Supplementary Material

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

Yinqiu He\textsuperscript{1}, Tiefeng Jiang\textsuperscript{2}, Jiyang Wen\textsuperscript{3}, Gongjun Xu\textsuperscript{1}

\textsuperscript{1}Department of Statistics, University of Michigan
\textsuperscript{2}School of Statistics, University of Minnesota
\textsuperscript{3}Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

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 $-2 \log L_n \xrightarrow{D} \chi_{mr}^2$ as $n \to \infty$. Note that $\mathbb{E}(\chi_{mr}^2) = mr$, $\text{var}(\chi_{mr}) = 2mr$, and when $mr \to \infty$, $(\chi_{mr}^2 - mr)/\sqrt{2mr} \xrightarrow{D} \mathcal{N}(0, 1)$. It follows that $P(\chi_{mr}^2 > \sqrt{2mr}z_\alpha + mr) \to \alpha$ and

$$\chi_{mr}^2(\alpha) = \sqrt{2mr} \times \{z_\alpha + o(1)\} + mr,$$

where $z_\alpha$ denotes the upper $\alpha$-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

$$P\{-2 \log L_n > \chi_{mr}^2(\alpha)\} = P\left\{\frac{-2 \log L_n + \mu_n}{n\sigma_n} > \frac{\chi_{mr}^2(\alpha) + \mu_n}{n\sigma_n}\right\}.$$

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

$$\frac{\chi_{mr}^2(\alpha) + \mu_n}{n\sigma_n} \to z_\alpha, \ \text{as} \ n \to \infty.$$ (S1.2)
When $mr \to \infty$, by (S1.1), we know (S1.2) is equivalent to

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

(S1.3) holds for any significance level $\alpha$ 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 $\mu_n$ in the second step, and obtain the conclusion in the third step.

**Step 1.** Note that

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

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

$$\log \frac{n + r - p - m}{n - p - m}$$
$$= \log \left( 1 + \frac{r}{n - p - m} \right)$$
$$= \frac{r}{n - p - m} - \frac{1}{2} \frac{r^2}{(n - p - m)^2} + \frac{1}{3} \frac{r^3}{(n - p - m)^3} + O\left( \frac{r^4}{n^4} \right), \quad (S1.4)$$
and similarly,

\[- \log \frac{n - p + r}{n - p} = - \log \left( 1 + \frac{r}{n - p} \right) = - \frac{r}{n - p} + \frac{1}{2} \frac{r^2}{(n - p)^2} - \frac{1}{3} \frac{r^3}{(n - p)^3} + O\left( \frac{r^4}{n^4} \right). \quad (S1.5)\]

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

\[
(S1.4) + (S1.5) = r \frac{n}{n - p} - p - m - r \frac{n}{n - p} - 1 + \frac{1}{2} \left\{ \frac{r^2}{(n - p - m)^2} - \frac{r^2}{(n - p)^2} \right\} + O\left( \frac{r^4}{n^4} \right) + \frac{1}{3} \left\{ \frac{r^3}{(n - p - m)^3} - \frac{r^3}{(n - p)^3} \right\} + O\left( \frac{r^4}{n^4} \right) \\
= \frac{rm}{(n - p - m)(n - p)} - \frac{1}{2} r^2 \times \frac{m(2n - 2p - m)}{(n - p - m)^2(n - p)^2} + O\left( \frac{r^3(m + r)}{n^4} \right). \quad (S1.6)
\]

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

\[
\frac{rm}{(n - p - m)(n - p)} = \frac{rm}{n^2 \{1 - (p + m)/n\} (1 - p/n)} = \frac{rm}{n^2} \left\{ 1 + \frac{2p + m}{n} + O\left( \frac{p^2 + m^2}{n^2} \right) \right\}. \quad (S1.7)
\]
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

\[
-\frac{1}{2} r^2 \times \frac{m(2n - 2p - m)}{(n - p - m)^2(n - p)^2} = -nmr^2 \left\{ 1 - \frac{p + m/2}{n} \right\} \frac{1}{n^4} \left\{ 1 + \frac{2(p + m)}{n} + \frac{2p}{n} + O\left(\frac{p^2 + m^2}{n^2}\right) \right\} = -\frac{mr^2}{n^3} \left\{ 1 + \frac{3p + 3m/2}{n} + O\left(\frac{p^2 + m^2}{n^2}\right) \right\}.
\]

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

\[
(S1.6) = (S1.7) + (S1.8) + O\left\{ \frac{r^3(m + r)}{n^4} \right\} = \frac{rm}{n^2} + \frac{rm}{n^2} \left\{ \frac{2p + m - r}{n} \right\} + O\left\{ \frac{mr(m^2 + r^2 + p^2)}{n^4} \right\}.
\]

We then know that $\sigma_n^2 = 2 \times (S1.6) = (2mr/n^2) \times \{ 1 + o(1) \}$, and thus $n\sigma_n = \sqrt{2mr} \{ 1 + o(1) \}$.
Step 2. In this step, we derive the form of $\mu_n$. Under the asymptotic region $R_A$, we know that by Lemma \[ ] and Taylor expansion,

$$
\mu_n = -mr \{ 1 + \frac{p - r}{n} + O\left( \frac{p^2 + r^2}{n^2} \right) \}
- \frac{1}{2} \times \frac{mr(m + r)}{n} \left\{ 1 + O\left( \frac{p + r}{n} \right) \right\} + o(1) m \times \frac{p + m + r}{n} 
= -mr - mr \frac{p + m/2 - r/2}{n} + o(1) m \times \frac{p + m + r}{n}.
$$

Step 3. As discussed, under $R_A$, (S1.3) holds for any level $\alpha$, 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 $R_A$. In the second step, we obtain the form of $\mu_n$. Thus we have

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

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 \(\phi_1(t) = E\{\exp(-2it \log L_n)\}\) 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

\[
\varsigma_l = (-1)^{l+1} \frac{l(l+1)}{2} \left[ \sum_{k=1}^{m} \left\{ \frac{B_{l+1}(1-k-p)/2}{(n/2)^l} - \frac{B_{l+1}(1-k+r-p)/2}{(n/2)^l} \right\} \right],
\]

and \(B_{l+1}(\cdot)\) is the Bernoulli polynomials which takes the form \(B_{l+1}(z) = \sum_{v=0}^{l+1} c_v z^v\).

We next estimate the order of \(\varsigma_l\) with respect to \(n\). We note that for any \(z_1\) and \(z_2\),

\[
B_{l+1}(z_1) - B_{l+1}(z_2) = \sum_{v=0}^{l+1} c_v (z_1^v - z_2^v) = \sum_{v=0}^{l+1} c_v \sum_{w=1}^{v} \binom{v}{w} (z_1 - z_2)^{w-1} z_2^{v-w}.
\]

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 \(\varsigma_l\) 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
\[-2 \log L_n \xrightarrow{D} \chi_{mr}^2 \text{ as } n \to \infty. \] 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 \(-2 \log L_n \xrightarrow{D} \chi_{mr}^2\) fails.

S1.3 Lemma used in Section S1.1

**Lemma 1.** Under the asymptotic regime \( R_A \),

\[
\mu_n = - \frac{nmr}{n + r - p} - \frac{1}{2} \frac{nmr(m + r)}{(n + r - p)^2} - \frac{nmr(m^2/3 + mr/2 + r^2/3)}{(n + r - p)^3} + O(1) \frac{mr(m^3 + r^3)}{n^3} + O(mr/n).
\]

**Proof.** By the definition of \( \mu_n \) in Theorem 3,

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

Note that

\[
\log \frac{(n + r - p - m)(n - p)}{(n - p - m)(n + r - p)} = \log \frac{n + r - p - m}{n + r - p} + \log \frac{n - p}{n + r - p} - \log \frac{n - p - m}{n + r - p} = \log \left(1 - \frac{m}{n + r - p}\right) + \log \left(1 - \frac{r}{n + r - p}\right) - \log \left(1 - \frac{m + r}{n + r - p}\right).
\]
It follows that

\[ \mu_n = n(n - m - p + r - 1/2) \log \left( 1 - \frac{m}{n + r - p} \right) \]  
(S1.12)

\[ + n(n - p - 1/2) \log \left( 1 - \frac{r}{n + r - p} \right) \]  
(S1.13)

\[ - n(n - m - p - 1/2) \log \left( 1 - \frac{m + r}{n + r - p} \right), \]  
(S1.14)

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

By the Taylor expansion, we have

\[ \log \left( 1 - \frac{m}{n + r - p} \right) = - \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m}{n + r - p} \right)^k, \]

\[ \log \left( 1 - \frac{r}{n + r - p} \right) = - \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{r}{n + r - p} \right)^k, \]

\[ \log \left( 1 - \frac{m + r}{n + r - p} \right) = - \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m + r}{n + r - p} \right)^k. \]

Then

\[ \frac{(\text{S1.12})}{n} = -(n + r - p) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m}{n + r - p} \right)^k + (m + 1/2) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m}{n + r - p} \right)^k \]

\[ = - \sum_{k=1}^{\infty} \frac{m^k}{k (n + r - p)^{k-1}} + \sum_{k=1}^{\infty} \frac{m^{k+1}}{k (n + r - p)^{k}} + \sum_{k=1}^{\infty} \frac{m^k/2}{k (n + r - p)^{k}}, \]
\[(S1.13)/n = -(n + r - p) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{r}{n + r - p} \right)^k + (r + 1/2) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{r}{n + r - p} \right)^k \]
\[= - \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^{k-1}} + \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^k} + \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^{k+1}} \]

and

\[(S1.14)/n = (n + r - p) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m + r}{n + r - p} \right)^k - (m + r + 1/2) \sum_{k=1}^{\infty} \frac{1}{k} \left( \frac{m + r}{n + r - p} \right)^k \]
\[= \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^{k-1}} - \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^k} - \sum_{k=1}^{\infty} \frac{1}{k (n + r - p)^{k+1}} \]

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

\[\sum_{k=1}^{\infty} \frac{1}{k} \frac{(m + r)^k - m^k - r^k}{(n + r - p)^{k-1}} - \sum_{k=1}^{\infty} \frac{1}{k} \frac{(m + r)^{k+1} - m^{k+1} - r^{k+1}}{(n + r - p)^{k+1}}, \quad (S1.15)\]
\[- \frac{1}{2} \sum_{k=1}^{\infty} \frac{1}{k} \frac{(m + r)^k - m^k - r^k}{(n + r - p)^{k}}. \quad (S1.16)\]
As \((m + r)^1 - m^1 - r^1 = 0\), we know

\[
\text{(S1.15)} = \sum_{k=2}^{\infty} \frac{1}{k} \frac{(m + r)^k - m^k - r^k}{(n + r - p)^{k-1}} - \sum_{k=1}^{\infty} \frac{1}{k} \frac{(m + r)^{k+1} - m^{k+1} - r^{k+1}}{(n + r - p)^k}
\]

\[
= \sum_{k=1}^{\infty} \left( \frac{1}{k+1} - \frac{1}{k} \right) \frac{(m + r)^{k+1} - m^{k+1} - r^{k+1}}{(n + r - p)^k}
\]

\[
= -\frac{2mr}{n + r - p} - \frac{1}{6} \frac{(m + r)^3 - m^3 - r^3}{(n + r - p)^2} - \frac{1}{12} \frac{(m + r)^4 - m^4 - r^4}{(n + r - p)^3} - \sum_{k=4}^{\infty} \frac{1}{k(k+1)} \frac{(m + r)^{k+1} - m^{k+1} - r^{k+1}}{(n + r - p)^k},
\]

