \leq& C_0\left\{o_p(1) + 2\lambda_{\max}(\mathbf{\Sigma}_{S,S})\right\}, \end{aligned}$$

for some constant C_0. Therefore, we have

$$n^{-1}\|\mathbf{\Theta} w\|_2^2 \leq 2C_0(o_p(1) + 2\Lambda_{\min}^{-2})O_p(\|a_S\|^2) = O_p(1).$$

□

Proof of Theorem 2. The arguments below are conditional on the sets $\mathcal{A}_S^{(1)}$ and $\mathcal{A}_S^{(2)}$ which have nonrandom limits by Proposition 1. Let $\breve{u}_{a1} = \max_{k\in\mathcal{A}_j^{(1)}} n^{-1}|\breve{v}_a^\top X_k|$ and $\breve{u}_{a2} = \max_{k\in\mathcal{A}_j^{(2)}} n^{-1}|\breve{v}_a^\top X_k|$, where $\breve{v}_a$ is as in Lemma 10. Then, $(\breve{u}_{a1}, \breve{u}_{a2}, \breve{v}_a)$ is a feasible point to problem (4.2). By the definition of $\tilde{v}_a$,

$$C_1\frac{n}{\log p}\tilde{u}_{a1}^2 + C_2\frac{n}{\log p}\tilde{u}_{a2}^2 + n^{-1}\|\tilde{v}_a\|^2 \le C_1\frac{n}{\log p}\breve{u}_{a1}^2 + C_2\frac{n}{\log p}\breve{u}_{a2}^2 + n^{-1}\|\breve{v}_a\|^2.$$

Then, for $i = 1, 2$, we must have

$$\begin{aligned}\sqrt{C_i\frac{n}{\log p}}\tilde{u}_{ai} &\le \max\{C_1, C_2\}\frac{n}{\log p}\max_{k\notin S} n^{-1}|w^\top\boldsymbol{\Theta}^\top X_k| + n^{-1}\|\breve{v}_a\|^2\\ &\le \max\{C_1, C_2\}\|w\|(\xi_0')^2 + M_a\end{aligned}$$

with probability tending to 1 for $0 < \xi_0' \le \min_l \kappa_{0l}\sqrt{sn(\log p)^{-1}}$ and some constant M_a according to Corollary 3 and Lemma 10. Then, by Assumptions 3 and 6,

$$|\sqrt{n}R(\tilde{v}_a, \beta_{-S})| = n^{-1/2}|\tilde{v}_a^\top\mathbf{X}_{-S}(\beta_{-S} - \hat{\beta}_{-S})| \le n^{-1}\max_{k\notin S}|\tilde{v}_a^\top X_k|\sqrt{n}\|\hat{\beta}_{-S} - \beta_{-S}\|_1 = o_p(1).$$

Hence, we obtain

$$\sqrt{n}(\tilde{\beta}_S(\tilde{v}_a) - a_S^\top\beta_S) = \frac{1}{\sqrt{n}}\tilde{v}_a^\top\epsilon + o_p(1). \tag{S1}$$

Finally we can apply the central limit theorem as in the proof of Theorem 1, which completes the proof. □

