Likelihood Ratio Test in Multivariate Linear Regression: from Low to High Dimension

Yinqiu He[1], Tiefeng Jiang[2], Jiyang Wen[3], Gongjun Xu[1]

[1]*Department of Statistics, University of Michigan*
[2]*School of Statistics, University of Minnesota*
[3]*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health*

Supplementary Material

We give proofs of main results and additional simulations in the Supplementary Material. Specifically, in Section S1–S5, we prove Theorems 1–5, respectively. We present the proof of Proposition 6 in Section S6 and provide additional simulations in Section S7.

S1. Theorem 1

Theorem 1 has two parts of conclusions, with $mr \to \infty$ and mr is finite respectively. We next prove the two parts in Sections S1.1 and S1.2 respectively. A lemma used in Section S1.1 is given and proved in Section S1.3.

S1.1 Proof of the part for $mr \to \infty$ in Theorem 1

In this section, we consider $mr \to \infty$ and $\max\{p, m, r\}/n \to 0$. We prove the conclusion for $mr \to \infty$ in Theorem 1 based on the result of Theorem 3, which is proved in Section S3.

When (p, m, r) are all fixed, we know that $-2\log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. Note that $\mathrm{E}(\chi^2_{mr}) = mr$, $\mathrm{var}(\chi^2_{mr}) = 2mr$, and when $mr \to \infty$, $(\chi^2_{mr} - mr)/\sqrt{2mr} \xrightarrow{D} \mathcal{N}(0, 1)$. It follows that $P(\chi^2_{mr} > \sqrt{2mr}z_\alpha + mr) \to \alpha$ and

$$\chi^2_{mr}(\alpha) = \sqrt{2mr} \times \{z_\alpha + o(1)\} + mr, \tag{S1.1}$$

where z_α denotes the upper α-quantile of $\mathcal{N}(0, 1)$.

We define the asymptotic regime $\mathcal{R}_A = \{(p, m, r, n) : n > p + m,\ p \geq r,\ mr \to \infty,\ \text{and}\ \max\{p, m, r\}/n \to 0\ \text{as}\ n \to \infty\}$. Under the asymptotic regime $\mathcal{R}_A$, Theorem 3 shows that $(-2\log L_n + \mu_n)/(n\sigma_n) \xrightarrow{D} \mathcal{N}(0, 1)$. Note that

$$P\{-2\log L_n > \chi^2_{mr}(\alpha)\} = P\Big\{\frac{-2\log L_n + \mu_n}{n\sigma_n} > \frac{\chi^2_{mr}(\alpha) + \mu_n}{n\sigma_n}\Big\}.$$

Thus when $n \to \infty$, $P\{-2\log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$ is equivalent to

$$\frac{\chi^2_{mr}(\alpha) + \mu_n}{n\sigma_n} \to z_\alpha, \quad \text{as } n \to \infty. \tag{S1.2}$$

When $mr \to \infty$, by (S1.1), we know (S1.2) is equivalent to

$$\frac{\sqrt{2mr} \times \{z_\alpha + o(1)\} + mr + \mu_n}{n\sigma_n} \to z_\alpha, \quad \text{as } n \to \infty. \tag{S1.3}$$

(S1.3) holds for any significance level α if and only if $n\sigma_n = \sqrt{2mr}\{1 + o(1)\}$ and $(\mu_n + mr)/\sqrt{2mr} = o(1)$.

Next we will prove that under $\mathcal{R}_A$, $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$ in the first step, derive the form of μ_n in the second step, and obtain the conclusion in the third step.

Step 1. Note that

$$\sigma_n^2 = 2\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}.$$

By the Taylor expansion, $\log(1-a) = -a - a^2/2 - a^3/3 + O(a^4)$ for $a = o(1)$. Under $\mathcal{R}_A$, we know that $p/n, m/n, r/n \to 0$ and $r/(n-p-m) \to 0$. Then we have

$$\begin{aligned}
&\log\frac{n+r-p-m}{n-p-m} \\
=\ &\log\left(1 + \frac{r}{n-p-m}\right) \\
=\ &\frac{r}{n-p-m} - \frac{1}{2}\frac{r^2}{(n-p-m)^2} + \frac{1}{3}\frac{r^3}{(n-p-m)^3} + O\left(\frac{r^4}{n^4}\right),
\end{aligned} \tag{S1.4}$$

and similarly,

$$
\begin{aligned}
-\log\frac{n-p+r}{n-p} &= -\log\Big(1+\frac{r}{n-p}\Big)\\
&= -\frac{r}{n-p}+\frac{1}{2}\frac{r^2}{(n-p)^2}-\frac{1}{3}\frac{r^3}{(n-p)^3}+O\Big(\frac{r^4}{n^4}\Big). \qquad \text{(S1.5)}
\end{aligned}
$$

Since for any numbers a and b, $a^2-b^2=(a-b)(a+b)$ and $a^3-b^3=(a-b)(a^2+b^2+ab)$, we then know

$$
\begin{aligned}
&\text{(S1.4)}+\text{(S1.5)}\\
&=\frac{r}{n-p-m}-\frac{r}{n-p}-\frac{1}{2}\Big\{\frac{r^2}{(n-p-m)^2}-\frac{r^2}{(n-p)^2}\Big\}\\
&\quad+\frac{1}{3}\Big\{\frac{r^3}{(n-p-m)^3}-\frac{r^3}{(n-p)^3}\Big\}+O\Big(\frac{r^4}{n^4}\Big)\\
&=\frac{rm}{(n-p-m)(n-p)}-\frac{1}{2}r^2\times\frac{m(2n-2p-m)}{(n-p-m)^2(n-p)^2}+O\Big\{\frac{r^3(m+r)}{n^4}\Big\}. \qquad \text{(S1.6)}
\end{aligned}
$$

We next examine the first two terms in (S1.6). Note that for $a=o(1)$ and $b=o(1)$, $1/(1-a)=1+a+O(a^2)$ and $1/\{(1-a)(1-b)\}=1+a+b+O(a^2+b^2)$. Then for the first term in (S1.6), we have

$$
\begin{aligned}
\frac{rm}{(n-p-m)(n-p)} &= \frac{rm}{n^2\{1-(p+m)/n\}(1-p/n)} \qquad \text{(S1.7)}\\
&= \frac{rm}{n^2}\Big\{1+\frac{2p+m}{n}+O\Big(\frac{p^2+m^2}{n^2}\Big)\Big\}.
\end{aligned}
$$

In addition, note that for $a = o(1)$ and $b = o(1)$, $1/\{(1-a)^2(1-b)^2\} = 1 + 2a + 2b + O(a^2 + b^2)$. Then for the second term in (S1.6), we have

$$\begin{aligned} & -\frac{1}{2}r^2 \times \frac{m(2n-2p-m)}{(n-p-m)^2(n-p)^2} \qquad \text{(S1.8)} \\ = \; & -nmr^2\Big\{1 - \frac{p+m/2}{n}\Big\}\frac{1}{n^4}\Big\{1 + \frac{2(p+m)}{n} + \frac{2p}{n} + O\Big(\frac{p^2+m^2}{n^2}\Big)\Big\} \\ = \; & -\frac{mr^2}{n^3}\Big\{1 + \frac{3p+3m/2}{n} + O\Big(\frac{p^2+m^2}{n^2}\Big)\Big\}. \end{aligned}$$

Combining (S1.7) and (S1.8), we obtain

$$\begin{aligned} \text{(S1.6)} &= \text{(S1.7)} + \text{(S1.8)} + O\Big\{\frac{r^3(m+r)}{n^4}\Big\} \\ &= \frac{rm}{n^2} + \frac{rm}{n^2}\Big\{\frac{2p+m-r}{n}\Big\} + O\Big\{\frac{mr(m^2+r^2+p^2)}{n^4}\Big\}. \qquad \text{(S1.9)} \end{aligned}$$

We then know that $\sigma_n^2 = 2 \times \text{(S1.6)} = (2mr/n^2) \times \{1 + o(1)\}$, and thus $n\sigma_n = \sqrt{2mr}\{1 + o(1)\}$.

Step 2. In this step, we derive the form of μ_n. Under the asymptotic region $\mathcal{R}_A$, we know that by Lemma 1 and Taylor expansion,

$$\begin{aligned}\mu_n &= -mr\Big\{1+\frac{p-r}{n}+O\Big(\frac{p^2+r^2}{n^2}\Big)\Big\}\\ &\quad -\frac{1}{2}\times\frac{mr(m+r)}{n}\Big\{1+O\Big(\frac{p+r}{n}\Big)\Big\}+o(1)mr\times\frac{p+m+r}{n}\\ &= -mr-mr\frac{p+m/2-r/2}{n}+o(1)mr\times\frac{p+m+r}{n}.\end{aligned}$$

Step 3. As discussed, under $\mathcal{R}_A$, (S1.3) holds for any level α, if and only if $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$ and $(\mu_n+mr)/\sqrt{2mr}=o(1)$. In the first step, we have shown that $n\sigma_n = \sqrt{2mr}\{1+o(1)\}$ under $\mathcal{R}_A$. In the second step, we obtain the form of μ_n. Thus we have

$$\frac{\mu_n+mr}{\sqrt{2mr}} = -\frac{\sqrt{mr}}{\sqrt{2}}\Big(\frac{p+m/2-r/2}{n}\Big)\times\{1+o(1)\},$$

which converges to 0, if and only if $\lim_{n\to\infty}\sqrt{mr}(p+m/2-r/2)n^{-1}=0$.

S1.2 Proof of the part for finite mr in Theorem 1

By Muirhead (2009), the characteristic function of $-2\log L_n$ is $\phi_1(t) = \mathrm{E}\{\exp(-2it\log L_n)\}$ and

$$\log\phi_1(t) = -\frac{mr}{2}\log(1-2it) + \sum_{l=1}^{\infty}\varsigma_l\{(1-2it)^{-l}-1\}, \tag{S1.10}$$

where

$$\varsigma_l = \frac{(-1)^{l+1}}{l(l+1)}\Big[\sum_{k=1}^{m}\Big\{\frac{\mathbb{B}_{l+1}\{(1-k-p)/2\}}{(n/2)^l} - \frac{\mathbb{B}_{l+1}\{(1-k+r-p)/2\}}{(n/2)^l}\Big\}\Big],$$

and $\mathbb{B}_{l+1}(\cdot)$ is the Bernoulli polynomials which takes the form $\mathbb{B}_{l+1}(z) = \sum_{v=0}^{l+1} c_v z^v$. We next estimate the order of ς_l with respect to n. We note that for any z_1 and z_2,

$$\begin{aligned}
&\mathbb{B}_{l+1}(z_1) - \mathbb{B}_{l+1}(z_2) \qquad\qquad (\text{S1.11})\\
=&\sum_{v=0}^{l+1} c_v(z_1^v - z_2^v) = \sum_{v=1}^{l+1} c_v \sum_{w=1}^{v}\binom{v}{w}(z_1-z_2)^w z_2^{v-w}\\
=&(z_1-z_2)\sum_{v=1}^{l+1}\sum_{w=1}^{v} c_v\binom{v}{w}(z_1-z_2)^{w-1}z_2^{v-w}.
\end{aligned}$$

Let $z_1 = (1-k-p)/2$ and $z_2 = (1-k+r-p)/2$. Then we have $z_1 - z_2 = (-r)/2$. When m and r are finite, the order of ς_l with respect to n is $O\{(p/n)^l\}$. When $p/n \to 0$, by the expansion (S1.10), we have $\phi_1(t) = (1-2it)^{-mr/2}\{1+o(1)\}$. Then

$-2\log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. When p/n is bounded from 0 below, (S1.10) does not converge to $-2^{-1}mr\log(1-2it)$ generally for all t. Then the approximation $-2\log L_n \xrightarrow{D} \chi^2_{mr}$ fails.

S1.3 Lemma used in Section S1.1

Lemma 1. *Under the asymptotic regime $\mathcal{R}_A$,*

$$\begin{aligned}\mu_n &= -\frac{nmr}{n+r-p} - \frac{1}{2}\frac{nmr(m+r)}{(n+r-p)^2} - \frac{nmr(m^2/3+mr/2+r^2/3)}{(n+r-p)^3} \\ &\quad +O(1)\frac{mr(m^3+r^3)}{n^3} + O(mr/n).\end{aligned}$$

Proof. By the definition of μ_n in Theorem 3,

$$\begin{aligned}\mu_n &= n(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + nr\log\frac{(n+r-p-m)}{(n+r-p)} \\ &\quad +nm\log\frac{(n-p)}{(n+r-p)}.\end{aligned}$$

Note that

$$\begin{aligned}&\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} \\ =\ &\log\frac{n+r-p-m}{n+r-p} + \log\frac{n-p}{n+r-p} - \log\frac{n-p-m}{n+r-p} \\ =\ &\log\left(1-\frac{m}{n+r-p}\right) + \log\left(1-\frac{r}{n+r-p}\right) - \log\left(1-\frac{m+r}{n+r-p}\right).\end{aligned}$$

It follows that

$$\begin{aligned}
\mu_n &= n(n-m-p+r-1/2)\log\Big(1-\frac{m}{n+r-p}\Big) && \text{(S1.12)}\\
&\quad +n(n-p-1/2)\log\Big(1-\frac{r}{n+r-p}\Big) && \text{(S1.13)}\\
&\quad -n(n-m-p-1/2)\log\Big(1-\frac{m+r}{n+r-p}\Big), && \text{(S1.14)}
\end{aligned}$$

which gives $\mu_n = \text{(S1.12)} + \text{(S1.13)} + \text{(S1.14)}$. We next analyze (S1.12)$/n$, (S1.13)$/n$ and (S1.14)$/n$ respectively.

By the Taylor expansion, we have

$$\begin{aligned}
\log\Big(1-\frac{m}{n+r-p}\Big) &= -\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m}{n+r-p}\Big)^k,\\
\log\Big(1-\frac{r}{n+r-p}\Big) &= -\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{r}{n+r-p}\Big)^k,\\
\log\Big(1-\frac{m+r}{n+r-p}\Big) &= -\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m+r}{n+r-p}\Big)^k.
\end{aligned}$$

Then

$$\begin{aligned}
\text{(S1.12)}/n &= -(n+r-p)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m}{n+r-p}\Big)^k + (m+1/2)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m}{n+r-p}\Big)^k\\
&= -\sum_{k=1}^{\infty}\frac{1}{k}\frac{m^k}{(n+r-p)^{k-1}} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{m^{k+1}}{(n+r-p)^k} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{m^k/2}{(n+r-p)^k},
\end{aligned}$$

$$
\begin{aligned}
(\text{S1.13})/n &= -(n+r-p)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{r}{n+r-p}\Big)^k + (r+1/2)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{r}{n+r-p}\Big)^k \\
&= -\sum_{k=1}^{\infty}\frac{1}{k}\frac{r^k}{(n+r-p)^{k-1}} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{r^{k+1}}{(n+r-p)^k} + \sum_{k=1}^{\infty}\frac{1}{k}\frac{r^k/2}{(n+r-p)^k},
\end{aligned}
$$

and

$$
\begin{aligned}
(\text{S1.14})/n &= (n+r-p)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m+r}{n+r-p}\Big)^k - (m+r+1/2)\sum_{k=1}^{\infty}\frac{1}{k}\Big(\frac{m+r}{n+r-p}\Big)^k \\
&= \sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k}{(n+r-p)^{k-1}} - \sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^{k+1}}{(n+r-p)^k} - \sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k/2}{(n+r-p)^k}.
\end{aligned}
$$

It follows that $\{(\text{S1.12}) + (\text{S1.13}) + (\text{S1.14})\}/n = (\text{S1.15}) + (\text{S1.16})$, where

$$
\sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k - m^k - r^k}{(n+r-p)^{k-1}} - \sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k}, \tag{S1.15}
$$

$$
-\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k}\frac{(m+r)^k - m^k - r^k}{(n+r-p)^k}. \tag{S1.16}
$$

As $(m+r)^1 - m^1 - r^1 = 0$, we know

$$
\begin{aligned}
\text{(S1.15)} &= \sum_{k=2}^{\infty} \frac{1}{k} \frac{(m+r)^k - m^k - r^k}{(n+r-p)^{k-1}} - \sum_{k=1}^{\infty} \frac{1}{k} \frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k} \\
&= \sum_{k=1}^{\infty} \Big(\frac{1}{k+1} - \frac{1}{k} \Big) \frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k} \\
&= -\frac{1}{2} \frac{2mr}{n+r-p} && \text{(S1.17)} \\
&\quad -\frac{1}{6} \frac{(m+r)^3 - m^3 - r^3}{(n+r-p)^2} && \text{(S1.18)} \\
&\quad -\frac{1}{12} \frac{(m+r)^4 - m^4 - r^4}{(n+r-p)^3} && \text{(S1.19)} \\
&\quad -\sum_{k=4}^{\infty} \frac{1}{k(k+1)} \frac{(m+r)^{k+1} - m^{k+1} - r^{k+1}}{(n+r-p)^k}, && \text{(S1.20)}
\end{aligned}
$$

which gives (S1.15) = (S1.17) + (S1.18) + (S1.19) + (S1.20). We have $n \times$ (S1.17) $= -nmr(n+r-p)^{-1}$, $n \times$ (S1.18) $= -2^{-1}nmr(m+r)(n+r-p)^{-2}$, and $n \times$ (S1.19) $= -nmr(m^2/3 + mr/2 + r^2/3) \times (n+r-p)^{-3}$. In addition,

$$
\begin{aligned}
|\text{(S1.20)}| &= \sum_{k=4}^{\infty} \frac{1}{k(k+1)} \frac{\sum_{q=1}^{k} \binom{k+1}{q} m^q r^{k+1-q}}{(n+r-p)^k} \\
&\leq \frac{mr}{n+r-p} \sum_{k=4}^{\infty} \frac{1}{k} \times \frac{2^{k+1} (\max\{m,r\})^{k-1}}{(n+r-p)^{k-1}} \\
&= \frac{mr}{n+r-p} O\Big\{ \Big(\frac{\max\{m,r\}}{n+r-p} \Big)^3 \Big\} = O(1) \frac{mr(m^3+r^3)}{n^4},
\end{aligned}
$$

where in the last two equations, we use the property of Taylor expansion and the condition that $\max\{p,m,r\} = o(n)$. Therefore, $n \times$ (S1.20) $= mr \times O\{(m^3+r^3)/n^3\}$.

Moreover,

$$
\begin{aligned}
\text{(S1.16)} &= \sum_{k=1}^{\infty} \frac{1}{k} \frac{(m+r)^k - m^k - r^k}{(n+r-p)^k} \\
&= \sum_{k=2}^{\infty} \frac{1}{k} \frac{\sum_{q=1}^{k-1} \binom{k}{q} m^q r^{k-q}}{(n+r-p)^k} \\
&\leq \frac{mr}{(n+r-p)^2} \sum_{k=2}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m,r\})^{k-2}}{(n+r-p)^{k-2}} \\
&= O(mr/n^2),
\end{aligned}
$$

where in the last equation we use the fact that

$$
\begin{aligned}
\sum_{k=2}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m,r\})^{k-2}}{(n+r-p)^{k-2}} &\leq 2 + \sum_{k=3}^{\infty} \frac{1}{k-2} \times \frac{2^k (\max\{m,r\})^{k-2}}{(n+r-p)^{k-2}} \\
&= 2 + 4 \sum_{k=1}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m,r\})^{k}}{(n+r-p)^{k}} \\
&= 2 + 4 \log[1 - \{2 \max\{m,r\}/(n+r-p)\}].
\end{aligned}
$$

In summary,

$$
\begin{aligned}
\mu_n &= \text{(S1.12)} + \text{(S1.13)} + \text{(S1.14)} \\
&= n \times \{\text{(S1.15)} + \text{(S1.16)}\} \\
&= -\frac{nmr}{n+r-p} - \frac{1}{2} \frac{nmr(m+r)}{(n+r-p)^2} - \frac{nmr(m^2/3 + mr/2 + r^2/3)}{(n+r-p)^3} \\
&\quad + O(1) \frac{mr(m^3+r^3)}{n^3} + O(mr/n).
\end{aligned}
$$

□

S2. Theorem 2

Similarly to Section S1, we prove Theorem 2 when $mr \to \infty$ and mr is finite in Sections S2.1 and S2.2 respectively.

S2.1 Proof of the part for $mr \to \infty$ in Theorem 2

When (p, m, r) are all fixed, by Bartlett correction, we know that with $\rho = 1 - (p - r/2 + m/2 + 1/2)/n$, $-2\rho \log L_n \xrightarrow{D} \chi^2_{mr}$ as $n \to \infty$. Note that under $\mathcal{R}_A = \{(p, m, r, n) : n > p + m,\ p \geq r,\ mr \to \infty,\ \text{and}\ \max\{p, m, r\}/n \to 0\ \text{ as }\ n \to \infty\}$, $\rho = 1 + o(1)$. Then similarly to the proof of Theorem 1 in Section S1.1, we know that under $\mathcal{R}_A$, $P\{-2\rho \log L_n > \chi^2_{mr}(\alpha)\} \to \alpha$ holds for any given significance level α if and only if $n\sigma_n = \sqrt{2mr}\{1 + o(1)\}$ and $(\mu_n + mr/\rho)/\sqrt{2mr} = o(1)$.

Following the same argument as in Section S1.1, we know that under $\mathcal{R}_A$, $n\sigma_n =$

$\sqrt{2mr}\{1+o(1)\}$. In addition, by the Taylor expansion,

$$\begin{aligned}
\frac{mr}{\rho} &= \frac{mr}{1-(p+m/2-r/2+1/2)/n} \\
&= \frac{nmr}{n-p+r-(m+r)/2} + mrO\Big(\frac{1}{n}\Big) \\
&= \frac{nmr}{n-p+r}\sum_{k=0}^{\infty}\Big\{\frac{m+r}{2(n-p+r)}\Big\}^k + mrO\Big(\frac{1}{n}\Big) \\
&= \frac{nmr}{n-p+r} + \frac{nmr(m+r)}{2(n+r-p)^2} + \frac{nmr(m+r)^2}{4(n-p+r)^3} \\
&\quad + mr\times O\Big(\frac{m^3+r^3}{n^3}\Big) + mrO\Big(\frac{1}{n}\Big),
\end{aligned}$$

where in the last equation, we use the fact that $\sum_{k=3}^{\infty}[(m+r)/\{2(n-p+r)\}]^k = O\{(m^3+r^3)/n^3\}$ as $\max\{p,m,r\}=o(n)$. It follows that under $\mathcal{R}_A$, by Lemma 1,

$$\begin{aligned}
&\frac{\mu_n-(-mr/\rho)}{\sqrt{2mr}} \qquad\qquad (\text{S2.21}) \\
=\ & \frac{1}{\sqrt{2mr}}\Big\{-\frac{nmr}{n+r-p} - \frac{nmr(m+r)}{2(n+r-p)^2} - \frac{nmr(m^2/3+mr/2+r^2/3)}{(n+r-p)^3} \\
&\qquad + \frac{nmr}{n-p+r} + \frac{nmr(m+r)}{2(n+r-p)^2} + \frac{nmr(m+r)^2}{4(n-p+r)^3}\Big\} \\
&+\sqrt{mr}\times O\Big(\frac{m^3+r^3}{n^3}\Big) + O(\sqrt{mr}/n) \\
=\ & -\frac{\sqrt{mr}}{\sqrt{2}}\frac{1}{12}\frac{n(m^2+r^2)}{(n+r-p)^3} + \sqrt{mr}\times O\Big(\frac{m^3+r^3}{n^3}\Big) + O(\sqrt{mr}/n) \\
=\ & -\frac{\sqrt{mr}(m^2+r^2)}{12\sqrt{2}n^2} + o(1)\times\frac{\sqrt{mr}(m^2+r^2)}{n^2} + O(\sqrt{mr}/n),
\end{aligned}$$

where in the last equation, we use the fact that $\max\{p,m,r\}=o(n)$. We thus know

that (S2.21) $= 0$ if and only if $\sqrt{mr}(m^2 + r^2)/n^2 \to 0$.