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

\[
|\text{(S1.20)}| = \sum_{k=4}^{\infty} \frac{1}{k(k+1)} \frac{\sum_{q=1}^{k} \binom{k+1}{q} m^q r^{k+1-q}}{(n + r - p)^k}
\]

\[
\leq \frac{mr}{n + r - p} \sum_{k=1}^{\infty} \frac{1}{k} \times \frac{2^{k+1}(\max\{m, r\})^{k-1}}{(n + r - p)^{k-1}}
\]

\[
= \frac{mr}{n + r - p} O\left\{ \left( \frac{\max\{m, r\}}{n + r - p} \right)^{3} \right\} = O(1) \frac{mr(m^3 + r^3)}{n^4},
\]

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

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

\[= \sum_{k=2}^{\infty} \frac{1}{k} \sum_{q=1}^{k-1} \frac{(k)^q m^q r^{k-q}}{(n + r - p)^k} \]

\[\leq \frac{mr}{(n + r - p)^2} \sum_{k=2}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m, r\})^{k-2}}{(n + r - p)^{k-2}} \]

\[= O(mr/n^2), \]

where in the last equation we use the fact that

\[
\sum_{k=2}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m, r\})^{k-2}}{(n + r - p)^{k-2}} \leq 2 + \sum_{k=3}^{\infty} \frac{1}{k-2} \times \frac{2^k (\max\{m, r\})^{k-2}}{(n + r - p)^{k-2}} \]

\[= 2 + 4 \sum_{k=1}^{\infty} \frac{1}{k} \times \frac{2^k (\max\{m, r\})^k}{(n + r - p)^k} \]

\[= 2 + 4 \log[1 - \{2 \max\{m, r\}/(n + r - p)\}]. \]

In summary,

\[\mu_n = (S1.12) + (S1.13) + (S1.14) \]

\[= n \times \{(S1.15) + (S1.16)\} \]

\[= -\frac{n m r}{n + r - p} - \frac{1}{2} \frac{n m r (m + r)}{(n + r - p)^2} - \frac{n m r (m^2/3 + m r/2 + r^2/3)}{(n + r - p)^3} \]

\[+O(1) \frac{m r (m^3 + r^3)}{n^3} + O(m r/n). \]
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 \(\mathcal{R}_A = \{(p, m, r, n) : n > p + m, \ p \geq r, \ mr \to \infty, \ and \ \max\{p, m, r\}/n \to 0 \ as \ n \to \infty\}, \rho = 1 + o(1)\). 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 \(\alpha\) 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,

\[
\frac{mr}{\rho} = \frac{mr}{1 - (p + m/2 - r/2 + 1/2)/n} \\
= \frac{nmr}{n - p + r - (m + r)/2} + mrO\left(\frac{1}{n}\right) \\
= \frac{nmr}{n - p + r} \sum_{k=0}^{\infty} \left\{ \frac{m + r}{2(n - p + r)} \right\}^k + mrO\left(\frac{1}{n}\right) \\
= \frac{nmr}{n - p + r} + \frac{nmr(m + r)}{2(n + r - p)^2} + \frac{nmr(m + r)^2}{4(n - p + r)^3} \\
+ mr \times O\left(\frac{m^3 + r^3}{n^3}\right) + mrO\left(\frac{1}{n}\right),
\]

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]

\[
\frac{\mu_n - (-mr/\rho)}{\sqrt{2mr}} = \frac{1}{\sqrt{2mr}} \left\{ - \frac{nmr}{n + r - p} - \frac{nmr(m + r)}{2(n + r - p)^2} - \frac{nmr(m^2/3 + mr/2 + r^2/3)}{(n + r - p)^3} \\
+ \frac{nmr}{n - p + r} + \frac{nmr(m + r)}{2(n - p + r)^2} + \frac{nmr(m + r)^2}{4(n - p + r)^3} \right\} \\
+ \sqrt{mr} \times O\left(\frac{m^3 + r^3}{n^3}\right) + O(\sqrt{mr}/n) \\
= - \frac{\sqrt{mr}}{\sqrt{2}} \frac{1}{12(n + r - p)^3} + \sqrt{mr} \times O\left(\frac{m^3 + r^3}{n^3}\right) + O(\sqrt{mr}/n) \\
= - \frac{\sqrt{mr}(m^2 + r^2)}{12\sqrt{2}n^2} + o(1) \times \frac{\sqrt{mr}(m^2 + r^2)}{n^2} + O(\sqrt{mr}/n),
\]

where in the last equation, we use the fact that \( \max\{p, m, r\} = o(n) \). We thus know
that \( (S.21) = 0 \) if and only if 
\[
\sqrt{mr} \left( m^2 + r^2 \right) / n^2 \to 0.
\]

**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 \( \phi_2(t) = E\{\exp(-2it\rho \log L_n)\} \). 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
\[
\tilde{\varsigma}_l = \frac{(-1)^{l+1}}{l(l + 1)} \left[ \sum_{k=1}^{m} \left\{ \frac{B_{l+1}(\tilde{z}_{k,1})}{(\rho n/2)^l} - \frac{B_{l+1}(\tilde{z}_{k,2})}{(\rho n/2)^l} \right\} \right],
\]
\( \tilde{z}_{k,1} = (1 - \rho)n/2 + (1 - k - p)/2 \) and \( \tilde{z}_{k,2} = (1 - \rho)n/2 + (1 - k + r - p)/2 \). Since \( \rho = 1 - (p - r/2 + m/2 + 1/2)/n, \)
\[
\tilde{z}_{k,1} = (p - r/2 + m/2 + 1/2)/2 + (1 - k - p)/2 = (3 - r + m)/4,
\]
\[
\tilde{z}_{k,2} = (p - r/2 + m/2 + 1/2)/2 + (1 - k + r - p)/2 = (3 + r + m)/4.
\]

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
\[
\log \phi_2(t) = -2^{-1}mr \log(1 - 2it) + O\{(n - p)^{-1}\} \text{ and } \phi_2(t) = (1 - 2it)^{-mr/2}[1 + O\{(n - p)^{-1}\}].
\]
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 $H_n \coloneqq \{-2\log L_n + \mu_n\}/(n\sigma_n) \xrightarrow{D} \mathcal{N}(0, 1)$, it is sufficient to show

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

(S3.22)

as $n \to \infty$ and $|s| < 1$, where $\sigma_n^2$ and $\mu_n$ are defined in Theorem 3. Equivalently, it suffices to show that for any subsequence $\{n_k\}$, there is a further subsequence $\{n_{kj}\}$ such that $H_{n_{kj}}$ 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 \( h \)th moment can be written as

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

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

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

The above integration is taken over the space of positive definite \( m \times m \) matrices, i.e., \( \{A_{m \times m} : A > 0\} \); and \( \text{tr}(A) \) 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
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 \( n_{kj} \) such that \( m_{kj}/(n_{kj} + r_{kj} - p_{kj}) \) and \( m_{kj}/(n_{kj} - p_{kj}) \) 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
\[
\sigma_n^2 = 2 \log \left( 1 - \frac{m}{n + r - p} \right) - 2 \log \left( 1 - \frac{m}{n - p} \right), \quad (S3.25)
\]
and \( m/(n + r - p) \to \gamma \in (0, 1/2] \), we know \( \sigma_n^2 \) 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

\[
\log \frac{\Gamma_m\left\{\frac{1}{2}n(1 + h) - \frac{1}{2}p\right\}}{\Gamma_m\left\{\frac{1}{2}(n - p)\right\}} = \log \frac{\Gamma_m\left\{\frac{1}{2}(n - p) + \frac{1}{2}nh\right\}}{\Gamma_m\left\{\frac{1}{2}(n - p)\right\}}
\]

\[
= - \left\{\frac{n^2h^2}{4} + \frac{nh}{2}\left(n - m - p - \frac{1}{2}\right)\right\} \log \left(1 - \frac{m}{n - p}\right)
\]

\[
+ \frac{mnh}{2}\left\{\log(n - p) - \log 2e\right\} + o(1),
\]

(S3.26)

and similarly, we can obtain

\[
\log \frac{\Gamma_m\left\{\frac{1}{2}(n + r - p)\right\}}{\Gamma_m\left\{\frac{1}{2}n(1 + h) + \frac{1}{2}(r - p)\right\}} = \log \frac{\Gamma_m\left\{\frac{1}{2}(n + r - p)\right\}}{\Gamma_m\left\{\frac{1}{2}(n + r - p) + \frac{1}{2}nh\right\}}
\]

\[
= \left\{\frac{n^2h^2}{4} + \frac{nh}{2}\left(n + r - m - p - \frac{1}{2}\right)\right\} \log \left(1 - \frac{m}{n + r - p}\right)
\]

\[
- \frac{mnh}{2}\left\{\log(n + r - p) - \log 2e\right\} + o(1).
\]

(S3.27)

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

\[
\log E \exp \left\{\frac{\log L_n}{n\sigma_n^2} s\right\} = \frac{n^2h^2}{4} \log \frac{(n + r - p - m)(n - p)}{(n - m)(n + r - p)} + \frac{h\mu_n}{2} + o(1)
\]

\[
= \frac{s^2}{2} + \frac{h\mu_n}{2} + o(1);
\]
where

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

Therefore, \( \log E \exp \left\{ \frac{\log L_n - \mu_n/2}{n^{\alpha_n/2}} s \right\} = s^2/2 + o(1) \) 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 \( r \to \infty \),

\[
\log \frac{\Gamma_m \left\{ \frac{1}{2}n(1 + h) - \frac{1}{2}p \right\}}{\Gamma_m \left\{ \frac{1}{2}(n - p) \right\}} = - \left\{ 2m + \left( n - p - m - \frac{1}{2} \right) \log \left( 1 - \frac{m}{n - p} \right) \right\} n h \frac{1}{2} \\
- \left\{ \frac{m}{n - p} + \log \left( 1 - \frac{m}{n - p} \right) \right\} n^2 h^2 \frac{1}{4} \\
+ m \left\{ \frac{(n - p + nh)}{2} \log \left( \frac{n - p + nh}{2} \right) - \frac{(n - p)}{2} \log \left( \frac{n - p}{2} \right) \right\} + o(1),
\]