S4 Additional numerical results

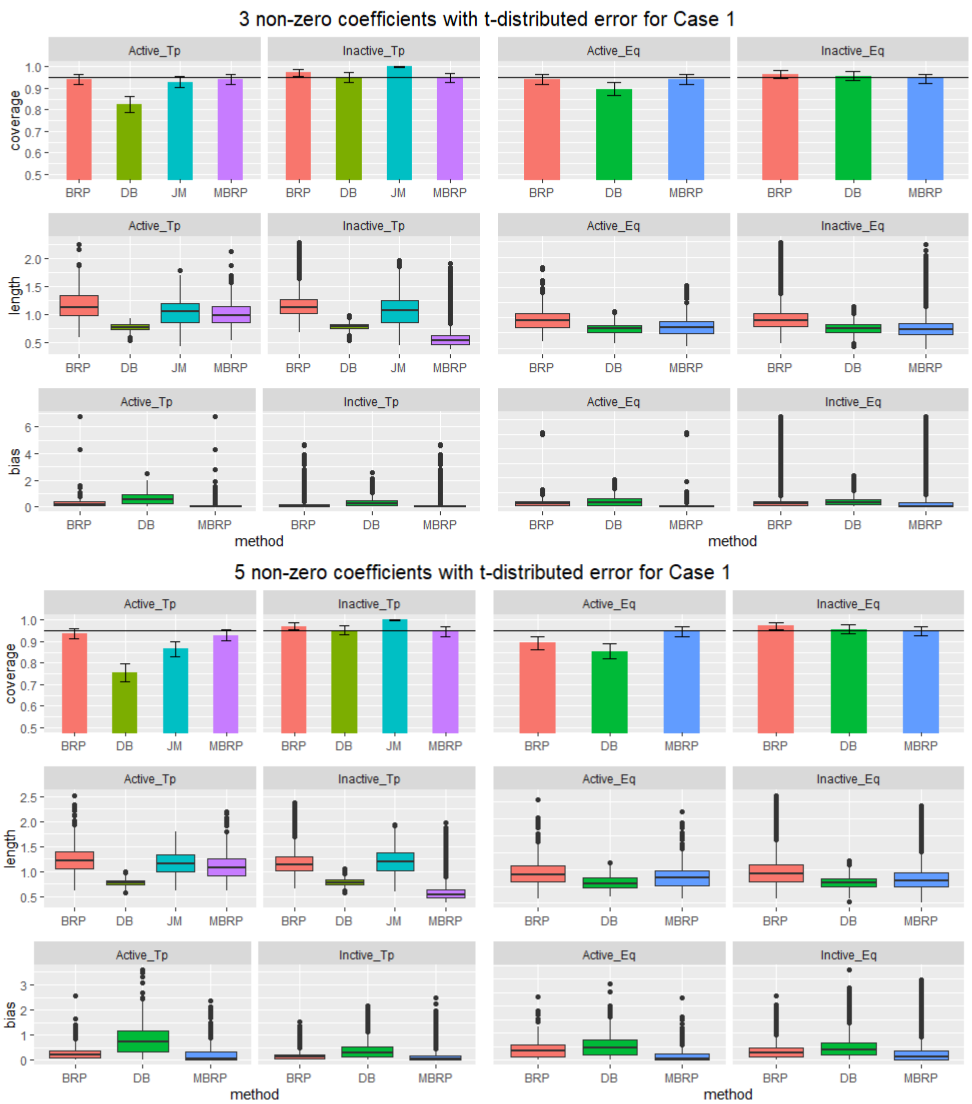

Figure 12: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 3, 5$ and t-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

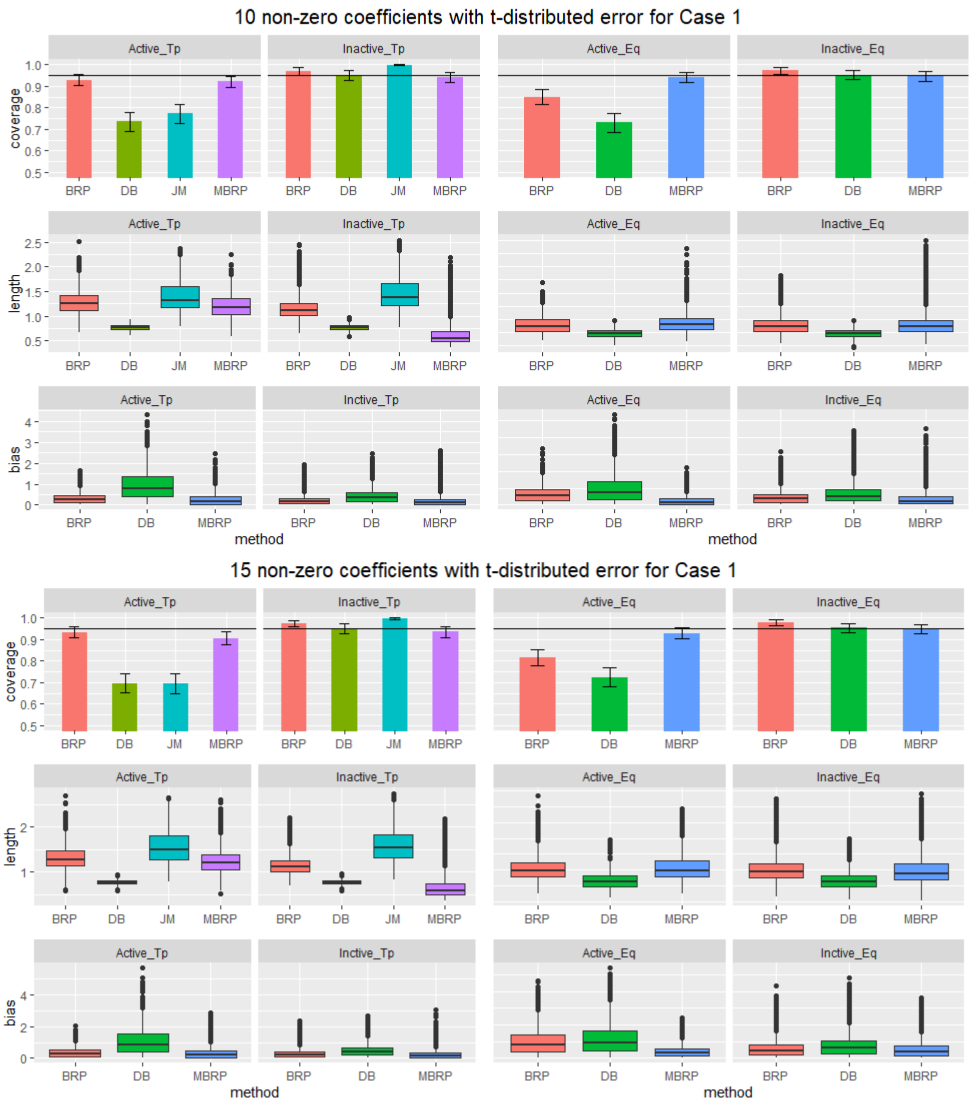

Figure 13: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 10, 15$ and t-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