S2.2 Proof of the part for finite mr in Theorem 2

By Muirhead (2009), for the LRT with Bartlett correction, the characteristic function of $-2\rho \log L_n$ is $\phi_2(t) = \mathrm{E}\{\exp(-2it\rho \log L_n)\}$. Moreover, we have

$$\log \phi_2(t) = -\frac{mr}{2} \log(1 - 2it) + \sum_{l=1}^{\infty} \tilde{\varsigma}_l \{(1 - 2it)^{-l} - 1\},$$

where

$$\tilde{\varsigma}_l = \frac{(-1)^{l+1}}{l(l+1)} \Big[\sum_{k=1}^{m} \Big\{ \frac{\mathbb{B}_{l+1}(\tilde{z}_{k,1})}{(\rho n/2)^l} - \frac{\mathbb{B}_{l+1}(\tilde{z}_{k,2})}{(\rho n/2)^l} \Big\} \Big],$$

$\tilde{z}_{k,1} = (1 - \rho)n/2 + (1 - k - p)/2$ and $\tilde{z}_{k,2} = (1 - \rho)n/2 + (1 - k + r - p)/2$. Since $\rho = 1 - (p - r/2 + m/2 + 1/2)/n$,

$$\begin{aligned} \tilde{z}_{k,1} =& (p - r/2 + m/2 + 1/2)/2 + (1 - k - p)/2 = (3 - r + m)/4, \\ \tilde{z}_{k,2} =& (p - r/2 + m/2 + 1/2)/2 + (1 - k + r - p)/2 = (3 + r + m)/4. \end{aligned}$$

In addition, $\rho n = n - (p - r/2 + m/2 + 1/2)$. Therefore, by the expansion in (S1.11), when m and r are fixed and $n - p \to \infty$, we have $\log \phi_2(t) = -2^{-1} mr \log(1 - 2it) + O\{(n - p)^{-1}\}$ and $\phi_2(t) = (1 - 2it)^{-mr/2}[1 + O\{(n - p)^{-1}\}]$. It follows that when

m and r are fixed and $n-p \to \infty$, $-2\rho \log L_n \xrightarrow{D} \chi^2_{mr}$. On the other hand, when $n-p$ is fixed, by the expansion in (S1.11), we know $\tilde{\varsigma}_l$ is of constant order in n, and thus $\sum_{l=1}^{\infty} \tilde{\varsigma}_l \{(1-2it)^{-l} - 1\}$ is not ignorable generally for all t. We then know the approximation $-2\rho \log L_n \xrightarrow{D} \chi^2_{mr}$ fails.

S3. Theorem 3

In this section, we give the proof of Theorem 3, where the main proof is in Section S3.1 and some lemmas used are provided and proved in Section S3.2.

S3.1 Proof of Theorem 3

Proof. To prove the central limit theorem that $H_n := \{-2\log L_n + \mu_n\}/(n\sigma_n) \xrightarrow{D} \mathcal{N}(0,1)$, it is sufficient to show

$$E \exp\left\{ \frac{\log L_n - \mu_n/2}{n\sigma_n/2} s \right\} \to \exp\{s^2/2\}, \tag{S3.22}$$

as $n \to \infty$ and $|s| < 1$, where σ_n^2 and μ_n are defined in Theorem 3. Equivalently, it suffices to show that for any subsequence $\{n_k\}$, there is a further subsequence $\{n_{k_j}\}$ such that $H_{n_{k_j}}$ converges to $\mathcal{N}(0,1)$ in distribution as $j \to \infty$. In the following, the further subsequence is selected in a way such that the subsequential limits of some bounded quantities (to be specified in the proof below) exist, which is guaranteed

by Bolzano-Weierstrass Theorem. Therefore, we only need to verify the theorems by assuming that the limits for these bounded quantities exist. In the following, we give the proof by discussing two settings $r \geq m$ and $m \geq r$ separately.

Case 1. When $r \geq m$ and $r \to \infty$. By Lemma 3, under the null hypothesis, the distribution of L_n can be reexpressed as the distribution of a product of independent beta random variables. Let $h = 2s/(n\sigma_n)$, by Lemma 2, then under the null hypothesis, L_n's hth moment can be written as

$$E \exp\left\{\frac{\log L_n}{n\sigma_n/2} s\right\} = E(L_n^h) = \frac{\Gamma_m\{\frac{1}{2}n(1+h) - \frac{1}{2}p\}\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n-p)\}\Gamma_m\{\frac{1}{2}n(1+h) + \frac{1}{2}(r-p)\}}, \quad \text{(S3.23)}$$

where $\Gamma_m(a)$, $a \in \mathbb{C}$ and $\mathrm{Re}(a) > (m-1)/2$, is the multivariate Gamma function defined to be

$$\Gamma_m(a) = \int_{A>0} e^{-\mathrm{tr}(A)} \det A^{a-(m+1)/2}(dA). \quad \text{(S3.24)}$$

The above integration is taken over the space of positive definite $m \times m$ matrices, i.e., $\{A_{m\times m} : A \succ 0\}$; and $\mathrm{tr}(A)$ is the trace of A. Note that when $m = 1$, $\Gamma_m(a)$ becomes the usual definition of Gamma function. By Lemma 4, $\Gamma_m(a)$ can be written as a

product of ordinary Gamma functions as

$$\Gamma_m(a) = \pi^{m(m-1)/4} \prod_{j=1}^{m} \Gamma\{a - (j-1)/2\}.$$

Note that $n > m + p$ and $r \geq 1$. Thus the limits of $m/(n+r-p)$ and $m/(n-p)$ are in $[0, 1]$ for all n. Applying the subsequence argument above, for any subsequence $\{n_k\}$, we take a further subsequence n_{k_j} such that $m_{k_j}/(n_{k_j}+r_{k_j}-p_{k_j})$ and $m_{k_j}/(n_{k_j}-p_{k_j})$ converge to some constants in $[0, 1]$. Thus without loss of generality, we consider the cases when $m/(n+r-p)$ and $m/(n-p)$ converge to some constants in $[0, 1]$. Next we give the proof by discussing different cases below.

Case 1.1 If $m/(n+r-p) \to \gamma > 0$, this implies that $m \to \infty$ as $n \to \infty$. And as $r \geq m$ and $n > p + m$, we know $m/(n+r-p) \leq 1/2$, then $\gamma \in (0, 1/2]$. Since $1 \geq m/(n-p) \geq m/(n+r-p)$, then $m/(n-p) \to \gamma' \in (0, 1]$.

If $\gamma' \in (0, 1)$, $nh \times [-\log\{1-m/(n-p)\}]^{1/2} = O(1)$, which satisfies the assumption of Lemma 5.4 in Jiang and Yang (2013). If $\gamma' = 1$, as

$$\sigma_n^2 = 2\log\left(1 - \frac{m}{n+r-p}\right) - 2\log\left(1 - \frac{m}{n-p}\right), \tag{S3.25}$$

and $m/(n+r-p) \to \gamma \in (0, 1/2]$, we know σ_n^2 has leading order $\log\{1 - m/(n-p)\}$. Then as $nh\sigma_n = O(1)$ by definition, we also know $nh \times [-\log\{1 - m/(n-p)\}]^{1/2} =$

$O(1)$, which satisfies the assumption of Lemma 5.4 in Jiang and Yang (2013). Following the lemma, we have

$$\begin{aligned}\log\frac{\Gamma_m\{\frac{1}{2}n(1+h)-\frac{1}{2}p\}}{\Gamma_m\{\frac{1}{2}(n-p)\}} &= \log\frac{\Gamma_m\{\frac{1}{2}(n-p)+\frac{1}{2}nh\}}{\Gamma_m\{\frac{1}{2}(n-p)\}}\\ &= -\left\{\frac{n^2h^2}{4}+\frac{nh}{2}\left(n-m-p-\frac{1}{2}\right)\right\}\log\left(1-\frac{m}{n-p}\right)\\ &\quad+\frac{mnh}{2}\{\log(n-p)-\log 2e\}+o(1), \end{aligned} \tag{S3.26}$$

and similarly, we can obtain

$$\begin{aligned}\log\frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}n(1+h)+\frac{1}{2}(r-p)\}} &= \log\frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n+r-p)+\frac{1}{2}nh\}}\\ &= \left\{\frac{n^2h^2}{4}+\frac{nh}{2}\left(n+r-m-p-\frac{1}{2}\right)\right\}\log\left(1-\frac{m}{n+r-p}\right)\\ &\quad-\frac{mnh}{2}\{\log(n+r-p)-\log 2e\}+o(1). \end{aligned} \tag{S3.27}$$

Combining (S3.23), (S3.26) and (S3.27), we have

$$\begin{aligned}\log E\exp\left\{\frac{\log L_n}{n\sigma_n/2}s\right\} &= \frac{n^2h^2}{4}\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}+\frac{h\mu_n}{2}+o(1)\\ &= \frac{s^2}{2}+\frac{h\mu_n}{2}+o(1),\end{aligned}$$

where

$$\mu_n = n(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + nr\log\frac{(n+r-p-m)}{(n+r-p)} + nm\log\frac{(n-p)}{(n+r-p)}.$$

Therefore, $\log E\exp\left\{\frac{\log L_n-\mu_n/2}{n\sigma_n/2}s\right\} = s^2/2+o(1)$ is proved.

Case 1.2 We discuss the case when $m/(n+r-p)\to 0$ and $m/(n-p)\to 0$ below. By Lemma 7, we know that when $n-p\to\infty$ and $r\to\infty$,

$$\begin{aligned}
&\log\frac{\Gamma_m\{\frac{1}{2}n(1+h)-\frac{1}{2}p\}}{\Gamma_m\{\frac{1}{2}(n-p)\}}\\
=&-\left\{2m+\left(n-p-m-\frac{1}{2}\right)\log\left(1-\frac{m}{n-p}\right)\right\}\frac{nh}{2}\\
&-\left\{\frac{m}{n-p}+\log\left(1-\frac{m}{n-p}\right)\right\}\frac{n^2h^2}{4}\\
&+m\left\{\frac{(n-p+nh)}{2}\log\frac{(n-p+nh)}{2}-\frac{(n-p)}{2}\log\frac{(n-p)}{2}\right\}+o(1), \qquad \text{(S3.28)}
\end{aligned}$$

and

$$
\begin{aligned}
&\log \frac{\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}n(1+h)+\frac{1}{2}(r-p)\}} \\
&= \Big\{2m + \Big(n+r-p-m-\frac{1}{2}\Big)\log\Big(1-\frac{m}{n+r-p}\Big)\Big\}\frac{nh}{2} \\
&\quad - m\Big\{\frac{(n+r-p+nh)}{2}\log\frac{(n+r-p+nh)}{2} - \frac{(n+r-p)}{2}\log\frac{(n+r-p)}{2}\Big\} \\
&\quad + \Big\{\frac{m}{n+r-p} + \log\Big(1-\frac{m}{n+r-p}\Big)\Big\}\frac{n^2h^2}{4} + o(1). \qquad \text{(S3.29)}
\end{aligned}
$$

By Taylor expansion of the log function, we have

$$
\begin{aligned}
\sigma_n^2 &= 2\log\left(1-\frac{m}{n+r-p}\right) - 2\log\left(1-\frac{m}{n-p}\right) \\
&= \frac{2mr}{(n-p)(n+r-p)}\{1+o(1)\}, \qquad \text{(S3.30)}
\end{aligned}
$$

where the second order terms of Taylor expansion of the log functions is ignorable as $m = o(n-p)$. Also, as $r \to \infty$,

$$
h = \frac{s}{n\sigma_n/2} = \frac{s\sqrt{2(n-p)(n+r-p)}}{n\sqrt{mr}}\{1+o(1)\} \to 0. \qquad \text{(S3.31)}
$$

Therefore, combining (S3.23), (S3.28) and (S3.29), we obtain

$$\begin{aligned}
&\log E\exp\left\{\frac{\log L_n}{n\sigma_n/2}s\right\} \qquad \text{(S3.32)}\\
=\ &\frac{n^2h^2}{4}\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}+\frac{n^2h^2}{4}\left(\frac{m}{n+r-p}-\frac{m}{n-p}\right)\\
&+\frac{nh}{2}(n-m-p-1/2)\log\frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}\\
&+\frac{nh}{2}r\log\frac{(n+r-p-m)}{(n+r-p)}+\frac{nh}{2}m\log\frac{n-p+nh}{n+r-p+nh}\\
&+\frac{m(n+r-p)}{2}\log\frac{n+r-p}{n+r-p+nh}\\
&+\frac{m(n-p)}{2}\log\frac{n-p+nh}{n-p}+o(1).
\end{aligned}$$

We then analyze the terms in (S3.32) separately. By (S3.31),

$$\begin{aligned}
&\frac{n^2h^2}{4}\left(\frac{m}{n+r-p}-\frac{m}{n-p}\right)\\
=\ &-\frac{s^2(n-p)(n+r-p)}{2mr}\times\frac{mr}{(n-p)(n+r-p)}\{1+o(1)\}\\
=\ &-\frac{s^2}{2}+o(1). \qquad \text{(S3.33)}
\end{aligned}$$

In addition, as $nh/(n-p)\to 0$ and $nh/(n+r-p)\to 0$, we have

$$\begin{aligned}
&\frac{m(n+r-p)}{2}\log\frac{n+r-p}{n+r-p+nh} \qquad \text{(S3.34)}\\
=\ &-\frac{m(n+r-p)}{2}\left\{\frac{nh}{n+r-p}-\frac{n^2h^2}{2(n+r-p)^2}+R_{n,1}\right\},
\end{aligned}$$

and

$$\begin{aligned} & \frac{m(n-p)}{2}\log\frac{n-p+nh}{n-p} \\ = \; & \frac{m(n-p)}{2}\Big\{\frac{nh}{n-p}-\frac{n^2h^2}{2(n-p)^2}+R_{n,2}\Big\}, \end{aligned} \tag{S3.35}$$

where the remainder terms

$$R_{n,1}=\sum_{k=3}^{\infty}\frac{1}{k}(-1)^{k+1}\frac{(nh)^k}{(n+r-p)^k}, \quad R_{n,2}=\sum_{k=3}^{\infty}\frac{1}{k}(-1)^{k+1}\frac{(nh)^k}{(n-p)^k}. \tag{S3.36}$$

Then we have

$$\begin{aligned} & \text{(S3.34)}+\text{(S3.35)} \\ = \; & \frac{mn^2h^2}{4(n+r-p)}-\frac{mn^2h^2}{4(n-p)}-\frac{m(n+r-p)}{2}R_{n,1}+\frac{m(n-p)}{2}R_{n,2} \\ = \; & -\frac{s^2}{2}+o(1), \end{aligned} \tag{S3.37}$$

where in the last equation, we use (S3.33) and Lemma 8. Furthermore, by $nh/(n+$

$r-p) \to 0$ and (S3.31), we have

$$\begin{aligned}
&\frac{nh}{2} m \log \frac{n-p+nh}{n+r-p+nh} \\
&= \frac{nh}{2} m \log \left\{ \frac{n-p}{n+r-p} + \frac{nhr}{(n+r-p)(n+r-p+nh)} \right\} \\
&= \frac{nh}{2} m \log \frac{n-p}{n+r-p} + \frac{nmh}{2} \frac{n+r-p}{n-p} \frac{nhr}{(n+r-p)(n+r-p+nh)} + o(1) \\
&= \frac{nh}{2} m \log \frac{n-p}{n+r-p} + s^2 + o(1). \qquad \text{(S3.38)}
\end{aligned}$$

Combining (S3.32), (S3.33), (S3.37) and (S3.38), we obtain $\log E \exp\left\{ \frac{\log L_n - \mu_n/2}{n\sigma_n/2} s \right\} = s^2/2 + o(1)$.

Case 1.3 When $m/(n+r-p) \to 0$ and $m/(n-p) \to \gamma \in (0,1]$, we know (S3.26) still holds following similar analysis to Case 1.1. And (S3.29) also holds following similar analysis to Case 1.2. To establish (S3.22), we next show that under this case,

the difference between the result of (S3.27) and (S3.29) is ignorable.

$$
\begin{aligned}
&\text{(S3.29)} - \text{(S3.27)} \\
&= 2m\frac{nh}{2} + \frac{mnh}{2}\{\log(n+r-p) - \log 2e\} + \frac{n^2h^2}{4} \times \frac{m}{n+r-p} \\
&\quad - \frac{m(n+r-p)}{2}\log\frac{n+r-p+nh}{n+r-p} - \frac{mnh}{2}\log\frac{n+r-p+nh}{2} + o(1) \\
&= mnh + \frac{mnh}{2}\left\{\log\left(\frac{n+r-p}{2}\right) - 1\right\} + \frac{n^2h^2}{4} \times \frac{m}{n+r-p} \\
&\quad - \frac{m(n+r-p)}{2}\log\left(1 + \frac{nh}{n+r-p}\right) - \frac{mnh}{2}\log\left(\frac{n+r-p}{2} + \frac{nh}{2}\right) + o(1).
\end{aligned}
\tag{S3.39}
$$

We then analyze the terms in (S3.39) separately.

Since $m/(n-p) \to \gamma \in (0,1]$, similarly to (S3.25), we know that $nh = 2s/\sigma_n = O(s)$. As $m/(n+r-p) \to 0$, it follows that $n^2h^2m/(n+r-p) \to 0$. Applying Taylor expansion, we then have

$$
\begin{aligned}
& \frac{mnh}{2}\log\left(\frac{n+r-p}{2} + \frac{nh}{2}\right) \\
= \; & \frac{mnh}{2}\left\{\log\left(\frac{n+r-p}{2}\right) + O\Big(\frac{nh}{n+r-p}\Big)\right\} \\
= \; & \frac{mnh}{2}\log\left(\frac{n+r-p}{2}\right) + O\Big(\frac{mn^2h^2}{n+r-p}\Big) \\
= \; & \frac{mnh}{2}\log\left(\frac{n+r-p}{2}\right) + o(1).
\end{aligned}
\tag{S3.40}
$$

Similarly, by $nh = O(s)$, $m/(n+r-p) \to 0$, and Taylor expansion, we have

$$\begin{aligned}
&\frac{m(n+r-p)}{2}\log\left(1+\frac{nh}{n+r-p}\right)\\
=&\frac{m(n+r-p)}{2}\left\{\frac{nh}{n+r-p}+O\Big(\frac{n^2h^2}{(n+r-p)^2}\Big)\right\}\\
=&\frac{mnh}{2}+o(1).
\end{aligned} \tag{S3.41}$$

In summary, combining (S3.40) and (S3.41), we have (S3.39) = (S3.29) − (S3.27) = $o(1)$. Then by the results in Case 1.1, we get the same conclusion as in Case 1.1.

Case 2. When $m > r$, $m \to \infty$. According to Lemma 3, we can make the following substitution

$$m \to r, \quad r \to m, \quad n-p \to n+r-p-m.$$

Then the substituted mean and variance are

$$\begin{aligned}
\mu_n = n(n-p-m-1/2)\log\frac{(n-p)(n-p+r-m)}{(n-p-m)(n+r-p)} + nm\log\frac{(n-p)}{(n+r-p)}\\
+ nr\log\frac{(n-p+r-m)}{(n+r-p)},
\end{aligned}$$

and

$$\sigma_n^2 = 2\log\frac{(n-p)(n-p+r-m)}{(n-p-m)(n+r-p)},$$

which take the same forms as those in the setting when $r \geq m$. And the theorem can be proved following similar analysis when $m \to \infty$, $n-p+r-m \to \infty$. $\square$

S3.2 Lemmas in the proof of Theorem 3

Lemma 2 (Corollary 10.5.2 in Muirhead (2009)). *Under the null hypothesis, L_n's h-th moment can be written as*

$$\mathrm{E}(L_n^h) = \frac{\Gamma_m\{\frac{1}{2}n(1+h) - \frac{1}{2}p\}\Gamma_m\{\frac{1}{2}(n+r-p)\}}{\Gamma_m\{\frac{1}{2}(n-p)\}\Gamma_m\{\frac{1}{2}n(1+h) + \frac{1}{2}(r-p)\}}.$$

Lemma 3 (Theorem 10.5.3 in Muirhead (2009)). *Under the null hypothesis, when $n-p \geq m$ and $r \geq m$, $\frac{2}{n}\log L_n$ has the same distribution as $\sum_{i=1}^m \log V_i$, where V_i's are independent random variables and $V_i \sim$ beta$(\frac{1}{2}(n-p-i+1), \frac{1}{2}r)$; when $n-p \geq m \geq r$, $\frac{2}{n}\log L_n$ has the same distribution as $\sum_{i=1}^r \log V_i$, where V_i's are independent and $V_i \sim$ beta$(\frac{1}{2}(n+r-p-m-i+1), \frac{1}{2}m)$.*

Lemma 4 (Theorem 2.1.12 in Muirhead (2009)). *The multivariate Gamma function defined in* (S3.24) *can be written as*

$$\Gamma_m(a) = \pi^{m(m-1)/4}\prod_{j=1}^m \Gamma(a-(j-1)/2).$$

Lemma 5. *Consider m is fixed and $a \to \infty$. We have*

$$\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-i} = \Big\{\frac{1}{2}\Big(\sigma_a^2 - \frac{m}{(a-1)^2}\Big)\Big\}\{1+O(1/a)\}, \quad \text{(S3.42)}$$

$$\sum_{i=1}^{m}\{\log(a-1)-\log(a-i)\} = -\mu_a + O\Big(\frac{m^2}{a^2}\Big), \quad \text{(S3.43)}$$

where $\mu_a = -(m-a+3/2)\log\{1-m/(a-1)\} + (a-1)m/a$ *and* $\sigma_a^2 = -2[m/(a-1)+\log\{1-m/(a-1)\}]$.

Proof. We first prove (S3.42). As m is fixed and $a \to \infty$, we have

$$\sigma_a^2 = -2\Big[\frac{m}{a-1} + \log\Big(1-\frac{m}{a-1}\Big)\Big] = \Big(\frac{m}{a-1}\Big)^2\{1+O(m/a)\},$$

and

$$\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-i} = \frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-1} + \epsilon_a = \frac{m(m-1)}{2(a-1)^2} + \epsilon_a,$$

where $|\epsilon_a| \le 2(a-1)^{-3}\sum_{i=1}^{m}(i-1)^2 \le 3(m/a)^3$. Therefore,

$$\begin{aligned}\frac{1}{a-1}\sum_{i=1}^{m}\frac{i-1}{a-i} &= \frac{m(m-1)}{2(a-1)^2}\Big\{1+O\Big(\frac{m}{a}\Big)\Big\} \\ &= \Big[\frac{1}{2}\Big\{\sigma_a^2 - \frac{m}{(a-1)^2}\Big\}\Big]\{1+O(1/a)\},\end{aligned}$$

where in the last equation, we use the fact that $O(m/a) = O(1/a)$ as m is fixed. Then (S3.42) is proved.

