Supplemental material for "Empirical Likelihood Ratio Tests for Coefficients in High Dimensional Heteroscedastic Linear Models"

Honglang Wang[†], Ping-Shou Zhong[‡1], Yuehua Cui[‡]

[†] Indiana University-Purdue University Indianapolis and [‡]Michigan State University

This supplemental material provides technical proofs to the theoretical results presented in the paper, and more simulation results with different covariance structures for covariates and different distributions for the error term.

1 Technical assumptions

For a symmetric matrix $\mathbf{M} = ((M_{jk}))$, $\lambda_{\min}(\mathbf{M})$ and $\lambda_{\max}(\mathbf{M})$ are the minimal and maximal eigenvalues of $\mathbf{M}$. For any matrix $\mathbf{M} = ((M_{jk}))$, let $\|\mathbf{M}\|_{\max} = \max_{j,k} |M_{jk}|$, $\|\mathbf{M}\|_1 = \max_k \sum_j |M_{jk}|$, $\|\mathbf{M}\|_2 = \sqrt{\lambda_{\max}(\mathbf{M}^\intercal\mathbf{M})}$, and $\|\mathbf{M}\|_\infty = \max_j \sum_k |M_{jk}|$.

Assumption 1. *(1) Assume the initial estimator $\hat{\boldsymbol{\beta}}$ satisfying $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 = O_p(s\sqrt{\log p/n})$.*

(2) Suppose the initial estimators $\hat{\mathbf{w}}_j$ satisfy $\max_{1\le j\le p} \|\hat{\mathbf{w}}_j - \mathbf{w}_j^0\|_1 = O_p(a_n)$, where $a_n = o(1/\sqrt{\log p})$.

(3) The prediction errors satisfy $\|\mathbb{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2/n = O_p(s\log p/n)$ and $\max_{1\le j\le p} \|\mathbb{X}_{\backslash j}(\hat{\mathbf{w}}_j - \mathbf{w}_j^0)\|_2^2/n = O_p(b_n)$, where $\mathbb{X}_{\backslash j}$ is the design matrix $\mathbb{X}$ with the j-th column deleted and $b_n = o(1/\sqrt{n})$.

(4) $\mathbf{X}_i$ and ϵ_i are all sub-Gaussian.

(5) $s\log p/\sqrt{n} = o(1)$.

Remark 1. *1. With (4) that $\mathbf{X}_i$ and ϵ_i are all sub-Gaussian, we have $X_{ik}\epsilon_i$ sub-exponential with $E(\epsilon_i X_{ik}) = 0$. By Bernstein inequality Vershynin (2010) and union bound inequality, we have*

$$P(\|\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\epsilon_i\|_\infty \ge t) \le C_1 p \exp(-C\min(t^2/C_2, t/C_3)n).$$

By taking $t = C'\sqrt{\log p/n}$ for some positive constant C' such that $CC'^2 > C_2$, we have

$$\|\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\epsilon_i\|_\infty = O_p(\sqrt{\log p/n}). \quad (1.1)$$

[1]Corresponding author: Ping-Shou Zhong, C418 Wells Hall, 619 Red Cedar Road, Michigan State University, East Lansing, MI 48824.

2. *For* $\eta_{ij} = X_{ij} - E(X_{ij}|\mathbf{X}_{i,\setminus j})$, *we have* η_{ij} *sub-gaussian since* $\mathbf{X}_i$ *is sub-gaussian. And for any* $k \neq j$, *we have* $E(X_{ik}\eta_{ij}) = E\{X_{ik}[X_{ij} - E(X_{ij}|\mathbf{X}_{i,\setminus j})]\} = E\{X_{ik}X_{ij} - E[X_{ik}X_{ij}|\mathbf{X}_{i,\setminus j}]\} = 0$. *Similarly, we have for any* $t > 0$ *and* $1 \leq j \neq k \leq p$,

$$P(\big|\frac{1}{n}\sum_{i=1}^{n} X_{ik}\eta_{ij}\big| \geq t) \leq C_1 p \exp(-C \min(t^2/C_2, t/C_3)n),$$

which leads to

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}\eta_{ij}\mathbf{X}_{i,\setminus j}\Big\|_\infty = O_p(\sqrt{\log p/n}). \tag{1.2}$$

3. *For the properties of the initial estimators in (1), (2) and (3) under the heteroscedasitic noise case, we can use the* $\sqrt{Lasso}$ *estimator as in Belloni et al. (2014). According to Theorem 7 in Belloni et al. (2014), we have that the* $\sqrt{Lasso}$ *estimators under certain conditions have these properties satisfied.*

Assumption 2. *(1) Assume the same assumption as Lasso projection case for the initial estimator* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 = O_p(s\sqrt{\log p/n})$.

(2) Assume similar assumption as Lasso projection case for the initial estimators $\hat{\boldsymbol{\gamma}}_j$, *i.e.,* $\max_{1\leq j\leq p}\|\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j^0\|_1 = O_p(a_n)$, *where* $a_n = o(1/\sqrt{\log p})$.

(3) Assume similar assumption as Lasso projection case for the prediction errors, i.e., $\|\mathbb{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)\|_2^2/n = O_p(s\log p/n)$ *and* $\max_{1\leq j\leq p}\|(\mathbb{Y}, \mathbb{X}_{\setminus j})(\hat{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_j^0)\|_2^2/n = O_p(b_n)$ *and* $b_n = o(1/\sqrt{n})$.

(4) $(\mathbf{X}_i^\intercal, \epsilon_i)^\intercal$ *is sub-Gaussian.*

(5) $s\log p/\sqrt{n} = o(1)$.

Remark 2. *For the condition (2) above, if we assume* $a = \max_{1\leq j\leq p} s_j$ *with* $s_j = \|\boldsymbol{\gamma}_j^0\|_0$ *and then the* $\sqrt{Lasso}$ *estimators for* $\boldsymbol{\gamma}_j^0$ *satisfy this condition with* $a_n = a\sqrt{\log p/n}$. *For the condition (3) above, since we assume that* $(\mathbf{X}_i^\intercal, \epsilon_i)^\intercal$ *is sub-Gaussian (which makes* $\boldsymbol{\beta}^{0\intercal}\mathbf{X}_i$ *also sub-Gaussian), then due to* $Cov(\boldsymbol{\beta}^{0\intercal}\mathbf{X}_i, \epsilon_i) = E(\epsilon_i\boldsymbol{\beta}^{0\intercal}\mathbf{X}_i) = 0$, *we have* $\epsilon_i\boldsymbol{\beta}^{0\intercal}\mathbf{X}_i$ *sub-exponential and by the Bernstein inequality, we have for any* $t > 0$,

$$P(\big|\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^\intercal\boldsymbol{\beta}^0\epsilon_i\big| \geq t) \leq 2\exp\{-C_1 n \min(t^2/C_2^2, t/C_2)\}.$$

This also leads to

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^\intercal\boldsymbol{\beta}^0\epsilon_i = O_p(\sqrt{\log p/n}), \tag{1.3}$$

as long as $\log p/n \to 0$. *And with the same argument, we have*

$$\frac{1}{n}\sum_{i=1}^{n} X_{ik}\eta_{ij,y} = O_p(\sqrt{\log p/n}), \tag{1.4}$$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i, \mathbf{X}_{i,\setminus j}^{\intercal})\boldsymbol{\gamma}_j^0\eta_{ij,y} = O_p(\sqrt{\log p/n}). \tag{1.5}$$

Assumption 3. *(1) For the eigenvalues of* $\boldsymbol{\Sigma}$*, there exist some constants* $\lambda_{\min}$ *and* $\lambda_{\max}$ *such that* $0 < \lambda_{\min} < \lambda_{\min}(\boldsymbol{\Sigma}) \leq \lambda_{\max}(\boldsymbol{\Sigma}) < \lambda_{\max} < \infty$.

(2) Assume $\mathbf{X}_i \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ *and* ϵ_i *to be sub-Gaussian.*

(3) The initial estimator $\hat{\boldsymbol{\beta}}$ *(e.g., LASSO) satisfies* $\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0\|_1 = O_p(s\sqrt{\log p/n})$.

(4) $s\sqrt{(\log p)^2 m^3/n} = o(1)$ *and* $s\sqrt{(\log p)^3 m^2/n^2} = o(1)$ *where* m *is the upper bound of the size of KFC set* $|\mathcal{S}|$.

(5) Assume $s\sqrt{\log p}\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}\left|\sigma_{jk} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}k}\right| = o(1)$.

Remark 3. *Condition (1) is a mild condition that assures the asymptotic identifiability of the model (Fan and Lv, 2008; Wang, 2009, 2012). Condition (2) is a common condition used for simplification of theoretical proofs in high dimensional setup; see for example, Wang (2009) and Zhang and Zhang (2014). Condition (3) was also used in Assumptions 1 and 2. Condition (4) is for controlling the size of the KFC set* $|\mathcal{S}|$*, and Condition (5) controls the partial correlation between the target covariate* X_{ij} *and* $\mathbf{X}_{i\mathbf{S}^*}$.

2 Technical Proofs

We first proves the Theorems and postpone the proof of Propositions 1-3 about the asymptotic normality in the end since they use the results from the proof of the corresponding Theorems.

2.1 Proof of Theorems

Proof of Theorem 1. As in Owen (2001), by (C0), with probability tending to 1, $-2\log \mathrm{EL}_n(\beta_j^0) = 2\sum_{i=1}^n \log(1+\lambda m_{ni})$ where λ satisfies

$$\sum_{i=1}^{n}\frac{m_{ni}}{1+\lambda m_{ni}} = 0. \tag{2.1}$$

The next step is to bound the magnitude of λ. Let $\lambda = |\lambda|u$ where $u = \mathrm{sign}(\lambda) \in \{-1, 1\}$. Now by $\sum_{i=1}^n m_{ni}/(1+\lambda m_{ni}) = 0$, we have

$$0 = \sum_{i=1}^{n}\frac{u m_{ni}}{1+\lambda m_{ni}} = \sum_{i=1}^{n} u m_{ni}\big\{1 - \frac{\lambda m_{ni}}{1+\lambda m_{ni}}\big\},$$

which implies

$$\sum_{i=1}^n u m_{ni} = \sum_{i=1}^n \frac{u\lambda m_{ni}^2}{1+\lambda m_{ni}} = \sum_{i=1}^n \frac{|\lambda| m_{ni}^2}{1+\lambda m_{ni}} \geq |\lambda| \sum_{i=1}^n \frac{m_{ni}^2}{1+|\lambda| \max_{1\leq i\leq n} |m_{ni}|}.$$

Thus we have

$$u\frac{1}{n}\sum_{i=1}^n m_{ni} \geq \frac{|\lambda|}{1+|\lambda|\max_{1\leq i\leq n}|m_{ni}|}\frac{1}{n}\sum_{i=1}^n m_{ni}^2.$$

which implies

$$|\lambda|\{\frac{1}{n}\sum_{i=1}^n m_{ni}^2 - (\max_{1\leq i\leq n}|m_{ni}|)u\frac{1}{n}\sum_{i=1}^n m_{ni}\} \leq u\frac{1}{n}\sum_{i=1}^n m_{ni}. \tag{2.2}$$

From (C1), by Lemma 3 in Owen (1990), we have $\max_{1\leq i\leq n}|W_{ni}| = o_p(n^{1/2})$, and together with (C2), we have

$$\max_{1\leq i\leq n}|m_{ni}| = o_p(n^{1/2}). \tag{2.3}$$

And since for any $\epsilon > 0$, obviously $W_{ni}^2 \mathbb{1}(|W_{ni}| > \epsilon s_n) \xrightarrow{p} 0$ due to $\mathrm{P}(|W_{ni}| > \epsilon s_n) \to 0$, we have by Dominated Convergence Theorem,

$$\frac{1}{s_n^2}\sum_{i=1}^n \mathrm{E}\big\{W_{ni}^2\mathbb{1}(|W_{ni}| > \epsilon s_n)\big\} \to 0.$$

Thus by Lindeberg-Feller Central Limit Theorem, we have $\sum_{i=1}^n W_{ni}/s_n \xrightarrow{d} \mathrm{N}(0,1)$, that is,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni} \xrightarrow{d} \mathrm{N}(0,\sigma_w^2). \tag{2.4}$$

By (2.4) and together with (C2), we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n m_{ni} = \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni} + \frac{1}{\sqrt{n}}\sum_{i=1}^n R_{ni} = \frac{1}{\sqrt{n}}\sum_{i=1}^n W_{ni} + o_p(1) \xrightarrow{d} \mathrm{N}(0,\sigma_w^2). \tag{2.5}$$

And by (C1) and (C2) we have

$$\frac{1}{n}\sum_{i=1}^n m_{ni}^2 = \frac{1}{n}\sum_{i=1}^n W_{ni}^2 + \frac{1}{n}\sum_{i=1}^n R_{ni}^2 + 2\frac{1}{n}\sum_{i=1}^n W_{ni}R_{ni} = \frac{1}{n}\sum_{i=1}^n W_{ni}^2 + o_p(1) \to \sigma_w^2. \tag{2.6}$$

Actually the above follows from checking the WLLN for triangular arrays. First of all $\sum_{i=1}^n \mathrm{P}(W_{ni}^2 > n) = n\mathrm{P}(W_{n1}^2 > n) \leq \mathrm{E}\big\{W_{n1}^2\mathbb{1}(W_{n1}^2 > n)\big\} \to 0$; and

$$\begin{aligned} n^{-2}\sum_{i=1}^n \mathrm{E}\big\{W_{ni}^4\mathbb{1}(W_{ni}^2 \leq n)\big\} &= n^{-1}\mathrm{E}\big\{W_{n1}^4\mathbb{1}(W_{n1}^2 \leq n)\big\} \\ &= n^{-1}\int_0^n 2y\mathrm{P}(W_{n1}^2 > y)dy \to 0 \end{aligned}$$

since $y\mathrm{P}(W_{n1}^2 > y) \le \mathrm{E}\{W_{n1}^2 \mathbb{1}(W_{n1}^2 > y)\} \to 0$ as $y \to \infty$.

Thus by (2.2), (2.3), (2.5) and (2.6), we have

$$|\lambda|(\frac{1}{n}\sum_{i=1}^n m_{ni}^2 + o_p(1)) = O_p(n^{-1/2})$$

and hence

$$|\lambda| = O_p(n^{-1/2}). \tag{2.7}$$

Then it follows from (2.3), we have $\max_{1\le i\le n}\left|\frac{\lambda m_{ni}}{1+\lambda m_{ni}}\right| = o_p(1)$. Therefore, from (2.1), we have

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n \frac{\lambda m_{ni}}{1+\lambda m_{ni}} = \frac{1}{n}\sum_{i=1}^n \lambda m_{ni}\Big\{1 - \lambda m_{ni} + \frac{[\lambda m_{ni}]^2}{1+\lambda m_{ni}}\Big\} \\ &= \frac{1}{n}\sum_{i=1}^n \lambda m_{ni} - \frac{[1+o_p(1)]}{n}\sum_{i=1}^n [\lambda m_{ni}]^2, \end{aligned}$$

which leads to

$$\frac{1}{n}\sum_{i=1}^n \lambda m_{ni} = \frac{[1+o_p(1)]}{n}\sum_{i=1}^n [\lambda m_{ni}]^2. \tag{2.8}$$

Again by using (2.1) and together with (2.5), we have

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^n \frac{m_{ni}}{1+\lambda m_{ni}} = \frac{1}{n}\sum_{i=1}^n m_{ni}\Big\{1 - \lambda m_{ni} + \frac{[\lambda m_{ni}]^2}{1+\lambda m_{ni}}\Big\} \\ &= \frac{1}{n}\sum_{i=1}^n m_{ni} - \frac{\lambda}{n}\sum_{i=1}^n m_{ni}^2 + \frac{1}{n}\sum_{i=1}^n \frac{m_{ni}[\lambda m_{ni}]^2}{1+\lambda m_{ni}} \\ &= \frac{1}{n}\sum_{i=1}^n m_{ni} - \frac{\lambda}{n}\sum_{i=1}^n m_{ni}^2 + O_p\Big\{\max_{1\le i\le n}\Big|\frac{m_{ni}}{1+\lambda m_{ni}}\Big|\frac{1}{n}\sum_{i=1}^n [\lambda m_{ni}]^2\Big\} \\ &= \frac{1}{n}\sum_{i=1}^n m_{ni} - \frac{\lambda}{n}\sum_{i=1}^n m_{ni}^2 + o_p\Big\{n^{1/2}\lambda^2\frac{1}{n}\sum_{i=1}^n m_{ni}^2\Big\} \\ &= \frac{1}{n}\sum_{i=1}^n m_{ni} - \frac{\lambda}{n}\sum_{i=1}^n m_{ni}^2 + o_p(n^{-1/2}), \end{aligned}$$

which leads to

$$\lambda = \{\frac{1}{n}\sum_{i=1}^n m_{ni}^2\}^{-1}\frac{1}{n}\sum_{i=1}^n m_{ni} + o_p(n^{-1/2}). \tag{2.9}$$

Finally, by Taylor expansion together with (2.5), (2.6), (2.8) and (2.9), we have

$$\begin{aligned}
-2\log \mathrm{EL}_n(\beta_j^0) &= 2\sum_{i=1}^n \log(1+\lambda m_{ni}) \\
&= 2\sum_{i=1}^n \lambda m_{ni} - [1+o_p(1)]\sum_{i=1}^n [\lambda m_{ni}]^2 \\
&= [1+o_p(1)]\sum_{i=1}^n [\lambda m_{ni}]^2 = [1+o_p(1)]\lambda^2 \sum_{i=1}^n m_{ni}^2 \\
&= [1+o_p(1)]\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_{ni}\Big)\Big(\frac{1}{n}\sum_{i=1}^n m_{ni}^2\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n m_{ni}\Big) + o_p(1) \\
&\xrightarrow{d} \chi_1^2, \text{ as } n \to \infty.
\end{aligned}$$

This completes the proof of the theorem. □

Proof of Theorem 2. We only need to control the term R_{ni}, which will be controlled one by one.