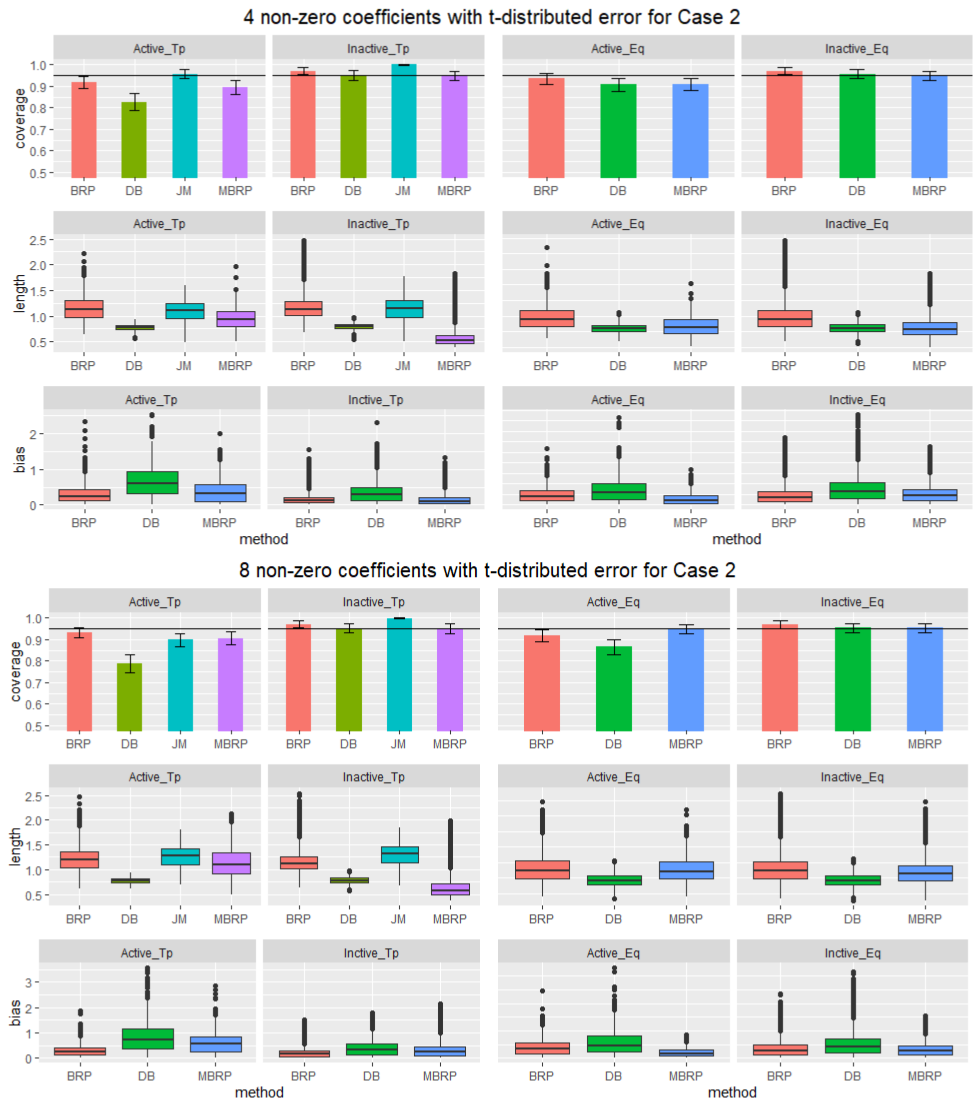

Figure 14: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 4, 8$ and t-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