(S3.28)
\begin{align*}
\log \frac{\Gamma_m \{ \frac{1}{2}(n + r - p) \}}{\Gamma_m \{ \frac{1}{2}n(1 + h) + \frac{1}{2}(r - p) \}} &= \left\{ 2m + \left( n + r - p - m - \frac{1}{2} \right) \log \left( 1 - \frac{m}{n + r - p} \right) \right\} \frac{nh}{2} \\
&\quad - m \left\{ \frac{(n + r - p + nh)}{2} \log \frac{n + r - p + nh}{2} - \frac{(n + r - p)}{2} \log \frac{n + r - p}{2} \right\} \\
&\quad + \left\{ \frac{m}{n + r - p} + \log \left( 1 - \frac{m}{n + r - p} \right) \right\} \frac{n^2 h^2}{4} + o(1). \tag{S3.29}
\end{align*}

By Taylor expansion of the log function, we have

\begin{align*}
\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)\}. \tag{S3.30}
\end{align*}

where the second order terms of Taylor expansion of the log functions is ignorable as

\( m = o(n - p). \) Also, as \( r \to \infty, \)

\begin{align*}
h &= \frac{s}{n \sigma_n^2/2} = \frac{s \sqrt{2(n - p)(n + r - p)}}{n \sqrt{mr}} \{1 + o(1)\} \to 0. \tag{S3.31}
\end{align*}
Therefore, combining (S3.23), (S3.28) and (S3.29), we obtain

$$\log E \exp \left\{ \frac{\log L_n}{n\sigma_n^2} s \right\} = \frac{n^2h^2}{4} \log \frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)} + \frac{n^2h^2}{4} \left( \frac{m}{n+r-p} - \frac{m}{n-p} \right)$$

$$+ \frac{nh}{2} (n-m-p-1/2) \log \frac{(n+r-p-m)(n-p)}{(n-p-m)(n+r-p)}$$

$$+ \frac{nh}{2} r \log \frac{n+r-p-m}{n+r-p} + \frac{nh}{2} m \log \frac{n-p+nh}{n+r-p+nh}$$

$$+ \frac{m(n+r-p)}{2} \log \frac{n+r-p}{n+r-p+nh} + \frac{m(n-p)}{2} \log \frac{n-p+nh}{n-p} + o(1).$$

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

$$\frac{n^2h^2}{4} \left( \frac{m}{n+r-p} - \frac{m}{n-p} \right)$$

$$= -s^2(n-p)(n+r-p) \times \frac{mr}{2mr} \frac{1+o(1)}{(n-p)(n+r-p)}$$

$$= -\frac{s^2}{2} + o(1).$$

(S3.33)

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

$$\frac{m(n+r-p)}{2} \log \frac{n+r-p}{n+r-p+nh}$$

$$= -\frac{m(n+r-p)}{2} \left\{ \frac{nh}{n+r-p} - \frac{n^2h^2}{2(n+r-p)^2} + R_{n,1} \right\}.$$

(S3.34)
and

\[
\frac{m(n-p)}{2} \log \frac{n-p+nh}{n-p} = \frac{m(n-p)}{2} \left\{ \frac{nh}{n-p} - \frac{n^2h^2}{2(n-p)^2} + R_{n,2} \right\}, \tag{S3.35}
\]

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

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

where in the last equation, we use \(\text{(S3.33)}\) and Lemma 8. Furthermore, by \(nh/(n+\)
Combining (S3.32), (S3.33), (S3.37) and (S3.38), we obtain
\[ \log E \exp \left\{ \frac{\log L_n - \mu_n}{n\sigma_n/2} s \right\} = s^2/2 + o(1). \]

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.

\[ \text{(S3.29)} - \text{(S3.27)} \]

\[
= 2m \frac{nh}{2} + m \frac{nh}{2} \{ \log (n + r - p) - \log 2e \} + \frac{n^2 h^2}{4} \times \frac{m}{n + r - p} \\
- \frac{m(n + r - p)}{2} \log \frac{n + r - p + nh}{n + r - p} - \frac{mnh}{2} \log \frac{n + r - p + nh}{2} + o(1) \\
= mnh + \frac{mn}{2} \left\{ \log \left( \frac{n + r - p}{2} \right) - 1 \right\} + \frac{n^2 h^2}{4} \times \frac{m}{n + r - p} \\
- \frac{m(n + r - p)}{2} \log \left( 1 + \frac{nh}{n + r - p} \right) - \frac{mnh}{2} \log \left( \frac{n + r - p}{2} + \frac{nh}{2} \right) + o(1). \\
\text{(S3.39)}
\]

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 \( n^2 h^2 m/(n + r - p) \to 0 \). Applying Taylor expansion, we then have

\[
\frac{mn}{2} \log \left( \frac{n + r - p}{2} + \frac{nh}{2} \right) \\
= \frac{mn}{2} \left\{ \log \left( \frac{n + r - p}{2} \right) + O \left( \frac{nh}{n + r - p} \right) \right\} \\
= \frac{mn}{2} \log \left( \frac{n + r - p}{2} \right) + O \left( \frac{mn^2 h^2}{n + r - p} \right) \\
= \frac{mn}{2} \log \left( \frac{n + r - p}{2} \right) + o(1). \\
\text{(S3.40)}
\]
Similarly, by \( nh = O(s) \), \( m/(n + r - p) \to 0 \), and Taylor expansion, we have

\[
\frac{m(n + r - p)}{2} \log \left( 1 + \frac{nh}{n + r - p} \right) = \frac{m(n + r - p)}{2} \left\{ \frac{nh}{n + r - p} + O\left( \frac{n^2h^2}{(n + r - p)^2} \right) \right\} = \frac{mnh}{2} + o(1).
\]

(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

\[
\mu_n = n(n - p - m - 1/2) \log \left( \frac{n - p)(n - p + r - m)}{(n - p - m)(n + r - p)} + nm \log \frac{(n - p)}{(n + r - p)} + nr \log \frac{(n - p + r - m)}{(n + r - p)}.
\]
and
\[ \sigma^2_n = 2 \log \frac{(n - p)(n - p + r - m)}{(n - p - m)(n + r - p)}, \]
which take the same forms as those in the setting when \( r \geq m \). And the theorem can be proved following similar analysis when \( m \to \infty, n - p + r - m \to \infty \).

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

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 \( V_i \sim \text{beta}(\frac{1}{2}(n - p - i + 1), \frac{1}{2}r) \); 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 \( V_i \sim \text{beta}(\frac{1}{2}(n + r - p - m - i + 1), \frac{1}{2}m) \).

Lemma 4 (Theorem 2.1.12 in Muirhead (2009)). The multivariate Gamma function defined in \( (\text{S3.24}) \) 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

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

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

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

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

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

$$\frac{1}{a-1} \sum_{i=1}^{m} \frac{i-1}{a-i} = \frac{m(m-1)}{2(a-1)^2} \left\{ 1 + O \left( \frac{m}{a} \right) \right\}$$

$$= \left[ \frac{1}{2} \left( \sigma_a^2 - \frac{m}{(a-1)^2} \right) \right] \left\{ 1 + O(1/a) \right\},$$
where in the last equation, we use the fact that $O(m/a) = O(1/a)$ as $m$ is fixed. Then \(\text{(S3.42)}\) is proved.

We then prove \(\text{(S3.43)}\). 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,

$$
\log \Gamma(a - 1) - \log \Gamma(a - m - 1) \\
= (a - 3/2) \log(a - 1) - (a - m - 3/2) \log(a - m - 1) - m \\
+ \frac{1}{12} \left( \frac{1}{a - 1} - \frac{1}{a - m - 1} \right) + O(a^{-3}) \\
= (a - 3/2) \log(a - 1) - (a - m - 3/2) \log(a - m - 1) - m + O(ma^{-2}).
$$
Since for integers $k \geq 1$, $\Gamma(k) = (k-1)! = \Pi_{i=1}^{k-1} i$. Then we have

$$\sum_{i=1}^{m} \{ \log(a - 1) - \log(a - i) \}$$

$$= m \log(a - 1) - \{ \log \Gamma(a - 1) - \log \Gamma(a - m - 1) \} + \log \left( 1 - \frac{m}{a - 1} \right)$$

$$= m \log(a - 1) - (a - 3/2) \log(a - 1) + (a - m - 3/2) \log(a - m - 1)$$

$$+ m + \log \left( 1 - \frac{m}{a - 1} \right) + O(ma^{-2})$$

$$= -(m - a + 3/2) \log \left( 1 - \frac{m}{a - 1} \right) + m - \frac{m}{a - 1} + O \left( \frac{m^2}{a^2} \right)$$

$$= -(m - a + 3/2) \log \left( 1 - \frac{m}{a - 1} \right) + \frac{a - 1}{a} m + O \left( \frac{m^2}{a^2} \right)$$

$$= -\mu_a + O \left( \frac{m^2}{a^2} \right),$$

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

\[ \square \]

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

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

for $1 \leq i \leq m$ and $x > -(a - m)/2$. Let $\mu_a$ and $\sigma_a$ 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 $1 \leq i \leq m$,

$$
g'_i(x) = \log \left( \frac{a - i}{2} + x \right) - \log \left( \frac{a - 1}{2} + x \right),
$$

$$
g''_i(x) = \frac{1}{a - i + x} - \frac{1}{a - 1 + x} = \frac{i - 1}{(a - i + x)(a - 1 + x)},
$$

$$
g^{(3)}_i(x) = -\frac{1}{(a - i + x)^2} + \frac{1}{(a - 1 + x)^2}
$$

$$
= -\frac{i - 1}{(a - i + x)^2} \cdot \frac{2a - i - 1}{2} + (i - 1)x
$$

$$
\frac{1}{(a - 1 + x)^2}.
$$

By Taylor expansion,

$$
g_i(t) - g_i(0) = g'_i(0)t + \frac{t^2}{2}g''_i(0) + \frac{t^3}{6}g^{(3)}_i(\xi_i)
$$

$$
= \{\log(a - i) - \log(a - 1)\}t + \frac{i - 1}{(a - 1)(a - i)}t^2 + \frac{t^3}{6}g^{(3)}_i(\xi_i).
$$

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

$$
\sum_{i=1}^{m} \{g_i(t) - g_i(0)\} = \mu_a t + \left[ \frac{1}{2} \left( \sigma_a^2 - \frac{m}{(a - 1)^2} \right) \right] t^2 + o(1)
$$

$$
= \mu_a t + \frac{\sigma_a^2}{2} t^2 + o(1).
$$
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

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

where

$$v_a = -[2m + (a - m - 3/2)\log\{1 - m/(a - 1)\}]; \quad \vartheta_a = -[m/(a - 1) + \log\{1 - m/(a - 1)\}];$$

$$\gamma_a(t) = m\left\{\left(\frac{a - 1}{2} + t\right)\log\left(\frac{a - 1}{2} + t\right) - \frac{a - 1}{2}\log\left(\frac{a - 1}{2}\right)\right\}. $$

Proof. By Lemma 4, we know

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

(S3.44)$$

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

\[
\sigma_n^2 = \frac{1}{2} \log \left( 1 - \frac{m}{n+r-p} \right) - \frac{1}{2} \log \left( 1 - \frac{m}{n-p} \right)
= \frac{mr}{2(n-p)(n+r-p)} \left\{ 1 + o(1) \right\}.
\]

Thus we have

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

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

\[
\frac{t}{a-i} \leq \frac{t}{n-p-m} = O(s) \sqrt{\frac{(n-p)(n-p+r)}{mr(n-p-m)^2}}
= O(s) \sqrt{\frac{1 + r/(n-p)}{mr(1 - m/(n-p))^2}}
= O(s) \sqrt{\left\{ \frac{1}{mr} + \frac{1}{m(n-p)} \right\} \{ 1 + o(1) \}} = o(1),
\]

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 \( 1 \leq i \leq m \), \( t \in [-\delta(a-i)/2, \delta(a-i)/2] \) for any given \( \delta \in (0, 1) \).