By (3) in Assumption 1, we have (1.1) and (1.2), which leads to

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^n R_{ni,1}\Big| &= \Big|\frac{1}{n}\sum_{i=1}^n (Y_i - \mathbf{X}_i^\intercal \boldsymbol{\beta}^0)(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\backslash j}\Big| \\
&= \Big|(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \frac{1}{n}\sum_{i=1}^n \mathbf{X}_{i,\backslash j}\epsilon_i\Big| \le \|\mathbf{w}_j^0 - \hat{\mathbf{w}}_j\|_1 \|\frac{1}{n}\sum_{i=1}^n \mathbf{X}_{i,\backslash j}\epsilon_i\|_\infty \\
&= O_p(a_n)O_p(\sqrt{\frac{\log p}{n}}) = O_p(a_n\sqrt{\frac{\log p}{n}}).
\end{aligned}$$

In order to have $\left|\frac{1}{n}\sum_{i=1}^n R_{ni,1}\right| = o_p(n^{-1/2})$ we need to have $a_n = o(1/\sqrt{\log p})$, which is true according to (2) in Assumption 1.

For $R_{ni,2}$, we have

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^n R_{ni,2}\Big| = \Big|\frac{1}{n}\sum_{i=1}^n (X_{ij} - \hat{\mathbf{w}}_j^\intercal \mathbf{X}_{i,\backslash j})\mathbf{X}_{i,\backslash j}^\intercal(\boldsymbol{\beta}_{\backslash j}^0 - \hat{\boldsymbol{\beta}}_{\backslash j})\Big| \\
=&\Big|\frac{1}{n}\sum_{i=1}^n \eta_{ij}\mathbf{X}_{i,\backslash j}^\intercal(\boldsymbol{\beta}_{\backslash j}^0 - \hat{\boldsymbol{\beta}}_{\backslash j}) + \frac{1}{n}\sum_{i=1}^n (\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\backslash j}\mathbf{X}_{i,\backslash j}^\intercal(\boldsymbol{\beta}_{\backslash j}^0 - \hat{\boldsymbol{\beta}}_{\backslash j})\Big| \\
\le&\Big|\frac{1}{n}\sum_{i=1}^n \eta_{ij}\mathbf{X}_{i,\backslash j}^\intercal(\boldsymbol{\beta}_{\backslash j}^0 - \hat{\boldsymbol{\beta}}_{\backslash j})\Big| + \Big|\frac{1}{n}\sum_{i=1}^n (\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\backslash j}\mathbf{X}_{i,\backslash j}^\intercal(\boldsymbol{\beta}_{\backslash j}^0 - \hat{\boldsymbol{\beta}}_{\backslash j})\Big|
\end{aligned}$$

$$\le \Big\| \frac{1}{n}\sum_{i=1}^n \eta_{ij}\mathbf{X}_{i,\setminus j}^\intercal \Big\|_\infty \Big\| \boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j} \Big\|_1 + \sqrt{\frac{1}{n}\sum_{i=1}^n \big\{ (\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\setminus j} \big\}^2} \sqrt{\frac{1}{n}\sum_{i=1}^n \big\{ \mathbf{X}_{i,\setminus j}^\intercal (\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j}) \big\}^2}$$
$$= O_p(\sqrt{\log p/n}) O_p(s\sqrt{\log p/n}) + O_p(\sqrt{s\log p/n}) O_p(\sqrt{b_n})$$
$$= O_P\big(s\log p/n + \sqrt{b_n s \log p/n}\big).$$

In order to have $\left|\frac{1}{n}\sum_{i=1}^n R_{ni,2}\right| = o_p(n^{-1/2})$ we need to have $s\log p/\sqrt{n} = o(1)$ and $b_n = o(1/\sqrt{n})$. Thus with (3) and (5) in Assumption 1, we have verified the first half condition in (C2), $\frac{1}{n}\sum_{i=1}^n R_{ni} = o_p(n^{-1/2})$.

Now for the second half of the condition in (C2),

$$\begin{aligned}\max_{1\le i\le n} |R_{ni,1}| &= \max_{1\le i\le n} \big|(Y_i - \mathbf{X}_i^\intercal \boldsymbol{\beta}^0)(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\setminus j}\big| = \max_{1\le i\le n} \big|(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\setminus j}\epsilon_i\big| \\ &\le \big\|\mathbf{w}_j^0 - \hat{\mathbf{w}}_j\big\|_1 \max_{1\le i\le n} \big\|\mathbf{X}_{i,\setminus j}\epsilon_i\big\|_\infty = \big\|\mathbf{w}_j^0 - \hat{\mathbf{w}}_j\big\|_1 \max_{1\le i\le n}\max_{1\le k\le p} \big|X_{ik}\epsilon_i\big|.\end{aligned}$$

Now since $\mathbf{X}_i$ and ϵ_i are all sub-Gaussian and then we have $X_{ik}\epsilon_i$ sub-exponential, and then by the union bound, we have

$$\mathrm{P}\big(\max_{1\le i\le n}\max_{1\le k\le p} \big|X_{ik}\epsilon_i\big| > t\big) \le \sum_{1\le i\le n}\sum_{1\le k\le p} \mathrm{P}(\big|X_{ik}\epsilon_i\big| > t) \le pnC_1 e^{-C_2 t}.$$

By taking $t = \log(pn)/C$ with $C < C_2$, we have $\max_{1\le i\le n}\max_{1\le k\le p}\big|X_{ik}\epsilon_i\big| = O_p(\log(pn))$. Hence we have

$$\max_{1\le i\le n} |R_{ni,1}| = \big\|\mathbf{w}_j^0 - \hat{\mathbf{w}}_j\big\|_1 \max_{1\le i\le n}\max_{1\le k\le p} \big|X_{ik}\epsilon_i\big| = O_p(a_n \log(pn)).$$

In order to make $\max_{1\le i\le n}|R_{ni,1}| = o_p(n^{1/2})$, we need $a_n\log(pn)/\sqrt{n} = o(1)$, which is true under assumption (2) for a_n in Assumption 1 since $a_n \log(pn)/\sqrt{n} = o(\log(pn)/\sqrt{n\log p}) = o(\sqrt{\log p/n}) = o(1)$.

$$\begin{aligned}\text{Note that } \max_{1\le i\le n} |R_{ni,2}| &= \max_{1\le i\le n} |(X_{ij} - \hat{\mathbf{w}}_j^\intercal \mathbf{X}_{i,\setminus j})\mathbf{X}_{i,\setminus j}^\intercal(\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j})| \\ &\le \max_{1\le i\le n} |(X_{ij} - \mathbf{w}_j^{0\intercal} \mathbf{X}_{i,\setminus j})\mathbf{X}_{i,\setminus j}^\intercal(\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j})| \\ &\quad + \max_{1\le i\le n} |(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)^\intercal \mathbf{X}_{i,\setminus j}\mathbf{X}_{i,\setminus j}^\intercal(\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j})| \\ &\le \|\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j}\|_1 \max_{1\le i\le n}\max_{1\le k\le p} |\eta_{ij}X_{ik}| \\ &\quad + \|(\mathbf{w}_j^0 - \hat{\mathbf{w}}_j)\|_1 \|(\boldsymbol{\beta}^0_{\setminus j} - \hat{\boldsymbol{\beta}}_{\setminus j})\|_1 \big(\max_{1\le i\le n}\max_{1\le k\le p} |X_{ik}|\big)^2.\end{aligned}$$

Now since η_{ij}'s and $\mathbf{X}_i$ are all sub-Gaussian, and then by similar analysis as above we have

$$\begin{aligned}\max_{1\le i\le n} |R_{ni,2}| &= O_p(s\sqrt{\log p/n})O_p(\log(pn)) + O_p(a_n s\sqrt{\log p/n})O_p(\log(pn)) \\ &= O_p(s\sqrt{\log p/n}\log(pn)).\end{aligned}$$

In order to make $\max_{1\le i\le n}|R_{ni,2}| = o_p(n^{1/2})$, we need

$$s\sqrt{\log p}\log(pn)/n = o_p(1),$$

which is true under assumption (5) in Assumption 1 since $s\sqrt{\log p}\log(pn)/n = o(\sqrt{\log p/n}) = o(1)$. Thus we have $\max_{1\le i\le n}|R_{ni}| = o_p(n^{1/2})$, which verifies the second half in the condition (C2).

Now we need to check out condition (C0). From the above analysis, we have $\max_{1\le i\le n}|R_{ni}| = o_p(\max_{1\le i\le n}|W_{ni}|)$. Thus we only need to prove that

$$\mathrm{P}(\min_{1\le i\le n} W_{ni} < 0 < \max_{1\le i\le n} W_{ni}) \to 1,$$

which just follows from the Gilvenko-Gantelli theorem over half-spaces as on page 219 in Owen (2001). $\square$

Proof of Theorem 3. Recall that $\sum_{i=1}^n R_{ni}/\sqrt{n} = R_{1n} + R_{2n} + R_{3n} + R_{4n}$ where

$$R_{1n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n -\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\big\{X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\},$$

$$R_{2n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \big\{\epsilon_i - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\big\}\big\{\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbf{X}_{i\mathcal{S}}\big\},$$

$$R_{3n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \big\{X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\big\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\big\}[\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}],$$

$$R_{4n} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \big\{\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\big\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\big\}[\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}].$$

Now for R_{1n}, we have

$$\begin{aligned} R_{1n} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n \big\{X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon} \\ &= -\Big\{\frac{1}{n}\sum_{i=1}^n \big\{X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal\Big\}\Big\{\sqrt{n}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\Big\}. \end{aligned}$$

Now we need to bound the two terms $n^{-1}\sum_{i=1}^n\big\{X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}$ and $\sqrt{n}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}$. In fact, for every $k \in \mathcal{S}$, we have that the two Gaussian random variables $X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}$ and X_{ik} have the following properties:

$$\begin{aligned} &\mathrm{E}(X_{ik}) = \mathrm{E}(X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}) = 0; \\ &\mathrm{E}(X_{ik}^2) = \sigma_{kk},\ \mathrm{E}[(X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}})^2] = \sigma_{jj} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}j}; \\ &\mathrm{Cov}(X_{ik}, X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}) = \mathrm{E}[X_{ik}(X_{ij} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}})] = \sigma_{kj} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}k} \\ &\qquad = \sigma_{kj} - \boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}\mathbf{e}_k = \sigma_{kj} - \boldsymbol{\Sigma}_{j\mathcal{S}}\mathbf{e}_k = \sigma_{kj} - \sigma_{jk} = 0. \end{aligned}$$

Thus we have

$$\begin{pmatrix} X_{ik} \\ X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}} \end{pmatrix} \sim \mathrm{N}\Big(\mathbf{0}, \begin{pmatrix} \sigma_{kk} & 0 \\ 0 & \sigma_{jj} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}j} \end{pmatrix}\Big). \tag{2.10}$$

Under (1) in Assumption 3, by Lemma A.3 from Bickel and Levina (2008), we have there exists constants $C, C_1, C_2 > 0$ such that

$$\mathrm{P}\Big\{\Big|\frac{1}{n}\sum_{i=1}^{n}\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{ij}\Big| \geq t\Big\} \leq C_1\exp(-C_2nt^2), \text{ for } 0 \leq t \leq C.$$

By union inequality, we then have

$$\mathrm{P}\Big\{\max_{\mathcal{S}:|\mathcal{S}|\leq m}\Big\|\frac{1}{n}\sum_{i=1}^{n}\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}\Big\|_\infty \geq t\Big\} \leq C_1mp^m\exp(-C_2nt^2),$$

for $0 \leq t \leq C$, where $|\{\mathcal{S} \subseteq \{1, 2, , \cdots, p\} : |\mathcal{S}| \leq m\}| \leq p^m$. For $mp^m\exp(-C_2nt^2) = \exp(-C_2nt^2 + m\log p + \log m)$, take $t = \sqrt{m\log p + \log m + C\log p/(C_2n)} \sim \sqrt{m\log p/n}$, and then we have $\max_{\mathcal{S}:|\mathcal{S}|\leq m}\big\|n^{-1}\sum_{i=1}^{n}\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}\big\|_\infty = O_p(\sqrt{m\log p/n})$.

Now in order to control $\sqrt{n}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}$, first notice that by the following matrix equality Henderson and Searle (1981)

$$\begin{aligned}&(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n)^{-1} = \big\{\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}} + (\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n - \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})\big\}^{-1} \\ =&\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \underbrace{\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\big\{\mathbf{I} + (\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n - \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\big\}^{-1}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n - \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}}_{\mathbf{\Delta}_\mathcal{S}},\end{aligned} \tag{2.11}$$

we have

$$\begin{aligned}\|\sqrt{n}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_1 = \|(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n)^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_1 &\leq \|\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_1 + \|\mathbf{\Delta}_\mathcal{S}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_1 \\ &\leq \sqrt{|\mathcal{S}|}\|\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2 + \sqrt{|\mathcal{S}|}\|\mathbf{\Delta}_\mathcal{S}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2 \\ &\leq \sqrt{|\mathcal{S}|}\|\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2 + \sqrt{|\mathcal{S}|}\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2.\end{aligned}$$

One of the most important results in matrix analysis is the Cauchy (eigenvalue) interlacing theorem. It asserts that the eigenvalues of any principal submatrix of a symmetric matrix interlace those of the symmetric matrix. For example, if an $n \times n$ symmetric matrix $\mathbf{S}$ can be partitioned as

$$\mathbf{S} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\intercal & \mathbf{C} \end{pmatrix},$$

in which $\mathbf{A}$ is an $r \times r$ principle submatrix, then for each $i \in 1, 2, \cdots, r$, we have

$$\lambda_i(\mathbf{S}) \leq \lambda_i(\mathbf{A}) \leq \lambda_{n-r+i}(\mathbf{S}).$$

In particular, we have $\lambda_{\min}(\mathbf{\Sigma}) \leq \lambda_{\min}(\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})$ and $\lambda_{\max}(\mathbf{\Sigma}) \geq \lambda_{\max}(\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}})$. Thus by the definition of maximum eigenvalue, we have

$$\|\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2 \leq \lambda_{\min}^{-1}\|\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\|_2.$$

So

$$\begin{aligned}\|\sqrt{n}(\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}\|_1 &\le \sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2 + \sqrt{|\mathcal{S}|}\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2 \\ &= \sqrt{|\mathcal{S}|}\big\{\lambda_{\min}^{-1} + \|\boldsymbol{\Delta}_{\mathcal{S}}\|_2\big\}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2.\end{aligned}$$

Now we have to control $\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2$ and $\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2$. In order to control the first one, by the sub-Gaussian tailed condition (2) in Assumption 3,

$$\mathrm{P}(\max_{\mathcal{S}:|\mathcal{S}|\le m}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2 \ge t\sqrt{n}) \le \mathrm{P}(\max_{\mathcal{S}:|\mathcal{S}|\le m}\max_{j\in\mathcal{S}}|\frac{1}{n}\sum_{i=1}^n X_{ij}\epsilon_i| \ge t/\sqrt{m}) \le p^m m\exp(-Cnt^2/m),$$

followed from the Bernstein inequality for t small. For $p^m m\exp(-Cnt^2/m) = \exp(m\log p + \log m - Cnt^2/m)$, take $t = \sqrt{m}\sqrt{m\log p + \log m + C_1\log p/Cn} \sim \sqrt{m^2\log p/n}$. Then we have the following order

$$\max_{\mathcal{S}:|\mathcal{S}|\le m}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2 = O_p(m\sqrt{\log p}).$$

Now for $\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2$ with $\boldsymbol{\Delta}_{\mathcal{S}} = \boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\big\{\mathbf{I} + (\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}})\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\big\}^{-1}(\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}})\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}$, we have to control $\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}}$ first. Note that

$$\begin{aligned}&\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}}\|_2 \ge \epsilon\Big) \le \mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{j,k}|\mathbb{X}_j^{\intercal}\mathbb{X}_k/n - \sigma_{jk}| \ge \epsilon/m\Big) \\ \le& m^2p^m\mathrm{P}\Big(|\mathbb{X}_j^{\intercal}\mathbb{X}_k/n - \sigma_{jk}| \ge \epsilon/m\Big) \le C_1m^2p^m\exp(-C_2n\epsilon^2/m^2)\end{aligned}$$

where the last inequality is also followed from Lemma A.3 in Bickel and Levina (2008) with constants $C_1, C_2 > 0$. For $m^2p^m\exp(-C_2n\epsilon^2/m^2) = \exp(2\log m + m\log p - C_2n\epsilon^2/m^2)$, by taking $\epsilon = m\sqrt{m\log p + 2\log m + C_1\log p/C_2n} \sim \sqrt{m^3\log p/n}$, we have $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}}\|_2 = O_p(\sqrt{m^3\log p/n})$. It follows then

$$\begin{aligned}\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2 &= \|\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\big\{\mathbf{I} + (\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}})\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\big\}^{-1}(\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}})\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\|_2 \\ &\le \|\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\|_2^2\|\mathbf{I} + (\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}})\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\big\}^{-1}\|_2\|\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{\mathcal{SS}}\|_2 \\ &= O_p(\sqrt{m^3\log p/n}),\end{aligned}$$

since $\|\boldsymbol{\Sigma}_{\mathcal{SS}}^{-1}\|_2 = \lambda_{\max}^{1/2}(\boldsymbol{\Sigma}_{\mathcal{SS}}^{-2}) \le \lambda_{\min}^{-1}$.

Thus we have

$$\begin{aligned}\|\sqrt{n}(\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}\|_1 &\le \sqrt{|\mathcal{S}|}\big\{\lambda_{\min}^{-1} + \|\boldsymbol{\Delta}_{\mathcal{S}}\|_2\big\}\|\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}/\sqrt{n}\|_2 \\ &= O_p(\sqrt{m^3\log p/n}),\end{aligned}$$

i.e., $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\|\sqrt{n}(\mathbb{X}_{\mathcal{S}}^{\intercal}\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^{\intercal}\boldsymbol{\epsilon}\|_1 = O_p(\sqrt{m^3\log p/n})$.

In summary, we then have

$$
\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\Big|\Big\{\frac{1}{n}\sum_{i=1}^n\big\{X_{ij}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal\Big\}\Big\{\sqrt{n}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\Big\}\Big|\\
\le&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\Big\|\frac{1}{n}\sum_{i=1}^n\big\{X_{ij}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal\Big\|_\infty\sup_{\mathcal{S}:|\mathcal{S}|\le m}\Big\|\sqrt{n}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\Big\|_1\\
=&O_p(\sqrt{m\log p/n})O_p(\sqrt{m^3\log p/n})=O_p(m^2\log p/n).
\end{aligned}
$$

And hence $R_{1n}=o_p(1)$.