We then prove (S3.43). Recall Stirling formula, (see, e.g., p. 368 Gamelin, 2001)

$$\log \Gamma(x) = (x - 1/2)\log x - x + \log\sqrt{2\pi} + \frac{1}{12x} + O(x^{-3})$$

as $x \to \infty$. Therefore,

$$\begin{aligned}
& \log\Gamma(a-1) - \log\Gamma(a-m-1) \\
= \; & (a-3/2)\log(a-1) - (a-m-3/2)\log(a-m-1) - m \\
& + \frac{1}{12}\Big(\frac{1}{a-1} - \frac{1}{a-m-1}\Big) + O(a^{-3}) \\
= \; & (a-3/2)\log(a-1) - (a-m-3/2)\log(a-m-1) - m + O(ma^{-2}).
\end{aligned}$$

Since for integers $k \geq 1$, $\Gamma(k) = (k-1)! = \Pi_{i=1}^{k-1} i$. Then we have

$$\begin{aligned}
& \sum_{i=1}^{m} \{\log(a-1) - \log(a-i)\} \\
= \; & m\log(a-1) - \{\log\Gamma(a-1) - \log\Gamma(a-m-1)\} + \log\Big(1 - \frac{m}{a-1}\Big) \\
= \; & m\log(a-1) - (a-3/2)\log(a-1) + (a-m-3/2)\log(a-m-1) \\
& + m + \log\Big(1 - \frac{m}{a-1}\Big) + O(ma^{-2}) \\
= \; & -(m-a+3/2)\log\Big(1 - \frac{m}{a-1}\Big) + m - \frac{m}{a-1} + O\Big(\frac{m^2}{a^2}\Big) \\
= \; & -(m-a+3/2)\log\Big(1 - \frac{m}{a-1}\Big) + \frac{a-1}{a}m + O\Big(\frac{m^2}{a^2}\Big) \\
= \; & -\mu_a + O\Big(\frac{m^2}{a^2}\Big),
\end{aligned}$$

where in the last two equations, we use the fact that $\frac{a-2}{a-1}m = \frac{a-1}{a}m + O(ma^{-2})$. □

Lemma 6. *Consider m is fixed and $a \to \infty$. Define*

$$g_i(x) = \Big(\frac{a-i}{2} + x\Big)\log\Big(\frac{a-i}{2} + x\Big) - \Big(\frac{a-1}{2} + x\Big)\log\Big(\frac{a-1}{2} + x\Big)$$

for $1 \leq i \leq m$ and $x > -(a-m)/2$. Let μ_a and σ_a be as in Lemma 5. If $t = o(a)$ and $mt^2/a^2 = o(1)$, we have that as $a \to \infty$,

$$\sum_{i=1}^{m} \{g_i(t) - g_i(0)\} = \mu_a t + \frac{\sigma_a^2}{2}t^2 + o(1).$$

Proof. We know for $1 \le i \le m$,

$$\begin{aligned}
g_i'(x) &= \log(\frac{a-i}{2}+x) - \log(\frac{a-1}{2}+x), \\
g_i''(x) &= \frac{1}{\frac{a-i}{2}+x} - \frac{1}{\frac{a-1}{2}+x} = \frac{\frac{i-1}{2}}{(\frac{a-i}{2}+x)(\frac{a-1}{2}+x)}, \\
g_i^{(3)}(x) &= -\frac{1}{(\frac{a-i}{2}+x)^2} + \frac{1}{(\frac{a-1}{2}+x)^2} \\
&= -\frac{\frac{i-1}{2}\cdot\frac{2a-i-1}{2}+(i-1)x}{(\frac{a-i}{2}+x)^2(\frac{a-1}{2}+x)^2}.
\end{aligned}$$

By Taylor expansion,

$$\begin{aligned}
g_i(t) - g_i(0) &= g_i'(0)t + \frac{t^2}{2}g_i''(0) + \frac{t^3}{6}g_i^{(3)}(\xi_i) \\
&= \{\log(a-i) - \log(a-1)\}t + \frac{i-1}{(a-1)(a-i)}t^2 + \frac{t^3}{6}g_i^{(3)}(\xi_i).
\end{aligned}$$

For $1 \le i \le m$, fixed m and $0 \le \xi_i \le t = o(a)$, we have $\sup_{|\xi_i|\le|t|,1\le i\le m}|g_i^{(3)}(\xi_i)| \le ca^{-3}$, where c denotes an universal constant. Therefore, as $t = o(a)$, $|t^3 g_i^{(3)}(\xi_i)| \le ct^3a^{-3} = o(1)$. In addition, by Lemma 5, and the fact that $mt^2/(a-1)^2 = o(1)$, we have as $a \to \infty$,

$$\begin{aligned}
\sum_{i=1}^{m}\{g_i(t) - g_i(0)\} &= \mu_a t + \Big[\frac{1}{2}\Big(\sigma_a^2 - \frac{m}{(a-1)^2}\Big)\Big]t^2 + o(1) \\
&= \mu_a t + \frac{\sigma_a^2}{2}t^2 + o(1).
\end{aligned}$$

□

Lemma 7. *Consider $n-p \to \infty$, $r \to \infty$, $m/(n-p) \to 0$ and $m/(n-p+r) \to 0$. For $t = nh/2$, $a = n-p+r$ or $a = n-p$, we have*

$$\log \frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} = \upsilon_a t + \vartheta_a t^2 + \gamma_a(t) + o(1),$$

where

$$\upsilon_a = -[2m + (a-m-3/2)\log\{1-m/(a-1)\}];\ \vartheta_a = -[m/(a-1) + \log\{1-m/(a-1)\}];$$
$$\gamma_a(t) = m\Big\{\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big) - \frac{a-1}{2}\log\Big(\frac{a-1}{2}\Big)\Big\}.$$

Proof. By Lemma 4, we know

$$\log \frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} = \sum_{i=1}^{m} \log \frac{\Gamma(\frac{a-i}{2}+t)}{\Gamma(\frac{a-i}{2})}. \tag{S3.44}$$

To prove the lemma, we expand each summed term in (S3.44), $\log\{\Gamma(\frac{a-i}{2}+t)/\Gamma(\frac{a-i}{2})\}$, by Lemma A.1. in Jiang and Qi (2015). To apply the lemma, we first need to check the condition that for each $1 \le i \le m$, $t \in [-\delta(a-i)/2, \delta(a-i)/2]$ for any given $\delta \in (0,1)$.

Recall that we previously define $nh = 2s/\sigma_n$ in Section S3.1. Then $t = nh/2 =$

$s\sigma_n^{-1}$. Note that when $m/(n-p)$ and $m/(n-p+r) \to 0$,

$$\begin{aligned}\sigma_n^2 &= \frac{1}{2}\log\left(1-\frac{m}{n+r-p}\right)-\frac{1}{2}\log\left(1-\frac{m}{n-p}\right)\\ &= \frac{mr}{2(n-p)(n+r-p)}\{1+o(1)\}.\end{aligned}$$

Thus we have

$$t = O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr}}. \tag{S3.45}$$

For $a = n-p+r$ or $a = n-p$, and $1 \le i \le m$, by (S3.45), we then have

$$\begin{aligned}\frac{t}{a-i} &\le \frac{t}{n-p-m} = O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr(n-p-m)^2}}\\ &= O(s)\sqrt{\frac{1+r/(n-p)}{mr\{1-m/(n-p)\}^2}}\\ &= O(s)\sqrt{\left\{\frac{1}{mr}+\frac{1}{m(n-p)}\right\}\{1+o(1)\}} = o(1),\end{aligned}$$

where the last two equations follow from the condition that $m/(n-p) \to 0, r \to \infty$ and $n-p \to \infty$. Then we know that for each $1 \le i \le m$, $t \in [-\delta(a-i)/2, \delta(a-i)/2]$ for any given $\delta \in (0,1)$.

Therefore, the condition of Lemma A.1. in Jiang and Qi (2015) is satisfied. By

that lemma, we know when $a \to \infty$, for uniformly $1 \le i \le m$,

$$\begin{aligned}\log \frac{\Gamma(\frac{a-i}{2}+t)}{\Gamma(\frac{a-i}{2})} &= \Big(\frac{a-i}{2}+t\Big)\log\Big(\frac{a-i}{2}+t\Big) - \frac{a-i}{2}\log\frac{a-i}{2} \\ &\quad -t - \frac{t}{a-i} + O\Big(\frac{t^2}{a^2}\Big).\end{aligned}$$

Write $\frac{t}{a-i} = \frac{t}{a} + \frac{t}{a} \times \frac{i}{a-i}$. Then similarly to Lemma 5, we have

$$\sum_{i=1}^{m} \frac{t}{a-i} = \frac{mt}{a} + \frac{tm(m+1)}{2a(a-1)} + O\Big\{\frac{t}{a} \times \Big(\frac{m}{a}\Big)^3\Big\}. \tag{S3.46}$$

For $a = n-p$, by (S3.45), $m/(n-p) \to 0$ and $m \le r$,

$$\begin{aligned}\frac{tm(m+1)}{a(a-1)} &= O(s)\sqrt{\frac{(n-p)(n-p+r)}{mr}}\frac{m^2}{(n-p)^2} \\ &= O(s)\sqrt{\frac{m(n-p+r)}{r(n-p)}}\frac{m}{n-p} \\ &= O(s)\sqrt{\frac{m}{\min\{n-p,r\}}}\frac{m}{n-p} = o(1).\end{aligned}$$

For $a = n-p+r$, similar conclusion, $tm(m+1)/\{a(a-1)\} = o(1)$, holds by substituting $n-p$ with $n-p+r$. In addition, for $a = n-p$ or $a = n-p+r$, by

(S3.45),

$$\begin{aligned}\frac{t}{a} \leq \frac{t}{n-p} &= O(s)\sqrt{\frac{\max\{n-p,r\}}{mr(n-p)}} \\ &= O(s)\sqrt{\frac{1}{m \times \min\{n-p,r\}}} = o(1). \end{aligned} \tag{S3.47}$$

Then based on (S3.46) and (S3.47), we obtain

$$\sum_{i=1}^{m}\Big\{-t-\frac{t}{a-i}+O(t^2/a^2)\Big\} = -mt-\frac{mt}{a}+o(1).$$

Therefore, from (S3.44), we have

$$\begin{aligned} &\log\frac{\Gamma_m(\frac{a-1}{2}+t)}{\Gamma_m(\frac{a-1}{2})} \\ =\ & -\frac{(a+1)mt}{a}+\sum_{i=1}^{m}\Big\{\Big(\frac{a-i}{2}+t\Big)\log\Big(\frac{a-i}{2}+t\Big)-\frac{a-i}{2}\log\frac{a-i}{2}\Big\}+o(1). \end{aligned} \tag{S3.48}$$

For $1 \leq i \leq m$, define the function

$$g_i(x) = \Big(\frac{a-i}{2}+x\Big)\log\Big(\frac{a-i}{2}+x\Big)-\Big(\frac{a-1}{2}+x\Big)\log\Big(\frac{a-1}{2}+x\Big),$$

and $x > -(a-m)/2$. We then know that the summation term "$\sum$" in (S3.48) equals

to

$$m\Big[\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big)-\frac{a-1}{2}\log\frac{a-1}{2}\Big]+\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}. \quad \text{(S3.49)}$$

We then examine the function $\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}$ in (S3.49). Note that by (S3.47), we know $t=o(a)$, $mt^2/a^2=o(1)$ and $mt/a=O(1)$ as $m<n-p$ and $m\le r$. Thus the conditions of Lemma 6 and Lemma A.3. in Jiang and Qi (2015) are satisfied when m is fixed and $m\to\infty$ respectively. When m is fixed, we apply Lemma 6; when $m\to\infty$, we apply Lemma A.3. in Jiang and Qi (2015). Then we obtain

$$\sum_{i=1}^{m}\{g_i(t)-g_i(0)\}=\mu_a t+\frac{1}{2}\sigma_a^2t^2+o(1),$$

where

$$\begin{aligned}\mu_a &= (m-a+3/2)\log\Big(1-\frac{m}{a-1}\Big)-m\frac{a-1}{a},\\ \sigma_a^2 &= -2\Big[\frac{m}{a-1}+\log\Big(1-\frac{m}{a-1}\Big)\Big].\end{aligned}$$

Therefore, the proposition can be proved by noticing

$$\upsilon_a = -\frac{(a+1)m}{a} + \mu_a; \quad \vartheta_a = \sigma_a^2/2;$$
$$\gamma_a(t) = m\Big[\Big(\frac{a-1}{2}+t\Big)\log\Big(\frac{a-1}{2}+t\Big) - \frac{a-1}{2}\log\frac{a-1}{2}\Big].$$

□

Lemma 8. *Under Case 1 in Section S3.1, $R_{n,1}$ and $R_{n,2}$ defined in* (S3.36) *satisfy*

$$-\frac{m(n+r-p)}{2}R_{n,1} + \frac{m(n-p)}{2}R_{n,2} = o(1).$$

Proof. Note that

$$\begin{aligned}
& -\frac{m(n+r-p)}{2}R_{n,1} + \frac{m(n-p)}{2}R_{n,2} \\
= \; & \frac{m}{2}\Big[\sum_{k=3}^{\infty}\frac{1}{k}(-nh)^k\Big\{\frac{1}{(n+r-p)^{k-1}} - \frac{1}{(n-p)^{k-1}}\Big\}\Big] \\
= \; & \frac{m}{2}\Big\{\sum_{k=3}^{\infty}\frac{1}{k}(-nh)^k\frac{-\sum_{q=1}^{k-1}\binom{k-1}{q}r^q(n-p)^{k-1-q}}{(n+r-p)^{k-1}(n-p)^{k-1}}\Big\} \\
= \; & \frac{mnh}{2}\sum_{k=3}^{\infty}\frac{1}{k}\Big(\frac{-nh}{n+r-p}\Big)^{k-1}\sum_{q=1}^{k-1}\binom{k-1}{q}\Big(\frac{r}{n-p}\Big)^q. \qquad \text{(S3.50)}
\end{aligned}$$

If $r/(n-p)=1$,

$$|(\text{S3.50})| \leq mnh \sum_{k=3}^{\infty} \Big(\frac{2nh}{n+r-p}\Big)^{k-1} = O\Big\{mnh \frac{n^2h^2}{(n+r-p)^2}\Big\},$$

where

$$\begin{aligned} \frac{mnh \times n^2h^2}{(n+r-p)^2} &= O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}} \times \frac{(n-p)(n+r-p)}{mr(n+r-p)^2}\Big\} \\ &= O\Big\{\frac{m\sqrt{r^2}}{\sqrt{mr}} \times \frac{r^2}{mr \times r^2}\Big\} = o(1), \end{aligned}$$

as $r \to \infty$.

If $r/(n-p) > 1$, as $\{nh/(n+r-p)\} \times \{r/(n-p)\} = O\{\sqrt{r}/\sqrt{m(n-p)(n+r-p)}\} = o(1)$,

$$|(\text{S3.50})| \leq mnh \sum_{k=3}^{\infty} \Big(\frac{2nh}{n+r-p} \times \frac{r}{n-p}\Big)^{k-1} = O\Big\{mnh\Big(\frac{2nh}{n+r-p}\Big)^2\Big(\frac{r}{n-p}\Big)^2\Big\},$$

where

$$
\begin{aligned}
& mnh\Big(\frac{nh}{n+r-p}\Big)^2\Big(\frac{r}{n-p}\Big)^2 \\
= \;& O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}} \times \frac{(n-p)(n+r-p)}{mr(n+r-p)^2} \times \frac{r^2}{(n-p)^2}\Big\} \\
= \;& O\Big\{\frac{\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}}\frac{r}{(n-p)(n+r-p)}\Big\} \\
= \;& O\Big\{\frac{r}{\sqrt{mr(n+r-p)(n-p)}}\Big\} = o(1),
\end{aligned}
$$

as $n+r-p \geq r$ and $n-p \to \infty$.

If $r/(n-p) < 1$,

$$
|(\text{S3.50})| \leq mnh\sum_{k=3}^{\infty}\Big(\frac{nh}{n+r-p}\Big)^{k-1}\frac{r}{(n-p)} = O\Big\{mnh\frac{(nh)^2}{(n+r-p)^2} \times \frac{r}{(n-p)}\Big\},
$$

where

$$
\begin{aligned}
& mnh\frac{(nh)^2}{(n+r-p)^2} \times \frac{r}{(n-p)} \\
= \;& O\Big\{\frac{m\sqrt{(n-p)(n+r-p)}}{\sqrt{mr}} \times \frac{(n-p)(n+r-p)}{mr(n+r-p)^2} \times \frac{r}{(n-p)}\Big\} \\
= \;& O\Big\{\frac{\sqrt{n-p}}{\sqrt{mr(n+r-p)}}\Big\} = o(1).
\end{aligned}
$$

□

S4. Theorem 4

We give the main proof of Theorem 4 in Section S4.1, where we use some concepts of hypergeometric function, which is introduced in Section S4.2, and the lemmas we use are given and proved in Section S4.3.

S4.1 Proof of Theorem 4

As $p/n = \rho_p$, $r/n = \rho_r$, $m/n = \rho_m$ with $\rho_p, \rho_r, \rho_m \in (0,1)$ and $\rho_p + \rho_m < 1$, we know that σ_n^2 in Theorem 3 satisfies

$$\sigma_n^2 = 2\log\Big\{\Big(1 - \frac{\rho_m}{1+\rho_r-\rho_p}\Big)\Big(1 - \frac{\rho_m}{1-\rho_p}\Big)^{-1}\Big\},$$

which is a positive constant, and we write the constant as σ^2. Then $T_1 = \{-2\log L_n + \mu_n\}/(n\sigma)$, and we examine the moment generating function $\mathrm{E}\{2s\log L_n/(n\sigma)\}$. Let $h = 2s/(n\sigma)$. By Lemma 9, we have

$$\begin{aligned} & \mathrm{E}\{2s\log L_n/(n\sigma)\} = \mathrm{E}\{\exp(h\log L_n)\} \\ = \ & \mathrm{E}L_n^h = \mathrm{E}_0 L_n^h \times {}_1F_1\Big(\frac{nh}{2}; \frac{1}{2}(n+r-p) + \frac{nh}{2}; -\frac{1}{2}\Omega\Big), \end{aligned} \tag{S4.51}$$

where $\mathrm{E}_0 L_n^h$ is the moment generating function of $\log L_n$ under H_0, and ${}_1F_1$ is the hypergeometric function, which depends on Ω only through its eigenvalues symmet-

rically.

As ${}_1F_1$ only depends on Ω via its eigenvalues symmetrically, without loss of generality, we consider the alternative with $\Omega = \mathrm{diag}(w_1, \cdots, w_m)$ and $w_1 \geq \cdots \geq w_m \geq 0$. Let $\xi_a = nh/2$, $\xi_b = (n + r - p + nh)/2$ and $Q = -\Omega/2 = \mathrm{diag}(-w_1/2, \cdots, -w_m/2)$, then we write ${}_1F_1(nh/2, (n + r - p + nh)/2; -\Omega/2)$ as ${}_1F_1(\xi_a; \xi_b; Q)$. Note that we assume that Ω has fixed rank m_0 in Theorem 4, then $\omega_1 \geq \ldots \geq \omega_{m_0} > 0$ are m_0 nonzero eigenvalues of Ω. Further define $\tilde{Q} = \mathrm{diag}(-\omega_1/2, \ldots, -\omega_{m_0}/2)$. By Lemma 11, we know ${}_1F_1(\xi_a; \xi_b; Q) = {}_1F_1(\xi_a; \xi_b; \tilde{Q})$. Then to evaluate ${}_1F_1(\xi_a; \xi_b; Q)$ when Q has fixed rank, without loss of generality, we consider ${}_1F_1(\xi_a; \xi_b; \tilde{Q})$.

Let $W = \log {}_1F_1(\xi_a; \xi_b; \tilde{Q})$ and $\tilde{Q} = -n\tilde{\Delta}/2$ with $\tilde{\Delta} = \mathrm{diag}(\delta_1, \ldots, \delta_{m_0})$. From Lemma 12, we know that $W(\tilde{\Delta})$ is the unique solution of each of the m_0 partial differential equations

$$\begin{aligned}&\Big[\frac{1}{2}(n + r - p - m_0 + 1) + \frac{nh}{2} + \frac{1}{2}n\delta_j + \frac{1}{2}\sum_{i \neq j}^{m_0} \frac{\delta_j}{\delta_j - \delta_i}\Big]\frac{\partial W}{\partial \delta_j} \\ &\quad + \delta_j\Big[\frac{\partial^2 W}{\partial \delta_j^2} + \Big(\frac{\partial W}{\partial \delta_j}\Big)^2\Big] - \frac{1}{2}\sum_{i \neq j}^{m_0} \frac{\delta_i}{\delta_j - \delta_i}\frac{\partial W}{\partial \delta_i} = -\frac{nh}{2} \times n,\end{aligned} \tag{S4.52}$$

for $j = 1, \ldots, m_0$, subject to the conditions that $W(\tilde{\Delta})$ is (a) a symmetric function of $\delta_1, \ldots, \delta_{m_0}$, and (b) analytic at $\tilde{\Delta} = \mathbf{0}_{m_0 \times m_0}$ with $W(\mathbf{0}_{m_0 \times m_0}) = 0$. As $r/n = \rho_r$,

$p/n = \rho_p$, m_0 is a fixed number and $nh = 2s/\sigma$, we can write (S4.52) into

$$\Big[\frac{1}{2}(1+\rho_r-\rho_p)n + \frac{1}{2}(2s/\sigma - m_0 + 1) + \frac{1}{2}n\delta_j + \frac{1}{2}\sum_{i\neq j}^{m_0}\frac{\delta_j}{\delta_j-\delta_i}\Big]\frac{\partial W}{\partial \delta_j}$$
$$+\delta_j\Big[\frac{\partial^2 W}{\partial \delta_j^2} + \Big(\frac{\partial W}{\partial \delta_j}\Big)^2\Big] - \frac{1}{2}\sum_{i\neq j}^{m_0}\frac{\delta_i}{\delta_j-\delta_i}\frac{\partial W}{\partial \delta_i} = -\frac{s}{\sigma}\times n. \tag{S4.53}$$

Similarly to Theorem 10.5.6 in Muirhead (2009), we write $W(\tilde{\Delta}) = P_0(\tilde{\Delta}) + P_1(\tilde{\Delta})/n + \ldots$. Note that $nh = 2s/\sigma$. Matching n on both sides of (S4.53), we obtain

$$\Big[\frac{1}{2}(1+\rho_r-\rho_p)n + \frac{1}{2}n\delta_j\Big]\frac{\partial P_0}{\partial \delta_j} = -\frac{sn}{\sigma}.$$

Solving this subject to conditions (a) and (b), we obtain

$$P_0(\tilde{\Delta}) = -\frac{2s}{\sigma}\sum_{j=1}^{m_0}\log\Big(1+\frac{\delta_j}{1+\rho_r-\rho_p}\Big).$$

Then we have $W(\tilde{\Delta}) = P_0(\tilde{\Delta}) + O(n^{-1})$. From (S4.51), we know

$$\mathrm{E}L_n^h = \mathrm{E}_0 L_n^h \times e^{\log {}_1F_1} = \mathrm{E}_0 e^{\frac{s}{n\sigma/2}\log(L_n)+W}. \tag{S4.54}$$

Write $W_\Delta = \sum_{j=1}^{m_0}\log[1+\delta_j(1+\rho_r-\rho_p)^{-1}]$ and $A_1 = 2/\sigma$. (S3.22) and (S4.54) show that $\{\log L_n - \mu_n/2\}/(n\sigma_n/2) \xrightarrow{D} \mathcal{N}(-A_1W_\Delta, 1)$, and thus $\{-2\log L_n + \mu_n\}/(n\sigma) \xrightarrow{D} \mathcal{N}(A_1W_\Delta, 1)$. Then the power $P(T_1 > z_\alpha) \to \bar{\Phi}(z_\alpha - A_1W_\Delta)$.

S4.2 Brief review of hypergeometric function

We rephrase some related definitions and results about hypergeometric function, where the details can be found in Chapter 7 in Muirhead (2009).