Therefore, the condition of Lemma A.1. in \( \text{Jiang and Qi (2015)} \) is satisfied. By
that lemma, we know when $a \to \infty$, for uniformly $1 \leq i \leq m$,

$$\log \frac{\Gamma(\frac{a-i}{2} + t)}{\Gamma(\frac{a-i}{2})} = \left(\frac{a-i}{2} + t\right) \log \left(\frac{a-i}{2} + t\right) - \frac{a-i}{2} \log \frac{a-i}{2} - t - \frac{t}{a-i} + O\left(\frac{t^2}{a^2}\right).$$

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

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

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

$$\frac{tm(m+1)}{a(a-1)} = O(s) \sqrt{\frac{(n-p)(n-p+r)}{mr}} \frac{m^2}{(n-p)^2}$$

$$= O(s) \sqrt{\frac{m(n-p+r)}{r(n-p)}} \frac{m}{n-p}$$

$$= O(s) \sqrt{\frac{m}{\min\{n-p,r\}}} \frac{m}{n-p} = o(1).$$

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),

$$\frac{t}{a} \leq \frac{t}{n - p} = O(s) \sqrt{\frac{\max\{n - p, r\}}{mr(n - p)}}$$

$$= O(s) \sqrt{\frac{1}{m \times \min\{n - p, r\}}} = o(1). \quad \text{(S3.47)}$$

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

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

Therefore, from (S3.44), we have

$$\log \frac{\Gamma_m(\frac{a-1}{2} + t)}{\Gamma_m(\frac{a-1}{2})}$$

$$= -\frac{(a + 1)mt}{a} + \sum_{i=1}^{m} \left\{ \left( \frac{a - i}{2} + t \right) \log \left( \frac{a - i}{2} + t \right) - \frac{a - i}{2} \log \frac{a - i}{2} \right\} + o(1). \quad \text{(S3.48)}$$

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

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

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

\[ m \left[ \left( \frac{a - 1}{2} + t \right) \log \left( \frac{a - 1}{2} + t \right) - \frac{a - 1}{2} \log \frac{a - 1}{2} \right] + \sum_{i=1}^{m} \{ g_i(t) - g_i(0) \}. \tag{S3.49} \]

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 \( m \leq r \). 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

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

where

\[ \mu_a = (m - a + 3/2) \log \left( 1 - \frac{m}{a - 1} \right) - m \frac{a - 1}{a}, \]

\[ \sigma_a^2 = -2 \left[ \frac{m}{a - 1} + \log \left( 1 - \frac{m}{a - 1} \right) \right]. \]
Therefore, the proposition can be proved by noticing

\[ \nu_a = -\frac{(a + 1)m}{a} + \mu_a; \quad \vartheta_a = \sigma_a^2/2; \]

\[ \gamma_a(t) = m\left[\left(\frac{a - 1}{2} + t\right)\log\left(\frac{a - 1}{2} + t\right) - \frac{a - 1}{2} \log \frac{a - 1}{2}\right]. \]

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

\[ -\frac{m(n + r - p)}{2} R_{n,1} + \frac{m(n - p)}{2} R_{n,2} \]

\[ = \frac{m}{2} \left[ \sum_{k=3}^{\infty} \frac{1}{k} (-nh)^k \left\{ \frac{1}{(n + r - p)^{k-1}} - \frac{1}{(n - p)^{k-1}} \right\} \right] \]

\[ = \frac{m}{2} \left\{ \sum_{k=3}^{\infty} \frac{1}{k} (-nh)^k \frac{\sum_{q=1}^{k-1} \binom{k-1}{q} r^q (n - p)^{k-1-q}}{(n + r - p)^{k-1}(n - p)^{k-1}} \right\} \]

\[ = \frac{mnh}{2} \sum_{k=3}^{\infty} \frac{1}{k} \left(\frac{nh}{n + r - p}\right)^{k-1} \sum_{q=1}^{k-1} \binom{k-1}{q} \left(\frac{r}{n - p}\right)^q. \]
If $r/(n - p) = 1$,

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

where

$$mnh \times n^2h^2 \over (n + r - p)^2 = O\left\{ \frac{m\sqrt{(n - p)(n + r - p)}}{\sqrt{mr}} \times \frac{(n - p)(n + r - p)}{mr(n + r - p)^2} \right\} = O\left\{ \frac{m \sqrt{r} \times \sqrt{r}}{mr \times r^2} \right\} = o(1),$$

as $r \to \infty$.

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

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

\[
mnh \left( \frac{nh}{n + r - p} \right)^2 \left( \frac{r}{n - p} \right)^2
= O\left\{ \frac{m \sqrt{(n - p)(n + r - p)}}{\sqrt{mr}} \times \frac{(n - p)(n + r - p)}{mr(n + r - p)^2} \times \frac{r^2}{(n - p)^2} \right\}
= O\left\{ \frac{\sqrt{(n - p)(n + r - p)}}{\sqrt{mr}} \frac{r}{(n - p)(n + r - p)} \right\}
= O\left\{ \frac{r}{\sqrt{mr(n + r - p)(n - p)}} \right\} = o(1),
\]

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

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

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

where

\[
mnh \frac{(nh)^2}{(n + r - p)^2} \times \frac{r}{(n - p)}
= O\left\{ \frac{m \sqrt{(n - p)(n + r - p)}}{\sqrt{mr}} \times \frac{(n - p)(n + r - p)}{mr(n + r - p)^2} \times \frac{r}{(n - p)} \right\}
= O\left\{ \frac{\sqrt{n - p}}{\sqrt{mr(n + r - p)}} \right\} = o(1).
\]
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 \( \frac{p}{n} = \rho_p, \frac{r}{n} = \rho_r, \frac{m}{n} = \rho_m \) with \( \rho_p, \rho_r, \rho_m \in (0, 1) \) and \( \rho_p + \rho_m < 1 \), we know that \( \sigma_n^2 \) in Theorem 3 satisfies

\[
\sigma_n^2 = 2 \log \left\{ \left( 1 - \frac{\rho_m}{1 + \rho_r - \rho_p} \right) \left( 1 - \frac{\rho_m}{1 - \rho_p} \right)^{-1} \right\},
\]

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

\[
E\{2s \log L_n/(n \sigma)\} = E\{\exp(h \log L_n)\} = E L_n^h = E_0 L_n^h \times \mathbf{1}_F \left( \frac{nh}{2}; \frac{1}{2} (n + r - p) + \frac{nh}{2}, -\frac{1}{2} \Omega \right),
\]  

(S4.51)

where \( E_0 L_n^h \) is the moment generating function of \( \log L_n \) under \( H_0 \), and \( \mathbf{1}_F \) is the hypergeometric function, which depends on \( \Omega \) only through its eigenvalues symmet-
rically.

As $1 \mathcal{F}_1$ only depends on $\Omega$ via its eigenvalues symmetrically, without loss of generality, we consider the alternative with $\Omega = \text{diag}(w_1, \cdots, w_m)$ and $w_1 \geq \cdots \geq w_m \geq 0$. Let $\xi_a = nh/2$, $\xi_b = (n + r - p + nh)/2$ and $Q = -\Omega/2 = \text{diag}(-w_1/2, \cdots, -w_m/2)$, then we write $1 \mathcal{F}_1(nh/2, (n + r - p + nh)/2; -\Omega/2)$ as $1 \mathcal{F}_1(\xi_a; \xi_b; Q)$. Note that we assume that $\Omega$ has fixed rank $m_0$ in Theorem 4, then $\omega_1 \geq \ldots \geq \omega_{m_0} > 0$ are $m_0$ nonzero eigenvalues of $\Omega$. Further define $\tilde{Q} = \text{diag}(-\omega_1/2, \ldots, -\omega_{m_0}/2)$. By Lemma 11, we know $1 \mathcal{F}_1(\xi_a; \xi_b; Q) = 1 \mathcal{F}_1(\xi_a; \xi_b; \tilde{Q})$. Then to evaluate $1 \mathcal{F}_1(\xi_a; \xi_b; Q)$ when $Q$ has fixed rank, without loss of generality, we consider $1 \mathcal{F}_1(\xi_a; \xi_b; \tilde{Q})$.

Let $W = \log 1 \mathcal{F}_1(\xi_a; \xi_b; \tilde{Q})$ and $\tilde{Q} = -n\tilde{\Delta}/2$ with $\tilde{\Delta} = \text{diag}(\delta_1, \ldots, \delta_{m_0})$. From Lemma 12, we know that $W(\tilde{\Delta})$ is the unique solution of each of the $m_0$ partial differential equations

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

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 $\tilde{\Delta} = 0_{m_0 \times m_0}$ with $W(0_{m_0 \times m_0}) = 0$. As $r/n = \rho_r$,
\[ p/n = \rho_p, \ m_0 \text{ is a fixed number and } nh = 2s/\sigma, \text{ we can write (S4.52) into } \]
\[ \left[ \frac{1}{2}(1 + \rho_r - \rho_p)n + \frac{1}{2}(2s/\sigma - m_0 + 1) + \frac{1}{2}n\delta_j + \frac{1}{2} \sum_{i \neq j}^{m_0} \frac{\delta_j - \delta_i}{\delta_j - \delta_i} \right] \frac{\partial W}{\partial \delta_j} \]
\[ + \delta_j \left[ \frac{\partial^2 W}{\partial \delta_j^2} + \left( \frac{\partial W}{\partial \delta_j} \right)^2 \right] - \frac{1}{2} \sum_{i \neq j}^{m_0} \frac{\delta_i}{\delta_j - \delta_i} \frac{\partial W}{\partial \delta_i} = -\frac{s}{\sigma} \times n. \quad (S4.53) \]

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

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

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

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

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

\[ E L_n^h = E_0 L_n^h \times e^{\log_1 F_1} = E_0 e^{-\frac{s}{\sigma} \log(L_n) + W}. \quad (S4.54) \]

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 \( \{\log L_n - \mu_n/2\}/(n\sigma_n/2) \xrightarrow{D} \mathcal{N}(-A_1 W_\Delta, 1) \), and thus \( \{-2 \log L_n + \mu_n\}/(n\sigma) \xrightarrow{D} \mathcal{N}(A_1 W_\Delta, 1) \). Then the power \( P(T_1 > z_\alpha) \rightarrow \Phi(z_\alpha - A_1 W_\Delta). \)
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 $\kappa$ of $k$ is written as $\kappa = (k_1, k_2, \ldots)$, where $\sum_i k_i = k$ 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 $\kappa$ as $C_\kappa(M)$. Then by the definition, we know the hypergeometric function $1F_1(\xi_a; \xi_b; Q)$ satisfies

$$1F_1(\xi_a; \xi_b; Q) = \sum_{k=0}^{\infty} \sum_{\kappa:k} (\xi_a)_\kappa \frac{C_\kappa(Q)}{(\xi_b)_\kappa k!},$$

(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 $\kappa$, 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 $N_\kappa(M) = \sum_{\{i_1, \ldots, i_t\}} l_{i_1}^{k_1} \cdots l_{i_t}^{k_t}$, where $t$ is the number of nonzero parts in the partition $\kappa$, 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 $\kappa$ and $\lambda$ 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.** $EL^h_n = E_0 L^h_n \times {}_1 F_1(nh/2; (n + r - p + nh)/2; -\Omega/2)$.