For R_{2n}, we have

$$
\begin{aligned}
R_{2n}&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\big\{\epsilon_i-\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\big\}\big\{\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}-\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\\
&=\big\{\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}-\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\big\}\frac{1}{\sqrt{n}}\sum_{i=1}^n\big\{\mathbf{X}_{i\mathcal{S}}\epsilon_i-\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\big\}\\
&=\big\{\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}-\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\big\}\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_{i\mathcal{S}}\epsilon_i-\big\{\frac{1}{n}\sum_{i=1}^n\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal\big\}\sqrt{n}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}\Big\}\\
&=\big\{\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}-\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\big\}\Big\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\mathbf{X}_{i\mathcal{S}}\epsilon_i-\mathbb{X}_{\mathcal{S}}^\intercal\boldsymbol{\epsilon}/\sqrt{n}\Big\}=0.
\end{aligned}
$$

Observe that we can rewrite R_{3n} as

$$
\begin{aligned}
R_{3n}=&\frac{1}{\sqrt{n}}\sum_{i=1}^n\big\{X_{ij}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\big\{\mathbf{X}_{i\mathcal{S}^*}^\intercal-\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\big\}[\boldsymbol{\beta}_{\mathcal{S}^*}^0-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}]\\
=&\frac{1}{\sqrt{n}}\mathbb{X}_j^\intercal\big\{\mathbf{I}-\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\big\}\mathbb{X}_{\mathcal{S}^*}[\boldsymbol{\beta}_{\mathcal{S}^*}^0-\hat{\boldsymbol{\beta}}_{\mathcal{S}^*}],
\end{aligned}
$$

where $\frac{1}{\sqrt{n}}\mathbb{X}_j^\intercal\big\{\mathbf{I}-\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\big\}\mathbb{X}_{\mathcal{S}^*}$ can be controlled as follows

$$
\begin{aligned}
&\|\frac{1}{\sqrt{n}}\mathbb{X}_j^\intercal\big\{\mathbf{I}-\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\big\}\mathbb{X}_{\mathcal{S}^*}\|_\infty=\max_{k\in\mathcal{S}^*}|\frac{1}{\sqrt{n}}\mathbb{X}_j^\intercal\big\{\mathbf{I}-\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\big\}\mathbb{X}_k|\\
\le&\sqrt{n}\max_{k\in\mathcal{S}^*}\Big\{\big|\mathbb{X}_j^\intercal\mathbb{X}_k/n-\sigma_{jk}\big|+\big|\sigma_{jk}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}\big|+\big|[\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n-\mathbf{\Sigma}_{j\mathcal{S}}]\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}\big|\\
&+\big|\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}[\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}]\big|+\big|\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Delta}_{\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}k}\big|\\
&+\big|\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Delta}_{\mathcal{S}}[\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}]\big|+\big|[\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n-\mathbf{\Sigma}_{j\mathcal{S}}]\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}[\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}]\big|\\
&+\big|[\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n-\mathbf{\Sigma}_{j\mathcal{S}}]\mathbf{\Delta}_{\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}k}\big|+\big|[\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n-\mathbf{\Sigma}_{j\mathcal{S}}]\mathbf{\Delta}_{\mathcal{S}}[\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}]\big|\Big\}
\end{aligned}
$$

$$
\begin{aligned}
\leq&\sqrt{n}\max_{k\in\mathcal{S}^*}\Big\{\big|\mathbb{X}_j^\intercal\mathbb{X}_k/n-\sigma_{jk}\big|+\big|\sigma_{jk}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}\big|+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_\infty\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\\
&+\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_\infty+\lambda_{\max}^2\|\mathbf{\Delta}_\mathcal{S}\|_2\\
&+\sqrt{|\mathcal{S}|}\lambda_{\max}\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_\infty+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_2\lambda_{\min}^{-1}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_2\\
&+\sqrt{|\mathcal{S}|}\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_\infty\|\mathbf{\Delta}_\mathcal{S}\|_2\lambda_{\max}+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_2\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_2\Big\}.
\end{aligned}
$$

And we have that

$$
\begin{aligned}
&\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}|\sigma_{jk}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_k|\geq\epsilon\Big)\\
\leq&p^{m+1}\mathrm{P}\Big(|\sigma_{jk}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_k|\geq\epsilon\Big)\leq C_1p^{m+1}\exp(-C_2n\epsilon^2)
\end{aligned}
$$

where the last inequality is also followed from Lemma A.3 in Bickel and Levina (2008) with constants $C_1, C_2>0$. For $p^{m+1}\exp(-C_2n\epsilon^2)=\exp((m+1)\log p-C_2n\epsilon^2)$, by taking $\epsilon=\sqrt{\{(m+1)\log p+C_1\log p\}/C_2n}\sim\sqrt{m\log p/n}$, we have $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}|\sigma_{jk}-\mathbb{X}_j^\intercal\mathbb{X}_k/n|=O_p(\sqrt{m\log p/n})$. Similarly, we have $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\|\mathbf{\Sigma}_{j\mathcal{S}}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}\|_\infty=O_p(\sqrt{m\log p/n})$ and $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_\infty=O_p(\sqrt{m\log p/n})$. By $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\|\mathbf{\Delta}_\mathcal{S}\|_2=O_p(\sqrt{m^3\log p/n})$, we have

$$
\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\|\frac{1}{\sqrt{n}}\mathbb{X}_j^\intercal\{\mathbf{I}-\mathbb{X}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\}\mathbb{X}_{\mathcal{S}^*}\|_\infty\\
\leq&\sqrt{n}\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}\Big\{\big|\mathbb{X}_j^\intercal\mathbb{X}_k/n-\sigma_{jk}\big|+\big|\sigma_{jk}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}\big|+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_\infty\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\\
&+\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_\infty+\lambda_{\max}^2\|\mathbf{\Delta}_\mathcal{S}\|_2\\
&+\sqrt{|\mathcal{S}|}\lambda_{\max}\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_\infty+\sqrt{|\mathcal{S}|}\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_\infty\|\mathbf{\Delta}_\mathcal{S}\|_2\lambda_{\max}\\
&+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_2\lambda_{\min}^{-1}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_2\\
&+\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n-\mathbf{\Sigma}_{j\mathcal{S}}\|_2\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n-\mathbf{\Sigma}_{\mathcal{S}k}\|_2\Big\}\\
=&\sqrt{n}\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\max_{k\in\mathcal{S}^*}\big|\sigma_{jk}-\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}\big|+O_p\{\sqrt{n}\sqrt{m^3\log p/n}\},
\end{aligned}
$$

since $\sqrt{m^3\log p/n}=o(1)$. Under condition (4) and (5) in Assumption 3, we have that $R_{3n}=o_p(1)$.

Note that

$$R_{4n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{ \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}} - \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbf{X}_{i\mathcal{S}} \big\} \big\{ \mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}^*} \big\} [\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}]$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{ \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}} \mathbf{X}_{i\mathcal{S}^*}^\intercal \big\} [\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}] \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}^*}/\sqrt{n}) [\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}]$$
$$- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{ \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbf{X}_{i\mathcal{S}} \mathbf{X}_{i\mathcal{S}^*}^\intercal \big\} [\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}]$$
$$+ \big\{ \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}^*}/\sqrt{n} \big\} [\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}] = 0.$$

Thus we have verified that $\frac{1}{n}\sum_{i=1}^n R_{ni} = o_p(n^{-1/2})$.

For $R_{ni,1}$, we have

$$\max_{1\le i\le n} |R_{ni,1}| = \|(\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \boldsymbol{\epsilon}\|_1 \max_{1\le i\le n} \|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}\big\} \mathbf{X}_{i\mathcal{S}}^\intercal\|_\infty$$
$$= \|(\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \boldsymbol{\epsilon}\|_1 \max_{1\le i\le n} \max_{k\in\mathcal{S}} \big|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}\big\} X_{ik}\big|$$

where $\sup_{\mathcal{S}:|\mathcal{S}|\le m} \|(\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \boldsymbol{\epsilon}\|_1 = O_p(\sqrt{m^3 \log p}/n)$. And since $X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}$ is Gaussian under the assumption that $\mathbf{X}$ is Gaussian, we have $\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}\big\} X_{ik}$ sub-exponential. So

$$\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m} \max_{1\le i\le n} \max_{k\in\mathcal{S}} \big|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}\big\} X_{ik}\big| > t \Big) \le p^m n m C_1 \exp(-C_2 t)$$

which leads to $\sup_{\mathcal{S}:|\mathcal{S}|\le m} \max_{1\le i\le n} \max_{k\in\mathcal{S}} \big|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \mathbf{X}_{i\mathcal{S}}\big\} X_{ik}\big| = O_p(m \log p)$. Thus we have

$$\sup_{\mathcal{S}:|\mathcal{S}|\le m} \max_{1\le i\le n} |R_{ni,1}| = O_p(m \log p \sqrt{m^3 \log p}/n) = o_p(n^{1/2})$$

since $(m \log p/n)\sqrt{m^3 \log p/n} = o(1)$.

For $R_{ni,2}$, we have

$$\max_{1\le i\le n} |R_{ni,2}| \le \|\mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1}\|_1 \max_{1\le i\le n} \|\mathbf{X}_{i\mathcal{S}} \epsilon_i - \mathbf{X}_{i\mathcal{S}} \mathbf{X}_{i\mathcal{S}}^\intercal (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1} \mathbb{X}_{\mathcal{S}}^\intercal \boldsymbol{\epsilon}\|_\infty,$$

where

$$\|\mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}})^{-1}\|_1 = \|\mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbb{X}_{\mathcal{S}}^\intercal \mathbb{X}_{\mathcal{S}}/n)^{-1}\|_1$$
$$= \|\mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} (\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \mathbf{\Delta}_{\mathcal{S}})\|_1$$
$$\le \|(\mathbf{\Sigma}_{j\mathcal{S}} - n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}}) \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1 + \|n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} \mathbf{\Delta}_{\mathcal{S}}\|_1$$
$$\le \|(\mathbf{\Sigma}_{j\mathcal{S}} - n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}}) \mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1 + \|(n^{-1} \mathbb{X}_j^\intercal \mathbb{X}_{\mathcal{S}} - \mathbf{\Sigma}_{j\mathcal{S}}) \mathbf{\Delta}_{\mathcal{S}}\|_1 + \|\mathbf{\Sigma}_{j\mathcal{S}} \mathbf{\Delta}_{\mathcal{S}}\|_1.$$

And by simple algebra, we have

$$\sup_{\mathcal{S}:|\mathcal{S}|\le m} \|(\mathbf{\Sigma}_{j\mathcal{S}} - n^{-1}\mathbb{X}_j^\intercal \mathbb{X}_\mathcal{S})\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1 = O_p(\sqrt{m^3 \log p/n}),$$
$$\sup_{\mathcal{S}:|\mathcal{S}|\le m} \|(n^{-1}\mathbb{X}_j^\intercal \mathbb{X}_\mathcal{S} - \mathbf{\Sigma}_{j\mathcal{S}})\mathbf{\Delta}_\mathcal{S}\|_1 = O_p(m^2 \log p/n),$$
$$\sup_{\mathcal{S}:|\mathcal{S}|\le m} \|\mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Delta}_\mathcal{S}\|_1 = O_p(m^2\sqrt{\log p/n}).$$

Now for

$$\begin{aligned}\max_{1\le i\le n} \|\mathbf{X}_{i\mathcal{S}}\epsilon_i - \mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_\infty &\le \max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\epsilon_i\|_\infty + \max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_\infty \\ &\le \max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\epsilon_i\|_\infty + \|(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_\infty \max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal\|_\infty,\end{aligned}$$

since $X_{ik}\epsilon_i$ is sub-exponential, we have

$$\begin{aligned}\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\epsilon_i\|_\infty > t\Big) &= \mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\max_{k\in\mathcal{S}}|X_{ik}\epsilon_i| > t\Big) \\ &\le p^m mnC_1e^{-C_2t}\end{aligned}$$

which leads to $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\epsilon_i\|_\infty = O_p(m\log p)$. And since $X_{ik}X_{il}$ is sub-exponential, we have

$$\begin{aligned}\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal\|_\infty > t\Big) &\le \mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\sqrt{m}\max_{k,l\in\mathcal{S}}|X_{ik}X_{il}| > t\Big) \\ &\le p^m m^2 nC_1e^{-C_2t}\end{aligned}$$

which leads to $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal\|_\infty = O_p(\sqrt{m}m\log p)$.

Since $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\|(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_1 = O_p(\sqrt{m^3\log p}/n)$, we have

$$\begin{aligned}&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\mathbf{X}_{i\mathcal{S}}\epsilon_i - \mathbf{X}_{i\mathcal{S}}\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\boldsymbol{\epsilon}\|_\infty \\ =&O_p(m\log p + \sqrt{m}m\log p\sqrt{m^3\log p}/n) = O_p(m\log p(1 + m^2\sqrt{\log p}/n)).\end{aligned}$$

In summary,

$$\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,2}| = O_p\{m^3\log p\sqrt{\log p/n}(1 + m^2\sqrt{\log p}/n)\},$$

since $\log p/n \to 0$. In order to have $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,2}| = o_p(n^{1/2})$, we need to have $m^3(\log p/\sqrt{n})\sqrt{\log p/n} = o(1)$, which is true under (4) in Assumption 3 since

$$m^3(\log p/\sqrt{n})\sqrt{\log p/n} = \sqrt{m^3\log p/n}\sqrt{(\log p)^2m^3/n} = o(1).$$

Observe that

$$\max_{1\le i\le n}|R_{ni,3}| \le \|\boldsymbol{\beta}_{\mathcal{S}^*}^0 - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1\max_{1\le i\le n}\|\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty.$$

Since

$$\|\mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_{\mathcal{S}^*}\|_\infty \le \max_{k\in\mathcal{S}^*}\big\{|\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}| + |\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n - \mathbf{\Sigma}_{\mathcal{S}k})| \\ + |\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Delta}_\mathcal{S}\mathbf{\Sigma}_{\mathcal{S}k}| + |\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Delta}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n - \mathbf{\Sigma}_{\mathcal{S}k})|\big\}, \tag{2.12}$$

we have

$$\begin{aligned}
&\max_{1\le i\le n}\|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\big\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_{\mathcal{S}^*}\big\}\|_\infty \\
\le& \max_{1\le i\le n}\max_{k\in\mathcal{S}^*}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{ik}| + \max_{1\le i\le n}\max_{k\in\mathcal{S}^*}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}| \\
&+ \max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}X_{i\mathcal{S}}\big\}X_{il}|\|\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n - \mathbf{\Sigma}_{\mathcal{S}k})\|_1 \\
&+ \max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{il}|\sqrt{m}\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbf{\Sigma}_{\mathcal{S}k}\|_2 \\
&+ \max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{il}|\|\mathbf{\Delta}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_k/n - \mathbf{\Sigma}_{\mathcal{S}k})\|_1.
\end{aligned}$$

Now since

$$\mathrm{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{ik}| > t\Big) \le p^{m+1}nC_1e^{-C_2t},$$

we have

$$\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{ik}| = O_p(m\log p).$$

Similarly, we have

$$\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\mathbf{X}_{i\mathcal{S}}^\intercal\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}k}| = O_p(m\log p), \\
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\max_{l\in\mathcal{S}}|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}X_{il}| = O_p(m\log p).
\end{aligned}$$

And then by simple algebra, we have

$$\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{k\in\mathcal{S}^*}\|(\mathbf{\Sigma}_{k\mathcal{S}} - n^{-1}\mathbb{X}_k^\intercal\mathbb{X}_\mathcal{S})\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1 = O_p(\sqrt{m^3\log p/n}), \\
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{k\in\mathcal{S}^*}\|(n^{-1}\mathbb{X}_k^\intercal\mathbb{X}_\mathcal{S} - \mathbf{\Sigma}_{k\mathcal{S}})\mathbf{\Delta}_\mathcal{S}\|_1 = O_p(m^2\log p/n).
\end{aligned}$$

Thus we have

$$\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}\|\big\{X_{ij} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\big\}\big\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_{\mathcal{S}^*}\big\}\|_\infty \\
=&O_p\{m\log p(1 + \sqrt{m^3\log p/n} + m^2\sqrt{\log p/n} + m^2\log p/n)\} \\
=&O_p\{m\log p(1 + \sqrt{m^3\log p/n} + m^2\log p/n)\},
\end{aligned}$$

which leads to

$$\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,3}| = O_p(s\sqrt{\log p/n}m\log p(1 + \sqrt{m^3\log p/n} + m^2\log p/n)).$$

In order to have $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,3}| = o_p(n^{1/2})$, we need

$$s\sqrt{\log p/n}(m\log p/\sqrt{n})(1+\sqrt{m^3\log p/n}+m^2\log p/n) = o(1),$$

which is true under (4) in Assumption 3.