Let k be a positive integer; a partition κ of k is written as $\kappa = (k_1, k_2, \ldots)$, where $\sum_t k_t = k$ and $k_1 \geq k_2 \geq \ldots$ are non-negative integers. In addition, let M be an $m \times m$ symmetric matrix with eigenvalues $l_1, \ldots, l_m$, and let $\kappa = (k_1, k_2 \ldots)$ be a partition of k into no more than m nonzero parts. We write the zonal polynomial of M corresponding to κ as $C_\kappa(M)$. Then by the definition, we know the hypergeometric function ${}_1F_1(\xi_a; \xi_b; Q)$ satisfies

$$ {}_1F_1(\xi_a; \xi_b; Q) = \sum_{k=0}^{\infty} \sum_{\kappa:k} \frac{(\xi_a)_\kappa}{(\xi_b)_\kappa} \frac{C_\kappa(Q)}{k!}, \tag{S4.55} $$

where $\sum_{\kappa:k}$ represents the summation over the partitions $\kappa = (k_1, \ldots, k_m)$, $k_1 \geq \ldots \geq k_m \geq 0$, of k, $C_\kappa(Q)$ is the zonal polynomial of Q corresponding to κ, and the generalized hypergeometric coefficient $(\xi)_\kappa$ is given by $(\xi)_\kappa = \prod_{i=1}^{t} (\xi - (i-1)/2)_{k_i}$ with $(a)_{k_i} = a(a+1)\ldots(a+k_i-1)$ and $(a)_0 = 1$.

We then characterize the zonal polynomials $C_\kappa(M)$. For given partition $\kappa = (k_1, k_2, \ldots)$ of k, define the monomial symmetric functions $N_\kappa(M) = \sum_{\{i_1,\ldots,i_t\}} l_{i_1}^{k_1} \ldots l_{i_t}^{k_t}$, where t is the number of nonzero parts in the partition κ, and the summation is over the distinct permutations $(i_1, \ldots, i_t)$ of t different integers from $1, \ldots, m$. For another

partition $\lambda = (\lambda_1, \lambda_2, \ldots)$, we write $\kappa > \lambda$ if $k_i > \lambda_i$ for the first index i for which the parts in κ and λ are unequal. Then we have $C_\kappa(M) = \sum_{\lambda \leq \kappa} c_{\kappa,\lambda} N_\lambda(M)$, where $c_{\kappa,\lambda}$ are constants.

S4.3 Lemmas in the proof of Theorem 4

Lemma 9. $\mathrm{E}L_n^h = \mathrm{E}_0 L_n^h \times {}_1F_1(nh/2; (n+r-p+nh)/2; -\Omega/2)$.

Proof. The result follows from Theorem 10.5.1 in Muirhead (2009). □

Lemma 10. *Suppose matrix M of size $m \times m$ has m eigenvalues $l_1, \ldots, l_m$, but only has m_0 positive eigenvalues $l_1, \ldots, l_{m_0}$ and $\tilde{M} = \mathrm{diag}(l_1, \ldots, l_{m_0})$. Then for given partition κ, the zonal polynomial functions satisfy $N_\kappa(M) = N_\kappa(\tilde{M})$.*

Proof. By the definition of monomial function $N_\lambda(M)$, we note that $\sum_{\{i_1,\ldots,i_t\}} l_{i_1}^{k_1} \ldots l_{i_t}^{k_t} = \sum_{\{\tilde{i}_1,\ldots,\tilde{i}_t\}} l_{\tilde{i}_1}^{k_1} \ldots l_{\tilde{i}_t}^{k_t}$, where $\sum_{\{\tilde{i}_1,\ldots,\tilde{i}_t\}}$ represents the summation over the distinct permutations $(\tilde{i}_1, \ldots, \tilde{i}_t)$ of t different integers from $1, \ldots, m_0$. It follows that $N_\lambda(M) = N_\lambda(\tilde{M})$, where $\tilde{M} = \mathrm{diag}(l_1, \ldots, l_{m_0})$. □

Lemma 11. *Suppose Q has fixed rank m_0, then ${}_1F_1(\xi_a; \xi_b; Q) = {}_1F_1(\xi_a; \xi_b; \tilde{Q})$.*

Proof. As Q has rank m_0, it only has m_0 nonzero eigenvalues. To prove the lemma, we note that the hypergeometric function can be expressed as the linear combination of the zonal polynomials of a matrix. We then state two properties of the zonal polynomial functions $C_\kappa(Q)$. First, by Corollary 7.2.4 in Muirhead (2009), we know

that when κ is a partition of k into more than m_0 nonzero parts, $C_\kappa(Q) = 0$. Second, when κ is a partition of k into fewer than m_0 nonzero parts, $C_\kappa(Q) = C_\kappa(\tilde{Q})$. To see this, we note that $C_\kappa(Q) = \sum_{\lambda \le \kappa} c_{\lambda,\kappa} N_\lambda(Q)$ and the constants $c_{\kappa,\lambda}$ do not depend on the eigenvalues of Q. Then by Lemma 10, we know that $C_\kappa(M) = C_\kappa(\tilde{M})$. Finally, by the definition in (S4.55), we have ${}_1F_1(\xi_a; \xi_b; Q) = {}_1F_1(\xi_a; \xi_b; \tilde{Q})$. □

Lemma 12. *$W = \log {}_1F_1(\xi_a; \xi_b; \tilde{Q})$ with $\tilde{Q} = -n\tilde{\Delta}/2$ discussed in Section S4.1 is the unique solution of each of the m_0 partial differential equations*

$$\Big[\frac{1}{2}(n + r - p - m_0 + 1) + \frac{nh}{2} + \frac{1}{2}n\delta_j + \frac{1}{2}\sum_{i \neq j}^{m_0} \frac{\delta_j}{\delta_j - \delta_i}\Big]\frac{\partial W}{\partial \delta_j}$$
$$+\delta_j\Big[\frac{\partial^2 W}{\partial \delta_j^2} + \left(\frac{\partial W}{\partial \delta_j}\right)^2\Big] - \frac{1}{2}\sum_{i \neq j}^{m_0} \frac{\delta_i}{\delta_j - \delta_i}\frac{\partial W}{\partial \delta_i} = -\frac{nh}{2} \times n,$$

for $j = 1, \ldots, m_0$, subject to the conditions that $W(\tilde{\Delta})$ is (a) a symmetric function of $\delta_1, \ldots, \delta_{m_0}$, and (b) analytic at $\tilde{\Delta} = \mathbf{0}_{m_0 \times m_0}$ with $W(\mathbf{0}_{m_0 \times m_0}) = 0$.

Proof. As m_0 is fixed, the result follows from Theorem 7.5.6 in Muirhead (2009) by changing of variables. □

S5. Theorem 5

We give the conditions of Theorem 5 in Section S5.1, and the main proof Theorem 5 is given in Section S5.2, while the lemmas we use in the proof are given and proved in Section S5.3.

S5.1 Conditions of Theorem 5

To derive Theorem 5, we need some regularity conditions. We use $\lambda_{\max}(\cdot)$ and $\lambda_{\min}(\cdot)$ to denote the largest and smallest eigenvalues of a matrix respectively; $\mathrm{diag}(\cdot)$ denotes the vector of diagonal elements of a matrix; $\max\mathrm{diag}(\cdot)$ and $\min\mathrm{diag}(\cdot)$ represent the maximum and minimum value of the diagonal elements of a matrix respectively; $\|\cdot\|$ denotes the ℓ_2-norm of a vector; and $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\intercal$ denotes the indicator vector with 1 on the ith entry.

Condition 1. *The rows of X and E independently follow multivariate Gaussian distribution with covariance matrices Σ_x and Σ respectively. There exist nonnegative constants t and τ and positive constants $(c_1, c_2, c_3, c_4, c_5)$ such that $\lambda_{\max}(\Sigma_x) \leq c_1 n^\tau$, $\lambda_{\min}(\Sigma) \geq c_2 n^{-t}$, $\min\mathrm{diag}(\Sigma_x) \geq c_3$, $\max\mathrm{diag}(\Sigma) \leq c_4$ and $\max\mathrm{diag}(B^\intercal \Sigma_x B) \leq c_5$.*

Condition 2. *For some constants κ, u, $c_6 > 0$ and $c_7 > 0$, and fixed $i \in \mathcal{M}_*$, there exists $\mathbf{a}_{0,i} \in \mathbb{R}^m$ with $\|\mathbf{a}_{0,i}\| = 1$ such that $\max\{\|\Sigma_x^{1/2} B \mathbf{a}_{0,i}\|, \|\Sigma^{1/2}\mathbf{a}_{0,i}\|\} \leq c_6 n^u$ and $|e_i^\intercal \Sigma_x B \mathbf{a}_{0,i}| \sigma_{x,i}^{-1} \geq c_7 n^{-\kappa}$, where $\sigma_{x,i}^2$ is the i-th diagonal element of Σ_x.*

Condition 3. *Assume $m = O(n^s)$ with $0 \leq s < 1$; $\iota + \tau < 1$, where $\iota = 2\kappa + 2u + t + s$; $p > c_9 n$ for some constant $c_9 > 1$; $\log p = O(n^\pi)$ for some constant $\pi \in (0, 1 - 2\kappa - 2u - t - s)$; and $\delta n^{1-\iota-\tau} \to \infty$ as $n \to \infty$.*

Remark 1. *In Condition 1, we assume that X and E follow the Gaussian distribution*

for the ease of theoretical developments. We allow the eigenvalues of Σ_x and Σ to diverge or degenerate as n grows, which is similarly assumed in Fan and Lv (2008) and Wang and Leng (2016) etc. in studying the linear regression with univariate response. The boundedness of the diagonal elements of Σ and $B^\intercal \Sigma_x B$ is satisfied when the variances of response variables are $O(1)$. Condition 2 implies that there exists a combination of the response variables whose absolute covariance with the i-th predictor is sufficiently large. In particular, suppose for each $i \in \mathcal{M}_*$, there exists $k_i \in \{1, \ldots, m\}$ such that $\mathrm{cov}(x_{1,i}, y_{1,k_i})\sigma_{x,i}^{-1} \geq c_7 n^{-\kappa}$. Then Condition 2 is satisfied under Condition 1. Condition 3 allows the number of predictors p grow exponentially with n. The requirement $2u + 2\kappa + \tau + t + s < 1$ is satisfied when the eigenvalues of Σ_x, $B^\intercal \Sigma_x B$ and Σ do not diverge or degenerate too fast with n, and the covariance between $x_{1,i}$ and $\mathbf{y}_1^\intercal \mathbf{a}_{0,i}$ is sufficiently large.

S5.2 Proof of Theorem 5

Before proceeding to the proof, we define some notations and provide some preliminary results. Note that by the form of ω_j, we could assume $\mathrm{E}(X) = \mathbf{0}$ with loss of generality. Let $Z = X\Sigma_x^{-1/2}$. We know that the entries in Z are i.i.d. $\mathcal{N}(0, 1)$ by Condition 1, and then with probability 1, the $n \times p$ matrix Z has full rank n. Let

$\mu_1^{1/2}, \ldots, \mu_n^{1/2}$ be the n singular values of Z. Then $Z^\intercal Z$ has the eigendecomposition

$$Z^\intercal Z = U^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n, 0, \ldots, 0) U, \tag{S5.56}$$

where U belongs to the orthogonal group $\mathcal{O}(p)$. We write $U^\intercal = (\mathbf{u}_1, \ldots, \mathbf{u}_p)$. It follows that the Moore-Penrose generalized inverse of (S5.56) is

$$(Z^\intercal Z)^+ = \sum_{i=1}^{n} \frac{1}{\mu_i} \mathbf{u}_i \mathbf{u}_i^\intercal.$$

Moreover, we have the decomposition

$$S := (Z^\intercal Z)^+ Z^\intercal Z = U^\intercal \mathrm{diag}(I_n, 0) U = \tilde{U}^\intercal \tilde{U}, \tag{S5.57}$$

where $\tilde{U} = (I_n, \mathbf{0})_{n\times p} U$ and $(I_n, \mathbf{0})_{n\times p}$ represents an $n \times p$ matrix with first n columns being I_n and 0 in the remaining columns. Since $X = Z\Sigma_x^{1/2}$, by (S5.56), we know that

$$X^\intercal X = \Sigma_x^{1/2} \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_x^{1/2}. \tag{S5.58}$$

In addition, define $P = I_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n$. We can then write ω_i equivalently as

$$\omega_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal Y^\intercal P^\intercal P\mathbf{x}^i}{\sqrt{(\mathbf{a}^\intercal Y^\intercal P^\intercal PY\mathbf{a})\{(\mathbf{x}^i)^\intercal P^\intercal P\mathbf{x}^i\}}}.$$

By the property of ω_i, we assume without loss of generality that X and E have mean zero. Furthermore, suppose $\mathrm{diag}(\Sigma_x) = \mathrm{diag}(\sigma_{x,1}, \ldots, \sigma_{x,p})$ and let

$$\zeta_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal Y^\intercal P^\intercal P\mathbf{x}^i}{\sigma_{x,i}\sqrt{n \times \mathbf{a}^\intercal Y^\intercal P^\intercal PY\mathbf{a}}}. \tag{S5.59}$$

Then by Lemma 18, we know $\omega_i = \zeta_i\{1+o(1)\}$ with probability $1-O\{\exp(-c_0 n/\log n)\}$ for some constant $c_0 > 0$. As $Y = XB + E$, we have

$$\zeta_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal B^\intercal X^\intercal P^\intercal P\mathbf{x}^i + \mathbf{a}^\intercal E^\intercal P^\intercal P\mathbf{x}^i}{\sigma_{x,i}\sqrt{n \times \mathbf{a}^\intercal (Y^\intercal P^\intercal PY)\mathbf{a}}} = \xi_i + \eta_i, \tag{S5.60}$$

where

$$\xi_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal B^\intercal X^\intercal P^\intercal P\mathbf{x}^i}{\sigma_{x,i}\sqrt{n\mathbf{a}^\intercal (Y^\intercal P^\intercal PY)\mathbf{a}}}, \quad \eta_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal E^\intercal P^\intercal P\mathbf{x}^i}{\sigma_{x,i}\sqrt{n\mathbf{a}^\intercal (Y^\intercal P^\intercal PY)\mathbf{a}}}.$$

Moreover, we write $B = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_m]$, where $\boldsymbol{\beta}_j$ represents the j-th column of B. We then study ξ_i and η_i separately.

Step 1: We first examine $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)^\intercal$.

Step 1.1 (bounding $\|\boldsymbol{\xi}\|$ from above) For $i = 1, \ldots, p$,

$$|\xi_i| \leq \{n\lambda_{\min}(Y^\intercal P^\intercal PY)\}^{-1/2}\sigma_{x,i}^{-1}\|B^\intercal X^\intercal P^\intercal P\mathbf{x}^i\|,$$

where $\|\cdot\|$ represents the ℓ_2-norm of a vector. Then we know

$$\|\boldsymbol{\xi}\|^2 = \sum_{i=1}^{p}\xi_i^2 \leq \{n\lambda_{\min}(Y^\intercal P^\intercal PY)\}^{-1}\sum_{i=1}^{p}\sigma_{x,i}^{-2}\|B^\intercal X^\intercal P^\intercal P\mathbf{x}^i\|^2 \tag{S5.61}$$

By Lemma 19, we know that there exist constants c_1 and c_0 such that $\lambda_{\min}(Y^\intercal P^\intercal PY) \geq c_1 n^{1-t}$ with probability $1 - O\{\exp(-c_0 n)\}$. To bound $\|\boldsymbol{\xi}\|$ from above, we then examine $\sum_{i=1}^{p}\sigma_{x,i}^{-2}\|B^\intercal X^\intercal P^\intercal P\mathbf{x}^i\|^2$. Since $\min_{1\leq i\leq p}\sigma_{x,i}^2 \geq c_3$ by Condition 1,

$$\begin{aligned}\sum_{i=1}^{p}\sigma_{x,i}^{-2}\|B^\intercal X^\intercal P^\intercal P\mathbf{x}^i\|^2 &\leq c_3^{-1}\sum_{i=1}^{p}\sum_{k=1}^{m}(\boldsymbol{\beta}_k^\intercal X^\intercal P^\intercal P\mathbf{x}^i)^2 \\ &= c_3^{-1}\sum_{k=1}^{m}\boldsymbol{\beta}_k^\intercal X^\intercal P^\intercal P\sum_{i=1}^{p}\mathbf{x}^i(\mathbf{x}^i)^\intercal P^\intercal PX\boldsymbol{\beta}_k. \end{aligned}\tag{S5.62}$$

As $\sum_{i=1}^{p}\mathbf{x}^i(\mathbf{x}^i)^\intercal = XX^\intercal$ and $P^\intercal P = I_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n$, we have

$$\begin{aligned}\text{(S5.62)} &= c_3^{-1}\sum_{k=1}^{m}\|\boldsymbol{\beta}_k^\intercal X^\intercal(I_n - \mathbf{1}_n\mathbf{1}_n^\intercal/n)X\|^2 \\ &\leq 2c_3^{-1}\times(A_{\boldsymbol{\xi},1} + A_{\boldsymbol{\xi},2}), \end{aligned}\tag{S5.63}$$

where $A_{\boldsymbol{\xi},1} = \sum_{k=1}^{m}\|\boldsymbol{\beta}_k^\intercal X^\intercal X\|^2$ and $A_{\boldsymbol{\xi},2} = \sum_{k=1}^{m}\|\boldsymbol{\beta}_k^\intercal X^\intercal(\mathbf{1}_n\mathbf{1}_n^\intercal/n)X\|^2$.

We next examine $A_{\boldsymbol{\xi},1}$ and $A_{\boldsymbol{\xi},2}$ separately. By (S5.58),

$$\begin{aligned} A_{\boldsymbol{\xi},1} &= \sum_{k=1}^{m} \boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_x^{1/2} \Sigma_x^{1/2} \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_x^{1/2} \boldsymbol{\beta}_k \\ &\leq p^2 \{\lambda_{\max}(p^{-1} Z Z^\intercal)\}^2 \lambda_{\max}(\Sigma_x) \sum_{k=1}^{m} \boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \tilde{U} \Sigma_x^{1/2} \boldsymbol{\beta}_k, \end{aligned} \tag{S5.64}$$

where in the last inequality, we use the fact that $\tilde{U} \Sigma_x \tilde{U}^\intercal \preceq \lambda_{\max}(\Sigma_x) I_n$, and $\mathrm{diag}(\mu_1, \ldots, \mu_n) \preceq p \lambda_{\max}(p^{-1} Z Z^\intercal) I_n$, as $\mu_1^{1/2}, \ldots, \mu_n^{1/2}$ are the singular values of Z. We then bound (S5.64) from above by examining $\boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \tilde{U} \Sigma_x^{1/2} \boldsymbol{\beta}_k$. For fixed $k = 1, \ldots, m$, let $Q \in \mathcal{O}(p)$ such that $\Sigma_x^{1/2} \boldsymbol{\beta}_k = \|\Sigma_x^{1/2} \boldsymbol{\beta}_k\|_2 Q e_1$. By (S5.57) and Lemma 14, we know

$$\boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \tilde{U} \Sigma_x^{1/2} \boldsymbol{\beta}_k = \|\Sigma_x^{1/2} \boldsymbol{\beta}_k\|^2 \langle Q^\intercal S Q e_1, e_1 \rangle \overset{(d)}{=} \|\Sigma_x^{1/2} \boldsymbol{\beta}_k\|^2 \langle S e_1, e_1 \rangle. \tag{S5.65}$$

By Condition 1, $\|\Sigma_x^{1/2} \boldsymbol{\beta}_k\|^2 = \boldsymbol{\beta}_k^\intercal \Sigma_x \boldsymbol{\beta}_k \leq c_5$ for some constant $c_5 > 0$. Then by (S5.65) and Lemma 15, we know for some positive constants c_0 and c_1,

$$P(\boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \tilde{U} \Sigma_x^{1/2} \boldsymbol{\beta}_k > c_1 n/p) \leq O\{\exp(-c_0 n)\}. \tag{S5.66}$$

Combining (S5.64), Lemma 16, Condition 1 and (S5.66), we then know for some positive constants c_0 and c, with probability $1 - O\{m \exp(-c_0 n)\}$, $A_{\boldsymbol{\xi},1} \leq c m p^2 n^\tau n/p = c m p n^{1+\tau}$.

For $A_{\boldsymbol{\xi},2}$, note that

$$A_{\boldsymbol{\xi},2} = \sum_{k=1}^{m} \boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \Sigma_x Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \Sigma_x^{1/2} \boldsymbol{\beta}_k. \tag{S5.67}$$

Similarly, considering fixed $k = 1, \ldots, m$, we let $Q \in \mathcal{O}(p)$ such that $\Sigma_x^{1/2}\boldsymbol{\beta}_k = \|\Sigma_x^{1/2}\boldsymbol{\beta}_k\| Q e_1$. Then

$$\begin{aligned}
& \boldsymbol{\beta}_k^\intercal \Sigma_x^{1/2} Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \Sigma_x Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \Sigma_x^{1/2} \boldsymbol{\beta}_k \\
= \; & \|\Sigma_x^{1/2}\boldsymbol{\beta}_k\|^2 e_1^\intercal Q^\intercal Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \Sigma_x Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z Q e_1 \\
\leq \; & \lambda_{\max}(\Sigma_x) \|\Sigma_x^{1/2}\boldsymbol{\beta}_k\|^2 e_1^\intercal Q^\intercal Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z Q Q^\intercal Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z Q e_1 \\
\overset{(d)}{=} \; & \lambda_{\max}(\Sigma_x) \|\Sigma_x^{1/2}\boldsymbol{\beta}_k\|^2 \|Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z e_1\|^2,
\end{aligned} \tag{S5.68}$$

where in the last equality, we use the fact that $ZQ \overset{(d)}{=} Z$. Since the entries in Z are i.i.d. $\mathcal{N}(0,1)$, we have $L = (L_1, \ldots, L_p)^\intercal = Z^\intercal \mathbf{1}_n / \sqrt{n} \sim \mathcal{N}(\mathbf{0}_{p\times 1}, I_p)$ with $L_1 = \mathbf{1}_n^\intercal Z e_1 / \sqrt{n}$. It follows that $\|Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z e_1\|^2 = \sum_{j=1}^p L_1^2 L_j^2$. Since $L_1^2 \sim \chi_1^2$, there exist constants c_0 and c_1 such that $P(|L_1^2 - 1| > c_1 n) \leq O\{\exp(-c_0 n)\}$. Moreover, note that L_j^2's are i.i.d. χ_1^2-distributed random variables. By Lemma 13, for some positive constants c_0 and c_1, when $p \geq n$,

$$P\Big(\sum_{j=2}^{p} L_j^2 / (p-1) > 1 + c_1 \Big) \leq O\{\exp(-c_0 p)\} \leq O\{\exp(-c_0 n)\}. \tag{S5.69}$$

Thus there exist constants c and c_0 such that with probability $1 - O\{\exp(-c_0 n)\}$,

$$\|Z^\intercal(\mathbf{1}_n\mathbf{1}_n^\intercal/n)Ze_1\|^2 = \sum_{j=1}^{p} L_1^2 L_j^2 = L_1^4 + L_1^2 \sum_{j=2}^{p} L_j^2 \leq cpn.$$

By Condition 1, $\lambda_{\max}(\Sigma_x) \leq c_1 n^\tau$ and $\|\Sigma_x^{1/2}\boldsymbol{\beta}_k\|^2 \leq c_2$ for some constant c_1 and c_2. From (S5.67) and (S5.68), we know that $A_{\boldsymbol{\xi},2} \leq cmpn^{\tau+1}$ with probability $1 - O\{m\exp(-c_0 n)\}$.