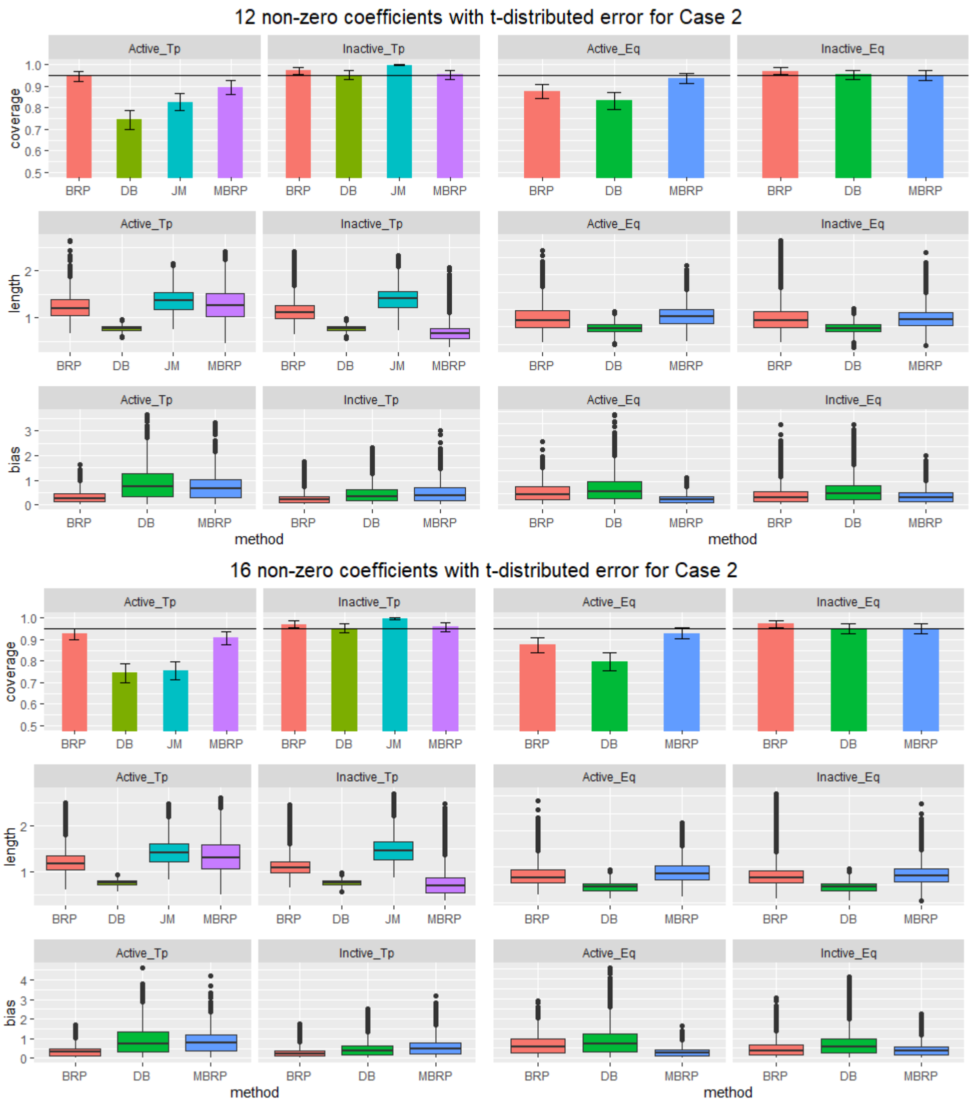

Figure 15: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 12, 16$ and t-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

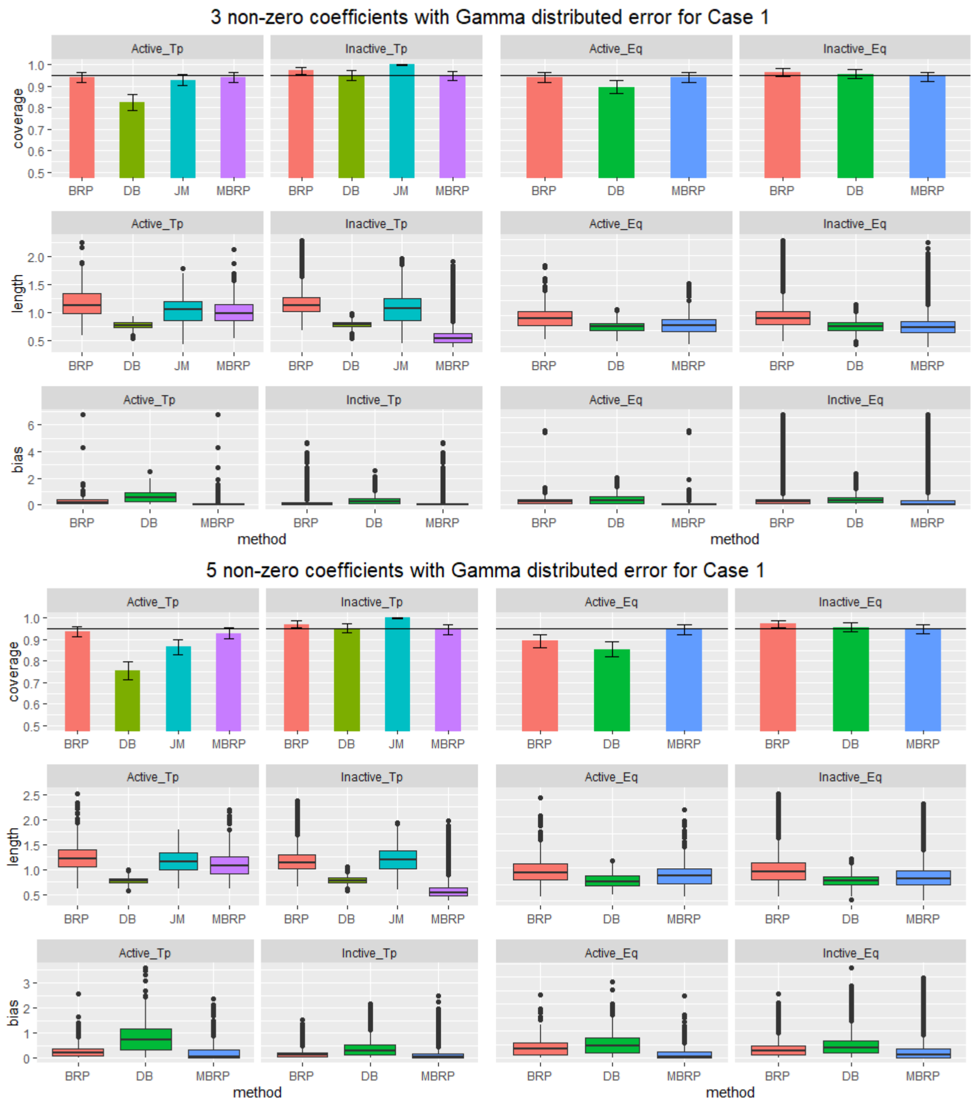

Figure 16: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 3, 5$ and Gamma-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