And for $\max_{1\le i\le n}|R_{ni,4}| = \|\boldsymbol{\beta}^0_{\mathcal{S}^*} - \hat{\boldsymbol{\beta}}_{\mathcal{S}^*}\|_1$
$\times \max_{1\le i\le n}\|\{\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty.$

And for

$$\begin{aligned}
&\max_{1\le i\le n}\|\{\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n(\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} - \boldsymbol{\Delta}_{\mathcal{S}})\}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
=&\max_{1\le i\le n}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
&+\max_{1\le i\le n}\|(\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{j\mathcal{S}})\boldsymbol{\Delta}_{\mathcal{S}}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
&+\max_{1\le i\le n}\|\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Delta}_{\mathcal{S}}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty,
\end{aligned}$$

by (2.12), we have

$$\begin{aligned}
&\max_{1\le i\le n}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
\le&\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1|X_{il}X_{ik}|\\
&+\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1X_{il}^2\|\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}k}\|_1\\
&+\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1X_{il}^2\|\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n - \boldsymbol{\Sigma}_{\mathcal{S}k})\|_1\\
&+\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1X_{il}^2\|\boldsymbol{\Delta}_{\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}k}\|_1\\
&+\max_{1\le i\le n}\max_{k\in\mathcal{S}^*}\max_{l\in\mathcal{S}}\|(\boldsymbol{\Sigma}_{j\mathcal{S}} - \mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_1X_{il}^2\|\boldsymbol{\Delta}_{\mathcal{S}}(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_k/n - \boldsymbol{\Sigma}_{\mathcal{S}k})\|_1\\
&=O_p(m^3\log p\sqrt{\log p/n}),
\end{aligned}$$

under the condition that $m^3\log p/n\to 0$. Similarly we have, if $m^3\log p/n\to 0$.

$$\begin{aligned}
\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}&\|(\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n - \boldsymbol{\Sigma}_{j\mathcal{S}})\boldsymbol{\Delta}_{\mathcal{S}}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
&=O_p\{m^{7/2}(\log p)^2/n\}\\
\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}&\|\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Delta}_{\mathcal{S}}\mathbf{X}_{i\mathcal{S}}\{\mathbf{X}_{i\mathcal{S}^*}^\intercal - \mathbf{X}_{i\mathcal{S}}^\intercal(\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}})^{-1}\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_{\mathcal{S}^*}\}\|_\infty\\
&=O_p\{m^{7/2}\log p\sqrt{\log p/n}\}.
\end{aligned}$$

In summary, if $m^3\log p/n\to 0$, $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,4}| = O_p\{sm^{7/2}(\log p)^2/n\}$. Thus in order to have $\sup_{\mathcal{S}:|\mathcal{S}|\le m}\max_{1\le i\le n}|R_{ni,4}| = o_p(n^{1/2})$, we need $sm^{7/2}(\log p)^2/n^{3/2} = o(1)$, which is true under the condition (4) in Assumption 3 since $sm^{7/2}(\log p)^2/n^{3/2} = s\sqrt{(\log p)^4 m^7/n^3} = s\sqrt{(\log p)^2 m^3/n} m^2 \log p/n = o(1)$.

From the above analysis, we have $\max_{1\le i\le n}|R_{ni}| = o_p(\max_{1\le i\le n}|W_{ni}|)$. Thus we only need to prove that $\mathrm{P}(\min_{1\le i\le n} W_{ni} < 0 < \max_{1\le i\le n} W_{ni}) \to 1$, which just follows from the Gilvenko-Gantelli theorem over half-spaces as on page 219 in Owen (2001). □

Proof of Theorem 4. Notice that

$$\frac{1}{n}\sum_{i=1}^{n} R_{ni,1} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2(\gamma_{j1}^0 - \hat{\gamma}_{j1}) + \frac{1}{n}\sum_{i=1}^{n}\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0(\gamma_{j1}^0 - \hat{\gamma}_{j1}) + \frac{1}{n}\sum_{i=1}^{n}\epsilon_i \mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1}).$$

By condition (2) in Assumption 2 and (1.3) implied from condition (4) in Assumption 2,

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0(\gamma_{j1}^0 - \hat{\gamma}_{j1})\Big| = \Big|(\gamma_{j1}^0 - \hat{\gamma}_{j1})\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i \boldsymbol{\beta}^{0\intercal}\mathbf{X}_i\Big| = O_p(a_n\sqrt{\frac{\log p}{n}}),$$

and

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i \mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1})\Big| \le \Big\|\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1}\Big\|_1\Big\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i \mathbf{X}_{i,\setminus j}\Big\|_\infty = O_p(a_n\sqrt{\frac{\log p}{n}}).$$

Thus we have $\frac{1}{n}\sum_{i=1}^{n} R_{ni,1} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2(\gamma_{j1}^0 - \hat{\gamma}_{j1}) + O_p(a_n\sqrt{\log p/n}) = O_p(a_n\sqrt{1/n})$. So in order to have $\frac{1}{n}\sum_{i=1}^{n} R_{ni,1} = o_p(n^{-1/2})$, we need $a_n = o_p(1)$. Note that

$$\begin{aligned}
\max_{1\le i\le n}|R_{ni,1}| \le & \max_{1\le i\le n}|\epsilon_i^2(\gamma_{j1}^0 - \hat{\gamma}_{j1})| + \max_{1\le i\le n}|\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0(\gamma_{j1}^0 - \hat{\gamma}_{j1})| \\
& + \max_{1\le i\le n}|\epsilon_i \mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1})| \\
=&|\gamma_{j1}^0 - \hat{\gamma}_{j1}|\{\max_{1\le i\le n}|\epsilon_i^2| + \max_{1\le i\le n}|\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0|\} + \|\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1}\|_1 \max_{1\le i\le n}\|\epsilon_i \mathbf{X}_{i,\setminus j}\|_\infty \\
=&|\gamma_{j1}^0 - \hat{\gamma}_{j1}|\{\max_{1\le i\le n}|\epsilon_i^2| + \max_{1\le i\le n}|\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0|\} + \|\boldsymbol{\gamma}_{j,\setminus 1}^0 - \hat{\boldsymbol{\gamma}}_{j,\setminus 1}\|_1 \max_{1\le i\le n}\max_{1\le k\le p}|\epsilon_i X_{ij}|.
\end{aligned}$$

And by the assumption that $\mathbf{X}_i$ and ϵ_i are sub-Gaussian, we have $\mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0$ is sub-Gaussian and ϵ_i^2, $\epsilon_i \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0$ and $X_{ij}\epsilon_i$ are all sub-exponential. Then we have

$$\mathrm{P}\Big(\max_{1\le i\le n}|\epsilon_i^2| > t\Big) \le n\mathrm{P}(|\epsilon_i^2| > t) \le nC_1 e^{-C_2 t}$$

which implies that $\max_{1\le i\le n}|\epsilon_i^2| = O_p(\log n)$. Thus we have $\max_{1\le i\le n}|R_{ni,1}| = O_p(a_n\log(pn))$. In order to achieve $\max_{1\le i\le n}|R_{ni,1}| = o_p(n^{1/2})$, we need $a_n\log(pn)/\sqrt{n} = o(1)$, which is true since $a_n = o(1/\sqrt{\log p})$.

$$\begin{aligned}
\text{For } R_{ni,2} &= \eta_{ij,y}\mathbf{X}_i^{\intercal}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}) = \eta_{ij,y}\{X_{ij}(\beta_j^0 - \hat{\beta}_j) + \mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\beta}_{\setminus j}^0 - \hat{\boldsymbol{\beta}}_{\setminus j})\} \\
&= \eta_{ij,y}\{[(Y_i, \mathbf{X}_{i,\setminus j}^{\intercal})\boldsymbol{\gamma}_j^0 + \eta_{ij,y}](\beta_j^0 - \hat{\beta}_j) + \mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\beta}_{\setminus j}^0 - \hat{\boldsymbol{\beta}}_{\setminus j})\} \\
&= \eta_{ij,y}^2(\beta_j^0 - \hat{\beta}_j) + \eta_{ij,y}(Y_i, \mathbf{X}_{i,\setminus j}^{\intercal})\boldsymbol{\gamma}_j^0(\beta_j^0 - \hat{\beta}_j) + \eta_{ij,y}\mathbf{X}_{i,\setminus j}^{\intercal}(\boldsymbol{\beta}_{\setminus j}^0 - \hat{\boldsymbol{\beta}}_{\setminus j}),
\end{aligned}$$

similarly as $R_{ni,1}$, by condition (1) and (1.4), (1.5), we have

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} R_{ni,2} &= \frac{1}{n}\sum_{i=1}^{n} \eta_{ij,y}^2(\beta_j^0 - \hat{\beta}_j) + O_p(s\sqrt{\log p/n}\sqrt{\log p/n}) \\
&= \frac{1}{n}\sum_{i=1}^{n} \eta_{ij,y}^2(\beta_j^0 - \hat{\beta}_j) + O_p(s\log p/n) \\
&= O_p(s\sqrt{\log p/n}\sqrt{1/n}) + O_p(s\log p/n) = O_p(s\sqrt{\log p}/n).
\end{aligned}$$

So in order to have $\frac{1}{n}\sum_{i=1}^{n} R_{ni,2} = o_p(n^{-1/2})$, we need to have $s\sqrt{\log p}/n = o_p(n^{-1/2})$, i.e., $s\sqrt{\log p/n} = o_p(1)$. Note that

$$\begin{aligned}
\max_{1\le i\le n} |R_{ni,2}| \le{}& \max_{1\le i\le n} |\eta_{ij,y}^2(\beta_j^0 - \hat{\beta}_j)| + \max_{1\le i\le n} |\eta_{ij,y}(Y_i, \mathbf{X}_{i,\setminus j}^\intercal)\boldsymbol{\gamma}_j^0(\beta_j^0 - \hat{\beta}_j)| \\
&+ \max_{1\le i\le n} |\eta_{ij,y}\mathbf{X}_{i,\setminus j}^\intercal(\boldsymbol{\beta}_{\setminus j}^0 - \hat{\boldsymbol{\beta}}_{\setminus j})| = O_p(s\sqrt{\log p/n}\log(pn)) = o_p(\sqrt{n})
\end{aligned}$$

since $s\sqrt{\log p/n}\log(pn)/\sqrt{n} = o(\sqrt{\log p/n}) = o(1)$.

Now for $R_{ni,3} = \mathbf{X}_i^\intercal(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})\big\{(Y_i, \mathbf{X}_{i,\setminus j}^\intercal)(\boldsymbol{\gamma}_j^0 - \hat{\boldsymbol{\gamma}}_j)\big\} = (\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})^\intercal\mathbf{X}_i(Y_i, \mathbf{X}_{i,\setminus j}^\intercal)(\boldsymbol{\gamma}_j^0 - \hat{\boldsymbol{\gamma}}_j)$, we have by (3) in Assumption 2

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^{n} R_{ni,3}\Big| &= \Big|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})^\intercal\mathbf{X}_i(Y_i, \mathbf{X}_{i,\setminus j}^\intercal)(\boldsymbol{\gamma}_j^0 - \hat{\boldsymbol{\gamma}}_j)\Big| \\
&\le \sqrt{\frac{1}{n}\sum_{i=1}^{n}[(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})^\intercal\mathbf{X}_i]^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}[(Y_i, \mathbf{X}_{i,\setminus j}^\intercal)(\boldsymbol{\gamma}_j^0 - \hat{\boldsymbol{\gamma}}_j)]^2} \\
&= O_p(\sqrt{s\log p/n})O_p(\sqrt{b_n}) = O_p(\sqrt{b_n s\log p/n}).
\end{aligned}$$

So in order to have $\frac{1}{n}\sum_{i=1}^{n} R_{ni,3} = o_p(n^{-1/2})$, we need to have $\sqrt{b_n s\log p/n} = o_p(n^{-1/2})$, i.e., $\sqrt{b_n s\log p} = o_p(1)$. And we also have

$$\begin{aligned}
\max_{1\le i\le n} |R_{ni,3}| &\le \|\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}}\|_1\|\boldsymbol{\gamma}_j^0 - \hat{\boldsymbol{\gamma}}_j\|_1 \max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|\big(\max_{1\le i\le n}|Y_i| + \max_{1\le i\le n}\max_{1\le j\le p}|X_{ij}|\big) \\
&= O_p(s\sqrt{\log p/n}a_n\log(pn)) = o_p(n^{1/2}).
\end{aligned}$$

Now we need to check out condition (C0). From the above analysis, we have $\max_{1\le i\le n}|R_{ni}| = o_p(\max_{1\le i\le n}|W_{ni}|)$. Thus we only need to prove that

$$\mathrm{P}(\min_{1\le i\le n} W_{ni} < 0 < \max_{1\le i\le n} W_{ni}) \to 1,$$

which just follows from the Gilvenko-Gantelli theorem over half-spaces as on page 219 in Owen (2001). □

2.2 Proof of Propositions

Proof of Proposition 4. With $W_{ni}^{(\text{lasso})} = \epsilon_i(X_{ij} - \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{X}_{i,\backslash j})$, simple calculation yields that $\mathrm{E}(W_{ni}^{(\text{lasso})}) = \mathrm{E}\big\{\epsilon_i(X_{ij} - \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{X}_{i,\backslash j})\big\} = 0$ and

$$\begin{aligned}
\mathrm{E}[(W_{ni}^{(\text{lasso})})^2] &= \mathrm{E}\big\{\epsilon_i^2(X_{ij} - \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{X}_{i,\backslash j})^2\big\} \\
&= \mathrm{E}\{\epsilon_i^2(X_{ij}^2 - 2X_{ij}\mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{X}_{i,\backslash j} + \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{X}_{i,\backslash j}\mathbf{X}_{i,\backslash j}^{\intercal}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Sigma}_{\backslash j,j})\} \\
&= \mathrm{E}\{Z_{ij}^2 - 2Z_{ij}\mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{Z}_{i,\backslash j} + \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{Z}_{i,\backslash j}\mathbf{Z}_{i,\backslash j}^{\intercal}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Sigma}_{\backslash j,j}\} \\
&= \theta_{i;jj} - 2\mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Theta}_{i;j,\backslash j} + \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Theta}_{i;\backslash j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Sigma}_{\backslash j,j}.
\end{aligned}$$

Note that if we assume the independence between the error term and the covariates and the homoscedasticity of the error terms, it easily follows that

$$\mathrm{E}[(W_{ni}^{(\text{lasso})})^2] = \sigma_\epsilon^2(\sigma_{jj} - \mathbf{\Sigma}_{j,\backslash j}\mathbf{\Sigma}_{\backslash j,\backslash j}^{-1}\mathbf{\Sigma}_{\backslash j,j}).$$

□

Proof of Proposition 5. First of all, simple algebra easily leads to $\mathrm{E}(W_{ni}) = 0$. Remember $X_{ij} = \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0 + \epsilon_i\gamma_{j1}^0 + \mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0 + \eta_{ij,y}$. Hence

$$\begin{aligned}
\mathrm{Var}(W_{ni}^{(\text{inv})}) &= \mathrm{Var}(\epsilon_i\eta_{ij,y}) = \mathrm{Var}(\epsilon_i(X_{ij} - \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0 - \epsilon_i\gamma_{j1}^0 - \mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0)) \\
&= \mathrm{Var}(\epsilon_iX_{ij}) + \mathrm{Var}(\epsilon_i\mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0) + \mathrm{Var}(\epsilon_i^2\gamma_{j1}^0) + \mathrm{Var}(\epsilon_i\mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0) \\
&- 2\mathrm{Cov}(\epsilon_iX_{ij}, \epsilon_i\mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0) - 2\mathrm{Cov}(\epsilon_iX_{ij}, \epsilon_i^2\gamma_{j1}^0) - 2\mathrm{Cov}(\epsilon_iX_{ij}, \epsilon_i\mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0) \\
&+ 2\mathrm{Cov}(\epsilon_i\mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0, \epsilon_i^2\gamma_{j1}^0) + 2\mathrm{Cov}(\epsilon_i\mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0, \epsilon_i\mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0) + 2\mathrm{Cov}(\epsilon_i^2\gamma_{j1}^0, \epsilon_i\mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0).
\end{aligned}$$

By the definition of Z_{ij}, we further simplify the variance of W_{ni} as follows.

$$\begin{aligned}
\mathrm{Var}(W_{ni}^{(\text{inv})}) &= \mathrm{Var}(Z_{ij}) + \mathrm{Var}(\gamma_{j1}^0\boldsymbol{\beta}^{0\intercal}\mathbf{Z}_i) + \mathrm{Var}(\gamma_{j1}^0\epsilon_i^2) + \mathrm{Var}(\boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{Z}_{i,\backslash j}) \\
&- 2\mathrm{Cov}(Z_{ij}, \gamma_{j1}^0\boldsymbol{\beta}^{0\intercal}\mathbf{Z}_i) - 2\mathrm{Cov}(Z_{ij}, \gamma_{j1}^0\epsilon_i^2) - 2\mathrm{Cov}(Z_{ij}, \boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{Z}_{i,\backslash j}) \\
&+ 2\mathrm{Cov}(\gamma_{j1}^0\boldsymbol{\beta}^{0\intercal}\mathbf{Z}_i, \gamma_{j1}^0\epsilon_i^2) + 2\mathrm{Cov}(\boldsymbol{\beta}^{0\intercal}\gamma_{j1}^0\mathbf{Z}_i, \boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{Z}_{i,\backslash j}) + 2\mathrm{Cov}(\gamma_{j1}^0\epsilon_i^2, \boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{Z}_{i,\backslash j}) \\
&= \theta_{i;jj} + (\gamma_{j1}^0)^2\boldsymbol{\beta}^{0\intercal}\mathbf{\Theta}_i\boldsymbol{\beta}^0 + (\gamma_{j1}^0)^2\kappa_i + \boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{\Theta}_{i;\backslash j,\backslash j}\boldsymbol{\gamma}_{j,\backslash 1}^0 \\
&\quad - 2\gamma_{j1}^0\boldsymbol{\beta}^{0\intercal}\mathbf{\Theta}_{i;\cdot,j} - 2\gamma_{j1}^0\varpi_{i;j} - 2\boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\mathbf{\Theta}_{i;\backslash j,j} + 2(\gamma_{j1}^0)^2\boldsymbol{\beta}^{0\intercal}\boldsymbol{\varpi}_i \\
&\quad + 2\gamma_{j1}^0\boldsymbol{\beta}^{0\intercal}\mathbf{\Theta}_{i;\cdot,\backslash j}\boldsymbol{\gamma}_{j,\backslash 1}^0 + 2\gamma_{j1}^0\boldsymbol{\gamma}_{j,\backslash 1}^{0\intercal}\boldsymbol{\varpi}_{i;\backslash j}.
\end{aligned}$$