In summary, we obtain that for some constants c and c_0, $A_{\boldsymbol{\xi},1}$ and $A_{\boldsymbol{\xi},2} \leq cmpn^{\tau+1}$ with probability $1 - O\{\exp(-c_0 n)\}$. Then by (S5.61), (S5.63) and Lemma 19, we have for some positive constants c_1 and c_0,

$$P\{\|\boldsymbol{\xi}\|^2 > c_1 n^{-(1+1-t)} pmn^{1+\tau}\} \leq O\{m\exp(-c_0 n)\} = O\{\exp(-c_0 n)\}, \quad \text{(S5.70)}$$

where the last equality is from Condition 3.

Step 1.2 (bounding $|\xi_i|$ for $i \in \mathcal{M}_*$ from below) Without loss of generality, we consider $B \neq \mathbf{0}_{p\times m}$. For fixed $i \in \mathcal{M}_*$,

$$\begin{aligned} \xi_i &= \max_{\mathbf{a}:\|\mathbf{a}\|_2=1} \frac{\mathbf{a}^\intercal B^\intercal X^\intercal P^\intercal P \mathbf{x}^i}{\sigma_{x,i}\sqrt{n \times \mathbf{a}^\intercal (Y^\intercal P^\intercal P Y)\mathbf{a}}} \\ &\geq \{n \times \mathbf{a}_{0,i}^\intercal Y^\intercal P^\intercal P Y \mathbf{a}_{0,i}\}^{-1/2} \sigma_{x,i}^{-1} |\mathbf{a}_{0,i}^\intercal B^\intercal X^\intercal P^\intercal P X e_i|, \end{aligned} \quad \text{(S5.71)}$$

where $\mathbf{a}_{0,i}$ in the last inequality is specified in Condition 2. To bound $|\xi_i|$ from below, we then examine (S5.71). By Lemma 19, there exist constants c_0 and c_2 such that with probability $1 - O\{\exp(-c_0 n)\}$,

$$\mathbf{a}_{0,i}^\intercal Y^\intercal P^\intercal P Y \mathbf{a}_{0,i} \leq c_2 n^{2u+1}. \tag{S5.72}$$

Moreover, as $P^\intercal P = I_n - \mathbf{1}_n \mathbf{1}_n^\intercal / n$, $\sigma_{x,i}^{-1} \mathbf{a}_{0,i}^\intercal B^\intercal X^\intercal P^\intercal P X e_i = \tilde{A}_{\xi,i,1} - \tilde{A}_{\xi,i,2}$, where $\tilde{A}_{\xi,i,1} = \sigma_{x,i}^{-1} \mathbf{a}_{0,i}^\intercal B^\intercal X^\intercal X e_i$ and $\tilde{A}_{\xi,i,2} = \sigma_{x,i}^{-1} \mathbf{a}_{0,i}^\intercal B^\intercal X^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) X e_i$.

We first consider $\tilde{A}_{\xi,i,1}$. From (S5.58),

$$\sigma_{x,i}^{-1} \mathbf{a}_{0,i}^\intercal B^\intercal X^\intercal X e_i = \sigma_{x,i}^{-1} \mathbf{a}_{0,i}^\intercal B^\intercal \Sigma_x^{1/2} \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_x^{1/2} e_i. \tag{S5.73}$$

Note that for fixed $i = 1, \ldots, p$, $\|\Sigma_x^{1/2} e_i\|^2 \sigma_{x,i}^{-2} = 1$. Then there exists $\tilde{Q} \in \mathcal{O}(p)$ such that $\Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} = \tilde{Q} e_1$, and

$$\begin{aligned} & \Sigma_x^{1/2} B \mathbf{a}_{0,i} - \langle \Sigma_x^{1/2} B \mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \\ = \; & (\|\Sigma_x^{1/2} B \mathbf{a}_{0,i}\|^2 - \langle \Sigma_x^{1/2} B \mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle^2)^{1/2} \tilde{Q} e_2. \end{aligned} \tag{S5.74}$$

By Condition 2, there exists constant c such that $\|\Sigma_x^{1/2} B \mathbf{a}_{0,i}\| \leq c n^u$. Thus

$$\Sigma_x^{1/2} B \mathbf{a}_{0,i} = \langle \Sigma_x^{1/2} B \mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle \tilde{Q} e_1 + O(n^u) \tilde{Q} e_2. \tag{S5.75}$$

Let $T_{\eta,1} = \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n)\tilde{U}\tilde{Q}e_1$. As $\tilde{Q}e_1 = \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1}$, it follows that

$$(\text{S5.73}) = \langle \Sigma_x^{1/2} B\mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle e_1^\intercal \tilde{Q}^\intercal T_{\eta,1} + O(n^u) e_2^\intercal \tilde{Q}^\intercal T_{\eta,1}. \tag{S5.76}$$

Since the uniform distribution on the orthogonal group $\mathcal{O}(p)$ is invariant under itself, $\tilde{U}\tilde{Q} \overset{(d)}{=} \tilde{U}$. Then as $(\mu_1, \ldots, \mu_n)^\intercal$ is independent of $\tilde{U}$ by Lemma 14, we know that $\tilde{Q}^\intercal T_{\eta,1} \overset{(d)}{=} \mathbf{R}$, where $\mathbf{R} = (R_1, \ldots, R_p)^\intercal = \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n)\tilde{U}e_1$. By (S5.76), we then have

$$(\text{S5.73}) \overset{(d)}{=} \xi_{i,1} + \xi_{i,2}, \tag{S5.77}$$

where $\xi_{i,1} = \langle \Sigma_x^{1/2} B\mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle R_1$ and $\xi_{i,2} = O(n^u) R_2$.

We next examine $\xi_{i,1}$ and $\xi_{i,2}$ separately. For $\xi_{i,1}$, as $\mu_1, \ldots, \mu_n \geq p\lambda_{\min}(p^{-1}ZZ^\intercal)$, and by (S5.57), we have

$$R_1 \geq p e_1^\intercal \tilde{U}^\intercal \lambda_{\min}(p^{-1}ZZ^\intercal) I_n \tilde{U} e_1 = p\lambda_{\min}(p^{-1}ZZ^\intercal)\langle Se_1, e_1 \rangle.$$

Thus, by Condition 1, Lemmas 15 and 16, and Bonferroni inequality, we have for some positive constants c_1 and c_0,

$$P(R_1 < c_1 p \times n/p) \leq O\{\exp(-c_0 n)\}. \tag{S5.78}$$

This, along with Condition 2, show that for some positive constants c_1 and c_0,

$$P(|\xi_{i,1}| < c_1 n^{1-\kappa}) \leq O\{\exp(-c_0 n)\}. \tag{S5.79}$$

We then consider $\xi_{i,2} = O(n^u)R_2$. By Lemma 17, we know that there exist positive constants c_1 and c_0 such that $P(|R_2| > c_1 n^{1/2}|W_1|) \leq O\{\exp(-c_0 n\}$, where W_1 is an independent $\mathcal{N}(0,1)$-distributed random variable. It follows that for some positive constants c_1 and c_0, we have

$$P(|\xi_{i,2}| > c_1 n^{u+1/2}|W_1|) \leq O\{\exp(-c_0 n)\}. \tag{S5.80}$$

For some constant $c_2 > 0$, let $x_n = \sqrt{2c_2} n^{1-\kappa-u}/\sqrt{\log n}$. Then by the classical Gaussian tail bound, we have

$$P(n^{1/2}|W| > x_n) \leq \sqrt{2/\pi}\frac{\exp\{-c_2 n^{1-2\kappa-2u}/\log n\}}{\sqrt{2c_2} n^{1/2-\kappa-u}/\sqrt{\log n}} \leq O\{\exp(-c_2 n^{1-2\kappa-2u}/\log n)\},$$

which, along with inequality (S5.80), show that for some positive constants c_1 and c_0,

$$P(|\xi_{i,2}| > c_1 n^u x_n) \leq O[\exp\{-c_0 n^{1-2\kappa-2u}/\log n\}], \tag{S5.81}$$

where $n^u x_n = \sqrt{2c_2} n^{1-\kappa}/\sqrt{\log n}$.

We then consider $\tilde{A}_{\xi,i,2}$. Similarly, we take $\tilde{Q} \in \mathcal{O}(p)$ satisfying $\Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} = \tilde{Q} e_1$ and (S5.74). As $Z\tilde{Q} \overset{(d)}{=} Z$ and by (S5.75), similarly to (S5.77), we have

$$\tilde{A}_{\xi,i,2} = \mathbf{a}_{0,i}^\intercal B^\intercal \Sigma_x^{1/2} Z^\intercal (\mathbf{1}_n \mathbf{1}_n^\intercal / n) Z \tilde{Q} e_1 \overset{(d)}{=} \tilde{\xi}_{i,1} + \tilde{\xi}_{i,2},$$

where $\tilde{\xi}_{i,1} = \langle \Sigma_x^{1/2} B \mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle e_1^\intercal Z^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal Z e_1 / n$ and $\tilde{\xi}_{i,2} = O(n^u) e_2^\intercal Z^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal Z e_1 / n$. Note that $\mathbf{1}_n^\intercal Z e_1 / \sqrt{n} \sim \mathcal{N}(0,1)$ and $\mathbf{1}_n^\intercal Z e_2 / \sqrt{n} \sim \mathcal{N}(0,1)$ independently. Then for some positive constants c_1 and c_0,

$$\begin{aligned} P(|e_1^\intercal Z^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal Z e_1 / n| > c_1 n^{1-\kappa-u}/\log n) &\leq O[\exp\{-c_0 n^{1-\kappa-u}/\log n\}], \\ P(|e_2^\intercal Z^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal Z e_1 / n| > c_1 n^{1-\kappa-u}/\log n) &\leq O[\exp\{-c_0 n^{1-\kappa-u}/\log n\}]. \end{aligned}$$

These, combined with (S5.78), show that there exist some constants c_1 and c_0 such that

$$\begin{aligned} P(|\tilde{\xi}_{i,1}| > c_1 |\xi_{i,1}| n^{-u-\kappa}/\log n) &\leq P(|\tilde{\xi}_{i,1}| > c_1 \langle \Sigma_x^{1/2} B \mathbf{a}_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \rangle n^{1-\kappa-u}/\log n) \\ &\quad + P(R_1 < c_1 n) \\ &\leq O\{\exp(-c_0 n^{1-\kappa-u}/\log n)\}, \\ P(|\tilde{\xi}_{i,2}| > c_1 n^{1-\kappa}/\log n) &\leq O\{\exp(-c_0 n^{1-\kappa-u}/\log n)\}. \end{aligned} \tag{S5.82}$$

In summary, by Bonferroni's inequality, combining (S5.71), (S5.72), (S5.77), (S5.79), (S5.81) and (S5.82), we have for some positive constants c_1 and c_0,

$$P(|\xi_i| < c_1(n \times n^{2u+1})^{-1/2} n^{1-\kappa}) \leq O[\exp\{-c_0 n^{1-2\kappa-2u}/\log n\}], \quad i \in \mathcal{M}_*. \tag{S5.83}$$

Step 2 We next examine $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)^\intercal$ defined in (S5.60).

Step 2.1 (bounding $\|\boldsymbol{\eta}\|_2$ from above) By Condition 1,

$$\eta_i = \max_{\mathbf{a}:\|\mathbf{a}\|=1} \frac{\mathbf{a}^\intercal E^\intercal P^\intercal P \mathbf{x}^i}{\sigma_{x,i}\sqrt{n \times \mathbf{a}^\intercal Y^\intercal P^\intercal P Y \mathbf{a}}} \leq \{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1/2} c_3^{-1} \|E^\intercal P^\intercal P \mathbf{x}^i\|. \tag{S5.84}$$

Let $\boldsymbol{\epsilon}^j$ denote the j-th column of E, then $E = [\boldsymbol{\epsilon}^1, \ldots, \boldsymbol{\epsilon}^m]$. As $P^\intercal P = I_n - \mathbf{1}_n \mathbf{1}_n^\intercal/n$, we have $\|E^\intercal P^\intercal P \mathbf{x}^i\|^2 = \sum_{j=1}^m \{(\boldsymbol{\epsilon}^j)^\intercal \mathbf{x}^i - (\boldsymbol{\epsilon}^j)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \mathbf{x}^i/n\}^2$. Note that $\sum_{i=1}^p \mathbf{x}^i (\mathbf{x}^i)^\intercal = XX^\intercal$. Then by (S5.84),

$$\begin{aligned}
\sum_{i=1}^p \eta_i^2 &\leq c_3^{-2}\{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1} \sum_{i=1}^p \sum_{j=1}^m 2 \times [\{(\boldsymbol{\epsilon}^j)^\intercal \mathbf{x}^i\}^2 + \{(\boldsymbol{\epsilon}^j)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \mathbf{x}^i/n\}^2] \\
&= 2c_3^{-2}\{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1} \\
&\quad \times \sum_{j=1}^m \{(\boldsymbol{\epsilon}^j)^\intercal \sum_{i=1}^p \mathbf{x}^i (\mathbf{x}^i)^\intercal \boldsymbol{\epsilon}^j + (\boldsymbol{\epsilon}^j)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \sum_{i=1}^p \mathbf{x}^i (\mathbf{x}^i)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \boldsymbol{\epsilon}^j/n^2\} \\
&= 2c_3^{-2}\{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1} \sum_{j=1}^m \{(\boldsymbol{\epsilon}^j)^\intercal XX^\intercal \boldsymbol{\epsilon}^j + (\boldsymbol{\epsilon}^j)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal XX^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \boldsymbol{\epsilon}^j/n^2\} \\
&\leq 2c_3^{-2}\{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1} \lambda_{\max}(\Sigma_x) \sum_{j=1}^m (A_{\eta,j,1} + A_{\eta,j,2}),
\end{aligned} \tag{S5.85}$$

where $A_{\eta,j,1} = (\boldsymbol{\epsilon}^j)^\intercal ZZ^\intercal \boldsymbol{\epsilon}^j$ and $A_{\eta,j,2} = (\boldsymbol{\epsilon}^j)^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal ZZ^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \boldsymbol{\epsilon}^j / n^2$.

Note that $A_{\eta,j,1} \leq p\lambda_{\max}(p^{-1}ZZ^\intercal)\|\boldsymbol{\epsilon}^j\|^2$. Suppose $\mathrm{diag}(\Sigma) = (\sigma_{\epsilon,1}^2, \ldots, \sigma_{\epsilon,m}^2)^\intercal$. Then by Condition 1 and Lemma 13, we know for some positive constants c and c_0,

$$P(\|\boldsymbol{\epsilon}^j\|_2^2 > cn\sigma_{\epsilon,j}^2/\log n) \leq \exp(-c_0 n/\log n). \tag{S5.86}$$

In addition, $A_{\eta,j,2} \leq p\lambda_{\max}(p^{-1}ZZ^\intercal) \times (\mathbf{1}_n^\intercal \boldsymbol{\epsilon}^j)^2/n$. Similarly to (S5.86), by Condition 1 and the tail bound of the Chi-squared distribution, there exist some positive constants c and c_0,

$$P\{(\mathbf{1}_n^\intercal \boldsymbol{\epsilon}^j)^2/n > cn\sigma_{\epsilon,j}^2/\log n\} \leq O\{\exp(-c_0 n/\log n)\}. \tag{S5.87}$$

Combining (S5.86) and (S5.87), we know that for some constants c_1, c_2 and c_0, with probability $1 - O\{m\exp(-c_0 n/\log n)\}$,

$$A_{\eta,j,1} + A_{\eta,j,2} \leq c_1 pn \sum_{j=1}^{m} \sigma_{\epsilon,j}^2/\log n \leq c_2 pnm/\log n, \tag{S5.88}$$

where the last inequality is from $\mathrm{diag}(\Sigma) \leq c_4$ for some constant $c_4 > 0$ by Condition 1.

Combining (S5.85), (S5.88), Lemma 19 and Conditions 1 and 3, we know for

some positive constants c_1, c_2 and c_0,

$$\begin{aligned} P(\|\boldsymbol{\eta}\|^2 > c_1\{n \times n^{1-t}\}^{-1} p n^{1+\tau} m/\log n) &\leq O\{m \exp(-c_2 n/\log n)\} \\ &= O\{\exp(-c_0 n/\log n)\}. \end{aligned} \tag{S5.89}$$

Step 2.2 (bounding $|\eta_i|$ from above) From Step 2.1, we know

$$\eta_i^2 \leq \{n\lambda_{\min}(Y^\intercal P^\intercal P Y)\}^{-1} \sigma_{x,i}^{-2} \sum_{j=1}^{m} (\boldsymbol{\epsilon}_j^\intercal P^\intercal P \mathbf{x}^i)^2. \tag{S5.90}$$

Then conditioning on X, $\sigma_{x,i}^{-1} \boldsymbol{\epsilon}_j^\intercal P^\intercal P \mathbf{x}^i \sim \mathcal{N}(0, \sigma_{\epsilon,j}^2 (\mathbf{x}^i)^\intercal P^\intercal P \mathbf{x}^i \sigma_{x,i}^{-2})$. Let $\mathcal{E}_1$ be the event $\{\mathrm{var}(\sigma_{x,i}^{-1} \boldsymbol{\epsilon}_j^\intercal P^\intercal P \mathbf{x}^i | X) \leq c_1 n\}$ for some constant $c_1 > 0$. Note that

$$\mathrm{var}(\boldsymbol{\epsilon}_{\epsilon,j}^\intercal P^\intercal P \mathbf{x}^i \sigma_{x,i}^{-1} | X) = \sigma_{\epsilon,j}^2 \{(\mathbf{x}^i)^\intercal \mathbf{x}^i - (\mathbf{x}^i)^\intercal \mathbf{1}_n^\intercal \mathbf{1}_n \mathbf{x}^i/n\} \sigma_{x,i}^{-2} \leq \sigma_{\epsilon,j}^2 (\mathbf{x}^i)^\intercal \mathbf{x}^i \sigma_{x,i}^{-2}.$$

Using the same argument as in Step 1.1, we can show that, there exist some positive constants c_1 and c_0,

$$P(\mathcal{E}_1^c) \leq P\{\sigma_{\epsilon,j}^2 (\mathbf{x}^i)^\intercal \mathbf{x}^i \sigma_{x,i}^{-2} > c_1 n\} \leq O\{\exp(-c_0 n)\}, \tag{S5.91}$$

where $\mathcal{E}_1^c$ represents the complement of the event $\mathcal{E}_1$. On the event $\mathcal{E}_1$, for any $a > 0$, by Condition 1, we have

$$P(|\boldsymbol{\epsilon}_j^\intercal P^\intercal P\mathbf{x}^i|\sigma_{x,i}^{-1} > a|X, \mathcal{E}_1) \leq P\{\sqrt{c_1 n}|W| > a\}, \tag{S5.92}$$

where W is an independent $\mathcal{N}(0,1)$-distributed random variable. Thus, combining (S5.91) and (S5.92), we have

$$P(|\boldsymbol{\epsilon}_j^\intercal P^\intercal P\mathbf{x}^i|\sigma_{x,i}^{-1} > a) \leq O\{\exp(-c_0 n)\} + P\{\sqrt{c_1 n}|W| > a\}. \tag{S5.93}$$

Let $x_n' = \sqrt{2c_0c_1}n^{1-\kappa-t/2-s/2-u}/\sqrt{\log n}$. Invoking the classical Gaussian tail bound, we have

$$P\{\sqrt{c_1 n}|W| > x_n'\} = O\{\exp(-c_0 n^{1-2\kappa-t-s-2u}/\log n)\}.$$

By (S5.90) and Lemma 19, we then have

$$\begin{aligned} P(|\eta_i| > (n^{1+1-t})^{-1/2}x_n'\sqrt{m}) \;\; \leq \;\; & \sum_{j=1}^{m} P(|\boldsymbol{\epsilon}_j^\intercal P^\intercal P\mathbf{x}^i| > x_n') \\ & + P\{\lambda_{\min}(Y^\intercal P^\intercal PY) < c_1 n^{1-t}\}, \end{aligned}$$

where $(n^{1+1-t})^{-1/2}x_n'\sqrt{m} = \sqrt{2c_0c_1m}n^{-\kappa-s/2-u}/\sqrt{\log n} \leq c_2 n^{-\kappa-u}/\sqrt{\log n}$ for some

constant $c_2 > 0$ by Condition 3. In summary, we have

$$\begin{aligned} & P[\max_{1\le i\le p} |\eta_i| > c_2 n^{-\kappa-u}/\sqrt{\log n}] \\ \le\ & O[p\exp\{-c_2 n^{1-2\kappa-t-s-2u}/\log n\}] \\ =\ & O[\exp\{-c_0 n^{1-2\kappa-t-s-2u}/\log n\}], \end{aligned} \tag{S5.94}$$

where the last equality is from Condition 3.

Step 3. We combine the results in Steps 1 and 2. By Bonferroni's inequlaity, it follows from (S5.70), (S5.83), (S5.89) and (S5.94) that, for some positive constants $\tilde{c}_1, \tilde{c}_2$ and $\tilde{c}$,

$$\begin{aligned} & P\{\min_{i\in\mathcal{M}_*} |\zeta_i| < \tilde{c}_1(n\times n^{2u+1})^{-1/2} n^{1-\kappa} \text{ or } \|\boldsymbol{\zeta}\|^2 > \tilde{c}_2(n\times n^{1-t})^{-1} n^{1+\tau} pm\} \\ \le\ & O[|\mathcal{M}_*|\exp\{-\tilde{c} n^{1-2\kappa-2u-t-s}/\log n\}]. \end{aligned} \tag{S5.95}$$

By Lemma 18 and (S5.95), we know that there exist some positive constants c_1, c_2 and c,

$$\begin{aligned} & P\{\min_{i\in\mathcal{M}_*} |\omega_i| < c_1(n\times n^{2u+1})^{-1/2} n^{1-\kappa} \text{ or } \|\boldsymbol{\omega}\|^2 > c_2(n\times n^{-t})^{-1} n^{1+\tau} pm\} \\ \le\ & O[|\mathcal{M}_*|\exp\{-c n^{1-2\kappa-2u-t-s}/\log n\}], \end{aligned}$$

which is $O[\exp\{-c_0 n^{1-2\kappa-2u-t-s}/\log n\}]$ for some constant $c_0 > 0$ by Condition 3 and $|\mathcal{M}_*| \leq p$. This shows that with overIwhelming probability $1-O[\exp\{-c_0 n^{1-2\kappa-2u-t-s}/\log n\}]$, the magnitudes of $\omega_i, i \in \mathcal{M}_*$ are uniformly at least of order $c_1(n \times n^{2u+1})^{-1/2} n^{1-\kappa}$, and for some positive constants c_1 and c_2,

$$\#\{1 \leq k \leq p : |\omega_k| \geq \min_{i \in \mathcal{M}_*} |\omega_i|\} \leq c_1 \frac{(n \times n^{2u+1}) \times pmn^{1+\tau}}{n \times n^{1-t} \times (n^{1-\kappa})^2} \leq c_2 p n^{s+2u+2\kappa+\tau+t-1},$$

where the last inequality is from Condition 3. Thus, if the proportion δ of features selected satisfies $\delta n^{1-2\kappa-2u-\tau-t-s} \to \infty$, then $\delta p \geq c_2 p n^{s+\tau+t+2\kappa+2u-1}$ when δ is sufficiently large, and we know with probability $1 - O[\exp\{-c_0 n^{1-2\kappa-2u-\tau-t-s}/\log n\}]$, $\mathcal{M}_* \subseteq \mathcal{M}_\delta$.