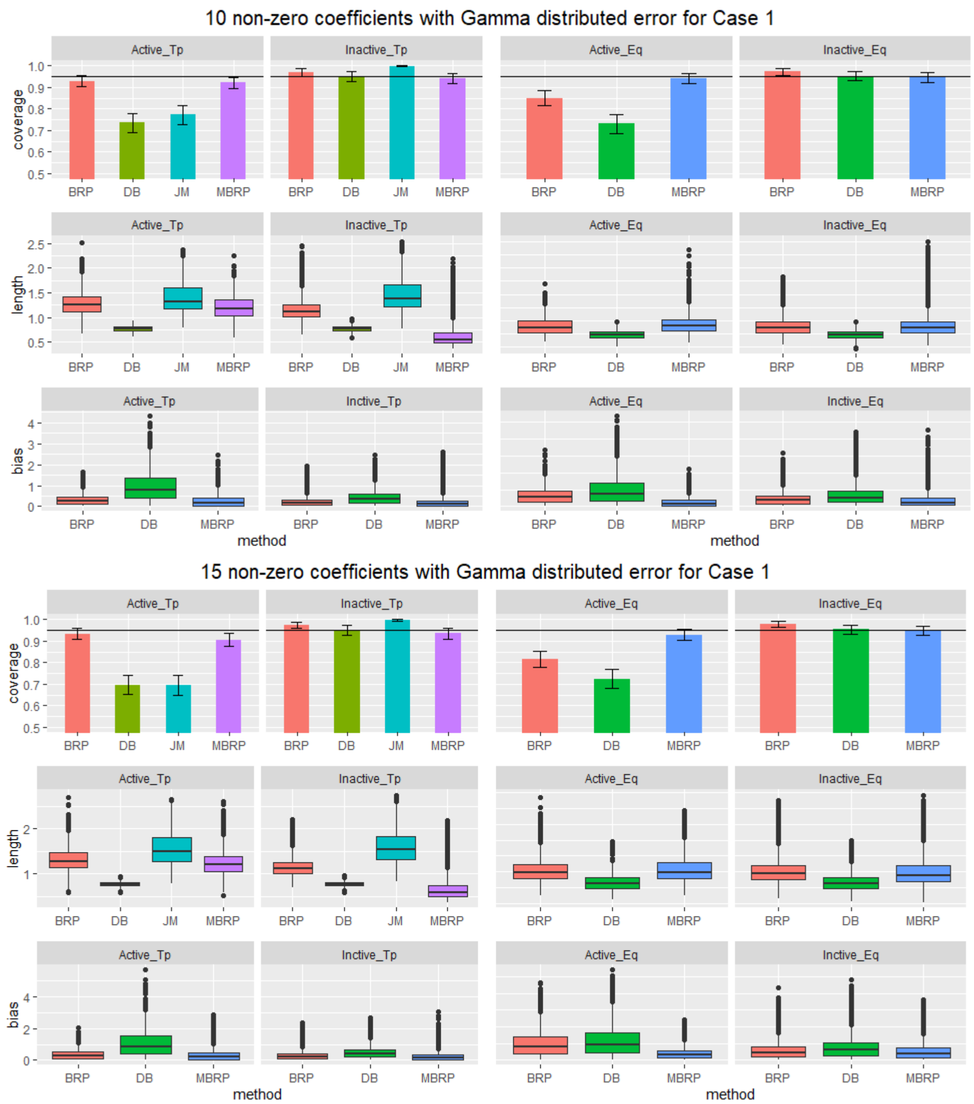

Figure 17: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 1 with $s_0 = 10, 15$ and Gamma-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