In addition, with the independence between ϵ_i and $\mathbf{X}_i$, and $\epsilon_i\gamma_{j1}^0 + \eta_{ij,y} = X_{ij} - \mathbf{X}_i^{\intercal}\boldsymbol{\beta}^0\gamma_{j1}^0 - \mathbf{X}_{i,\backslash j}^{\intercal}\boldsymbol{\gamma}_{j,\backslash 1}^0$, we have $\mathrm{Cov}(\epsilon_i, \epsilon_i\gamma_{j1}^0 + \eta_{ij,y}) = 0$, i.e., $-\gamma_{j1}^0\mathrm{Var}(\epsilon_i) = \mathrm{Cov}(\epsilon_i, \eta_{ij,y})$. Hence

$$\begin{aligned}
\mathrm{Var}(W_{ni}^{(\text{inv})}) &= \mathrm{Var}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0) - \epsilon_i^2\gamma_{j1}^0) \\
&= \mathrm{Var}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0)) + \mathrm{Var}(\epsilon_i^2\gamma_{j1}^0) - 2\mathrm{Cov}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0), \epsilon_i^2\gamma_{j1}^0) \\
&= \mathrm{Var}(\epsilon_i)\mathrm{Var}(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0) + \mathrm{Var}(\epsilon_i^2\gamma_{j1}^0) - 2\mathrm{Cov}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0), \epsilon_i^2\gamma_{j1}^0).
\end{aligned}$$

By the formula $\text{Cov}(X,Y) = \text{E}(XY) - \text{E}(X)\text{E}(Y)$ for any random variable X and Y, it yields $\text{Var}(W_{ni}^{(\text{inv})}) = \text{Var}(\epsilon_i)[\text{Var}(\eta_{ij,y}) + \text{Var}(\epsilon_i\gamma_{j1}^0) + 2\text{Cov}(\eta_{ij,y}, \epsilon_i\gamma_{j1}^0)] + \text{Var}(\epsilon_i^2\gamma_{j1}^0) - 2\text{E}(\gamma_{j1}^0\epsilon_i^3(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0)) + 2\text{E}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0))\text{E}(\epsilon_i^2\gamma_{j1}^0)$. Due to the independence between ϵ_i and $\epsilon_i\gamma_{j1}^0 + \eta_{ij,y}$, we have $\text{E}(\gamma_{j1}^0\epsilon_i^3(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0)) = \text{E}(\epsilon_i(\eta_{ij,y} + \epsilon_i\gamma_{j1}^0))\text{E}(\epsilon_i^2\gamma_{j1}^0) = 0$. It then follows that

$$\begin{aligned}
\text{Var}(W_{ni}^{(\text{inv})}) &= \text{Var}(\epsilon_i)[\text{Var}(\eta_{ij,y}) + \text{Var}(\epsilon_i\gamma_{j1}^0) - 2(\gamma_{j1}^0)^2\text{Var}(\epsilon_i)] + \text{Var}(\epsilon_i^2\gamma_{j1}^0) \\
&= \text{Var}(\epsilon_i)[\text{Var}(\eta_{ij,y}) - (\gamma_{j1}^0)^2\text{Var}(\epsilon_i)] + \text{Var}(\epsilon_i^2\gamma_{j1}^0) \\
&= \text{Var}(\epsilon_i)\text{Var}(\eta_{ij,y}) + (\gamma_{j1}^0)^2(\text{Var}(\epsilon_i^2) - \text{Var}^2(\epsilon_i)).
\end{aligned}$$

If furthermore we assume homoscedasticity and normality for the error term then we have $\text{Var}(\epsilon_i^2) - \text{Var}^2(\epsilon_i) = \text{E}(\epsilon_i^4) - 2[\text{E}(\epsilon_i^2)]^2 = 3\sigma_\epsilon^4 - 2\sigma_\epsilon^4 = \text{Var}^2(\epsilon_i)$, which leads to the same result in Theorem 3.1 from Liu and Luo (2014), i.e.,

$$\begin{aligned}
\text{Var}(W_{ni}^{(\text{inv})}) &= \text{Var}(\epsilon_i)\text{Var}(\eta_{ij,y}) + (\gamma_{j1}^0)^2(\text{Var}(\epsilon_i^2) - \text{Var}^2(\epsilon_i)) \\
&= \text{Var}(\epsilon_i)\text{Var}(\eta_{ij,y}) + [\text{Cov}(\epsilon_i, \eta_{ij,y})]^2 = \sigma_\epsilon^2\sigma_{\eta_j,y}^2 + (\gamma_{j1}^0)^2\sigma_\epsilon^4 \\
&= \sigma_\epsilon^2\sigma_{\eta_j,y}^2 + (\beta_j^0)^2\sigma_{\eta_j,y}^4.
\end{aligned}$$

□

For the proof of the three propositions about the asymptotic normality, they are followed from the proof of the corresponding theorems. We here just prove the Proposition 2.

Proof of Proposition 2. In order to get the asymptotic normality of $\hat{\beta}_j^{(\text{kfc-de})}$, we have to deal with $n^{-1}\sum_{i=1}^n \tilde{X}_{ij}^2$. Now since

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \tilde{X}_{ij}^2 &= \frac{1}{n}\sum_{i=1}^n \left\{X_{ij} - \mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbf{X}_{i\mathcal{S}}\right\}^2 \\
&= \frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_j - \frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_j = \frac{1}{n}\mathbb{X}_j^\intercal\left\{\mathbf{I} - \mathbb{X}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S})^{-1}\mathbb{X}_\mathcal{S}^\intercal\right\}\mathbb{X}_j,
\end{aligned}$$

we have

$$\begin{aligned}
&|\frac{1}{n}\sum_{i=1}^n \tilde{X}_{ij}^2 - (\sigma_{jj} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}j})| \\
=&|\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_j - \frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}(\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_\mathcal{S}/n)^{-1}\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_j/n - (\sigma_{jj} - \mathbf{\Sigma}_{j\mathcal{S}}\mathbf{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\mathbf{\Sigma}_{\mathcal{S}j})| \\
\le&\Big\{\big|\mathbb{X}_j^\intercal\mathbb{X}_j/n - \sigma_{jj}\big| + 2\|\mathbb{X}_j^\intercal\mathbb{X}_\mathcal{S}/n - \mathbf{\Sigma}_{j\mathcal{S}}\|_\infty\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max} \\
&+ \lambda_{\max}^2\|\mathbf{\Delta}_\mathcal{S}\|_2 + 2\sqrt{|\mathcal{S}|}\lambda_{\max}\|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_j/n - \mathbf{\Sigma}_{\mathcal{S}j}\|_\infty \\
&+ \lambda_{\min}^{-1}\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_j/n - \mathbf{\Sigma}_{\mathcal{S}j}\|_2^2 + \|\mathbf{\Delta}_\mathcal{S}\|_2\|\mathbb{X}_\mathcal{S}^\intercal\mathbb{X}_j/n - \mathbf{\Sigma}_{\mathcal{S}j}\|_2^2\Big\}.
\end{aligned}$$

And since

$$\text{P}\Big(\sup_{\mathcal{S}:|\mathcal{S}|\le m}|\sigma_{jj} - \frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_j| \ge \epsilon\Big) \le p^m\text{P}\Big(|\sigma_{jj} - \frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_j| \ge \epsilon\Big) \le C_1p^m\exp(-C_2n\epsilon^2)$$

we have $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}|\sigma_{jj}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_j|=O_p(\sqrt{m\log p/n})$.

Now for the term $\|\boldsymbol{\Sigma}_{j\mathcal{S}}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}\|_\infty$, we have proved above that

$$\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\|\big(\boldsymbol{\Sigma}_{j\mathcal{S}}-\frac{1}{n}\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}\big)\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\|_\infty=O_p(\sqrt{m\log p/n}).$$

By $\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2=O_p(\sqrt{m^3\log p/n})$, we have

$$\begin{aligned}
&\sup_{\mathcal{S}:|\mathcal{S}|\leq m}|\frac{1}{n}\sum_{i=1}^n\tilde{X}_{ij}^2-(\sigma_{jj}-\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}j})|\\
\leq&\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\Big\{\big|\mathbb{X}_j^\intercal\mathbb{X}_j/n-\sigma_{jj}\big|+2\|\mathbb{X}_j^\intercal\mathbb{X}_{\mathcal{S}}/n-\boldsymbol{\Sigma}_{j\mathcal{S}}\|_\infty\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\\
&+\lambda_{\max}^2\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2+2\sqrt{|\mathcal{S}|}\lambda_{\max}\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2\|\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_j/n-\boldsymbol{\Sigma}_{\mathcal{S}j}\|_\infty\\
&+\lambda_{\min}^{-1}\|\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_j/n-\boldsymbol{\Sigma}_{\mathcal{S}j}\|_2^2+\|\boldsymbol{\Delta}_{\mathcal{S}}\|_2\|\mathbb{X}_{\mathcal{S}}^\intercal\mathbb{X}_j/n-\boldsymbol{\Sigma}_{\mathcal{S}j}\|_2^2\Big\}\\
=&\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\Big\{O_p(\sqrt{m\log p/n})+O_p(\sqrt{m\log p/n})\sqrt{|\mathcal{S}|}\lambda_{\min}^{-1}\lambda_{\max}\\
&+\lambda_{\max}^2O_p(\sqrt{m^3\log p/n})+\sqrt{|\mathcal{S}|}\lambda_{\max}O_p(\sqrt{m^3\log p/n})O_p(\sqrt{m\log p/n})\\
&+|\mathcal{S}|O_p(\sqrt{m\log p/n})^2\lambda_{\min}^{-1}+|\mathcal{S}|O_p(\sqrt{m\log p/n})^2O_p(\sqrt{m^3\log p/n})\Big\}\\
=&O_p\{\sqrt{m^3\log p/n}\}.
\end{aligned}$$

Thus we have

$$\sup_{\mathcal{S}:|\mathcal{S}|\leq m}\Big|\frac{1}{n}\sum_{i=1}^n\tilde{X}_{ij}^2-(\sigma_{jj}-\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}j})\Big|=O_p\{\sqrt{m^3\log p/n}\}=o_p(1).\tag{2.13}$$

Hence we have the following asymptotic normality by Slutsky's theorem

$$\sqrt{n}(\hat{\beta}_j^{(\text{kfc-de})}-\beta_j^0)=\frac{\sum_{i=1}^n m_{ni}^{(\text{kfc})}(\beta_j^0)/\sqrt{n}}{\sum_{i=1}^n\tilde{X}_{ij}^2/n}\xrightarrow{d}\mathrm{N}(0,\sigma_{\text{kfc}}^2),$$

where $\sigma_{\text{kfc}}^2=\lim_{n\to\infty}n^{-1}\sum_{i=1}^n(\theta_{i;jj}-2\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Theta}_{i;j\mathcal{S}}+\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Theta}_{i;\mathcal{S}\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}j})/(\sigma_{jj}-\boldsymbol{\Sigma}_{j\mathcal{S}}\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1}\boldsymbol{\Sigma}_{\mathcal{S}j})$. □

3 Additional simulation results

In this Section, we provide more simulation results with different covariance structures for covariates and different distributions for the error term in Tables 3-30.

Table 1: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.2$, and the random error are generated by the standard normal distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.054	0.304	0.760	0.984	1.000	1.000
		400	0.052	0.482	0.964	1.000	1.000	1.000
	200	200	0.052	0.294	0.762	0.976	1.000	1.000
		400	0.044	0.460	0.980	1.000	1.000	1.000
	500	200	0.064	0.292	0.760	0.972	1.000	1.000
		400	0.040	0.488	0.972	1.000	1.000	1.000
EL-INV	100	200	0.040	0.296	0.748	0.984	1.000	1.000
		400	0.054	0.470	0.962	1.000	1.000	1.000
	200	200	0.044	0.290	0.774	0.976	1.000	1.000
		400	0.038	0.458	0.980	1.000	1.000	1.000
	500	200	0.048	0.276	0.784	0.978	1.000	1.000
		400	0.034	0.490	0.972	1.000	1.000	1.000
EL-LASSO	100	200	0.052	0.312	0.770	0.990	1.000	1.000
		400	0.054	0.490	0.970	1.000	1.000	1.000
	200	200	0.048	0.308	0.786	0.982	1.000	1.000
		400	0.038	0.462	0.978	1.000	1.000	1.000
	500	200	0.056	0.300	0.788	0.980	1.000	1.000
		400	0.042	0.512	0.976	1.000	1.000	1.000
Wald	100	200	0.048	0.266	0.748	0.964	1.000	1.000
		400	0.048	0.502	0.970	1.000	1.000	1.000
	200	200	0.064	0.270	0.742	0.972	1.000	1.000
		400	0.038	0.486	0.978	1.000	1.000	1.000
	500	200	0.052	0.284	0.794	0.968	0.998	1.000
		400	0.040	0.486	0.978	1.000	1.000	1.000
Score	100	200	0.050	0.264	0.746	0.962	1.000	1.000
		400	0.052	0.480	0.966	1.000	1.000	1.000
	200	200	0.062	0.268	0.740	0.970	1.000	1.000
		400	0.040	0.474	0.978	1.000	1.000	1.000
	500	200	0.062	0.272	0.794	0.970	0.998	1.000
		400	0.038	0.498	0.976	1.000	1.000	1.000

Table 2: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.2$, and the random error are generated by the standard normal distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.060	0.256	0.754	0.964	1.000	1.000
		400	0.046	0.474	0.970	1.000	1.000	1.000
	200	200	0.052	0.256	0.718	0.964	1.000	1.000
		400	0.048	0.506	0.972	1.000	1.000	1.000
	500	200	0.058	0.254	0.708	0.966	1.000	1.000
		400	0.048	0.522	0.980	1.000	1.000	1.000
EL-INV	100	200	0.056	0.260	0.766	0.970	1.000	1.000
		400	0.040	0.460	0.974	1.000	1.000	1.000
	200	200	0.048	0.236	0.712	0.976	1.000	1.000
		400	0.052	0.512	0.976	1.000	1.000	1.000
	500	200	0.044	0.254	0.724	0.968	1.000	1.000
		400	0.046	0.508	0.978	1.000	1.000	1.000
EL-LASSO	100	200	0.062	0.272	0.790	0.972	1.000	1.000
		400	0.046	0.486	0.974	1.000	1.000	1.000
	200	200	0.056	0.252	0.722	0.980	1.000	1.000
		400	0.052	0.510	0.980	1.000	1.000	1.000
	500	200	0.052	0.280	0.736	0.970	1.000	1.000
		400	0.044	0.526	0.984	1.000	1.000	1.000
Wald	100	200	0.042	0.260	0.754	0.968	0.998	1.000
		400	0.040	0.486	0.976	1.000	1.000	1.000
	200	200	0.058	0.278	0.766	0.970	1.000	1.000
		400	0.050	0.502	0.976	1.000	1.000	1.000
	500	200	0.054	0.284	0.764	0.984	1.000	1.000
		400	0.044	0.502	0.978	1.000	1.000	1.000
Score	100	200	0.042	0.262	0.758	0.966	1.000	1.000
		400	0.042	0.490	0.974	1.000	1.000	1.000
	200	200	0.058	0.280	0.760	0.968	1.000	1.000
		400	0.048	0.498	0.976	1.000	1.000	1.000
	500	200	0.058	0.268	0.764	0.988	1.000	1.000
		400	0.042	0.506	0.980	1.000	1.000	1.000

Table 3: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.2$, and the random error are generated by the standard normal distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.070	0.260	0.728	0.954	1.000	1.000
		400	0.044	0.508	0.976	1.000	1.000	1.000
	200	200	0.052	0.300	0.760	0.970	0.998	1.000
		400	0.042	0.514	0.978	1.000	1.000	1.000
	500	200	0.062	0.248	0.724	0.960	1.000	1.000
		400	0.086	0.486	0.958	1.000	1.000	1.000
EL-INV	100	200	0.072	0.256	0.740	0.952	1.000	1.000
		400	0.040	0.488	0.968	1.000	1.000	1.000
	200	200	0.038	0.296	0.750	0.982	1.000	1.000
		400	0.038	0.504	0.974	1.000	1.000	1.000
	500	200	0.062	0.232	0.734	0.966	1.000	1.000
		400	0.072	0.472	0.962	1.000	1.000	1.000
EL-LASSO	100	200	0.074	0.276	0.748	0.956	1.000	1.000
		400	0.048	0.512	0.974	1.000	1.000	1.000
	200	200	0.048	0.326	0.762	0.982	1.000	1.000
		400	0.040	0.532	0.974	1.000	1.000	1.000
	500	200	0.068	0.254	0.762	0.968	1.000	1.000
		400	0.088	0.486	0.964	1.000	1.000	1.000
Wald	100	200	0.040	0.228	0.734	0.976	1.000	1.000
		400	0.048	0.480	0.968	1.000	1.000	1.000
	200	200	0.042	0.238	0.776	0.972	0.998	1.000
		400	0.034	0.528	0.976	1.000	1.000	1.000
	500	200	0.052	0.228	0.730	0.976	1.000	1.000
		400	0.078	0.482	0.964	1.000	1.000	1.000
Score	100	200	0.042	0.226	0.738	0.974	1.000	1.000
		400	0.040	0.476	0.966	1.000	1.000	1.000
	200	200	0.042	0.232	0.770	0.974	0.998	1.000
		400	0.036	0.522	0.976	1.000	1.000	1.000
	500	200	0.058	0.236	0.726	0.978	1.000	1.000
		400	0.084	0.484	0.958	1.000	1.000	1.000

Table 4: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.2$, and the random error are generated by a mixture normal distribution $0.7\mathrm{N}(0,1) + 0.3\mathrm{N}(0,5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.062	0.086	0.188	0.324	0.482	0.632
		400	0.048	0.118	0.276	0.536	0.760	0.916
	200	200	0.072	0.118	0.202	0.316	0.468	0.642
		400	0.046	0.102	0.224	0.470	0.730	0.904
	500	200	0.052	0.070	0.152	0.262	0.428	0.602
		400	0.064	0.110	0.248	0.516	0.738	0.876
EL-INV	100	200	0.048	0.066	0.156	0.290	0.454	0.618
		400	0.038	0.100	0.256	0.510	0.738	0.904
	200	200	0.058	0.084	0.174	0.288	0.456	0.616
		400	0.034	0.096	0.212	0.458	0.714	0.888
	500	200	0.056	0.058	0.134	0.242	0.412	0.562
		400	0.062	0.098	0.238	0.500	0.736	0.872
EL-LASSO	100	200	0.060	0.086	0.172	0.316	0.486	0.652
		400	0.046	0.124	0.270	0.536	0.754	0.924
	200	200	0.078	0.106	0.198	0.302	0.474	0.658
		400	0.042	0.106	0.238	0.490	0.744	0.910
	500	200	0.062	0.064	0.156	0.280	0.446	0.602
		400	0.070	0.114	0.260	0.524	0.752	0.884
Wald	100	200	0.044	0.064	0.146	0.280	0.424	0.580
		400	0.046	0.112	0.278	0.548	0.790	0.918
	200	200	0.038	0.074	0.162	0.284	0.446	0.660
		400	0.040	0.094	0.252	0.514	0.768	0.934
	500	200	0.054	0.060	0.150	0.290	0.454	0.626
		400	0.064	0.102	0.242	0.514	0.766	0.916
Score	100	200	0.038	0.068	0.144	0.272	0.428	0.588
		400	0.048	0.122	0.260	0.546	0.792	0.920
	200	200	0.036	0.078	0.160	0.290	0.454	0.666
		400	0.044	0.088	0.248	0.502	0.752	0.930
	500	200	0.052	0.060	0.162	0.292	0.460	0.636
		400	0.062	0.100	0.242	0.518	0.760	0.908