S5.3 Lemmas in the proof of Theorem 5

Lemma 13 (Lemma 3 in Fan and Lv (2008))**.** *Let ϑ_i, $i = 1, 2, \ldots, n$ be i.i.d. χ_1^2-distributed random variables. Then for any $\epsilon > 0$, we have $P(n^{-1}\sum_{i=1}^n \vartheta_i > 1+\epsilon) \leq \exp(-A_\epsilon n)$, where $A_\epsilon = [\epsilon - \log(1+\epsilon)]/2 > 0$; for any $\epsilon \in (0,1)$, $P(n^{-1}\sum_{i=1}^n \vartheta_i < 1-\epsilon) \leq \exp(-B_\epsilon n)$,where $B_\epsilon = [-\epsilon - \log(1-\epsilon)]/2 > 0$.*

Lemma 14 (Lemma 1 in Fan and Lv (2008))**.** *For U and $(\mu_1, \ldots, \mu_n)^\intercal$ in* (S5.56),

and $\tilde{O}$ uniformly distributed on the orthogonal group $\mathcal{O}(p)$, we know that

$$(I_n, \mathbf{0})_{n\times p} U \overset{(d)}{=} (I_n, \mathbf{0})_{n\times p} \tilde{O}, \tag{S5.96}$$

and $(\mu_1, \ldots, \mu_n)^\intercal$ is independent of $(I_n, \mathbf{0})_{n\times p} U$.

Proof. As $\mu_1^{1/2}, \ldots, \mu_n^{1/2}$ are n singular values of Z, we know that Z has the singular value decomposition $Z = V_1 D_1 U$, where $V_1 \in \mathcal{O}(n)$, $U \in \mathcal{O}(p)$ is given in (S5.56), and D_1 is an $n \times p$ diagonal matrix whose diagonal elements are $\mu_1^{1/2}, \ldots, \mu_n^{1/2}$. Since the entries in Z are i.i.d. $\mathcal{N}(0,1)$, for any $\tilde{O} \in \mathcal{O}(p)$, $Z\tilde{O} \overset{(d)}{=} Z$. Thus, conditional on V_1 and $(\mu_1, \ldots, \mu_n)^\intercal$, the conditional distribution of $(I_n, \mathbf{0})_{n\times p} U$ is invariant under $\mathcal{O}(p)$. This shows that (S5.96) holds for $\tilde{O}$ uniformly distributed on the orthogonal group $\mathcal{O}(p)$, and $(\mu_1, \ldots, \mu_n)^\intercal$ is independent of $(I_n, \mathbf{0})_{n\times p} U$. □

Lemma 15 (Lemma 4 in Fan and Lv (2008))**.** *S defined in* (S5.57) *is uniformly distributed on the Grassmann manifold $\mathcal{G}_{p,n}$. For any constant $c_0 > 0$, there are constants c_1 and c_2 with $0 < c_1 < 1 < c_2$ such that*

$$P(\langle Se_1, e_1\rangle < c_1 n/p \text{ or } > c_2 n/p) \le 4\exp(-c_0 n).$$

Lemma 16. *The matrix Z is of size $n \times p$ and the matrix $\tilde{Z}$ is of size $n \times m$ with Condition 3 satisfied. The entries in Z and $\tilde{Z}$ are i.i.d. $\mathcal{N}(0,1)$. For some constants*

$c_1, c_0 > 0$,

$$P\{\lambda_{\max}(n^{-1}ZZ^\intercal) > c_1 \text{ or } \lambda_{\min}(n^{-1}Z^\intercal Z) < 1/c_1\} \leq \exp(-c_0 n). \tag{S5.97}$$

There exist some constants $c_1 > 1$, $c > 0$ *and* $c_0 > 0$, *when* $n > c$,

$$P[\lambda_{\max}\{n^{-1}(P\tilde{Z})^\intercal P\tilde{Z}\} > c_1 \text{ or } \lambda_{\min}\{n^{-1}(P\tilde{Z})^\intercal P\tilde{Z}\} < 1/c_1] \leq \exp(-c_0 n/\log n). \tag{S5.98}$$

Proof. As the entries in Z are i.i.d. $\mathcal{N}(0,1)$, by Appendix A.7 in Fan and Lv (2008), we know that (S5.97) holds. For $\tilde{Z}$, since its entries are also i.i.d. $\mathcal{N}(0,1)$ and $n > c_7 m$ for some $c_7 > 1$, symmetrically, we know there exist constants $\tilde{c}_1 > 1$ and $\tilde{c}_0 > 0$ such that

$$P\{\lambda_{\max}(n^{-1}\tilde{Z}^\intercal\tilde{Z}) > \tilde{c}_1 \text{ or } \lambda_{\min}(n^{-1}\tilde{Z}^\intercal\tilde{Z}) < 1/\tilde{c}_1\} \leq \exp(-\tilde{c}_0 n). \tag{S5.99}$$

Since $(P\tilde{Z})^\intercal P\tilde{Z} = \tilde{Z}^\intercal P\tilde{Z} = \tilde{Z}^\intercal\tilde{Z} - \tilde{Z}^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \tilde{Z}/n$, by Weyl's inequality, we have

$$\begin{aligned} \lambda_{\max}\{(P\tilde{Z})^\intercal P\tilde{Z}\} &\leq \lambda_{\max}(\tilde{Z}^\intercal\tilde{Z}) + \lambda_{\max}(-\tilde{Z}^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \tilde{Z}/n), \\ \lambda_{\min}\{(P\tilde{Z})^\intercal P\tilde{Z}\} &\geq \lambda_{\min}(\tilde{Z}^\intercal\tilde{Z}) + \lambda_{\min}(-\tilde{Z}^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \tilde{Z}/n). \end{aligned} \tag{S5.100}$$

Let $A_Z = \tilde{Z}^\intercal \mathbf{1}_n \mathbf{1}_n^\intercal \tilde{Z}/n$. As $\text{rank}(A_Z) = 1$ and $\text{tr}(A_Z) \geq 0$, we know $\lambda_{\max}(-A_Z) =$

$-\lambda_{\min}(A_Z) = 0$ and $\lambda_{\min}(-A_Z) = -\lambda_{\max}(A_Z) = -\mathrm{tr}(A_Z) = -\mathbf{1}_n^\intercal \tilde{Z}\tilde{Z}^\intercal \mathbf{1}_n/n$.

We then examine $\mathbf{1}_n^\intercal \tilde{Z}\tilde{Z}^\intercal \mathbf{1}_n/n$. For $z_{ij} \sim \mathcal{N}(0,1)$ independently,

$$\mathbf{1}_n^\intercal \tilde{Z}\tilde{Z}^\intercal \mathbf{1}_n/n = \sum_{j=1}^{m} \Big(\sum_{i=1}^{n} z_{ij}/\sqrt{n} \Big)^2 \sim \chi_m^2.$$

By Lemma 13, we know for the random variable $\tilde{W} \sim \chi_m^2$ and any constant $c_2 > 0$, there exists constant $c_3 > 0$ such that

$$P\{|\tilde{W}/m - 1| > c_2 n/(m \log n)\} \leq \exp\{-c_3 m \times n/(m \log n)\}, \tag{S5.101}$$

This implies that with probability $1 - O\{\exp(-c_3 n/\log n)\}$, $\lambda_{\max}(A_Z/n) = \mathbf{1}_n^\intercal \tilde{Z}\tilde{Z}^\intercal \mathbf{1}_n/n^2 \leq c_2/\log n$ for some constant $c_2 > 0$, as $m = O(n^s)$ with $s \in [0,1)$.

When n is sufficiently large, there exists constant c_1 such that $1 > c_1 > \tilde{c}_1$ and $1/c_1 + c_2/\log n < 1/\tilde{c}_1$. Thus by (S5.100) and (S5.101), we know there exists constant $c_0 > 0$ such that with probability $1 - \exp(-c_0 n/\log n)$,

$$\begin{aligned}
\{\lambda_{\min}(n^{-1}(P\tilde{Z})^\intercal P\tilde{Z}) < 1/c_1\} &\subseteq \{\lambda_{\min}(n^{-1}\tilde{Z}^\intercal \tilde{Z}) < 1/c_1 + c_2/\log n\} \\
&\subseteq \{\lambda_{\min}(n^{-1}\tilde{Z}^\intercal \tilde{Z}) < 1/\tilde{c}_1\}, \\
\{\lambda_{\max}(n^{-1}(P\tilde{Z})^\intercal P\tilde{Z}) > c_1\} &\subseteq \{\lambda_{\max}(n^{-1}\tilde{Z}^\intercal \tilde{Z}) > \tilde{c}_1\}.
\end{aligned}$$

By (S5.99), (S5.98) is then proved.

□

Lemma 17. *For* $\mathbf{R} = (R_1, \ldots, R_p)^\intercal = \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n)\tilde{U}e_1$*, there exist positive constants* c_1 *and* c_0 *such that* $P(|R_2| > c_1 n^{1/2}|W_1|) \leq O\{\exp(-c_0 n\}$*, where* $|W_1|$ *is an independent* $\mathcal{N}(0,1)$*-distributed random variable.*

Proof. Let $\tilde{\mathbf{R}} = (R_2, \ldots, R_p)^\intercal$. We first show that $\tilde{\mathbf{R}}$ is invariant under the orthogonal group $\mathcal{O}(p-1)$. For any $Q \in \mathcal{O}(p-1)$, let $\tilde{Q} = \mathrm{diag}(1, Q) \in \mathcal{O}(p)$. By Lemma 14, we know that $\tilde{U}$ is independent of $\mathrm{diag}(\mu_1, \ldots, \mu_n)$ and $\tilde{Q}\tilde{U} \overset{(d)}{=} \tilde{U}$. Thus

$$\tilde{Q}^\intercal \mathbf{R} = \tilde{Q}^\intercal \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n)\tilde{U}\tilde{Q}\tilde{Q}^\intercal e_1 \overset{(d)}{=} \tilde{U}^\intercal \mathrm{diag}(\mu_1, \ldots, \mu_n)\tilde{U}e_1,$$

where we use the fact that $\tilde{Q}^\intercal e_1 = e_1$. This implies that $\tilde{\mathbf{R}}$ is invariant under the orthogonal group $\mathcal{O}(p-1)$. It follows that $\tilde{\mathbf{R}} \overset{(d)}{=} \|\tilde{\mathbf{R}}\|\mathbf{W}/\|\mathbf{W}\|_2$, where $\mathbf{W} = (W_1, \ldots, W_{p-1})^\intercal \sim \mathcal{N}(0, I_{p-1})$, independent of $\|\tilde{\mathbf{R}}\|$.

In particular, we have $R_2 \overset{(d)}{=} \|\tilde{\mathbf{R}}\|W_1/\|\mathbf{W}\|_2$. Note that $\|\tilde{\mathbf{R}}\| \leq \|\mathbf{R}\|$ and $\|\mathbf{R}\|^2 = e_1^\intercal \tilde{U}^\intercal \mathrm{diag}(\mu_1^2, \ldots, \mu_n^2)\tilde{U}e_1$. Since $\mu_1, \ldots, \mu_n \leq p\lambda_{\max}(p^{-1}ZZ^\intercal)$,

$$\|\mathbf{R}\|^2 \leq \{\lambda_{\max}(p^{-1}ZZ^\intercal)\}^2 p^2 e_1^\intercal \tilde{U}^\intercal \tilde{U}e_1 = \{\lambda_{\max}(p^{-1}ZZ^\intercal)\}^2 p^2 \langle Se_1, e_1\rangle.$$

By Lemmas 14 and 16, we then know for some positive constants c_1 and c_0, $P(\|\mathbf{R}\| > c_1\sqrt{pn}) \leq O\{\exp(-c_0 n)\}$. Moreover, by Lemma 13, we know for some constants c_1

and c_0, $P\{\|\mathbf{W}\|^2 < c_1(p-1)\} \le \exp(-c_0 n)$ when $p > n$. Thus we obtain that for some constants c_1 and c_0, $P(|R_2| > c_1|W_1|n^{1/2}) \le O\{\exp(-c_0 n)\}$, where W_1 is an independent $\mathcal{N}(0,1)$-distributed random variable. □

Lemma 18. *For some constant $c_0 > 0$, $\omega_i = \zeta_i\{1+o(1)\}$ with probability $1 - O\{\exp(-c_0 n/\log n)\}$.*

Proof. By the definitions of ω_i and ζ_i and $P^\intercal P = P$, we know

$$\begin{aligned}
\omega_i - \zeta_i &\le \max_{\mathbf{a}} \frac{\mathbf{a}^\intercal Y^\intercal P\mathbf{x}^i(\sqrt{n}\sigma_{x,i} - \sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i})}{\sigma_{x,i}\sqrt{n((\mathbf{x}^i)^\intercal P\mathbf{x}^i)(\mathbf{a}^\intercal Y^\intercal PY\mathbf{a})}} = \zeta_i \frac{\sqrt{n}\sigma_{x,i} - \sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i}}{\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i}}, \\
\zeta_i - \omega_i &\le \max_{\mathbf{a}} \frac{\mathbf{a}^\intercal Y^\intercal P^\intercal\mathbf{x}^i(\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i} - \sqrt{n}\sigma_{x,i})}{\sigma_{x,i}\sqrt{n((\mathbf{x}^i)^\intercal P\mathbf{x}^i)(\mathbf{a}^\intercal Y^\intercal PY\mathbf{a})}} = \zeta_i \frac{\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i} - \sqrt{n}\sigma_{x,i}}{\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i}}.
\end{aligned}$$

Thus

$$|\omega_i - \zeta_i| \le \zeta_i \left| \frac{\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i} - \sqrt{n}\sigma_{x,i}}{\sqrt{(\mathbf{x}^i)^\intercal P\mathbf{x}^i}} \right| = \zeta_i \Big| 1 - \sigma_{x,i}\sqrt{n/\{(\mathbf{x}^i)^\intercal P\mathbf{x}^i\}} \Big|.$$

Let $\bar{x}^i = \sum_{k=1}^n x_{k,i}/n$, which is the mean of the entries in $\mathbf{x}^i$. It follows that $(\mathbf{x}^i)^\intercal P\mathbf{x}^i = \sum_{k=1}^n (x_{k,i} - \bar{x}^i)^2 = \sum_{k=1}^n x_{k,i}^2 - n(\bar{x}^i)^2$. Then with $\tilde{c}_1 = c_1/2$

$$\begin{aligned}
& P(|(\mathbf{x}^i)^\intercal P\mathbf{x}^i/(n\sigma_{x,i}^2) - 1| > c_1/\log n) \qquad \text{(S5.102)} \\
\le\ & P\{(\bar{x}^i)^2/\sigma_{x,i}^2 > \tilde{c}_1/\log n\} + P\Big\{\Big|\sum_{k=1}^n x_{k,i}^2/(n\sigma_{x,i}^2) - 1\Big| > \tilde{c}_1/\log n\Big\}.
\end{aligned}$$

By Condition 1, $(\sqrt{n}\bar{x}^i)^2/\sigma_{x,i}^2 \sim \chi_1^2$. Then by the tail of χ_1^2 distribution, for some constant $c_0 > 0$, $P\{(\bar{x}^i)^2/\sigma_{x,i}^2 > \tilde{c}_1/\log n\} \leq O\{\exp(-c_0 n/\log n)\}$. In addition, $x_{k,i}^2/\sigma_{x,i}^2$, $k = 1, \ldots, n$ are i.i.d. χ_1^2-distributed random variables. By Lemma 13, there exists some constant $c_0 > 0$,

$$P\Big\{\Big|\sum_{k=1}^{n} x_{k,i}^2/(n\sigma_{x,i}^2) - 1\Big| > \tilde{c}_1/\log n\Big\} \leq O\{\exp(c_0 n/\log n)\}.$$

In summary, we know for any constant $c_1 > 0$, there exists constant $c_0 > 0$ such that (S5.102) $\leq O\{\exp(-c_0 n/\log n)\}$. Thus, for $i = 1, \ldots, p$, $\omega_i = \zeta_i\{1 + O(1/\sqrt{\log n})\} = \zeta_i(1 + o(1))$ with probability $1 - O\{p\exp(-c_0 n/\log n)\} = 1 - O\{\exp(-c_0 n/\log n)\}$, where the last equality is from Condition 3. □

Lemma 19. *Consider $n > c$ for c in Lemma 16. There exist constants c_1, c_2 and c_0, with probability $1 - O\{\exp(-c_0 n/\log n)\}$,*

$$\lambda_{\min}(Y^\intercal P^\intercal P Y) \geq c_1 n^{1-t},$$

and for $\mathbf{a}_{0,i}$ in Condition 2,

$$\mathbf{a}_{0,i}^\intercal Y^\intercal P^\intercal P Y \mathbf{a}_{0,i} \leq c_2 n^{2u+1}. \tag{S5.103}$$

Proof. Since X and E follow independent Gaussian distributions by Condition 1,

the rows of Y are independent multivariate Gaussian with mean zero and covariance $\Sigma_y = B^\intercal \Sigma_x B + \Sigma$. Define $\tilde{Z} = Y\Sigma_y^{-1/2}$. Then $\tilde{Z}$ is of size $n \times m$ and the entries in $\tilde{Z}$ are i.i.d. $\mathcal{N}(0,1)$. Thus the concentration inequality (S5.98) in Lemma 16 holds. It follows that there exist constants c_1 and c_0, with probability $1 - O\{\exp(-c_0 n)\}$,

$$\begin{aligned} Y^\intercal P^\intercal P Y =& \Sigma_y^{1/2} \tilde{Z}^\intercal P^\intercal P \tilde{Z} \Sigma_y^{1/2} \\ \succeq& n\Sigma_y^{1/2} \lambda_{\min}(n^{-1}\tilde{Z}^\intercal P^\intercal P \tilde{Z}) I_m \Sigma_y^{1/2} \succeq c_1 n \Sigma_y. \end{aligned}$$

By Weyl's inequality and Condition 1, we then know

$$\lambda_{\min}(Y^\intercal P^\intercal P Y) \geq c_1 n \lambda_{\min}(\Sigma_y) \geq c_1 n \lambda_{\min}(\Sigma) \geq c_1 n^{1-t}. \tag{S5.104}$$

Similarly we know for some constant c_2, with probability $1 - O\{\exp(-c_0 n)\}$,

$$\begin{aligned} \mathbf{a}_{0,i}^\intercal Y^\intercal P^\intercal P Y \mathbf{a}_{0,i} &= \mathbf{a}_{0,i}^\intercal \Sigma_y^{1/2} \tilde{Z}^\intercal P^\intercal P \tilde{Z} \Sigma_y^{1/2} \mathbf{a}_{0,i} \\ &\leq c_2 n \mathbf{a}_{0,i}^\intercal \Sigma_y \mathbf{a}_{0,i} \\ &= c_2 n \mathbf{a}_{0,i}^\intercal (B^\intercal \Sigma_x B + \Sigma) \mathbf{a}_{0,i} \\ &\leq c_2 n^{2u+1}, \end{aligned}$$

where the last inequality is from Condition 2. □

S6. Proposition 6 (Meinshausen et al., 2009, Theorem 3.2)

The proof of Proposition 6 directly follows the proof in Meinshausen et al. (2009). For $z \in (0,1)$, define

$$\psi(z) = \frac{1}{J}\sum_{j=1}^{J} 1\{p^{(j)} \le z\}. \tag{S6.105}$$

Note that $\{Q(\gamma) \le \alpha\}$ and $\{\psi(\alpha\gamma) \ge \gamma\}$ are equivalent. For a random variable U taking values in $[0,1]$,

$$\sup_{\gamma\in(\gamma_{\min},1)} \frac{1\{U \le \alpha\gamma\}}{\gamma} = \begin{cases} 0 & U \ge \alpha \\ \alpha/U & \alpha\gamma_{\min} \le U < \alpha \\ 1/\gamma_{\min} & U < \alpha\gamma_{\min}. \end{cases}$$

Thus when U has a uniform distribution on $[0,1]$,

$$\mathrm{E}\Big[\sup_{\gamma\in(\gamma_{\min},1)} \frac{1\{U \le \alpha\gamma\}}{\gamma}\Big] = \int_0^{\alpha\gamma_{\min}} \gamma_{\min}^{-1} dx + \int_{\alpha\gamma_{\min}}^{\alpha} \alpha x^{-1} dx = \alpha(1 - \log\gamma_{\min}).$$

Hence, define the event $\mathcal{B}_{(j)}$ as $\mathcal{M}_* \subseteq \mathcal{M}_\delta$ for the jth split, then

$$\begin{aligned} \mathrm{E}\Big[\sup_{\gamma\in(\gamma_{\min},1)} 1\{p^{(j)} \le \alpha\gamma\}/\gamma\Big] &\le \mathrm{E}\Big\{\mathrm{E}\Big[\sup_{\gamma\in(\gamma_{\min},1)} 1\{p^{(j)} \le \alpha\gamma\}/\gamma \Big| \mathcal{B}^{(j)}\Big]\Big\} + \frac{1}{\gamma_{\min}} P\{\mathcal{B}^{(j)}\} \\ &\le \alpha(1 - \log\gamma_{\min}) + O[\exp\{-c_0 n^{1-\iota}/\log n\}], \end{aligned}$$

where the constant ι is given in Theorem 5. Averaging over J splits yields

$$\mathrm{E}\Big[\sup_{\gamma\in(\gamma_{\min},1)}\frac{1}{\gamma}\frac{1}{J}\sum_{j=1}^{J}1\{p^{(j)}/\gamma\le\alpha\}\Big]\le\alpha(1-\log\gamma_{\min})+O[\exp\{-c_0n^{1-\iota}/\log n\}].$$

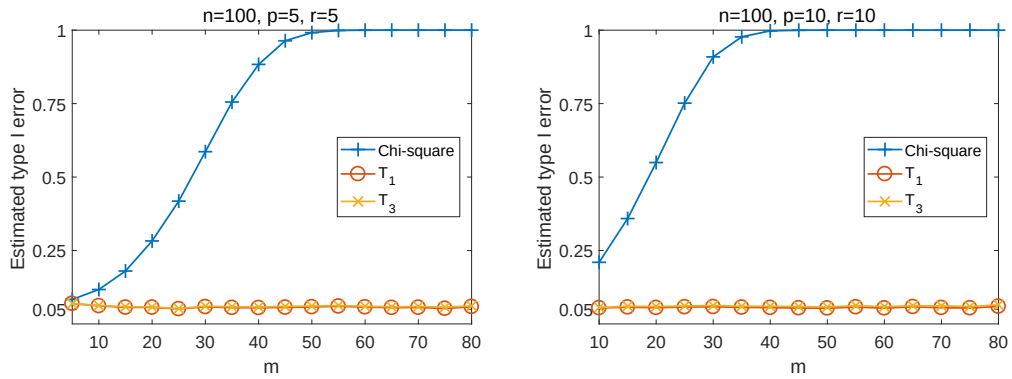
From Markov inequality and (S6.105), $\mathrm{E}[\sup_{\gamma\in(\gamma_{\min},1)}1\{\psi(\alpha\gamma)\ge\gamma\}]\le\alpha(1-\log\gamma_{\min})+O[\exp\{-c_0n^{1-\iota}/\log n\}]$. Since $\{Q(\gamma)\le\alpha\}$ and $\{\psi(\alpha\gamma)\ge\gamma\}$ are equivalent, it follows that $P[\inf_{\gamma\in(\gamma_{\min},1)}Q(\gamma)\le\alpha]\le\alpha(1-\log\gamma_{\min})+O[\exp\{-c_0n^{1-\iota}/\log n\}]$, which implies that $P[\inf_{\gamma\in(\gamma_{\min},1)}Q(\gamma)(1-\log\gamma_{\min})\le\alpha]\le\alpha+O[\exp\{-c_0n^{1-\iota}/\log n\}]$. By definition of p_t, $\limsup_{n\to\infty}P[p_t\le\alpha]\le\alpha$ is obtained.