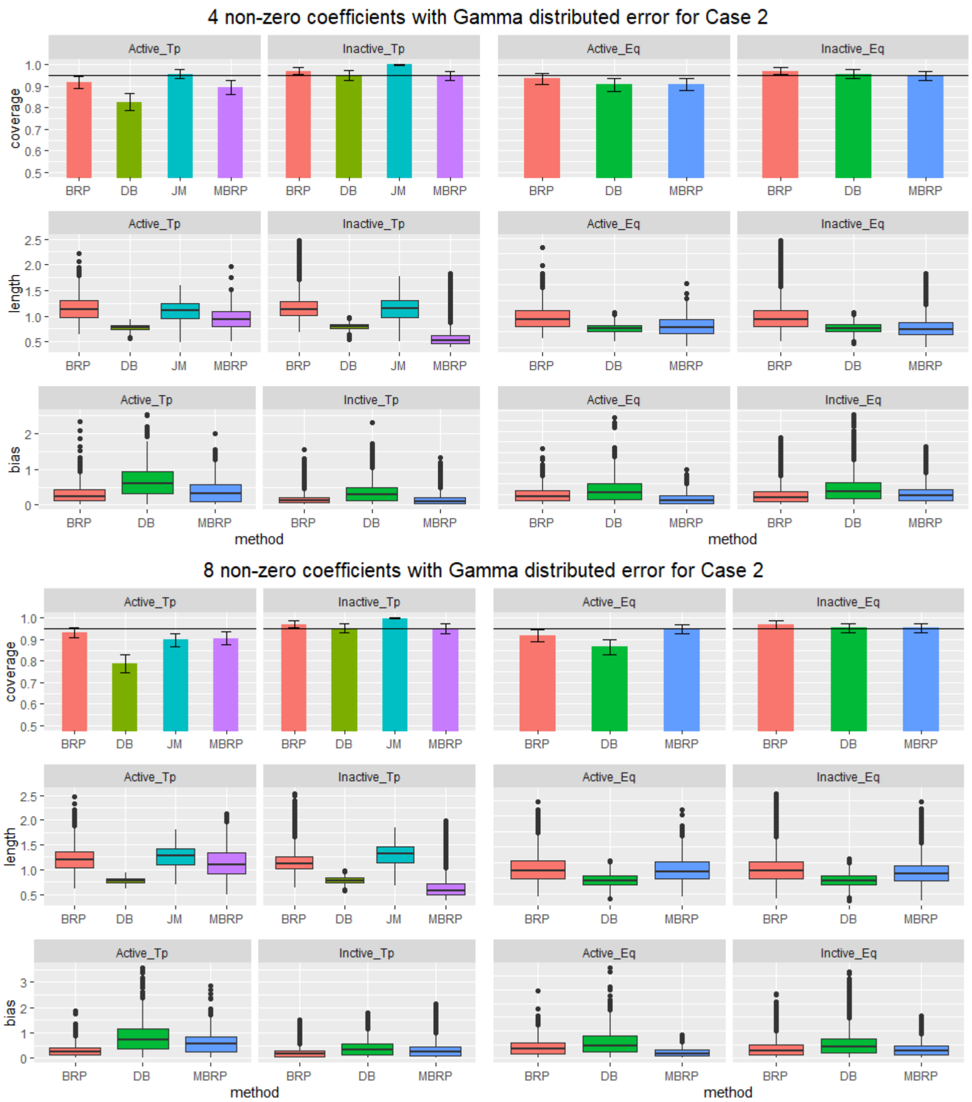

Figure 18: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 4, 8$ and Gamma-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

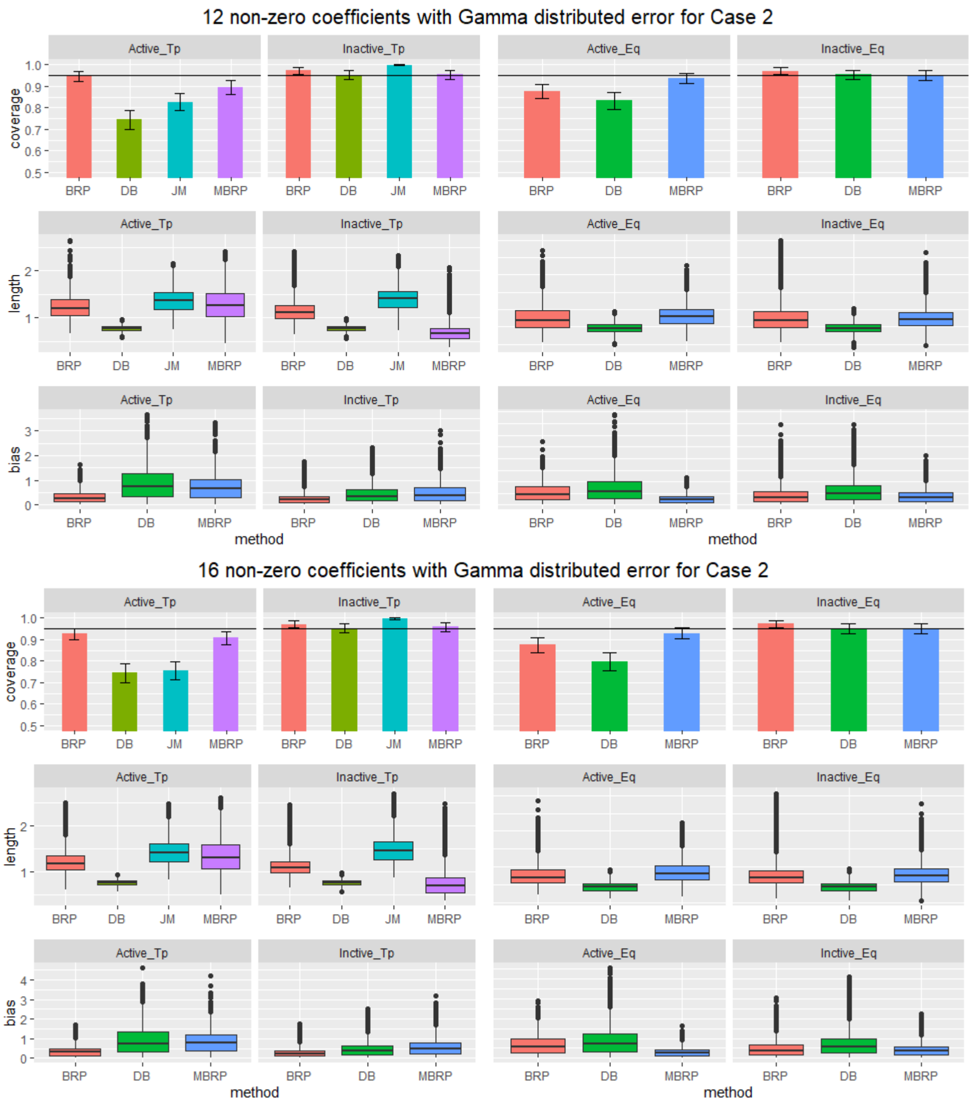

Figure 19: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for Case 2 with $s_0 = 12, 16$ and Gamma-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