Table 5: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.2$, and the random error are generated by a mixture normal distribution $0.7\mathrm{N}(0,1) + 0.3\mathrm{N}(0,5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.054	0.080	0.168	0.276	0.440	0.606
		400	0.056	0.106	0.278	0.538	0.768	0.894
	200	200	0.056	0.096	0.166	0.294	0.444	0.596
		400	0.070	0.110	0.274	0.532	0.762	0.906
	500	200	0.056	0.080	0.132	0.278	0.414	0.562
		400	0.048	0.096	0.254	0.506	0.724	0.872
EL-INV	100	200	0.036	0.064	0.142	0.260	0.414	0.586
		400	0.040	0.098	0.256	0.524	0.754	0.894
	200	200	0.054	0.078	0.146	0.264	0.408	0.598
		400	0.052	0.104	0.264	0.522	0.758	0.898
	500	200	0.054	0.070	0.118	0.248	0.426	0.562
		400	0.042	0.094	0.236	0.488	0.712	0.860
EL-LASSO	100	200	0.054	0.074	0.166	0.286	0.448	0.622
		400	0.054	0.110	0.288	0.562	0.768	0.908
	200	200	0.060	0.094	0.170	0.292	0.452	0.612
		400	0.058	0.118	0.286	0.554	0.780	0.914
	500	200	0.066	0.072	0.144	0.284	0.438	0.586
		400	0.048	0.100	0.266	0.516	0.734	0.888
Wald	100	200	0.054	0.090	0.168	0.314	0.474	0.678
		400	0.048	0.098	0.280	0.564	0.788	0.912
	200	200	0.058	0.090	0.170	0.294	0.460	0.666
		400	0.052	0.112	0.286	0.552	0.808	0.924
	500	200	0.054	0.096	0.192	0.292	0.452	0.638
		400	0.052	0.090	0.276	0.530	0.758	0.892
Score	100	200	0.054	0.084	0.166	0.306	0.494	0.668
		400	0.050	0.098	0.278	0.570	0.788	0.916
	200	200	0.050	0.084	0.168	0.286	0.468	0.656
		400	0.054	0.110	0.276	0.552	0.800	0.930
	500	200	0.052	0.090	0.182	0.288	0.454	0.638
		400	0.052	0.090	0.258	0.526	0.760	0.888

Table 6: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.2$, and the random error are generated by a mixture normal distribution $0.7\mathrm{N}(0,1)+0.3\mathrm{N}(0,5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.068	0.074	0.188	0.318	0.498	0.632
		400	0.064	0.084	0.222	0.470	0.682	0.858
	200	200	0.078	0.096	0.184	0.300	0.472	0.622
		400	0.056	0.094	0.262	0.490	0.710	0.872
	500	200	0.066	0.102	0.180	0.286	0.444	0.604
		400	0.054	0.078	0.228	0.482	0.698	0.868
EL-INV	100	200	0.056	0.068	0.162	0.308	0.462	0.628
		400	0.054	0.070	0.222	0.448	0.656	0.848
	200	200	0.070	0.062	0.152	0.278	0.456	0.598
		400	0.046	0.078	0.230	0.458	0.692	0.866
	500	200	0.052	0.088	0.168	0.272	0.440	0.606
		400	0.046	0.082	0.218	0.458	0.702	0.864
EL-LASSO	100	200	0.062	0.092	0.186	0.330	0.500	0.648
		400	0.066	0.084	0.228	0.486	0.692	0.860
	200	200	0.082	0.082	0.186	0.306	0.478	0.644
		400	0.058	0.098	0.264	0.498	0.732	0.876
	500	200	0.058	0.100	0.188	0.310	0.476	0.626
		400	0.056	0.090	0.242	0.476	0.712	0.884
Wald	100	200	0.056	0.074	0.138	0.280	0.452	0.618
		400	0.064	0.084	0.226	0.494	0.726	0.886
	200	200	0.054	0.084	0.166	0.330	0.488	0.680
		400	0.054	0.088	0.262	0.496	0.746	0.892
	500	200	0.046	0.062	0.134	0.260	0.434	0.622
		400	0.050	0.084	0.232	0.510	0.726	0.888
Score	100	200	0.054	0.070	0.134	0.270	0.442	0.624
		400	0.060	0.082	0.226	0.486	0.722	0.882
	200	200	0.052	0.086	0.168	0.318	0.490	0.668
		400	0.052	0.092	0.256	0.492	0.746	0.886
	500	200	0.046	0.060	0.138	0.258	0.434	0.620
		400	0.048	0.076	0.238	0.504	0.722	0.894

Table 7: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.2$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.060	0.154	0.422	0.682	0.856	0.930
		400	0.072	0.242	0.642	0.898	0.968	0.986
	200	200	0.076	0.184	0.414	0.656	0.808	0.914
		400	0.058	0.214	0.592	0.892	0.968	0.984
	500	200	0.052	0.168	0.382	0.668	0.854	0.950
		400	0.050	0.220	0.590	0.880	0.974	0.992
EL-INV	100	200	0.044	0.140	0.396	0.660	0.840	0.924
		400	0.060	0.234	0.622	0.898	0.968	0.986
	200	200	0.062	0.166	0.394	0.656	0.814	0.914
		400	0.048	0.214	0.578	0.884	0.970	0.986
	500	200	0.048	0.136	0.362	0.674	0.860	0.954
		400	0.040	0.212	0.584	0.882	0.976	0.988
EL-LASSO	100	200	0.054	0.148	0.440	0.690	0.864	0.932
		400	0.068	0.250	0.650	0.904	0.968	0.988
	200	200	0.078	0.176	0.416	0.670	0.832	0.920
		400	0.060	0.218	0.606	0.892	0.974	0.986
	500	200	0.054	0.162	0.394	0.694	0.870	0.954
		400	0.046	0.236	0.606	0.894	0.976	0.992
Wald	100	200	0.038	0.132	0.384	0.658	0.868	0.970
		400	0.066	0.240	0.640	0.926	0.988	0.994
	200	200	0.034	0.132	0.384	0.688	0.868	0.950
		400	0.052	0.206	0.602	0.916	0.988	0.996
	500	200	0.054	0.170	0.378	0.666	0.874	0.956
		400	0.044	0.214	0.606	0.918	0.992	0.998
Score	100	200	0.040	0.130	0.370	0.662	0.866	0.968
		400	0.060	0.232	0.636	0.922	0.988	0.994
	200	200	0.034	0.136	0.388	0.690	0.868	0.948
		400	0.052	0.198	0.604	0.916	0.988	0.994
	500	200	0.054	0.168	0.374	0.664	0.874	0.956
		400	0.042	0.210	0.598	0.916	0.992	0.998

Table 8: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.2$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.056	0.168	0.442	0.720	0.874	0.936
		400	0.060	0.238	0.620	0.888	0.970	0.986
	200	200	0.088	0.152	0.394	0.654	0.842	0.922
		400	0.070	0.236	0.596	0.900	0.976	0.996
	500	200	0.088	0.142	0.378	0.652	0.846	0.940
		400	0.052	0.230	0.626	0.896	0.974	0.996
EL-INV	100	200	0.046	0.142	0.436	0.716	0.878	0.950
		400	0.044	0.222	0.618	0.884	0.964	0.988
	200	200	0.068	0.130	0.384	0.654	0.834	0.924
		400	0.060	0.206	0.584	0.896	0.980	0.992
	500	200	0.064	0.118	0.382	0.650	0.852	0.938
		400	0.050	0.222	0.624	0.886	0.968	0.996
EL-LASSO	100	200	0.062	0.164	0.454	0.736	0.886	0.948
		400	0.056	0.238	0.642	0.896	0.968	0.988
	200	200	0.082	0.152	0.406	0.674	0.850	0.926
		400	0.070	0.234	0.616	0.908	0.982	0.994
	500	200	0.080	0.138	0.398	0.686	0.874	0.942
		400	0.056	0.244	0.644	0.906	0.974	0.996
Wald	100	200	0.050	0.148	0.396	0.700	0.896	0.958
		400	0.036	0.216	0.650	0.920	0.986	0.996
	200	200	0.048	0.142	0.362	0.664	0.856	0.944
		400	0.052	0.216	0.610	0.908	0.990	1.000
	500	200	0.046	0.130	0.340	0.668	0.862	0.960
		400	0.044	0.238	0.636	0.926	0.994	0.998
Score	100	200	0.046	0.142	0.384	0.702	0.900	0.960
		400	0.038	0.224	0.638	0.910	0.986	0.996
	200	200	0.054	0.134	0.364	0.652	0.862	0.942
		400	0.048	0.206	0.616	0.906	0.988	1.000
	500	200	0.044	0.132	0.348	0.668	0.866	0.960
		400	0.044	0.232	0.624	0.926	0.994	0.998

Table 9: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.2$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.058	0.136	0.366	0.654	0.838	0.920
		400	0.058	0.216	0.626	0.908	0.966	0.988
	200	200	0.046	0.156	0.366	0.652	0.844	0.928
		400	0.038	0.220	0.616	0.902	0.974	0.994
	500	200	0.066	0.144	0.388	0.678	0.852	0.944
		400	0.038	0.246	0.616	0.912	0.988	0.998
EL-INV	100	200	0.038	0.120	0.364	0.630	0.838	0.924
		400	0.048	0.202	0.616	0.902	0.966	0.988
	200	200	0.032	0.134	0.374	0.642	0.842	0.926
		400	0.036	0.198	0.584	0.906	0.972	0.992
	500	200	0.056	0.132	0.370	0.642	0.848	0.938
		400	0.032	0.230	0.608	0.902	0.984	0.998
EL-LASSO	100	200	0.054	0.130	0.386	0.664	0.848	0.928
		400	0.054	0.228	0.648	0.916	0.972	0.990
	200	200	0.042	0.162	0.398	0.660	0.860	0.930
		400	0.042	0.226	0.610	0.914	0.978	0.994
	500	200	0.068	0.160	0.396	0.672	0.866	0.942
		400	0.038	0.232	0.628	0.912	0.986	0.998
Wald	100	200	0.072	0.136	0.362	0.674	0.878	0.964
		400	0.050	0.200	0.664	0.936	0.988	0.998
	200	200	0.060	0.142	0.400	0.672	0.868	0.950
		400	0.026	0.166	0.604	0.930	0.996	1.000
	500	200	0.052	0.134	0.374	0.648	0.874	0.950
		400	0.040	0.206	0.632	0.938	0.994	0.998
Score	100	200	0.068	0.130	0.362	0.674	0.880	0.962
		400	0.050	0.194	0.666	0.936	0.986	0.998
	200	200	0.056	0.140	0.400	0.676	0.872	0.946
		400	0.032	0.176	0.602	0.932	0.994	1.000
	500	200	0.044	0.136	0.366	0.650	0.878	0.950
		400	0.040	0.208	0.638	0.942	0.994	0.998

Table 10: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.2$, and the random error are generated by a $0.7X_1\mathrm{N}(0,1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.072	0.232	0.632	0.908	0.990	1.000
		400	0.040	0.368	0.892	0.996	1.000	1.000
	200	200	0.072	0.224	0.654	0.928	0.994	1.000
		400	0.068	0.364	0.884	0.996	1.000	1.000
	500	200	0.088	0.252	0.652	0.924	0.986	1.000
		400	0.044	0.394	0.908	0.998	1.000	1.000
EL-INV	100	200	0.070	0.222	0.616	0.910	0.996	1.000
		400	0.038	0.352	0.892	1.000	1.000	1.000
	200	200	0.054	0.210	0.648	0.924	0.994	1.000
		400	0.066	0.338	0.876	0.996	1.000	1.000
	500	200	0.066	0.246	0.636	0.912	0.984	1.000
		400	0.038	0.388	0.912	1.000	1.000	1.000
EL-LASSO	100	200	0.076	0.242	0.626	0.914	0.996	1.000
		400	0.040	0.372	0.902	1.000	1.000	1.000
	200	200	0.068	0.220	0.664	0.934	0.996	1.000
		400	0.068	0.352	0.878	0.998	1.000	1.000
	500	200	0.076	0.258	0.648	0.920	0.988	1.000
		400	0.046	0.394	0.914	1.000	1.000	1.000
Wald	100	200	0.230	0.510	0.888	0.986	1.000	1.000
		400	0.216	0.682	0.988	1.000	1.000	1.000
	200	200	0.272	0.480	0.840	0.986	1.000	1.000
		400	0.258	0.672	0.976	1.000	1.000	1.000
	500	200	0.234	0.510	0.862	0.990	1.000	1.000
		400	0.210	0.714	0.988	1.000	1.000	1.000
Score	100	200	0.236	0.508	0.886	0.986	1.000	1.000
		400	0.212	0.688	0.988	1.000	1.000	1.000
	200	200	0.266	0.482	0.840	0.984	1.000	1.000
		400	0.258	0.664	0.978	1.000	1.000	1.000
	500	200	0.230	0.508	0.862	0.990	1.000	1.000
		400	0.214	0.716	0.986	1.000	1.000	1.000

Table 11: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.2$, and the random error are generated by a $0.7X_1\mathrm{N}(0,1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.088	0.222	0.616	0.894	0.984	1.000
		400	0.056	0.392	0.906	1.000	1.000	1.000
	200	200	0.080	0.248	0.646	0.922	0.986	1.000
		400	0.064	0.348	0.896	0.998	1.000	1.000
	500	200	0.062	0.254	0.660	0.932	0.990	1.000
		400	0.052	0.406	0.900	0.998	1.000	1.000
EL-INV	100	200	0.072	0.224	0.602	0.900	0.988	1.000
		400	0.048	0.372	0.894	1.000	1.000	1.000
	200	200	0.066	0.238	0.628	0.912	0.988	1.000
		400	0.060	0.334	0.892	0.998	1.000	1.000
	500	200	0.050	0.234	0.618	0.922	0.994	1.000
		400	0.048	0.384	0.896	0.998	1.000	1.000
EL-LASSO	100	200	0.084	0.230	0.616	0.906	0.988	1.000
		400	0.058	0.392	0.906	1.000	1.000	1.000
	200	200	0.074	0.246	0.656	0.928	0.988	1.000
		400	0.066	0.344	0.906	0.998	1.000	1.000
	500	200	0.052	0.252	0.634	0.930	0.994	1.000
		400	0.050	0.400	0.906	0.998	1.000	1.000
Wald	100	200	0.242	0.476	0.854	0.986	1.000	1.000
		400	0.244	0.680	0.992	1.000	1.000	1.000
	200	200	0.292	0.512	0.848	0.986	1.000	1.000
		400	0.242	0.674	0.984	1.000	1.000	1.000
	500	200	0.246	0.504	0.862	0.982	0.998	1.000
		400	0.214	0.730	0.984	1.000	1.000	1.000
Score	100	200	0.242	0.474	0.862	0.986	1.000	1.000
		400	0.244	0.680	0.992	1.000	1.000	1.000
	200	200	0.278	0.508	0.848	0.986	1.000	1.000
		400	0.234	0.670	0.982	1.000	1.000	1.000
	500	200	0.244	0.498	0.858	0.982	0.998	1.000
		400	0.212	0.736	0.984	1.000	1.000	1.000

Table 12: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.5$, and the random error are generated by a $N(0, 1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.060	0.254	0.698	0.948	0.998	1.000
		400	0.058	0.418	0.918	0.996	1.000	1.000
	200	200	0.030	0.252	0.690	0.958	0.998	1.000
		400	0.040	0.396	0.934	0.996	1.000	1.000
	500	200	0.042	0.232	0.664	0.944	0.996	1.000
		400	0.052	0.392	0.920	1.000	1.000	1.000
EL-INV	100	200	0.038	0.264	0.722	0.954	1.000	1.000
		400	0.044	0.470	0.938	0.998	1.000	1.000
	200	200	0.028	0.266	0.728	0.968	0.998	1.000
		400	0.050	0.438	0.966	0.996	1.000	1.000
	500	200	0.034	0.264	0.738	0.968	0.998	1.000
		400	0.050	0.446	0.956	1.000	1.000	1.000
EL-LASSO	100	200	0.046	0.304	0.744	0.966	1.000	1.000
		400	0.054	0.512	0.948	0.998	1.000	1.000
	200	200	0.036	0.306	0.764	0.970	0.998	1.000
		400	0.052	0.476	0.974	0.996	1.000	1.000
	500	200	0.040	0.288	0.764	0.970	0.998	1.000
		400	0.052	0.508	0.964	1.000	1.000	1.000
Wald	100	200	0.056	0.272	0.734	0.966	1.000	1.000
		400	0.050	0.520	0.930	1.000	1.000	1.000
	200	200	0.034	0.256	0.758	0.970	0.998	1.000
		400	0.058	0.476	0.978	1.000	1.000	1.000
	500	200	0.036	0.248	0.748	0.978	1.000	1.000
		400	0.058	0.476	0.962	1.000	1.000	1.000
Score	100	200	0.048	0.234	0.708	0.950	1.000	1.000
		400	0.030	0.498	0.942	1.000	1.000	1.000
	200	200	0.020	0.228	0.712	0.958	0.998	1.000
		400	0.042	0.438	0.974	1.000	1.000	1.000
	500	200	0.030	0.204	0.714	0.968	1.000	1.000
		400	0.040	0.424	0.944	1.000	1.000	1.000