*Proof.* The result follows from Theorem 10.5.1 in [Muirhead (2009)](Muirhead2009).

**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 $\tilde{M} = \text{diag}(l_1, \ldots, l_{m_0})$. Then for given partition $\kappa$, 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 $\sum_{\{i_1, \ldots, i_t\}} l_{i_1}^{k_1} \cdots l_{i_t}^{k_t} = \sum_{\{\tilde{i}_1, \ldots, \tilde{i}_t\}} l_{\tilde{i}_1}^{k_1} \cdots l_{\tilde{i}_t}^{k_t}$, 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 $\tilde{M} = \text{diag}(l_1, \ldots, l_{m_0})$.

**Lemma 11.** Suppose $Q$ has fixed rank $m_0$, then $\ {}_1 F_1(\xi_a; \xi_b; Q) = \ {}_1 F_1(\xi_a; \xi_b; \tilde{Q})$.

*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)](Muirhead2009), we know
that when $\kappa$ is a partition of $k$ into more than $m_0$ nonzero parts, $C_\kappa(Q) = 0$. Second, when $\kappa$ 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 $C_\kappa(Q) = \sum_{\lambda \leq \kappa} c_{\lambda, \kappa} N_\lambda(Q)$ 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

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

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 $\tilde{\Delta} = 0_{m_0 \times m_0}$ with $W(0_{m_0 \times m_0}) = 0$.

**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 $\lambda_{\text{max}}(\cdot)$ and $\lambda_{\text{min}}(\cdot)$ to denote the largest and smallest eigenvalues of a matrix respectively; $\text{diag}(\cdot)$ denotes the vector of diagonal elements of a matrix; $\max \text{ diag}(\cdot)$ and $\min \text{ diag}(\cdot)$ represent the maximum and minimum value of the diagonal elements of a matrix respectively; $\|\cdot\|$ denotes the $\ell_2$-norm of a vector; and $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)^\top$ denotes the indicator vector with 1 on the $i$th entry.

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

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

**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 $\pi \in (0, 1 - 2\kappa - 2u - t - s); \delta n^{1-\iota-\tau} \to \infty$ as $n \to \infty$.

**Remark 1.** In Condition 7, we assume that $X$ and $E$ follow the Gaussian distribution
for the ease of theoretical developments. We allow the eigenvalues of $\Sigma_x$ and $\Sigma$ to diverge or degenerate as $n$ grows, which is similarly assumed in Fan and Li (2008) and Wang and Leng (2016) etc. in studying the linear regression with univariate response. The boundedness of the diagonal elements of $\Sigma$ and $B^\top \Sigma_x B$ 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 $i \in M_*$, there exists $k_i \in \{1, \ldots, m\}$ such that $\text{cov}(x_{1,i}, y_{1,k_i}) \sigma_{x,i}^{-1} \geq c_7 n^{-\kappa}$. 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 $\Sigma_x$, $B^\top \Sigma_x B$ and $\Sigma$ do not diverge or degenerate too fast with $n$, and the covariance between $x_{1,i}$ and $y_{1}^\top a_{0,i}$ 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 $\omega_j$, we could assume $E(X) = 0$ 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 \( Z^\top Z \) has the eigendecomposition

\[
Z^\top Z = U^\top \text{diag}(\mu_1, \ldots, \mu_n, 0, \ldots, 0) U,
\]

where \( U \) belongs to the orthogonal group \( O(p) \). We write \( U^\top = (u_1, \ldots, u_p) \). It follows that the Moore-Penrose generalized inverse of \( Z^\top Z \) is

\[
(Z^\top Z)^+ = \sum_{i=1}^n \frac{1}{\mu_i} u_i u_i^\top.
\]

Moreover, we have the decomposition

\[
S := (Z^\top Z)^+ Z^\top Z = U^\top \text{diag}(I_n, 0) U = \tilde{U}^\top \tilde{U},
\]

where \( \tilde{U} = (I_n, 0)_{n \times p} U \) and \( (I_n, 0)_{n \times p} \) 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

\[
X^\top X = \Sigma_x^{1/2} \tilde{U}^\top \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_x^{1/2}.
\]
In addition, define $P = I_n - \frac{1_n 1_n^\top}{n}$. We can then write $\omega_i$ equivalently as

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

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

$$
\zeta_i = \max_{a : \|a\| = 1} \frac{a^\top Y^\top P^\top P x^i}{\sigma_{x,i} \sqrt{n} \times a^\top (Y^\top P^\top PY a)}
$$

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

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

where

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

Moreover, we write $B = [\beta_1, \ldots, \beta_m]$, where $\beta_j$ represents the $j$-th column of $B$. We then study $\xi_i$ and $\eta_i$ separately.

**Step 1:** We first examine $\xi = (\xi_1, \ldots, \xi_p)^\top$. 
Step 1.1 (bounding $\|\xi\|$ from above) For $i = 1, \ldots, p$,

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

where $\|\cdot\|$ represents the $\ell_2$-norm of a vector. Then we know

$$\|\xi\|^2 = \sum_{i=1}^{p} \xi_i^2 \leq \{n\lambda_{\min}(Y^T P^T P Y)\}^{-1} \sum_{i=1}^{p} \sigma_{x,i}^{-2} \|B^T X^T P^T P x^i\|^2$$ \hspace{2cm} (S5.61)

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

$$\sum_{i=1}^{p} \sigma_{x,i}^{-2} \|B^T X^T P^T P x^i\|^2 \leq c_3^{-1} \sum_{i=1}^{p} \sum_{k=1}^{m} (\beta_k^T X^T P^T P x^i)^2$$

$$= c_3^{-1} \sum_{k=1}^{m} \beta_k^T X^T P^T P \sum_{i=1}^{p} (x^i)^T P^T P X \beta_k.$$ \hspace{2cm} (S5.62)

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

$$\|\beta_k^T X^T (I_n - 1_n 1_n^T / n) X\|^2$$

$$\leq 2c_3^{-1} \times (A_{\xi,1} + A_{\xi,2}),$$ \hspace{2cm} (S5.63)

where $A_{\xi,1} = \sum_{k=1}^{m} \|\beta_k^T X^T X\|^2$ and $A_{\xi,2} = \sum_{k=1}^{m} \|\beta_k^T X^T (1_n 1_n^T / n) X\|^2$. 

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

$$
A_{\xi,1} = \sum_{k=1}^{m} \beta_k^\top \Sigma_{x}^{1/2} \tilde{U} \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_{x}^{1/2} \tilde{U} \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_{x}^{1/2} \beta_k
$$

$$
\leq p^2 \left\{ \lambda_{\max}(p^{-1} ZZ^\top) \right\}^2 \lambda_{\max}(\Sigma_x) \sum_{k=1}^{m} \beta_k^\top \Sigma_{x}^{1/2} \tilde{U} \Sigma_{x}^{1/2} \beta_k,
$$

(S5.64)

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

$$
\beta_k^\top \Sigma_{x}^{1/2} \tilde{U} \Sigma_{x}^{1/2} \beta_k = \|\Sigma_{x}^{1/2} \beta_k\|^2 \langle Q^\top SQ e_1, e_1 \rangle \equiv \|\Sigma_{x}^{1/2} \beta_k\|^2 \langle S e_1, e_1 \rangle.
$$

(S5.65)

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

$$
P(\beta_k^\top \Sigma_{x}^{1/2} \tilde{U} \Sigma_{x}^{1/2} \beta_k > c_1 n/p) \leq O\{\exp(-c_0 n)\}.
$$

(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 $1 - O\{m \exp(-c_0 n)\}$, $A_{\xi,1} \leq cmp^2 n^\top n/p = cmp n^{1+\tau}$. 
For $A_{\xi,2}$, note that

$$A_{\xi,2} = \sum_{k=1}^{m} \beta_k^\top \Sigma_x^{1/2} Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z \Sigma_x^{1/2} \beta_k.$$  \hspace{1cm} (S5.67)

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

$$\beta_k^\top \Sigma_x^{1/2} Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z \Sigma_x^{1/2} \beta_k = \|\Sigma_x^{1/2} \beta_k\|^2 e_1^\top Q^\top Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z Q e_1$$

$$\leq \lambda_{\max}(\Sigma_x) \|\Sigma_x^{1/2} \beta_k\|^2 e_1^\top Q^\top Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z Q^\top Z (1_n 1_n^\top / n) Z Q e_1$$

$$\overset{(d)}{=} \lambda_{\max}(\Sigma_x) \|\Sigma_x^{1/2} \beta_k\|^2 \| Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z e_1\|^2,$$  \hspace{1cm} (S5.68)

where in the last equality, we use the fact that $Z \overset{d}{=} Z$. Since the entries in $Z$ are i.i.d. $\mathcal{N}(0, 1)$, we have $L = (L_1, \ldots, L_p)^\top = Z \Sigma_x^{1/2} (1_n 1_n^\top / n) \sim \mathcal{N}(0_{p \times 1}, I_p)$ with $L_1 = 1_n^\top Z e_1 / \sqrt{n}$. It follows that $\|Z \Sigma_x^{1/2} (1_n 1_n^\top / n) Z e_1\|^2 = \sum_{j=1}^{p} L_j^2 = \sum_{j=1}^{p} L_j^2 = \chi_1^2$. 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. $\chi_1^2$-distributed random variables. By Lemma 13, for some positive constants $c_0$ and $c_1$, when $p \geq n$,

$$P\left(\sum_{j=2}^{p} L_j^2 / (p - 1) > 1 + c_1 \right) \leq O\{\exp(-c_0 p)\} \leq O\{\exp(-c_0 n)\}.  \hspace{1cm} (S5.69)$$
Thus there exist constants $c$ and $c_0$ such that with probability $1 - O\{\exp(-c_0n)\}$,

$$\|Z^\top(1_n^\top 1_n/n)Ze_1\|^2 = \sum_{j=1}^p L_1^2 L_2^2 = L_1^4 + L_1^3 \sum_{j=2}^p L_2^2 \leq cpn.$$ 

By Condition 1, $\lambda_{\max}(\Sigma_x) \leq c_1 n^\tau$ and $\|\Sigma_x^{1/2} \beta_k\|^2 \leq c_2$ for some constant $c_1$ and $c_2$. From (S5.67) and (S5.68), we know that $A_{\xi,2} \leq cmpn^{\tau+1}$ with probability $1 - O\{m \exp(-c_0n)\}$.