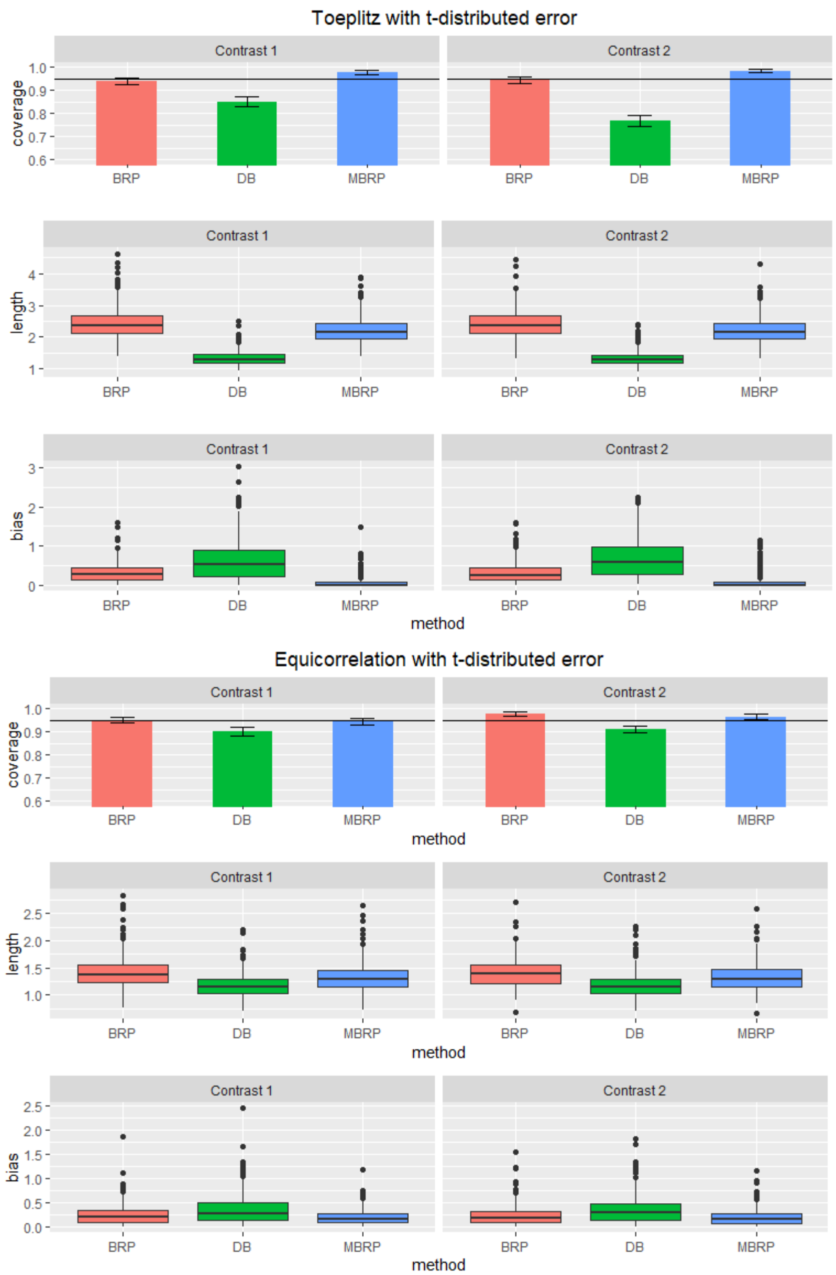

Figure 20: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for different contrasts with t-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