S7. Supplementary Simulations

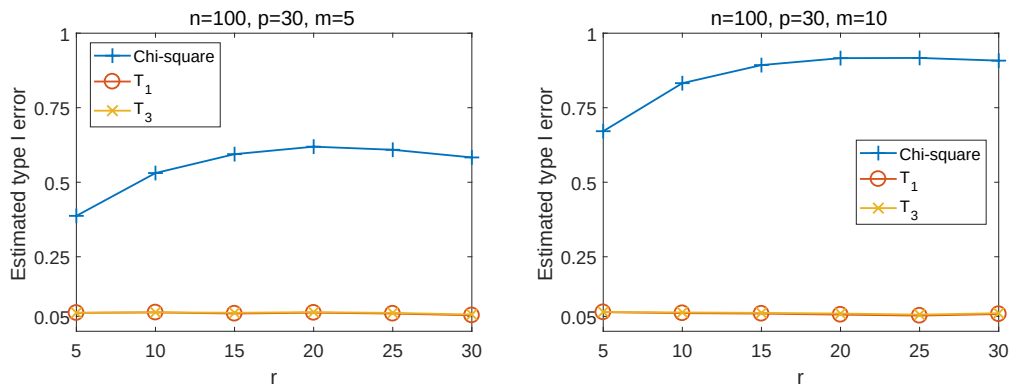
S7.1 Supplementary simulations when $n > p + m$

Estimated type I errors

We provide additional simulations under H_0 following the same set-up as in Figure 4. In Figure S1, we present the estimated type I errors of the χ^2 approximation and the normal approximations of T_1 and T_3 with varying m and r respectively. It exhibits similar pattern as in Figure 4, which shows that as (p, m, r) become larger, the χ^2 approximation performs poorly, while the normal approximations for T_1 and T_3 still control the type I error well.



(a) Estimated type I error versus m



(b) Estimated type I error versus r

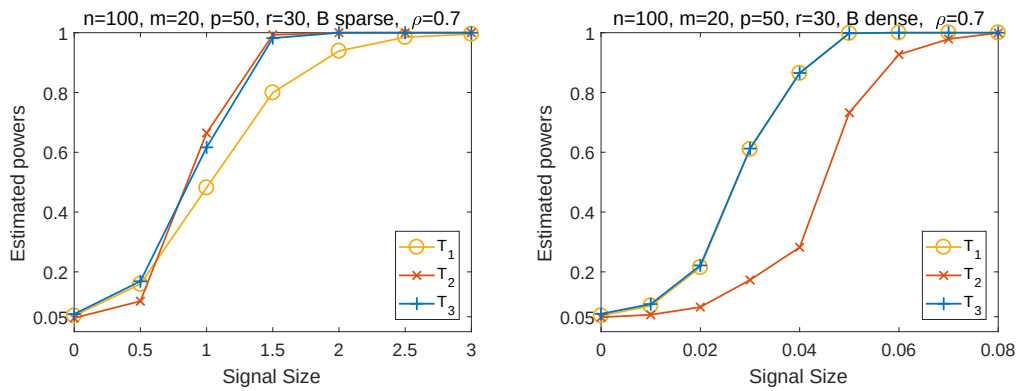
Figure S1: Estimated type I error

Additional simulations under alternative hypotheses

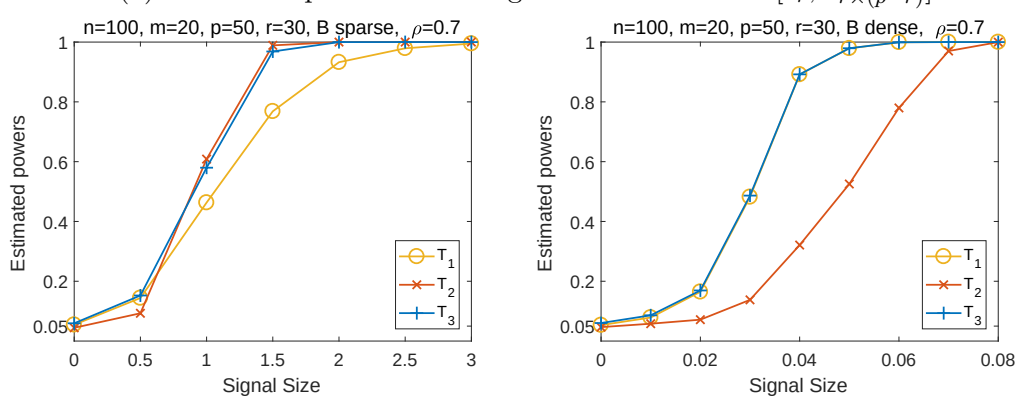
In this section, we generate data from the multivariate regression model $Y = XB + E$, where the rows of X and E are independent multivariate Gaussian with covariance matrices $\Sigma_x = (\rho^{|i-j|})_{p\times p}$ and $\Sigma = (\rho^{|i-j|})_{m\times m}$ respectively. We consider a sparse scenario when only the $(1,1)$-entry of B is nonzero with a value v_d. We also consider a dense scenario when all the entries of B are independently generated from $\mathcal{N}(0, \sigma_d^2)$.

For each scenario, we estimate the test powers for different v_d or σ_d^2 values, which are referred to as the signal sizes in the following. We take $n = 100, m = 20, p = 50, r = 30$ and conduct 10,000 simulations for two different C matrices. In the first case, we take $C = [I_r, \mathbf{0}_{r\times(p-r)}]$, where I_r is an identity matrix of dimension $r \times r$, $\mathbf{0}_{r\times(p-r)}$ is an all zero matrix of dimension $r \times (p-r)$. Then $H_0 : CB = \mathbf{0}_{r\times m}$ examines the relationship between Y and the first r predictors of X. In the second case, we take $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$, where $\mathbf{1}_r$ is an all 1 vector of length r, and $\mathbf{0}_{r\times(p-r-1)}$ is an all zero matrix of dimension $r \times (p-r-1)$. Then $H_0 : CB = \mathbf{0}_{r\times m}$ tests the equivalence of effects of the first r predictors and the last predictor. For two types of B and two types of C matrices, we plot the estimated powers of T_1, T_2, T_3 versus signal sizes with $\rho = 0.7$, $\rho = 0.5$ and $\rho = 0$ in Figures S2, S3 and S4 respectively, where similar results are observed.

Figures S2–S4 show that under the dense B scenario, T_1 is more powerful than T_2; but under the sparse B scenario, T_2 is more powerful than T_1. In addition, the combined statistic T_3 still maintains high power under both scenarios. These results demonstrate the good performance of the proposed statistic T_3. Note that the patterns we observe in Figures S2–S4 are similar to that in Figure 5, which indicates that the conclusion we obtain under the canonical form can be instructive when considering the linear form.

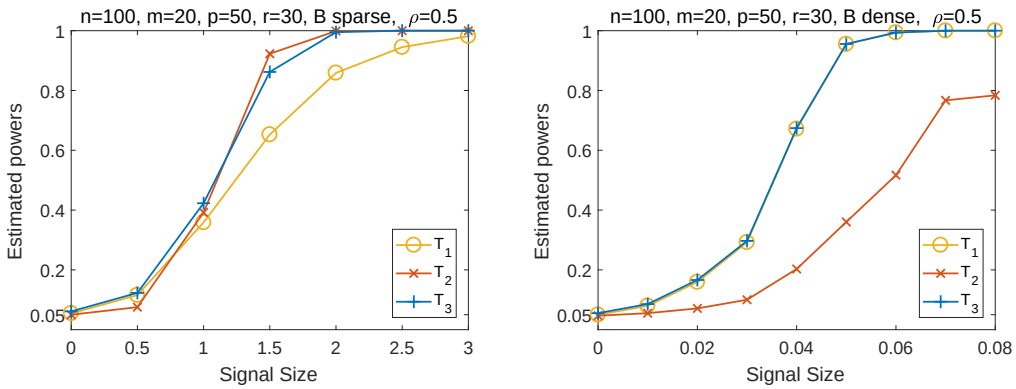


(a) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r)}]$

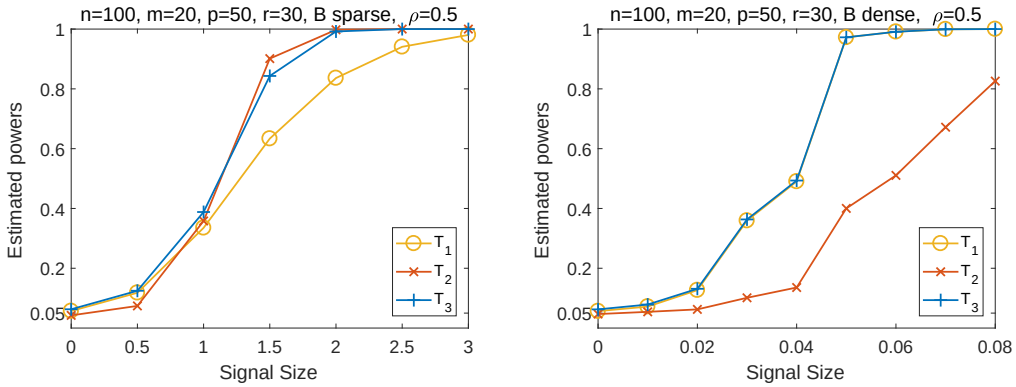


(b) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$

Figure S2: Estimated powers versus signal sizes with $\rho = 0.7$

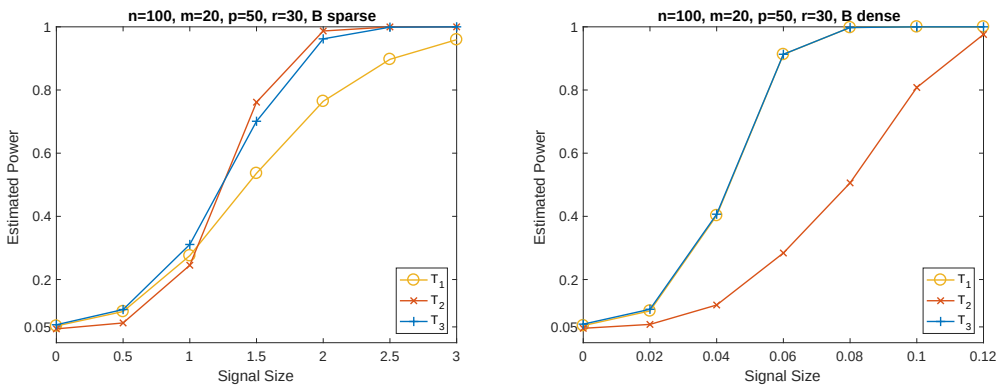


(a) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r)}]$

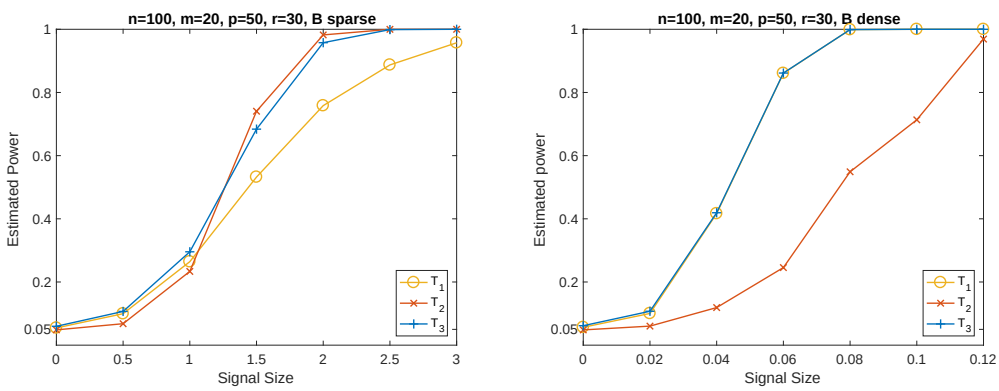


(b) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$

Figure S3: Estimated powers versus signal sizes with $\rho = 0.5$



(a) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r)}]$



(b) Estimated powers versus signal sizes when $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$

Figure S4: Estimated powers versus signal sizes with $\rho = 0$

Robustness with other distributions

We further conduct some simulations considering other distributions, which exhibit similar patterns as in Figure S2 and imply the robustness of the proposed methods.

(a) X and Y follow multinomial distributions For $i = 1, \ldots, n$ and $j = 1, \ldots, p$, we generate the entry $x_{i,j}$ in X independently and identically in the following way. In particular, we first generate $z_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$, and set the value of $x_{i,j}$ as below:

$$x_{i,j} = \begin{cases} -3 & z_{i,j} < -1, \\ -2 & z_{i,j} \in [-1, -0.4), \\ -1 & z_{i,j} \in [-0.4, 0), \\ 1 & z_{i,j} \in [0, 0.4), \\ 2 & z_{i,j} \in [0.4, 1), \\ 3 & z_{i,j} > 1. \end{cases}$$

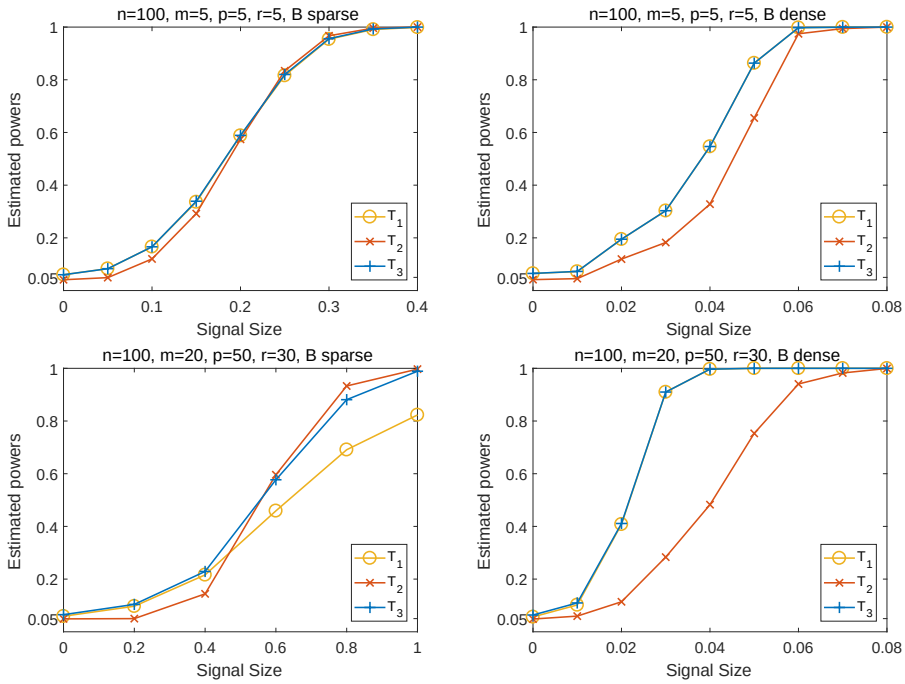
Given B and X, we generate $W = XB + E$, where the entries of E are i.i.d. $\mathcal{N}(0,1)$. For $i = 1, \ldots, n$ and $j = 1, \ldots, p$, let $w_{i,j}$ and $y_{i,j}$ denote the entries of W and Y

respectively. We then set

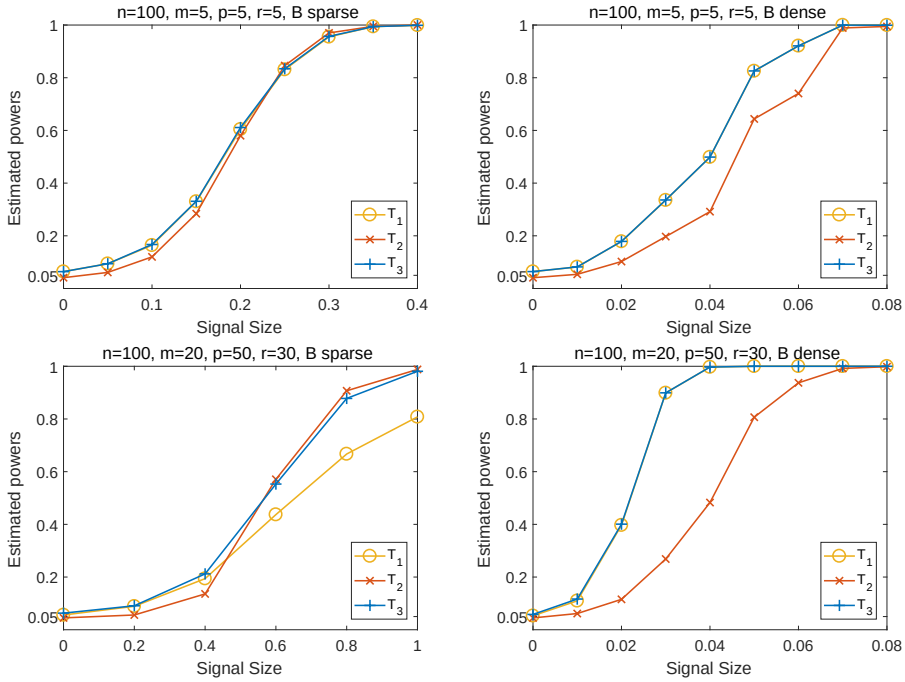
$$y_{i,j} = \begin{cases} -3 & w_{i,j} < -1, \\ -2 & w_{i,j} \in [-1, -0.4), \\ -1 & w_{i,j} \in [-0.4, 0), \\ 1 & w_{i,j} \in [0, 0.4), \\ 2 & w_{i,j} \in [0.4, 1), \\ 3 & w_{i,j} > 1. \end{cases}$$

We present the results in Figure S5, where "B sparse" and "B dense" represent two different types of B matrix, which are generated following the same method as in Section S7.1. Similarly, we also take $C = [I_r, \mathbf{0}_{r\times(p-r)}]$ and $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$ respectively. We can observe similar patterns to that in Figure S2.

(b) Errors follow t distribution In this part, we examine the case when the errors in matrix E independently and identically follows t distribution. In particular, we first generate the entries in X as i.i.d. $\mathcal{N}(0, 1)$. Then we generate the entries in E as i.i.d. t_{df} with $df \in \{3, 5\}$. The results are summarized in Figure S6, where similar patterns are observed as in Figure S2.

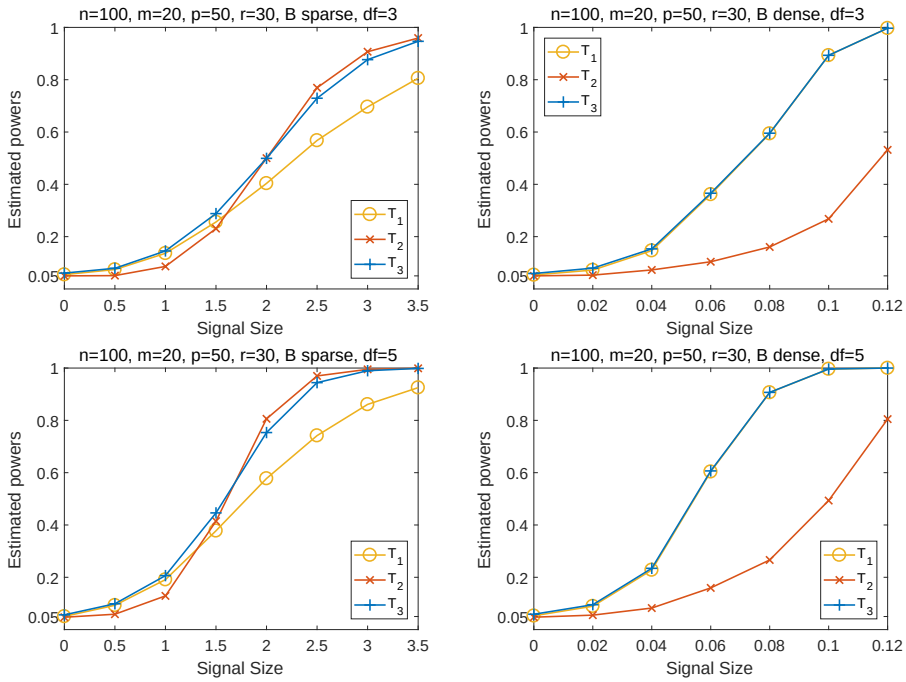


(a) When $C = [I_r, \mathbf{0}_{r\times(p-r)}]$

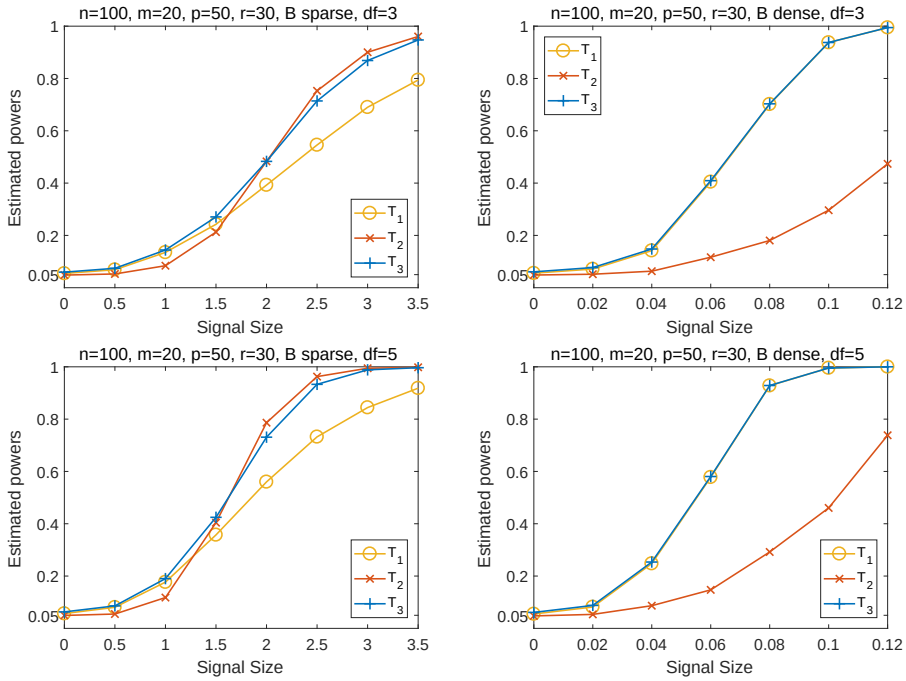


(b) When $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$

Figure S5: Power comparison when X and Y follow multinomial distribution



(a) When $C = [I_r, \mathbf{0}_{r\times(p-r)}]$



(b) When $C = [I_r, \mathbf{0}_{r\times(p-r-1)}, -\mathbf{1}_r]$

Figure S6: Power comparison when entries in E follow t distribution

S7.2 Supplementary simulations when $n < p + m$

Supplementary simulations with normal distribution

Under the similar set-up to that of Figure 6, we present additional results with $r_k = 5$ in Figure S7, where similar patterns are observed as in Figure 6.