In summary, we obtain that for some constants $c$ and $c_0$, $A_{\xi,1}$ and $A_{\xi,2} \leq cmpn^{\tau+1}$ with probability $1 - O\{\exp(-c_0n)\}$. Then by (S5.61), (S5.63) and Lemma 19 we have for some positive constants $c_1$ and $c_0$,

$$P\{\|\xi\|^2 > c_1 n^{-(1+1-t)} pmn^{1+\tau}\} \leq O\{m \exp(-c_0n)\} = O\{\exp(-c_0n)\}, \quad (S5.70)$$

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 $B \neq 0_{p \times m}$. For fixed $i \in \mathcal{M}_*$,

$$\xi_i = \max_{a: \|a\|_2 = 1} \frac{a^\top B^\top X^\top P^\top Px_i}{\sigma_{x,i} \sqrt{n \times a^\top (Y^\top P^\top PY)a}} \geq \{n \times a_{0,i}^\top Y^\top P^\top PYa_{0,i}\}^{-1/2} \sigma_{x,i}^{-1} |a_{0,i}^\top B^\top X^\top P^\top PXe_i|, \quad (S5.71)$$
where $a_{0,i}$ in the last inequality is specified in Condition 2. To bound $|\xi_i|$ from below, we then examine (5.71). By Lemma 19, there exist constants $c_0$ and $c_2$ such that with probability $1 - O\{\exp(-c_0n)\},$

$$a_{0,i}^\top Y^\top P^\top PYa_{0,i} \leq c_2 n^{2u+1}. \quad (5.72)$$

Moreover, as $P^\top P = I_n - n_1n_1^\top /n$, $\sigma^{-1}_{x,i}a_{0,i}^\top B^\top X^\top P^\top Px_i = \tilde{A}_{\xi,1,i} - \tilde{A}_{\xi,2,i}$, where

$$\tilde{A}_{\xi,1,i} = \sigma^{-1}_{x,i}a_{0,i}^\top B^\top X^\top Xe_i$$

and

$$\tilde{A}_{\xi,2,i} = \sigma^{-1}_{x,i}a_{0,i}^\top B^\top X^\top (1_n1_n^\top /n)Xe_i.$$

We first consider $\tilde{A}_{\xi,1,i}$. From (5.58),

$$\sigma^{-1}_{x,i}a_{0,i}^\top B^\top X^\top Xe_i = \sigma^{-1}_{x,i}a_{0,i}^\top B^\top \Sigma_{x,i}^{1/2} \tilde{U}^\top \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \Sigma_{x,i}^{1/2} e_i. \quad (5.73)$$

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

$$\Sigma_{x,i}^{1/2} B a_{0,i} - \langle \Sigma_{x,i}^{1/2} B a_{0,i}, \Sigma_{x,i}^{1/2} e_i \sigma_{x,i}^{-1} \rangle \Sigma_{x,i}^{1/2} e_i \sigma_{x,i}^{-1}$$

$$= (\|\Sigma_{x,i}^{1/2} B a_{0,i}\|^2 - \langle \Sigma_{x,i}^{1/2} B a_{0,i}, \Sigma_{x,i}^{1/2} e_i \sigma_{x,i}^{-1} \rangle^2)^{1/2} \tilde{Q} e_2. \quad (5.74)$$

By Condition 2, there exists constant $c$ such that $\|\Sigma_{x,i}^{1/2} B a_{0,i}\| \leq cn^u$. Thus

$$\Sigma_{x,i}^{1/2} B a_{0,i} = \langle \Sigma_{x,i}^{1/2} B a_{0,i}, \Sigma_{x,i}^{1/2} e_i \sigma_{x,i}^{-1} \rangle \tilde{Q} e_1 + O(n^u) \tilde{Q} e_2. \quad (5.75)$$
Let \( T_{\eta,1} = \tilde{U}^\top \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} \tilde{Q} e_1 \). As \( \tilde{Q} e_1 = \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \), it follows that

\[
\text{(S5.73)} = \left( \Sigma_x^{1/2} B a_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \right) e_1^\top \tilde{Q}^\top T_{\eta,1} + O(n^u) e_2^\top \tilde{Q}^\top T_{\eta,1}. \tag{S5.76}
\]

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

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

where \( \xi_{i,1} = \left( \Sigma_x^{1/2} B a_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1} \right) R_1 \) 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 \( \mu_1, \ldots, \mu_n \geq p \lambda_{\text{min}} (p^{-1} ZZ^\top) \), and by \( \text{(S5.57)} \), we have

\[
R_1 \geq p e_1^\top \tilde{U}^\top \lambda_{\text{min}} (p^{-1} ZZ^\top) I_n \tilde{U} e_1 = p \lambda_{\text{min}} (p^{-1} ZZ^\top) \langle S e_1, e_1 \rangle.
\]

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 $c_0$,

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

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

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

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

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

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

$$P(|\xi_{i,2}| > c_1 n^{u} x_n) \leq O\{\exp(-c_0 n^{1-2\kappa-2u}/\log n)\}, \quad (S5.81)$$
where \( n^u x_n = \sqrt{2c_2 n^{1-k}} / \sqrt{\log n} \).

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 (5.74). As \( Z \tilde{Q} \overset{(d)}{=} Z \) and by (5.75), similarly to (S5.77), we have

\[
\tilde{A}_{\xi,i,2} = a_{0,i}^\top B_1^\top \Sigma_x^{1/2} Z(1_n^\top 1_n/n)Z \tilde{Q} e_1 \overset{(d)}{=} \tilde{\xi}_{i,1} + \tilde{\xi}_{i,2},
\]

where \( \tilde{\xi}_{i,1} = (\Sigma_x^{1/2} B a_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1}) e_1^\top Z 1_n^\top 1_n Z e_1/n \) and \( \tilde{\xi}_{i,2} = O(n^u) e_2^\top Z 1_n^\top 1_n Z e_1/n \).

Note that \( 1_n^\top Z e_1/\sqrt{n} \sim \mathcal{N}(0, 1) \) and \( 1_n^\top Z e_2/\sqrt{n} \sim \mathcal{N}(0, 1) \) independently. Then for some positive constants \( c_1 \) and \( c_0 \),

\[
P(|e_1^\top 1_n^\top 1_n Z e_1/n| > c_1 n^{1-k-u} / \log n) \leq O\{\exp(-c_0 n^{1-k-u} / \log n)\},
\]

\[
P(|e_2^\top 1_n^\top 1_n Z e_1/n| > c_1 n^{1-k-u} / \log n) \leq O\{\exp(-c_0 n^{1-k-u} / \log n)\}.
\]

These, combined with (5.78), show that there exist some constants \( c_1 \) and \( c_0 \) such that

\[
P(|\tilde{\xi}_{i,1}| > c_1 |\xi_{i,1}| n^{-u-k} / \log n) \leq P(|\tilde{\xi}_{i,1}| > c_1 (\Sigma_x^{1/2} B a_{0,i}, \Sigma_x^{1/2} e_i \sigma_{x,i}^{-1}) n^{1-k-u} / \log n)
\]

\[+P(R_1 < c_1 n) \leq O\{\exp(-c_0 n^{1-k-u} / \log n)\},
\]

\[
P(|\tilde{\xi}_{i,2}| > c_1 n^{1-k} / \log n) \leq O\{\exp(-c_0 n^{1-k-u} / \log n)\}.
\]
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 $c_0$,

$$P(|\xi_i| < c_1(n \times n^{2u+1})^{-1/2}n^{1-\kappa}) \leq O\left[\exp\{-c_0n^{1-2\kappa-2u}/\log n\}\right], \quad i \in \mathcal{M}_\ast.$$  \hspace{1cm} (S5.83)

**Step 2** We next examine $\eta = (\eta_1, \ldots, \eta_p)\top$ defined in (S5.60).

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

$$\eta_i = \max_{a : \|a\|_2 = 1} \frac{a\top E\top P\top Px_i}{\sigma_{x,i}\sqrt{n} \times a\top Y\top P\top PYa} \leq \left\{n\lambda_{\min}(Y\top P\top PY)\right\}^{-1/2}c_3^{-1}\|E\top P\top Px_i\|.$$  \hspace{1cm} (S5.84)

Let $e_j$ denote the j-th column of $E$, then $E = [e^1, \ldots, e^m]$. As $P\top P = I_n - 1_n1_n\top/n$, we have $\|E\top P\top Px_i\|^2 = \sum_{j=1}^m \{(e_j)\top x_i - (e_j)\top 1_n 1_n\top x_j/n\}^2$. Note that $\sum_{i=1}^p x_i(x_i)\top = XX\top$. Then by (S5.84),

$$\sum_{i=1}^p \eta_i^2 \leq c_3^{-2}\left\{n\lambda_{\min}(Y\top P\top PY)\right\}^{-1} \sum_{i=1}^p \sum_{j=1}^m 2 \times \left[\{(e_j)\top x_i\}^2 + \{(e_j)\top 1_n 1_n\top x_j/n\}^2\right]$$

$$= 2c_3^{-2}\left\{n\lambda_{\min}(Y\top P\top PY)\right\}^{-1} \sum_{j=1}^m \{(e_j)\top X X\top e_j + (e_j)\top 1_n 1_n\top X X\top 1_n 1_n\top e_j/n^2\}$$

$$= 2c_3^{-2}\left\{n\lambda_{\min}(Y\top P\top PY)\right\}^{-1} \sum_{j=1}^m \{(e_j)\top X X\top e_j + (e_j)\top 1_n 1_n\top X X\top 1_n 1_n\top e_j/n^2\} \leq 2c_3^{-2}\left\{n\lambda_{\min}(Y\top P\top PY)\right\}^{-1}\lambda_{\max}(\Sigma_x) \sum_{j=1}^m (A_{\eta,j,1} + A_{\eta,j,2}),$$  \hspace{1cm} (S5.85)
where $A_{\eta,j,1} = (\epsilon^j)^\tau ZZ^\tau \epsilon^j$ and $A_{\eta,j,2} = (\epsilon^j)^\tau 1_n 1_n^\tau ZZ^\tau 1_n 1_n \epsilon^j / n^2$.

Note that $A_{\eta,j,1} \leq p \lambda_{\max}(p^{-1} ZZ^\tau) \| \epsilon^j \|^2$. Suppose $\text{diag}(\Sigma) = (\sigma_{\epsilon,1}^2, \ldots, \sigma_{\epsilon,m}^2)^\tau$.