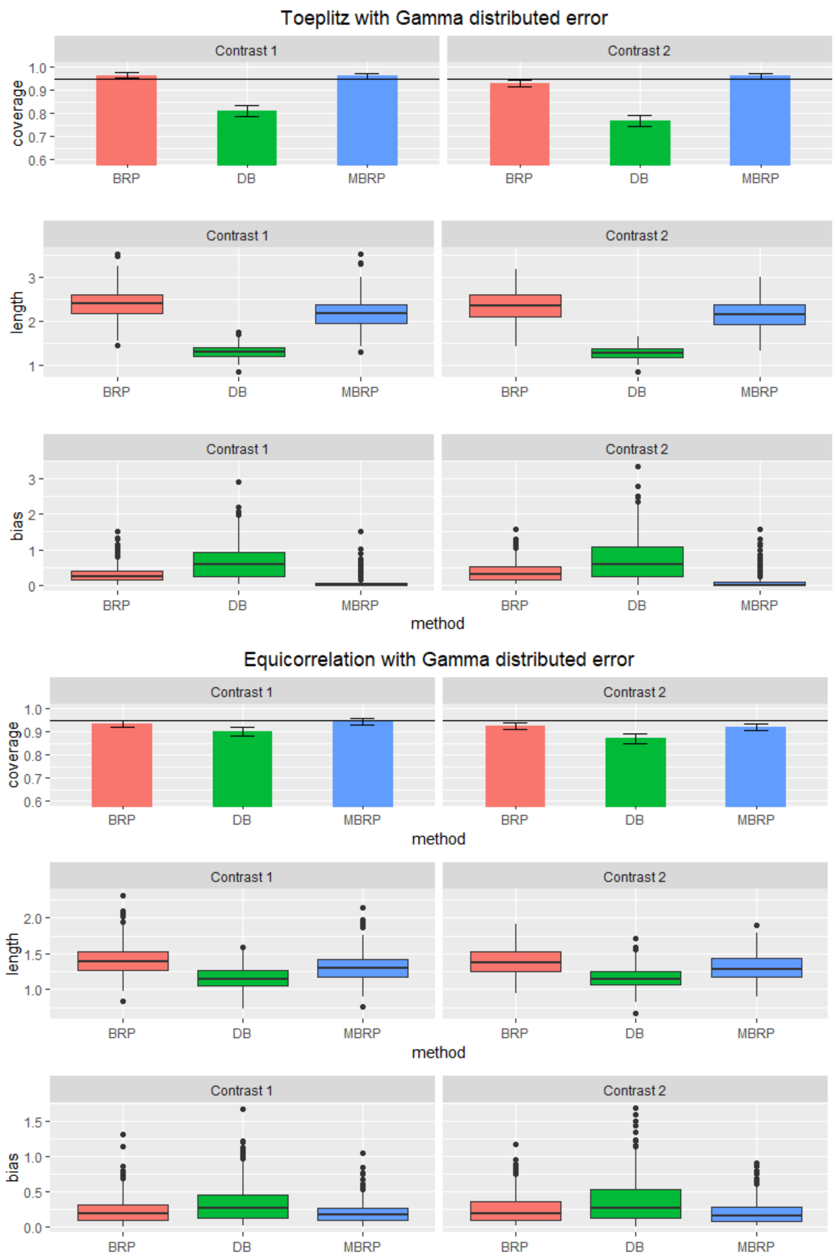

Figure 21: Barplots for the empirical coverage and boxplots for the lengths and biases of the 95% confidence intervals for different contrasts with Gamma-distributed error. The horizontal line in the barplots indicates the nominal level. Error bars in the barplots represent the interval within one standard deviation of the empirical coverage.

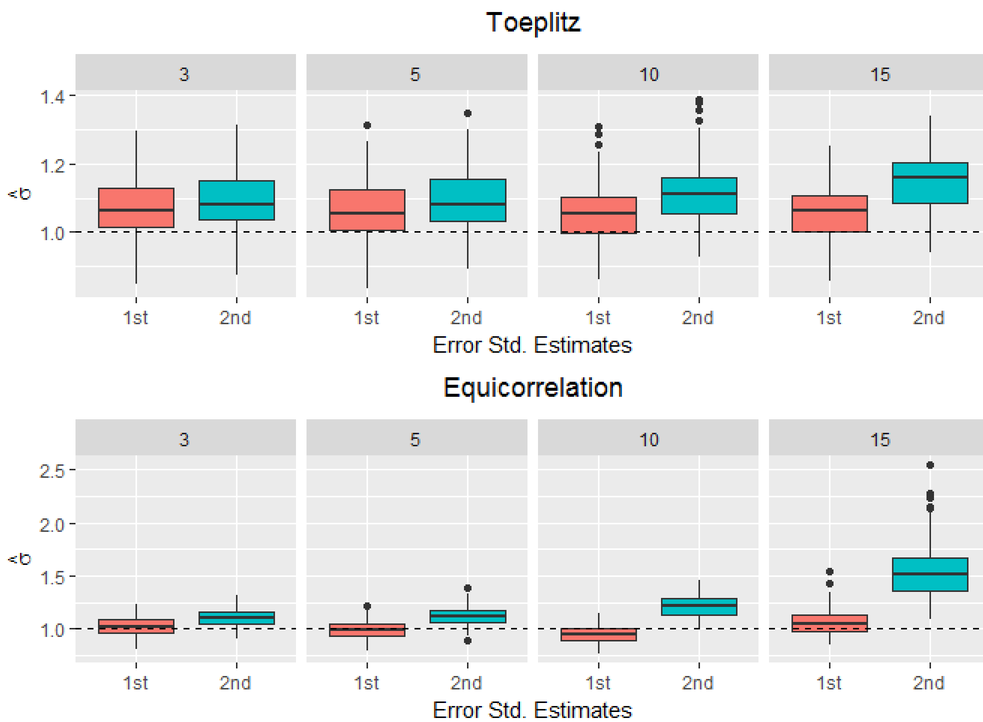

Figure 22: Boxplots of the two different error variance estimators. Data sets are generated by Case 1 with $s_0 = 3, 5, 10$ and 15. "1st" denotes the estimator $\|Y - X\hat{\beta}\|^2/n$ and "2nd" denotes the estimator $\|Y - X\hat{\beta}\|^2/(n - \|\hat{\beta}\|_0)$. The number on the top of each panel denotes the number of non-zero coefficients. The horizontal dashed line corresponds to the true error variance.

References

Vershynin, R. (2010). Introduction to the non-asymptotic analysis of random matrices. arXiv preprint arXiv:1011.3027.

Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, **25**, 655-686.