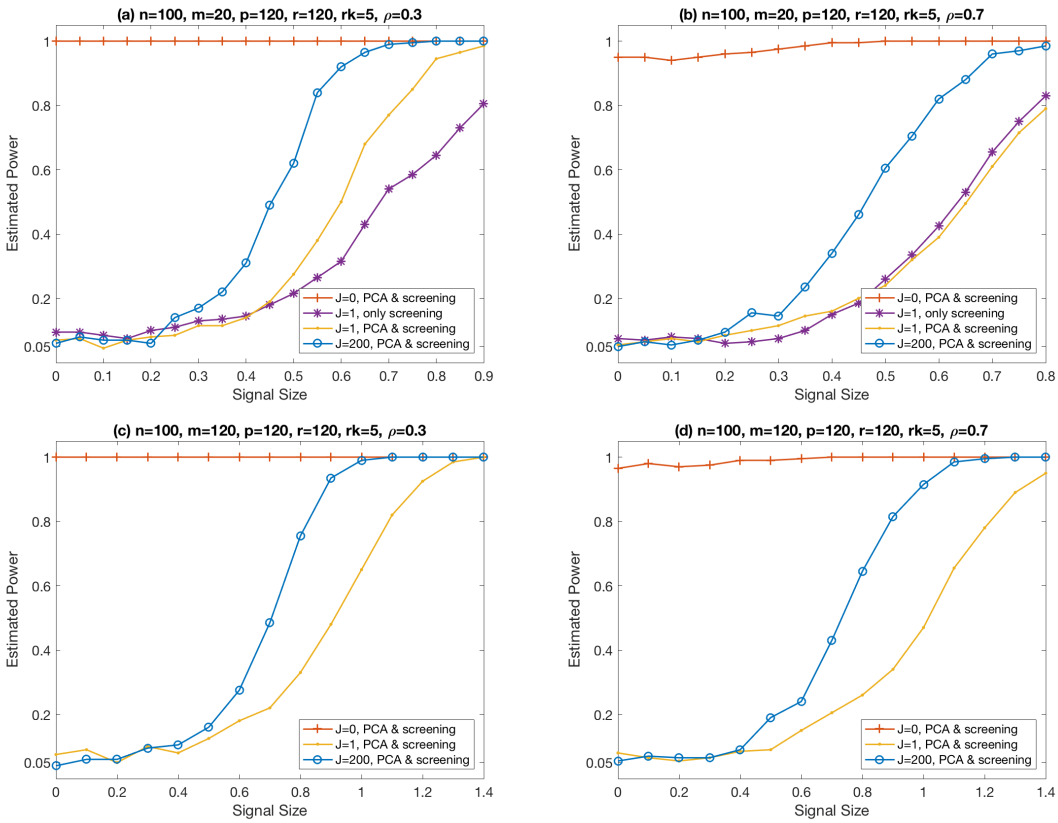


Figure S7: Estimated powers versus signal sizes when $n < m + p$

In addition, under the similar set-up to that of Figure 6, we conduct simulations when $\rho = 0$ and $r_k \in \{1, 5\}$. The results are presented in Figure S8, where similar patterns are observed as in Figure 6.

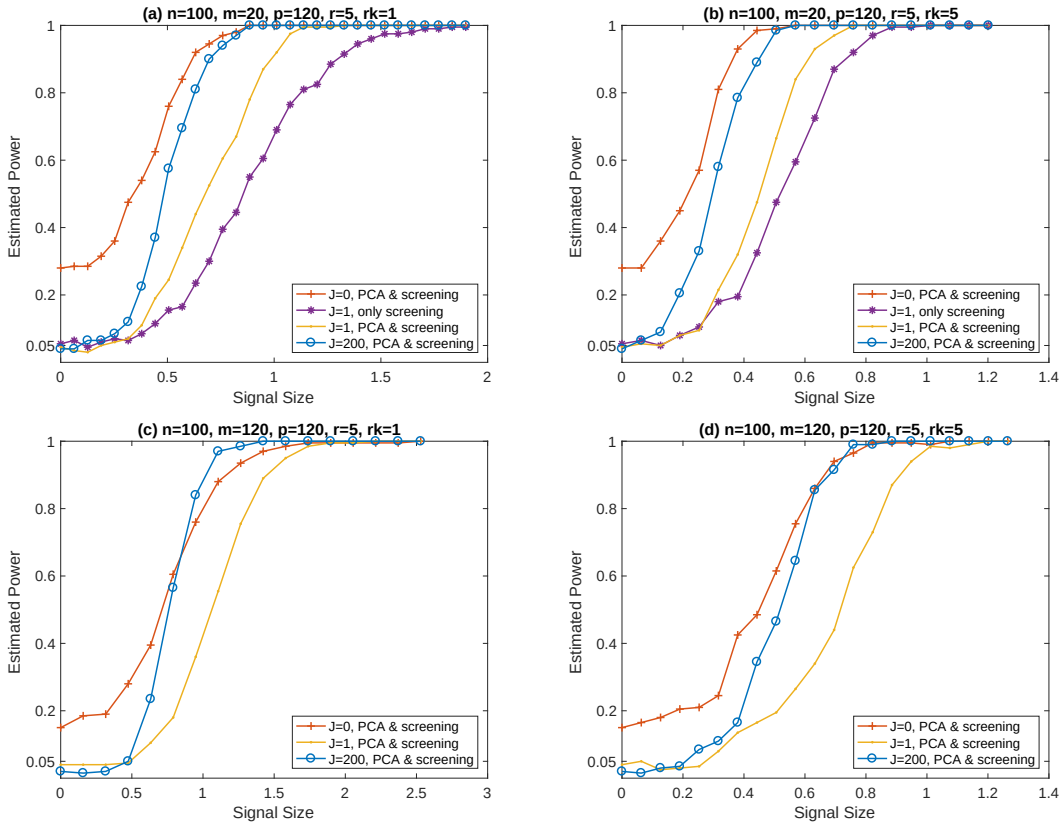


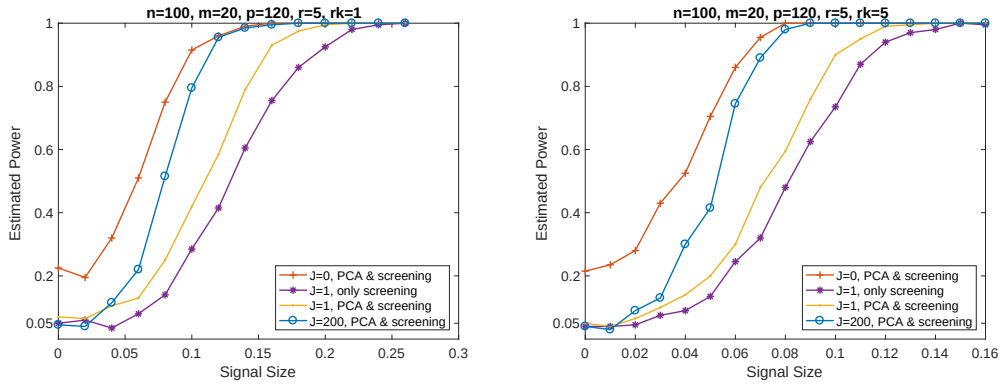
Figure S8: Estimated powers versus signal sizes when $n < m + p$

Robustness with other distributions

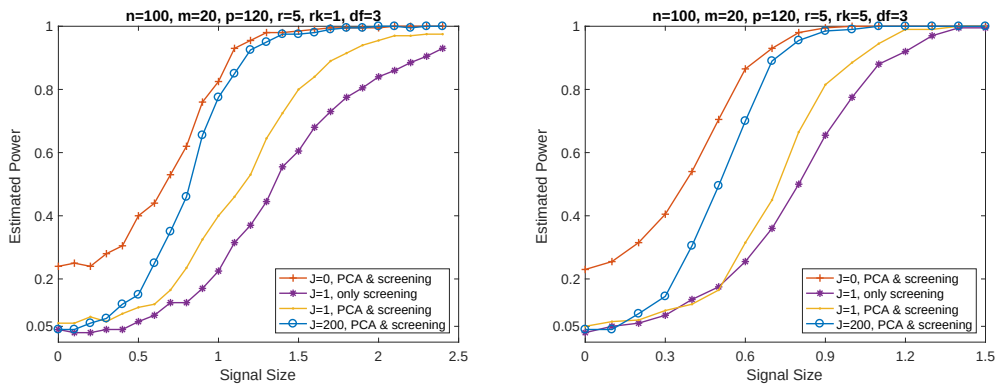
To examine the robustness of the two-step procedure, we generate X and Y following Section S7.1 with $n = 100$, $m = 20$, $p = 120$. We then generate B and apply the testing procedure similarly as in Section 5.2 with $r_k \in \{1, 5\}$. The results are presented in Figure S9, where part (a) gives the results when X and Y follow multinomial distribution, and parts (b) and (c) give the results when the error terms in E are i.i.d. t_3 or t_5. We note that similar patterns are observed as in Figure 6. This shows that the proposed two-step procedure is robust to the normal assumption.

S7.3 Simulations on $P\{\psi(\alpha\gamma) \geq \gamma\}$

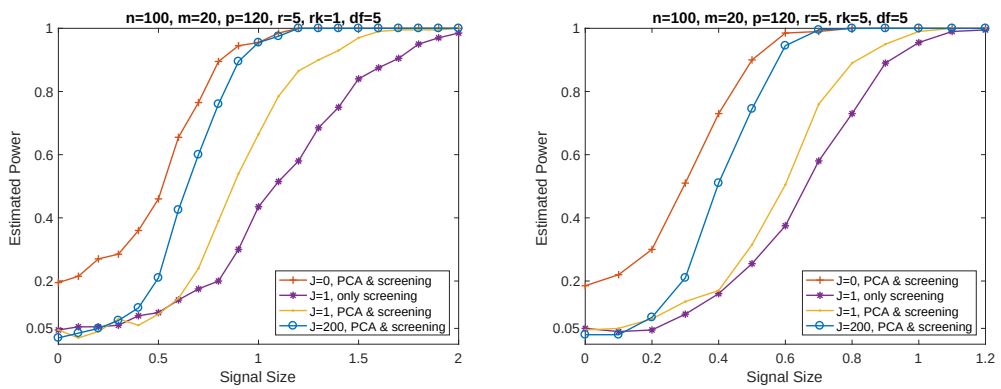
We conduct a simulation study to illustrate how the value of $P\{\psi(\alpha\gamma) \geq \gamma\}$ depends on the correlations of the p-values. We consider an "ideal" case with equal correlated p-values. Specifically we generate $p^{(j)} = 1 - \Phi(V_{J,j})$ for $j = 1, \ldots, J$, where $V_J = (V_{J,1}, \ldots, V_{J,J})^\intercal \sim \mathcal{N}(\mathbf{0}, \Sigma_J)$ with $\Sigma_J = (1-\rho)I_{J,J} + \rho \mathbf{1}_J^\intercal \mathbf{1}_J$. Note that larger ρ value implies larger correlations between $p^{(j)}$'s. We take $J = 200$ and use 10^6 Monte Carlo repetitions to estimate $P\{\psi(\alpha\gamma) \geq \gamma\}$. Figure S10 gives the simulation results for $\rho \in \{0, 0.2, 0.4, 0.6, 0.8, 0.9, 0.95, 1\}$, and $\gamma \in (0, 1)$ and $(0, 0.01)$ respectively. When ρ is small, the largest value of $P\{\psi(\alpha\gamma) \geq \gamma\}$ is attained at $5 \times 10^{-3} = J^{-1}$; when $\rho = 1$, the largest value is attained at $\gamma = 1$. These observations are consistent with the above theoretical argument.



(a) X and Y follow multinomial distributions



(b) Entries in E follow t_3 distribution



(c) Entries in E follow t_5 distribution

Figure S9: Estimated powers of two-step procedure with other distributions

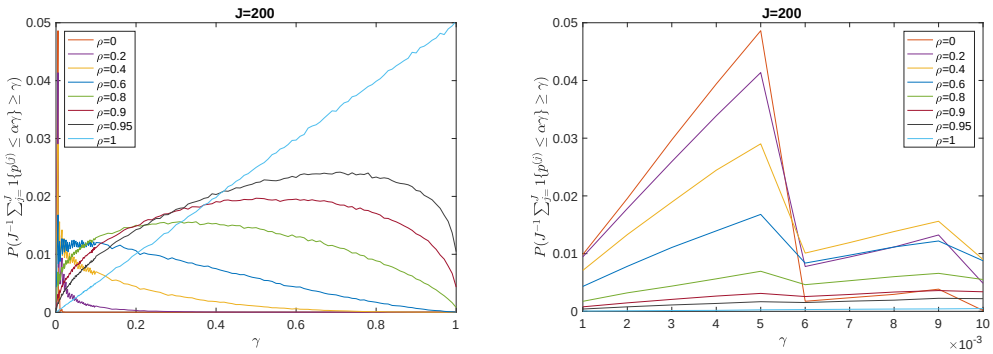


Figure S10: Estimated $P\{\psi(\alpha\gamma) \geq \gamma\}$ versus γ under different correlation levels

S7.4 Simulations compared with screening using lasso

In this paper, we propose the two-stage testing procedure using the screening with canonical correlations. Note that the proposed method aggregates the joint information of the response variables, and thus could be better than simply applying the marginal screening with respect to each response variable. To further study the effect of highly correlated predictors, we compare our method to using lasso with cross-validation, which is expected to account for the dependence in the predictors while not for the dependence in the responses.

In particular, for the screening with canonical correlations, 20% predictors are selected as in Section 5.2; for the screening with lasso, we select the predictors ($\leq 20\%$ of all predictors) that minimize the MSE in 10-fold cross-validation. In the simulations, we take $C = [I_r, \mathbf{0}_{r\times(p-r)}]$, and generate the rows of X and E as independent multivariate Gaussian with covariance matrices $\Sigma_x = (\rho^{|i-j|})_{p\times p}$ and $\Sigma = (\rho^{|i-j|})_{m\times m}$

respectively. For each setting considered, we choose $\rho \in \{0.7, 0.9\}$, which are the cases when the predictors are of large correlations.

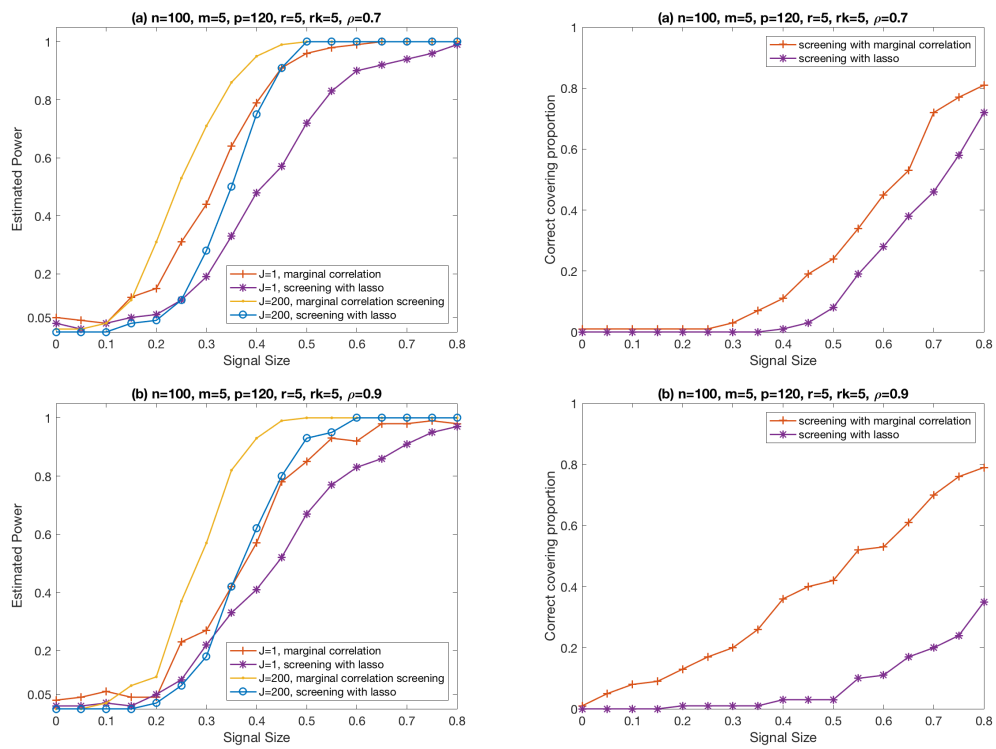


Figure S11: Screening Comparison: B is diagonal

We next consider two simulation settings, whose results are provided in the following Figures S11 and S12 respectively. In the first setting, we choose B to be a $p \times m$ diagonal matrix with σ_s in the first r_k diagonal entries, where σ_s represents the signal size that varies in simulations. We take $n = 100, p = 120, m = 5, r = 5$ and $r_k = 5$. In the second setting, we generate B with a nonzero submatrix of size $r_k \times m$ in the upper left corner, where the entries are randomly generated from $\mathcal{N}(0, \sigma_s^2)$.

We take $n = 100, p = 120, m = 5, r = 120$ and $r_k = 5$. In both Figures S11 and S12, we provide the estimated powers versus signal sizes in the left column, where J represents the number of splits similarly as in Figure 6. In addition, we provide the corresponding proportion of simulations that cover the true active set (correct covering proportion) versus signal sizes in the right column.

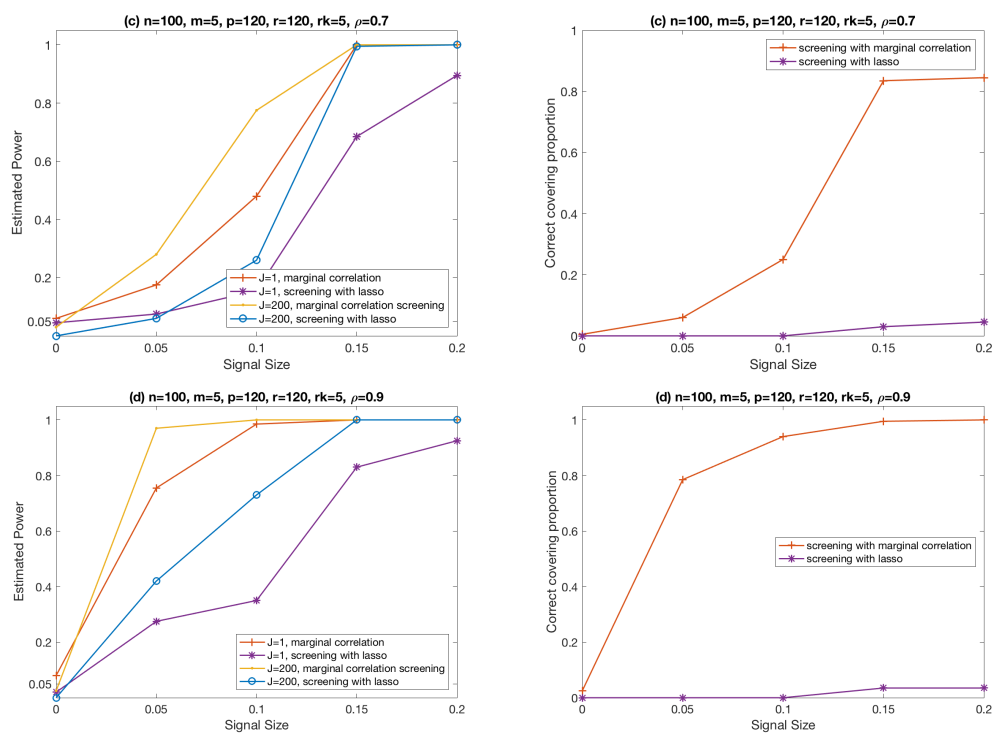


Figure S12: Screening Comparison: B has a nonzero submatrix

By the simulation results, we find that under the considered simulation settings, even though the correlations among predictors are large, using the canonical correlation in screening performs better than using lasso with cross-validation, in terms of both test power and correct covering proportion. The results suggest that the

correlation-based procedure can still account for the dependence among predictors reasonably under certain settings with correlated predictors. In addition, comparing the test power and corresponding correct covering proportion in Figures S11 and S12, we find that the under selection of the true active set generally leads to loss of power in testing. To further improve the test power, it is still of interest to develop a screening approach that could fit a wider range of scenarios and is also computationally efficient. Besides the two screening approaches compared here, we can also generalize other screening methods to the multivariate regression setting, as discussed in Remark 3 on Page 20. We will further study this in the follow-up research.

S8. Supplementary Results of Real Data Analysis

In this section, we present the analysis results of the regressions of GEPs on CNVs for the same dataset in Section 6. Then the m-variate response is the GEPs data and the p-variate predictor is the CNVs data, where now the dimension parameters are $(p, m) = (138, 673), (87, 1161), (18, 516)$ for the three chromosomes correspondingly. Similarly to Section 6, we apply the proposed procedure with $n_S = 26$, $n_T = 63$ and $J = 2000$. As m values are large in this case, we choose different fixed numbers of principal components when applying PCA on the response Y. The chosen number of principal components and predictors are denoted as m_0 and p_0 respectively, which are generally chosen as large as possible considering the sample size given. We next

	Chromosome pair					
(m_0, p_0)	$8 \to 8$	$17 \to 17$	$22 \to 22$	$8 \to 17$	$17 \to 22$	$8 \to 22$
(10,45)	×	×	×	×	×	✓
(15,45)	×	×	×	×	×	✓
(15,40)	×	×	×	×	×	✓
(20,40)	×	×	×	×	×	✓
(20,35)	×	×	×	×	×	✓

Table S1: Decision results

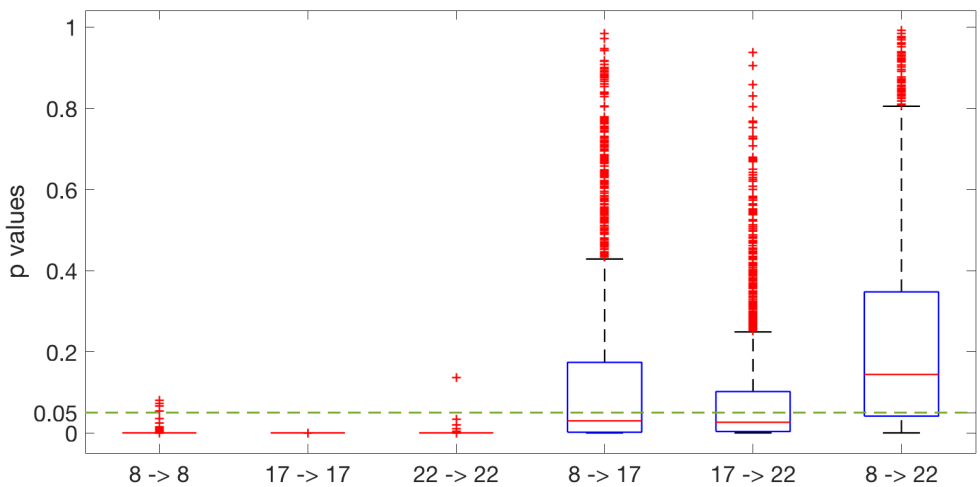


Figure S13: Boxplot of p values for regressions on different chromosome pairs

provide the decision results in Table S1, where the notations follow the same meaning as in Table 1. In addition, to further illustrate the results, we also report the boxplots of the p-values with respect to different chromosome pairs in Figure S13, where $(m_0, p_0) = (15, 40)$.

From the results, we can see that the p-values presented in Figure S13 support the test results in Table S1. Particularly, in the boxplots of the regressions on the same

chromosome pairs (the first three boxplots), the obtained p-values are significantly smaller than 0.05. For the regressions of the 17th on the 8th chromosomes and the 22nd on the 17th chromosomes (the 4th and 5th boxplots), the medians of the p-values are smaller 0.05. These observations are consistent with the rejections of the corresponding null hypotheses. Moreover, for the regression of the 22nd on the 8th chromosomes (the 6th boxplot), most of the p-values are greater than 0.05, which supports the decision that we accept the corresponding null hypothesis.

References

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

Gamelin, T. (2001). *Complex analysis*. Springer Science & Business Media.

Jiang, T. and Y. Qi (2015). Likelihood ratio tests for high-dimensional normal distributions. *Scandinavian Journal of Statistics 42*(4), 988–1009.

Jiang, T. and F. Yang (2013). Central limit theorems for classical likelihood ratio tests for high-dimensional normal distributions. *Ann. Statist. 41*(4), 2029–2074.

Meinshausen, N., L. Meier, and P. Bühlmann (2009). P-values for high-dimensional regression. *Journal of the American Statistical Association 104*(488), 1671–1681.

Muirhead, R. J. (2009). *Aspects of multivariate statistical theory*, Volume 197. John Wiley & Sons.

Wang, X. and C. Leng (2016). High dimensional ordinary least squares projection for screening variables.

Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78 (3), 589–611.