Then by Condition 1 and Lemma 13, we know for some positive constants $c$ and $c_0$,

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

In addition, $A_{\eta,j,2} \leq p \lambda_{\max}(p^{-1} ZZ^\tau) \times (1_n^\tau \epsilon^j)^2 / n$. Similarly to (S5.86), by Condition 1

and the tail bound of the Chi-squared distribution, there exist some positive constants $c$ and $c_0$,

$$P\{(1_n^\tau \epsilon^j)^2 / n > c n \sigma_{\epsilon,j}^2 / \log n\} \leq O\{\exp(-c_0 n / \log n)\}. \quad (S5.87)$$

Combining (S5.86) and (S5.87), we know that for some constants $c_1, c_2$ and $c_0$, with

probability $1 - O\{m \exp(-c_0 n / \log n)\}$,

$$A_{\eta,j,1} + A_{\eta,j,2} \leq c_1 p n \sum_{j=1}^m \sigma_{\epsilon,j}^2 / \log n \leq c_2 p n m / \log n, \quad (S5.88)$$

where the last inequality is from $\text{diag}(\Sigma) \leq c_4$ 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 $c_0$,  

$$
P(\|\eta\|^2 > c_1\{n \times n^{1-t}\}^{-1}p n^{1+r}m/ \log n) \leq O\{m \exp(-c_2n/ \log n)\} = O\{\exp(-c_0n/ \log n)\}.
$$  \text{(S5.89)}

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

$$
\eta_i^2 \leq \{n \lambda_{\min}(Y^T P^T P Y)\}^{-1} \sigma_{x,i}^{-2} \sum_{j=1}^m (\epsilon_j^T P^T P x_i)^2.
$$  \text{(S5.90)}

Then conditioning on $X$, $\sigma_{x,i}^{-1} \epsilon_j^T P^T P x_i \sim \mathcal{N}(0, \sigma_{\epsilon,i,j}^2 (x^i)^T P^T P x_i \sigma_{x,i}^{-2})$. Let $E_1$ be the event $\{\text{var}(\sigma_{x,i}^{-1} \epsilon_j^T P^T P x_i | X) \leq c_1 n\}$ for some constant $c_1 > 0$. Note that  

$$
\text{var}(\epsilon_j^T P^T P x_i \sigma_{x,i}^{-1} | X) = \sigma_{\epsilon,i,j}^2 \{(x^i)^T x_i - (x^i)^T 1_n 1_n x^i / n\} \sigma_{x,i}^{-2} \leq \sigma_{\epsilon,i,j}^2 (x^i)^T x_i \sigma_{x,i}^{-2}.
$$

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

$$
P(E_1) \leq P\{\sigma_{\epsilon,i,j}^2 (x^i)^T x_i \sigma_{x,i}^{-2} > c_1 n\} \leq O\{\exp(-c_0n)\},
$$  \text{(S5.91)}
where $E_1^c$ represents the complement of the event $E_1$. On the event $E_1$, for any $a > 0$, by Condition 1 we have

$$P(|\epsilon_j^T P x^i|\sigma_{x,i}^{-1} > a|X, E_1) \leq P\{\sqrt{c_1 n}|W| > a\}, \quad \text{(S5.92)}$$

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

$$P(|\epsilon_j^T P x^i|\sigma_{x,i}^{-1} > a) \leq O\{\exp(-c_0 n)\} + P\{\sqrt{c_1 n}|W| > a\}. \quad \text{(S5.93)}$$

Let $x'_n = \sqrt{2c_0 c_1 n^{1-\kappa-t/2-s/2-u}/\sqrt{\log n}}$. 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

$$P(|\eta_i| > (n^{1+1-t})^{-1/2} x'_n \sqrt{m}) \leq \sum_{j=1}^m P(|\epsilon_j^T P x^i| > x'_n) + P\{\lambda_{\min}(Y^T P Y) < c_1 n^{1-t}\},$$

where $(n^{1+1-t})^{-1/2} x'_n \sqrt{m} = \sqrt{2c_0 c_1 m n^{-\kappa-s/2-u}/\sqrt{\log n}} \leq c_2 n^{-\kappa-u}/\sqrt{\log n}$ for some
constant $c_2 > 0$ by Condition 3. In summary, we have

\[
P[\max_{1 \leq i \leq p} |\eta_i| > c_2 n^{-\kappa-u} / \sqrt{\log n}] \\
\leq O[p \exp\{-c_2 n^{1-2\kappa-t-s-2u} / \log n\}] \\
= O[\exp\{-c_0 n^{1-2\kappa-t-s-2u} / \log n\}, \quad (S5.94)
\]

where the last equality is from Condition 3.

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

\[
P\{\min_{i \in \mathcal{M}_s} |\zeta_i| < \tilde{c}_1 (n \times n^{2u+1})^{-1/2} n^{1-\kappa} \text{ or } \|\zeta\|^2 > \tilde{c}_2 (n \times n^{1-t})^{-1} n^{1+\tau} p m\} \\
\leq O[|\mathcal{M}_s| \exp\{-\tilde{c} n^{1-2\kappa-2u-t-s} / \log n\}]. \quad (S5.95)
\]

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

\[
P\{\min_{i \in \mathcal{M}_s} |\omega_i| < c_1 (n \times n^{2u+1})^{-1/2} n^{1-\kappa} \text{ or } \|\omega\|^2 > c_2 (n \times n^{-t})^{-1} n^{1+\tau} p m\} \\
\leq O[|\mathcal{M}_s| \exp\{-cn^{1-2\kappa-2u-t-s} / \log n\}],
\]
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 overwhelming 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 \( c_2 \),

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

where the last inequality is from Condition 3. Thus, if the proportion \( \delta \) of features selected satisfies \( \delta n^{1-2\kappa-2u-\tau-t-s} \to \infty \), then \( \delta p \geq c_2 pn^{s+\tau+t+2\kappa+2u-1} \) when \( \delta \) 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)]) \[ \text{Let } \vartheta_i, i = 1, 2, \ldots, n \text{ be i.i.d. } \chi^2_1 \text{-distributed random variables. Then for any } \epsilon > 0, \text{ we have } P(n^{-1} \sum_{i=1}^n \vartheta_i > 1 + \epsilon) \leq \exp(-A_\epsilon n), \text{ where } A_\epsilon = [\epsilon - \log(1 + \epsilon)]/2 > 0; \text{ for any } \epsilon \in (0, 1), \text{ } P(n^{-1} \sum_{i=1}^n \vartheta_i < 1 - \epsilon) \leq \exp(-B_\epsilon n), \text{ where } B_\epsilon = [-\epsilon - \log(1 - \epsilon)]/2 > 0. \]

**Lemma 14** (Lemma 1 in [Fan and Lv (2008)]) \[ \text{For } U \text{ and } (\mu_1, \ldots, \mu_n)^\top \text{ in } (S5.56), \]
and $\tilde{O}$ uniformly distributed on the orthogonal group $O(p)$, we know that

$$(I_n, 0)_{n \times p}^{(d)} U \equiv (I_n, 0)_{n \times p} \tilde{O}, \quad (S5.96)$$

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

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 $V_1 \in O(n)$, $U \in O(p)$ 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. $N(0, 1)$, for any $\tilde{O} \in O(p)$, $Z \tilde{O}^{(d)} = Z$. Thus, conditional on $V_1$ and $(\mu_1, \ldots, \mu_n)^\top$, the conditional distribution of $(I_n, 0)_{n \times p} U$ is invariant under $O(p)$. This shows that (S5.96) holds for $\tilde{O}$ uniformly distributed on the orthogonal group $O(p)$, and $(\mu_1, \ldots, \mu_n)^\top$ is independent of $(I_n, 0)_{n \times p} U$. \qed

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

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

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. $N(0, 1)$. For some constants
There exist some constants $c_1 > 1$, $c > 0$ and $c_0 > 0$, when $n > c$,

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

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

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

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

$$\lambda_{\text{max}}\{(P\tilde{Z})^\top P\tilde{Z}\} \leq \lambda_{\text{max}}(\tilde{Z}^\top \tilde{Z}) + \lambda_{\text{max}}(-\tilde{Z}^\top 1_n 1_n^\top \tilde{Z}/n),$$

$$\lambda_{\text{min}}\{(P\tilde{Z})^\top P\tilde{Z}\} \geq \lambda_{\text{min}}(\tilde{Z}^\top \tilde{Z}) + \lambda_{\text{min}}(-\tilde{Z}^\top 1_n 1_n^\top \tilde{Z}/n). \tag{S5.100}$$

Let $A_Z = \tilde{Z}^\top 1_n 1_n^\top \tilde{Z}/n$. As $\text{rank}(A_Z) = 1$ and $\text{tr}(A_Z) \geq 0$, we know $\lambda_{\text{max}}(-A_Z) =$
\[-\lambda_{\min}(A_Z) = 0 \quad \text{and} \quad \lambda_{\min}(-A_Z) = -\lambda_{\max}(A_Z) = -\text{tr}(A_Z) = -1^n_1 \tilde{Z}^\top 1_n/n.\]

We then examine $1^n_1 \tilde{Z}^\top 1_n/n$. For $z_{ij} \sim \mathcal{N}(0, 1)$ independently,

$$1^n_1 \tilde{Z}^\top 1_n/n = \sum_{j=1}^m \left( \sum_{i=1}^n z_{ij}/\sqrt{n} \right)^2 \sim \chi^2_m.$$  

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

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

This implies that with probability $1 - O\{\exp(-c_3 n/\log n)\}$, $\lambda_{\max}(A_Z/n) = 1^n_1 \tilde{Z}^\top 1_n/n^2 \leq c_2/\log n$ 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 $1 - \exp(-c_0 n/\log n)$,

$$\{\lambda_{\min}(n^{-1}(P\tilde{Z})^\top P\tilde{Z}) < 1/c_1\} \subseteq \{\lambda_{\min}(n^{-1}\tilde{Z}^\top \tilde{Z}) < 1/c_1 + c_2/\log n\} \subseteq \{\lambda_{\min}(n^{-1}\tilde{Z}^\top \tilde{Z}) < 1/\tilde{c}_1\},$$

$$\{\lambda_{\max}(n^{-1}(P\tilde{Z})^\top P\tilde{Z}) > c_1\} \subseteq \{\lambda_{\max}(n^{-1}\tilde{Z}^\top \tilde{Z}) > \tilde{c}_1\}.$$  

By (S5.99), (S5.98) is then proved.
Lemma 17. For $\mathbf{R} = (R_1, \ldots, R_p)^\top = \tilde{U}^\top \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U} e_1$, 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 $\tilde{\mathbf{R}} = (R_2, \ldots, R_p)^\top$. We first show that $\tilde{\mathbf{R}}$ is invariant under the orthogonal group $O(p - 1)$. For any $Q \in O(p - 1)$, let $\tilde{Q} = \text{diag}(1, Q) \in O(p)$. By Lemma 14, we know that $\tilde{U}$ is independent of $\text{diag}(\mu_1, \ldots, \mu_n)$ and $\tilde{Q} \tilde{U}^{(d)} \tilde{Q}^\top \tilde{U} = \tilde{U}^\top \text{diag}(\mu_1, \ldots, \mu_n) \tilde{U}$, where we use the fact that $\tilde{Q}^\top e_1 = e_1$. This implies that $\tilde{\mathbf{R}}$ is invariant under the orthogonal group $O(p - 1)$. It follows that $\tilde{\mathbf{R}}^{(d)} = \|\tilde{\mathbf{R}}\| W / \|W\|_2$, where $W = (W_1, \ldots, W_{p-1})^\top \sim \mathcal{N}(0, I_{p-1})$, independent of $\|\tilde{\mathbf{R}}\|$.