Table 13: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.5$, and the random error are generated by a $N(0,1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.054	0.256	0.670	0.940	0.996	1.000
		400	0.044	0.402	0.922	1.000	1.000	1.000
	200	200	0.044	0.218	0.614	0.940	0.996	1.000
		400	0.050	0.406	0.906	0.998	1.000	1.000
	500	200	0.046	0.156	0.610	0.938	0.998	1.000
		400	0.036	0.406	0.932	1.000	1.000	1.000
EL-INV	100	200	0.042	0.258	0.712	0.954	0.998	1.000
		400	0.044	0.424	0.938	1.000	1.000	1.000
	200	200	0.036	0.224	0.672	0.952	0.998	1.000
		400	0.052	0.422	0.940	1.000	1.000	1.000
	500	200	0.038	0.176	0.640	0.966	0.998	1.000
		400	0.028	0.438	0.958	1.000	1.000	1.000
EL-LASSO	100	200	0.052	0.304	0.730	0.962	0.998	1.000
		400	0.044	0.450	0.944	1.000	1.000	1.000
	200	200	0.048	0.252	0.704	0.970	0.998	1.000
		400	0.058	0.432	0.948	1.000	1.000	1.000
	500	200	0.042	0.210	0.676	0.972	0.998	1.000
		400	0.036	0.462	0.960	1.000	1.000	1.000
Wald	100	200	0.064	0.340	0.804	0.980	0.998	1.000
		400	0.042	0.528	0.974	1.000	1.000	1.000
	200	200	0.052	0.298	0.790	0.984	1.000	1.000
		400	0.074	0.522	0.974	1.000	1.000	1.000
	500	200	0.036	0.260	0.790	0.988	0.998	1.000
		400	0.034	0.564	0.990	1.000	1.000	1.000
Score	100	200	0.060	0.304	0.750	0.968	0.998	1.000
		400	0.030	0.448	0.948	1.000	1.000	1.000
	200	200	0.032	0.240	0.732	0.970	0.998	1.000
		400	0.064	0.446	0.956	1.000	1.000	1.000
	500	200	0.036	0.208	0.724	0.982	0.998	1.000
		400	0.026	0.486	0.966	1.000	1.000	1.000

Table 14: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.5$, and the random error are generated by a $\mathrm{N}(0, 1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.062	0.230	0.654	0.928	0.990	1.000
		400	0.052	0.416	0.934	1.000	1.000	1.000
	200	200	0.046	0.246	0.668	0.942	0.996	1.000
		400	0.050	0.380	0.938	1.000	1.000	1.000
	500	200	0.066	0.244	0.670	0.922	0.996	1.000
		400	0.052	0.392	0.918	0.998	1.000	1.000
EL-INV	100	200	0.056	0.226	0.676	0.954	0.990	1.000
		400	0.046	0.428	0.948	1.000	1.000	1.000
	200	200	0.026	0.244	0.712	0.972	0.998	1.000
		400	0.034	0.386	0.952	1.000	1.000	1.000
	500	200	0.046	0.252	0.712	0.950	1.000	1.000
		400	0.044	0.402	0.938	0.998	1.000	1.000
EL-LASSO	100	200	0.066	0.254	0.710	0.956	0.99	1.000
		400	0.052	0.448	0.954	1.000	1.00	1.000
	200	200	0.038	0.258	0.726	0.974	1.00	1.000
		400	0.044	0.410	0.956	1.000	1.00	1.000
	500	200	0.058	0.262	0.714	0.950	1.00	1.000
		400	0.054	0.410	0.948	0.998	1.00	1.000
Wald	100	200	0.060	0.250	0.718	0.964	0.998	1.000
		400	0.048	0.474	0.970	1.000	1.000	1.000
	200	200	0.032	0.260	0.750	0.976	1.000	1.000
		400	0.044	0.450	0.962	1.000	1.000	1.000
	500	200	0.058	0.272	0.728	0.960	0.998	1.000
		400	0.054	0.454	0.962	1.000	1.000	1.000
Score	100	200	0.050	0.226	0.672	0.954	0.994	1.000
		400	0.046	0.422	0.948	1.000	1.000	1.000
	200	200	0.022	0.226	0.696	0.964	1.000	1.000
		400	0.038	0.398	0.952	1.000	1.000	1.000
	500	200	0.050	0.238	0.680	0.942	0.998	1.000
		400	0.044	0.396	0.950	1.000	1.000	1.000

Table 15: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.5$, and the random error are generated by a mixture normal distribution $0.7\text{N}(0,1) + 0.3\text{N}(0,5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.052	0.078	0.168	0.250	0.390	0.534
		400	0.056	0.102	0.240	0.464	0.696	0.848
	200	200	0.050	0.074	0.148	0.276	0.426	0.540
		400	0.058	0.094	0.224	0.436	0.636	0.814
	500	200	0.044	0.080	0.146	0.238	0.386	0.530
		400	0.040	0.086	0.208	0.416	0.624	0.792
EL-INV	100	200	0.040	0.080	0.176	0.282	0.408	0.552
		400	0.058	0.130	0.276	0.516	0.732	0.874
	200	200	0.040	0.076	0.158	0.310	0.452	0.594
		400	0.064	0.120	0.254	0.486	0.718	0.850
	500	200	0.034	0.062	0.162	0.280	0.406	0.582
		400	0.042	0.104	0.246	0.490	0.686	0.842
EL-LASSO	100	200	0.048	0.100	0.202	0.306	0.442	0.598
		400	0.062	0.142	0.298	0.558	0.768	0.906
	200	200	0.052	0.082	0.182	0.332	0.476	0.626
		400	0.068	0.138	0.286	0.536	0.742	0.870
	500	200	0.038	0.076	0.176	0.302	0.440	0.610
		400	0.048	0.110	0.272	0.528	0.716	0.872
Wald	100	200	0.044	0.066	0.174	0.272	0.434	0.592
		400	0.052	0.134	0.310	0.552	0.766	0.904
	200	200	0.040	0.060	0.158	0.304	0.462	0.630
		400	0.054	0.128	0.282	0.532	0.754	0.892
	500	200	0.038	0.068	0.156	0.268	0.442	0.624
		400	0.042	0.088	0.252	0.504	0.726	0.900
Score	100	200	0.034	0.058	0.126	0.250	0.390	0.558
		400	0.048	0.112	0.270	0.526	0.752	0.890
	200	200	0.032	0.052	0.122	0.262	0.418	0.566
		400	0.048	0.102	0.244	0.488	0.716	0.878
	500	200	0.032	0.058	0.134	0.236	0.392	0.592
		400	0.032	0.080	0.212	0.456	0.696	0.856

Table 16: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.5$, and the random error are generated by a mixture normal distribution $0.7\mathrm{N}(0,1){+}0.3\mathrm{N}(0,5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.050	0.080	0.150	0.260	0.408	0.564
		400	0.050	0.078	0.214	0.400	0.652	0.814
	200	200	0.060	0.078	0.136	0.246	0.376	0.496
		400	0.056	0.100	0.238	0.450	0.668	0.824
	500	200	0.050	0.074	0.168	0.266	0.418	0.550
		400	0.072	0.088	0.238	0.454	0.662	0.818
EL-INV	100	200	0.038	0.082	0.142	0.260	0.426	0.574
		400	0.028	0.088	0.206	0.418	0.660	0.816
	200	200	0.052	0.080	0.142	0.242	0.386	0.534
		400	0.038	0.098	0.236	0.440	0.686	0.836
	500	200	0.046	0.082	0.162	0.284	0.440	0.594
		400	0.052	0.092	0.240	0.484	0.692	0.840
EL-LASSO	100	200	0.048	0.100	0.164	0.302	0.464	0.624
		400	0.042	0.094	0.236	0.456	0.696	0.844
	200	200	0.062	0.100	0.158	0.286	0.426	0.584
		400	0.060	0.116	0.266	0.498	0.704	0.860
	500	200	0.050	0.100	0.192	0.318	0.482	0.622
		400	0.068	0.100	0.274	0.528	0.716	0.860
Wald	100	200	0.054	0.098	0.208	0.362	0.560	0.720
		400	0.034	0.122	0.306	0.566	0.768	0.902
	200	200	0.056	0.106	0.194	0.340	0.534	0.690
		400	0.062	0.146	0.318	0.596	0.792	0.938
	500	200	0.062	0.124	0.230	0.380	0.554	0.712
		400	0.050	0.138	0.354	0.628	0.812	0.924
Score	100	200	0.040	0.068	0.166	0.302	0.486	0.656
		400	0.030	0.090	0.248	0.478	0.716	0.872
	200	200	0.044	0.092	0.150	0.284	0.448	0.626
		400	0.048	0.114	0.264	0.536	0.744	0.904
	500	200	0.050	0.084	0.202	0.332	0.500	0.646
		400	0.046	0.108	0.280	0.548	0.750	0.894

Table 17: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.5$, and the random error are generated by a mixture normal distribution $0.7\mathrm{N}(0, 1) + 0.3\mathrm{N}(0, 5^2)$.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.066	0.094	0.158	0.250	0.388	0.522
		400	0.068	0.112	0.264	0.450	0.682	0.854
	200	200	0.084	0.088	0.138	0.244	0.354	0.496
		400	0.064	0.112	0.242	0.464	0.660	0.834
	500	200	0.072	0.078	0.134	0.242	0.386	0.544
		400	0.044	0.092	0.204	0.412	0.650	0.820
EL-INV	100	200	0.052	0.070	0.134	0.234	0.374	0.516
		400	0.058	0.112	0.264	0.454	0.684	0.858
	200	200	0.068	0.072	0.128	0.236	0.380	0.520
		400	0.054	0.106	0.234	0.448	0.668	0.836
	500	200	0.070	0.064	0.122	0.250	0.418	0.568
		400	0.036	0.076	0.208	0.420	0.674	0.830
EL-LASSO	100	200	0.060	0.084	0.158	0.254	0.424	0.562
		400	0.070	0.128	0.282	0.476	0.720	0.876
	200	200	0.080	0.082	0.164	0.264	0.412	0.550
		400	0.054	0.116	0.264	0.498	0.696	0.854
	500	200	0.074	0.078	0.138	0.276	0.444	0.612
		400	0.046	0.094	0.226	0.452	0.708	0.846
Wald	100	200	0.056	0.078	0.144	0.282	0.428	0.618
		400	0.058	0.124	0.302	0.550	0.780	0.922
	200	200	0.066	0.084	0.152	0.264	0.434	0.598
		400	0.048	0.110	0.268	0.538	0.756	0.898
	500	200	0.060	0.080	0.150	0.292	0.450	0.626
		400	0.036	0.082	0.252	0.502	0.762	0.906
Score	100	200	0.042	0.062	0.130	0.250	0.378	0.562
		400	0.054	0.106	0.266	0.480	0.732	0.890
	200	200	0.054	0.068	0.128	0.236	0.378	0.548
		400	0.042	0.096	0.246	0.480	0.710	0.880
	500	200	0.050	0.064	0.136	0.250	0.404	0.592
		400	0.034	0.076	0.206	0.448	0.694	0.866

Table 18: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.5$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.050	0.138	0.354	0.630	0.838	0.936
		400	0.054	0.200	0.564	0.846	0.958	0.982
	200	200	0.064	0.124	0.348	0.556	0.766	0.888
		400	0.044	0.180	0.530	0.812	0.942	0.976
	500	200	0.072	0.116	0.314	0.564	0.774	0.894
		400	0.036	0.206	0.556	0.866	0.940	0.978
EL-INV	100	200	0.042	0.146	0.380	0.656	0.844	0.934
		400	0.054	0.218	0.604	0.866	0.950	0.986
	200	200	0.054	0.124	0.358	0.606	0.808	0.912
		400	0.044	0.216	0.572	0.826	0.950	0.978
	500	200	0.060	0.132	0.340	0.610	0.804	0.918
		400	0.038	0.250	0.622	0.894	0.956	0.982
EL-LASSO	100	200	0.056	0.176	0.432	0.704	0.864	0.952
		400	0.066	0.246	0.648	0.888	0.960	0.986
	200	200	0.066	0.154	0.398	0.640	0.826	0.926
		400	0.054	0.252	0.598	0.868	0.962	0.982
	500	200	0.070	0.148	0.374	0.640	0.828	0.932
		400	0.062	0.284	0.652	0.908	0.964	0.988
Wald	100	200	0.056	0.134	0.366	0.698	0.878	0.968
		400	0.050	0.222	0.622	0.902	0.984	0.998
	200	200	0.060	0.130	0.364	0.606	0.820	0.940
		400	0.044	0.208	0.606	0.894	0.972	0.994
	500	200	0.056	0.122	0.330	0.648	0.858	0.944
		400	0.034	0.250	0.620	0.938	0.980	0.996
Score	100	200	0.042	0.106	0.338	0.636	0.864	0.956
		400	0.036	0.188	0.622	0.904	0.978	0.998
	200	200	0.050	0.110	0.302	0.584	0.794	0.928
		400	0.036	0.172	0.570	0.886	0.970	0.988
	500	200	0.050	0.110	0.286	0.596	0.832	0.932
		400	0.028	0.206	0.592	0.920	0.974	0.996

Table 19: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.5$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.060	0.158	0.370	0.578	0.802	0.910
		400	0.066	0.238	0.586	0.854	0.948	0.986
	200	200	0.048	0.122	0.318	0.586	0.790	0.888
		400	0.052	0.190	0.518	0.826	0.958	0.986
	500	200	0.058	0.136	0.330	0.566	0.796	0.898
		400	0.036	0.186	0.526	0.822	0.956	0.990
EL-INV	100	200	0.046	0.158	0.376	0.598	0.816	0.916
		400	0.060	0.238	0.612	0.858	0.956	0.986
	200	200	0.048	0.112	0.336	0.612	0.816	0.918
		400	0.040	0.186	0.540	0.836	0.966	0.984
	500	200	0.060	0.142	0.356	0.616	0.812	0.918
		400	0.034	0.212	0.550	0.854	0.962	0.992
EL-LASSO	100	200	0.060	0.186	0.398	0.642	0.836	0.930
		400	0.062	0.254	0.618	0.864	0.960	0.986
	200	200	0.054	0.140	0.388	0.656	0.840	0.940
		400	0.054	0.210	0.578	0.860	0.970	0.986
	500	200	0.066	0.166	0.382	0.656	0.828	0.924
		400	0.036	0.236	0.590	0.868	0.970	0.996
Wald	100	200	0.054	0.208	0.430	0.700	0.888	0.962
		400	0.062	0.286	0.674	0.934	0.992	0.998
	200	200	0.062	0.166	0.462	0.754	0.912	0.966
		400	0.060	0.256	0.692	0.932	0.990	1.000
	500	200	0.064	0.200	0.454	0.718	0.896	0.962
		400	0.040	0.266	0.662	0.932	0.996	1.000
Score	100	200	0.046	0.160	0.380	0.634	0.848	0.952
		400	0.046	0.236	0.616	0.910	0.976	0.998
	200	200	0.048	0.126	0.374	0.674	0.870	0.954
		400	0.048	0.210	0.618	0.890	0.982	0.998
	500	200	0.052	0.142	0.398	0.672	0.860	0.952
		400	0.034	0.196	0.604	0.900	0.994	1.000

Table 20: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.5$, and the random error are generated by a t_3 distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.064	0.126	0.328	0.568	0.756	0.882
		400	0.066	0.202	0.546	0.822	0.944	0.978
	200	200	0.060	0.138	0.314	0.614	0.790	0.902
		400	0.046	0.188	0.522	0.822	0.944	0.986
	500	200	0.056	0.120	0.364	0.622	0.810	0.910
		400	0.046	0.172	0.500	0.842	0.948	0.992
EL-INV	100	200	0.054	0.126	0.322	0.576	0.772	0.884
		400	0.052	0.200	0.552	0.834	0.950	0.976
	200	200	0.062	0.132	0.314	0.634	0.806	0.902
		400	0.044	0.176	0.552	0.834	0.950	0.988
	500	200	0.040	0.112	0.394	0.646	0.826	0.916
		400	0.044	0.164	0.530	0.846	0.954	0.990
EL-LASSO	100	200	0.068	0.140	0.358	0.608	0.784	0.898
		400	0.060	0.222	0.566	0.848	0.958	0.980
	200	200	0.068	0.150	0.354	0.664	0.834	0.918
		400	0.050	0.204	0.570	0.852	0.956	0.988
	500	200	0.052	0.134	0.424	0.666	0.834	0.932
		400	0.052	0.182	0.576	0.860	0.964	0.994
Wald	100	200	0.048	0.120	0.342	0.646	0.848	0.952
		400	0.056	0.208	0.604	0.906	0.976	0.994
	200	200	0.044	0.108	0.344	0.670	0.858	0.946
		400	0.042	0.208	0.584	0.882	0.988	0.998
	500	200	0.038	0.110	0.388	0.674	0.860	0.948
		400	0.046	0.176	0.608	0.916	0.986	0.998
Score	100	200	0.040	0.094	0.308	0.590	0.804	0.932
		400	0.056	0.186	0.558	0.870	0.968	0.990
	200	200	0.036	0.096	0.314	0.626	0.832	0.938
		400	0.028	0.176	0.544	0.864	0.982	0.998
	500	200	0.036	0.090	0.334	0.644	0.848	0.928
		400	0.040	0.160	0.558	0.896	0.980	0.996

Table 21: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.5$, and the random error are generated by a $0.7X_1\mathrm{N}(0, 1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.078	0.204	0.588	0.874	0.976	1.000
		400	0.056	0.344	0.890	0.988	0.998	1.000
	200	200	0.058	0.232	0.588	0.900	0.988	0.998
		400	0.056	0.322	0.844	0.990	0.998	1.000
	500	200	0.076	0.252	0.604	0.890	0.980	1.000
		400	0.058	0.366	0.854	0.994	1.000	1.000
EL-INV	100	200	0.052	0.210	0.588	0.870	0.972	1.000
		400	0.054	0.330	0.878	0.992	1.000	1.000
	200	200	0.044	0.220	0.590	0.896	0.988	0.996
		400	0.044	0.346	0.864	0.984	1.000	1.000
	500	200	0.062	0.238	0.614	0.890	0.980	1.000
		400	0.050	0.386	0.854	0.994	1.000	1.000
EL-LASSO	100	200	0.076	0.226	0.644	0.894	0.984	1.000
		400	0.068	0.378	0.914	0.994	1.000	1.000
	200	200	0.060	0.260	0.634	0.920	0.992	0.998
		400	0.060	0.386	0.886	0.992	1.000	1.000
	500	200	0.070	0.272	0.662	0.914	0.990	1.000
		400	0.060	0.416	0.894	0.998	1.000	1.000
Wald	100	200	0.254	0.496	0.820	0.982	1.000	1.000
		400	0.208	0.666	0.978	0.998	1.000	1.000
	200	200	0.222	0.480	0.862	0.994	1.000	1.000
		400	0.222	0.650	0.972	1.000	1.000	1.000
	500	200	0.204	0.508	0.868	0.982	1.000	1.000
		400	0.240	0.644	0.984	1.000	1.000	1.000
Score	100	200	0.240	0.448	0.804	0.974	1.000	1.000
		400	0.198	0.650	0.978	0.998	1.000	1.000
	200	200	0.202	0.458	0.838	0.992	1.000	1.000
		400	0.228	0.628	0.964	1.000	1.000	1.000
	500	200	0.186	0.496	0.850	0.982	1.000	1.000
		400	0.228	0.618	0.978	1.000	1.000	1.000

Table 22: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.5$, and the random error are generated by a $0.7X_1\mathrm{N}(0,1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.070	0.228	0.606	0.864	0.982	0.992
		400	0.068	0.348	0.874	0.994	1.000	1.000
	200	200	0.074	0.230	0.616	0.870	0.986	0.996
		400	0.066	0.304	0.858	0.982	1.000	1.000
	500	200	0.056	0.198	0.604	0.892	0.984	1.000
		400	0.056	0.416	0.884	0.994	1.000	1.000
EL-INV	100	200	0.056	0.210	0.588	0.864	0.980	0.992
		400	0.066	0.344	0.872	0.994	1.000	1.000
	200	200	0.050	0.224	0.604	0.876	0.988	0.996
		400	0.062	0.306	0.850	0.984	1.000	1.000
	500	200	0.050	0.192	0.596	0.892	0.984	1.000
		400	0.046	0.414	0.880	0.994	1.000	1.000
EL-LASSO	100	200	0.068	0.238	0.636	0.900	0.984	0.996
		400	0.068	0.372	0.894	0.994	1.000	1.000
	200	200	0.066	0.250	0.630	0.892	0.988	0.996
		400	0.068	0.330	0.868	0.988	1.000	1.000
	500	200	0.052	0.224	0.632	0.918	0.988	1.000
		400	0.054	0.452	0.900	0.994	1.000	1.000
Wald	100	200	0.246	0.524	0.856	0.986	1.000	1.000
		400	0.216	0.710	0.984	1.000	1.000	1.000
	200	200	0.262	0.538	0.862	0.992	0.998	1.000
		400	0.212	0.726	0.980	1.000	1.000	1.000
	500	200	0.210	0.512	0.900	0.998	1.000	1.000
		400	0.268	0.754	0.984	1.000	1.000	1.000
Score	100	200	0.220	0.492	0.836	0.982	1.000	1.000
		400	0.194	0.672	0.974	1.000	1.000	1.000
	200	200	0.232	0.492	0.840	0.990	0.998	1.000
		400	0.196	0.652	0.972	0.998	1.000	1.000
	500	200	0.194	0.474	0.874	0.998	1.000	1.000
		400	0.232	0.720	0.978	1.000	1.000	1.000

Table 23: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.5$, and the random error are generated by a $0.7X_1\text{N}(0,1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.044	0.234	0.612	0.868	0.976	0.994
		400	0.046	0.376	0.876	0.988	1.000	1.000
	200	200	0.074	0.206	0.566	0.864	0.978	0.996
		400	0.070	0.344	0.878	0.988	1.000	1.000
	500	200	0.060	0.198	0.556	0.882	0.984	1.000
		400	0.060	0.384	0.870	0.984	1.000	1.000
EL-INV	100	200	0.038	0.212	0.602	0.878	0.972	0.994
		400	0.044	0.362	0.880	0.988	1.000	1.000
	200	200	0.044	0.190	0.560	0.864	0.976	0.994
		400	0.060	0.334	0.876	0.988	1.000	1.000
	500	200	0.060	0.184	0.546	0.864	0.982	1.000
		400	0.052	0.358	0.866	0.986	1.000	1.000
EL-LASSO	100	200	0.052	0.234	0.636	0.896	0.982	0.992
		400	0.050	0.386	0.894	0.988	1.000	1.000
	200	200	0.064	0.206	0.586	0.896	0.984	0.996
		400	0.066	0.358	0.888	0.988	1.000	1.000
	500	200	0.072	0.214	0.588	0.888	0.986	1.000
		400	0.058	0.400	0.888	0.992	1.000	1.000
Wald	100	200	0.226	0.508	0.864	0.982	0.994	1.000
		400	0.222	0.686	0.976	1.000	1.000	1.000
	200	200	0.234	0.440	0.826	0.980	1.000	1.000
		400	0.238	0.672	0.972	1.000	1.000	1.000
	500	200	0.232	0.454	0.836	0.984	1.000	1.000
		400	0.224	0.684	0.982	1.000	1.000	1.000
Score	100	200	0.202	0.474	0.832	0.978	0.994	1.000
		400	0.194	0.656	0.974	1.000	1.000	1.000
	200	200	0.202	0.410	0.792	0.978	0.998	1.000
		400	0.222	0.648	0.964	1.000	1.000	1.000
	500	200	0.218	0.418	0.816	0.980	1.000	1.000
		400	0.210	0.668	0.978	1.000	1.000	1.000

Table 24: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.2$, and the random error are generated by a $X_1 \sum_{j=2}^{p} X_{j-1} X_j \mathrm{N}(0,1)/(p-1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.086	0.910	0.998	1.000	1.000	1.000
		400	0.052	0.990	1.000	1.000	1.000	1.000
	200	200	0.090	0.948	1.000	1.000	1.000	1.000
		400	0.070	0.992	1.000	1.000	1.000	1.000
	500	200	0.096	0.954	1.000	1.000	1.000	1.000
		400	0.054	0.998	1.000	1.000	1.000	1.000
EL-INV	100	200	0.078	0.896	0.996	1.000	1.000	1.000
		400	0.050	0.990	1.000	1.000	1.000	1.000
	200	200	0.078	0.938	1.000	1.000	1.000	1.000
		400	0.066	0.992	1.000	1.000	1.000	1.000
	500	200	0.090	0.948	1.000	1.000	1.000	1.000
		400	0.050	0.998	1.000	1.000	1.000	1.000
EL-LASSO	100	200	0.078	0.896	0.996	1.000	1.000	1.000
		400	0.050	0.990	1.000	1.000	1.000	1.000
	200	200	0.084	0.946	1.000	1.000	1.000	1.000
		400	0.070	0.992	1.000	1.000	1.000	1.000
	500	200	0.090	0.954	1.000	1.000	1.000	1.000
		400	0.050	0.998	1.000	1.000	1.000	1.000
Wald	100	200	0.232	0.982	1.000	1.000	1.000	1.000
		400	0.262	1.000	1.000	1.000	1.000	1.000
	200	200	0.212	0.996	1.000	1.000	1.000	1.000
		400	0.236	1.000	1.000	1.000	1.000	1.000
	500	200	0.248	0.992	1.000	1.000	1.000	1.000
		400	0.218	1.000	1.000	1.000	1.000	1.000
Score	100	200	0.226	0.982	1.000	1.000	1.000	1.000
		400	0.262	1.000	1.000	1.000	1.000	1.000
	200	200	0.212	0.996	1.000	1.000	1.000	1.000
		400	0.236	1.000	1.000	1.000	1.000	1.000
	500	200	0.240	0.992	1.000	1.000	1.000	1.000
		400	0.218	1.000	1.000	1.000	1.000	1.000

Table 25: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.2$, and the random error are generated by a $X_1 \sum_{j=2}^{p} X_{j-1}X_j \mathrm{N}(0,1)/(p-1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.094	0.964	1.000	1.000	1.000	1.000
		400	0.066	1.000	1.000	1.000	1.000	1.000
	200	200	0.078	0.986	1.000	1.000	1.000	1.000
		400	0.058	1.000	1.000	1.000	1.000	1.000
	500	200	0.050	0.994	1.000	1.000	1.000	1.000
		400	0.060	1.000	1.000	1.000	1.000	1.000
EL-INV	100	200	0.098	0.958	1.000	1.000	1.000	1.000
		400	0.064	1.000	1.000	1.000	1.000	1.000
	200	200	0.078	0.986	1.000	1.000	1.000	1.000
		400	0.052	1.000	1.000	1.000	1.000	1.000
	500	200	0.054	0.996	1.000	1.000	1.000	1.000
		400	0.056	1.000	1.000	1.000	1.000	1.000
EL-LASSO	100	200	0.096	0.962	1.000	1.000	1.000	1.000
		400	0.072	1.000	1.000	1.000	1.000	1.000
	200	200	0.078	0.986	1.000	1.000	1.000	1.000
		400	0.054	1.000	1.000	1.000	1.000	1.000
	500	200	0.052	0.996	1.000	1.000	1.000	1.000
		400	0.060	1.000	1.000	1.000	1.000	1.000
Wald	100	200	0.284	1.000	1.000	1.000	1.000	1.000
		400	0.246	1.000	1.000	1.000	1.000	1.000
	200	200	0.230	1.000	1.000	1.000	1.000	1.000
		400	0.256	1.000	1.000	1.000	1.000	1.000
	500	200	0.210	1.000	1.000	1.000	1.000	1.000
		400	0.236	1.000	1.000	1.000	1.000	1.000
Score	100	200	0.284	1.000	1.000	1.000	1.000	1.000
		400	0.236	1.000	1.000	1.000	1.000	1.000
	200	200	0.230	1.000	1.000	1.000	1.000	1.000
		400	0.258	1.000	1.000	1.000	1.000	1.000
	500	200	0.214	1.000	1.000	1.000	1.000	1.000
		400	0.238	1.000	1.000	1.000	1.000	1.000

Table 26: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a banded matrix with $\rho = 0.5$, and the random error are generated by a $X_1 \sum_{j=2}^{p} X_{j-1}X_j \mathrm{N}(0,1)/(p-1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.076	0.314	0.770	0.962	0.996	0.998
		400	0.062	0.544	0.970	0.998	1.000	1.000
	200	200	0.062	0.366	0.832	0.990	0.998	1.000
		400	0.052	0.558	0.968	0.998	1.000	1.000
	500	200	0.076	0.392	0.850	0.980	1.000	1.000
		400	0.058	0.564	0.980	1.000	1.000	1.000
EL-INV	100	200	0.068	0.306	0.770	0.962	0.996	1.000
		400	0.054	0.548	0.960	0.998	1.000	1.000
	200	200	0.054	0.362	0.836	0.992	0.996	1.000
		400	0.046	0.558	0.970	0.998	1.000	1.000
	500	200	0.064	0.394	0.860	0.988	1.000	1.000
		400	0.054	0.580	0.986	1.000	1.000	1.000
EL-LASSO	100	200	0.076	0.356	0.800	0.968	0.996	1.000
		400	0.066	0.582	0.974	0.998	1.000	1.000
	200	200	0.066	0.388	0.862	0.994	0.998	1.000
		400	0.052	0.596	0.974	1.000	1.000	1.000
	500	200	0.074	0.430	0.878	0.990	1.000	1.000
		400	0.062	0.616	0.990	1.000	1.000	1.000
Wald	100	200	0.254	0.612	0.928	1.000	1.000	1.000
		400	0.214	0.828	0.998	1.000	1.000	1.000
	200	200	0.226	0.652	0.978	1.000	1.000	1.000
		400	0.228	0.844	1.000	1.000	1.000	1.000
	500	200	0.204	0.666	0.962	1.000	1.000	1.000
		400	0.238	0.846	1.000	1.000	1.000	1.000
Score	100	200	0.250	0.574	0.918	0.998	1.000	1.000
		400	0.206	0.824	0.998	1.000	1.000	1.000
	200	200	0.202	0.614	0.972	1.000	1.000	1.000
		400	0.220	0.808	1.000	1.000	1.000	1.000
	500	200	0.200	0.640	0.962	1.000	1.000	1.000
		400	0.234	0.832	1.000	1.000	1.000	1.000

Table 27: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a Toeplitz matrix with $\rho = 0.5$, and the random error are generated by a $X_1 \sum_{j=2}^{p} X_{j-1}X_j\mathrm{N}(0,1)/(p-1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.082	0.324	0.792	0.964	0.992	0.996
		400	0.068	0.542	0.968	0.998	1.000	1.000
	200	200	0.094	0.376	0.818	0.980	0.996	1.000
		400	0.068	0.520	0.968	1.000	1.000	1.000
	500	200	0.054	0.326	0.846	0.990	1.000	1.000
		400	0.060	0.620	0.976	1.000	1.000	1.000
EL-INV	100	200	0.070	0.318	0.790	0.960	0.990	0.996
		400	0.062	0.532	0.966	0.996	1.000	1.000
	200	200	0.068	0.370	0.822	0.980	0.996	1.000
		400	0.060	0.514	0.966	1.000	1.000	1.000
	500	200	0.054	0.346	0.856	0.990	1.000	1.000
		400	0.056	0.630	0.978	1.000	1.000	1.000
EL-LASSO	100	200	0.070	0.340	0.810	0.972	0.994	0.996
		400	0.068	0.554	0.972	0.998	1.000	1.000
	200	200	0.084	0.402	0.844	0.982	0.996	1.000
		400	0.064	0.550	0.978	1.000	1.000	1.000
	500	200	0.056	0.376	0.884	0.994	1.000	1.000
		400	0.064	0.656	0.982	1.000	1.000	1.000
Wald	100	200	0.250	0.652	0.960	0.996	1.000	1.000
		400	0.236	0.852	0.996	1.000	1.000	1.000
	200	200	0.256	0.680	0.976	0.998	1.000	1.000
		400	0.204	0.874	0.998	1.000	1.000	1.000
	500	200	0.198	0.688	0.992	1.000	1.000	1.000
		400	0.274	0.902	1.000	1.000	1.000	1.000
Score	100	200	0.228	0.620	0.944	0.994	1.000	1.000
		400	0.192	0.816	0.996	1.000	1.000	1.000
	200	200	0.222	0.644	0.970	0.998	1.000	1.000
		400	0.190	0.840	0.998	1.000	1.000	1.000
	500	200	0.188	0.650	0.986	1.000	1.000	1.000
		400	0.236	0.876	1.000	1.000	1.000	1.000

Table 28: Empirical size and power of the proposed EL-based test procedures and two existing procedures under the homogenous case. In this table, covariates are generated by a multivariate normal distribution with covariance given by a block diagonal matrix with $\rho = 0.5$, and the random error are generated by a $X_1 \sum_{j=2}^{p} X_{j-1}X_j \mathrm{N}(0,1)/(p-1)$ distribution.

			β_1^0					
Method	p	n	0	0.1	0.2	0.3	0.4	0.5
EL-KFC	100	200	0.064	0.592	0.952	0.992	1.000	1.000
		400	0.068	0.842	1.000	1.000	1.000	1.000
	200	200	0.076	0.556	0.962	0.998	1.000	1.000
		400	0.066	0.856	1.000	1.000	1.000	1.000
	500	200	0.070	0.602	0.992	1.000	1.000	1.000
		400	0.060	0.874	1.000	1.000	1.000	1.000
EL-INV	100	200	0.050	0.574	0.954	0.992	1.000	1.000
		400	0.060	0.844	1.000	1.000	1.000	1.000
	200	200	0.056	0.550	0.960	0.998	1.000	1.000
		400	0.060	0.862	1.000	1.000	1.000	1.000
	500	200	0.066	0.584	0.992	1.000	1.000	1.000
		400	0.062	0.878	1.000	1.000	1.000	1.000
EL-LASSO	100	200	0.060	0.600	0.958	0.994	1.000	1.000
		400	0.070	0.862	1.000	1.000	1.000	1.000
	200	200	0.068	0.568	0.966	0.998	1.000	1.000
		400	0.064	0.868	1.000	1.000	1.000	1.000
	500	200	0.070	0.612	0.996	1.000	1.000	1.000
		400	0.066	0.886	1.000	1.000	1.000	1.000
Wald	100	200	0.226	0.842	0.996	1.000	1.000	1.000
		400	0.208	0.970	1.000	1.000	1.000	1.000
	200	200	0.226	0.828	0.996	1.000	1.000	1.000
		400	0.232	0.966	1.000	1.000	1.000	1.000
	500	200	0.240	0.850	1.000	1.000	1.000	1.000
		400	0.220	0.982	1.000	1.000	1.000	1.000
Score	100	200	0.212	0.808	0.994	1.000	1.000	1.000
		400	0.182	0.962	1.000	1.000	1.000	1.000
	200	200	0.204	0.802	0.996	1.000	1.000	1.000
		400	0.208	0.954	1.000	1.000	1.000	1.000
	500	200	0.214	0.832	1.000	1.000	1.000	1.000
		400	0.204	0.972	1.000	1.000	1.000	1.000

References

Alexandre Belloni, Victor Chernozhukov, Lie Wang, et al. Pivotal estimation via square-root lasso in nonparametric regression. *The Annals of Statistics*, 42(2):757–788, 2014.

Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36(1):199–227, 2008.

Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5): 849–911, 2008.

Harold V Henderson and Shayle R Searle. On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60, 1981.

Weidong Liu and Shan Luo. Hypothesis testing for high-dimensional regression models. *manuscript*, 2014.

Art B Owen. Empirical likelihood ratio confidence regions. *The Annals of Statistics*, 18(1): 90–120, 1990.

Art B Owen. *Empirical likelihood*. CRC press, 2001.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. *arXiv preprint arXiv:1011.3027*, 2010.

Hansheng Wang. Forward regression for ultra-high dimensional variable screening. *Journal of the American Statistical Association*, 104(488):1512–1524, 2009.

Hansheng Wang. Factor profiled sure independence screening. *Biometrika*, 99(1):15–28, 2012.

Cun-Hui Zhang and Stephanie S Zhang. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1):217–242, 2014.