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

$$\|\mathbf{R}\|^2 \leq \{\lambda_{\max}(p^{-1} ZZ^\top)\}^2 p^2 e_1^\top \tilde{U}^\top \tilde{U} e_1 = \{\lambda_{\max}(p^{-1} ZZ^\top)\}^2 p^2 \langle S e_1, e_1 \rangle.$$
and $c_0$, $P\{\|W\|^2 < c_1(p-1)\} \leq \exp(-c_0 n)$ when $p > n$. Thus we obtain that for some constants $c_1$ and $c_0$, $P(|R_2| > c_1|W_1|n^{1/2}) \leq O\{\exp(-c_0 n)\}$, 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 $\omega_i$ and $\zeta_i$ and $P^\top P = P$, we know

$$\omega_i - \zeta_i \leq \max_a \frac{a^\top Y P x^i (\sqrt{n}\sigma_{x,i} - \sqrt{(x^i)^\top P x^i})}{\sqrt{(x^i)^\top P x^i}} = \zeta_i \frac{\sqrt{n}\sigma_{x,i} - \sqrt{(x^i)^\top P x^i}}{\sqrt{(x^i)^\top P x^i}},$$

$$\zeta_i - \omega_i \leq \max_a \frac{a^\top Y P^\top P x^i (\sqrt{(x^i)^\top P x^i} - \sqrt{n}\sigma_{x,i})}{\sqrt{(x^i)^\top P x^i}} = \zeta_i \frac{\sqrt{(x^i)^\top P x^i} - \sqrt{n}\sigma_{x,i}}{\sqrt{(x^i)^\top P x^i}}.$$

Thus

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

Let $\bar{x}^i = \sum_{k=1}^n x_{k,i} / n$, which is the mean of the entries in $x^i$. It follows that

$$(x^i)^\top P x^i = \sum_{k=1}^n (x_{k,i} - \bar{x}^i)^2 = \sum_{k=1}^n x_{k,i}^2 - n \bar{x}^i)^2.$$ Then with $\tilde{c}_1 = c_1 / 2$

$$P\{|(x^i)^\top P x^i / (n\sigma_{x,i}^2) - 1| > c_1 / \log n \} \leq P\{(\bar{x}^i)^2 / \sigma_{x,i}^2 > \tilde{c}_1 / \log n \} + P\left\{ \sum_{k=1}^n x_{k,i}^2 / (n\sigma_{x,i}^2) - 1 \right\} > \tilde{c}_1 / \log n \}.$$
By Condition 1, \((\sqrt{n}\bar{x})^2/\sigma^2_{x,i} \sim \chi^2_1\). Then by the tail of \(\chi^2_1\) distribution, for some constant \(c_0 > 0\), \(P\{\bar{x}^2/\sigma^2_{x,i} > \tilde{c}_1/\log n\} \leq O\{\exp(-c_0 n/\log n)\}\). In addition, \(x^2_{k,i}/\sigma^2_{x,i}, k = 1, \ldots, n\) are i.i.d. \(\chi^2_1\)-distributed random variables. By Lemma 13, there exists some constant \(c_0 > 0\),

\[
P\left\{ \left| \sum_{k=1}^{n} \frac{x^2_{k,i}}{n\sigma^2_{x,i}} - 1 \right| > \tilde{c}_1/\log n \right\} \leq O\{\exp(c_0 n/\log n)\}.
\]

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 \(1 - O\{\exp(-c_0 n/\log n)\}\),

\[
\lambda_{\min}(Y^T P^T P Y) \geq c_1 n^{1-t},
\]

and for \(a_{0,i}\) in Condition 4

\[
a_{0,i}^T Y^T P^T P Y a_{0,i} \leq c_2 n^{2u+1}. \quad \text{(S5.103)}
\]

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 $\Sigma_y = B^\top \Sigma_x B + \Sigma$. 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 $1 - O\{\exp(-c_0 n)\}$,

$$Y^\top P^\top P Y = \Sigma_y^{1/2} \tilde{Z}^\top P^\top P \tilde{Z} \Sigma_y^{1/2}$$

$$\succeq n \Sigma_y^{1/2} \lambda_{\min}(n^{-1} \tilde{Z}^\top P^\top P \tilde{Z}) I_m \Sigma_y^{1/2} \succeq c_1 n \Sigma_y.$$

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

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

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

$$a_{0,i}^\top Y^\top P^\top P Y a_{0,i} = a_{0,i}^\top \Sigma_y^{1/2} \tilde{Z}^\top P^\top P \tilde{Z} \Sigma_y^{1/2} a_{0,i}$$

$$\leq c_2 n a_{0,i}^\top \Sigma_y a_{0,i}$$

$$= c_2 n a_{0,i}^\top (B^\top \Sigma_x B + \Sigma) a_{0,i}$$

$$\leq c_2 n^{2u+1},$$

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

\[
\psi(z) = \frac{1}{J} \sum_{j=1}^{J} 1\{p^{(j)} \leq z\}. \tag{S6.105}
\]

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

\[
\sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1\{U \leq \alpha \gamma\}}{\gamma} = \begin{cases} 
0 & U \geq \alpha \\
\alpha/U & \alpha \gamma_{\min} \leq U < \alpha \\
1/\gamma_{\min} & U < \alpha \gamma_{\min}.
\end{cases}
\]

Thus when \( U \) has a uniform distribution on \([0, 1]\),

\[
\mathbb{E}\left[ \sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1\{U \leq \alpha \gamma\}}{\gamma} \right] = \int_{0}^{\alpha \gamma_{\min}} \gamma_{\min}^{-1} dx + \int_{\alpha \gamma_{\min}}^{\alpha} \alpha x^{-1} dx = \alpha (1 - \log \gamma_{\min}).
\]

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

\[
\mathbb{E}\left[ \sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1\{p^{(j)} \leq \alpha \gamma\}}{\gamma} \right] \leq \mathbb{E}\left[ \mathbb{E}\left[ \sup_{\gamma \in (\gamma_{\min}, 1)} \frac{1\{p^{(j)} \leq \alpha \gamma\}}{\gamma} \bigg| B^{(j)} \right] \right] + \frac{1}{\gamma_{\min}} \mathbb{P}\{B^{(j)}\} \\
\leq \alpha (1 - \log \gamma_{\min}) + O\{\exp\{-c_0 n^{-1} / \log n\}\},
\]
where the constant \( \iota \) is given in Theorem 5. Averaging over \( J \) splits yields

\[
E \left[ \sup_{\gamma \in (\gamma_{\min}, 1)} \mathbb{1}_{\frac{1}{J} \sum_{j=1}^{J} 1\{p^{(j)}/\gamma \leq \alpha\}} \right] \leq \alpha(1 - \log \gamma_{\min}) + O[\exp\{-c_0 n^{1-\iota}/\log n\}].
\]

From Markov inequality and (S6.105), \( E[\sup_{\gamma \in (\gamma_{\min}, 1)} 1\{\psi(\alpha \gamma) \geq \gamma\}] \leq \alpha(1-\log \gamma_{\min}) + O[\exp\{-c_0 n^{1-\iota}/\log n\}] \). Since \( \{Q(\gamma) \leq \alpha\} \) and \( \{\psi(\alpha \gamma) \geq \gamma\} \) are equivalent, it follows that \( P[\inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma) \leq \alpha] \leq \alpha(1 - \log \gamma_{\min}) + O[\exp\{-c_0 n^{1-\iota}/\log n\}] \), which implies that \( P[\inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma)(1 - \log \gamma_{\min}) \leq \alpha] \leq \alpha + O[\exp\{-c_0 n^{1-\iota}/\log n\}] \). By definition of \( p_t \), \( \limsup_{n \to \infty} P[p_t \leq \alpha] \leq \alpha \) 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 \( \chi^2 \) 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 \( \chi^2 \) approximation performs poorly, while the normal approximations for \( T_1 \) and \( T_3 \) still control the type I error well.
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 $\nu_d$ or $\sigma_d^2$ 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 $C = [I_r, 0_{r \times (p-r)}]$, where $I_r$ is an identity matrix of dimension $r \times r$, $0_{r \times (p-r)}$ is an all zero matrix of dimension $r \times (p-r)$. Then $H_0 : CB = 0_{r \times m}$ examines the relationship between $Y$ and the first $r$ predictors of $X$. In the second case, we take $C = [I_r, 0_{r \times (p-r-1)}, -1_r]$, where $1_r$ is an all 1 vector of length $r$, and $0_{r \times (p-r-1)}$ is an all zero matrix of dimension $r \times (p-r-1)$. Then $H_0 : CB = 0_{r \times m}$ 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 $C = [I_r, 0_{r \times (p-r)}]$.

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

Figure S2: Estimated powers versus signal sizes with $\rho = 0.7$.
Figure S3: Estimated powers versus signal sizes with $\rho = 0.5$

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

(b) Estimated powers versus signal sizes when $C = [I_r, 0_{r \times (p-r-1)}, -1_r]$
(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) \textit{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 \(z_{i,j} \sim \mathcal{N}(0,1)\), 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 $C = [I_r, 0 \times (p-r)]$ and $C = [I_r, 0 \times (p-r-1), -1_r]$ 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 $C = [I_r, 0_{r \times (p-r)}]$

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

Figure S5: Power comparison when $X$ and $Y$ follow multinomial distribution
(a) When \( C = \left[ I_r, \mathbf{0}_{r \times (p-r)} \right] \)

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

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](image)

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 \( V_J = (V_{J,1}, \ldots, V_{J,J})^T \sim \mathcal{N}(0, \Sigma_J) \) with \( \Sigma_J = (1 - \rho)I_{J,J} + \rho 1_J 1_J^T \). Note that larger \( \rho \) 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 \( \rho \) 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.
Figure S9: Estimated powers of two-step procedure with other distributions

(a) $X$ and $Y$ follow multinomial distributions

(b) Entries in $E$ follow $t_3$ distribution

(c) Entries in $E$ follow $t_5$ distribution
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 $C = [I_r, 0_{r 	imes (p - r)}]$, 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.

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 $\sigma_s$ in the first $r_k$ diagonal entries, where $\sigma_s$ 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 $\mathcal{N}(0, \sigma_s^2)$. 

Figure S11: Screening Comparison: $B$ is diagonal
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
provide the decision results in Table S1 where the notations follow the same meaning as in Table I. 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


