New Tests for High-Dimensional Linear Regression

Based on Random Projection

Changyu Liu, Xingqiu Zhao

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

Jian Huang

Department of Statistics and Actuarial Science, University of Iowa, Iowa, U.S.A.

Supplementary Material

S1 Lemmas

We first introduce some notation. For a matrix $B$, we denote the Frobenius norm of $B$ by $||B||_F = tr(B^T B)^{1/2}$ and the spectral norm of $B$ by $||B||_{sp} = \max_{||x||_2=1} ||Bx||_2$. If $B$ is symmetric, we use $B \succeq 0$ when $B$ is positive semi-definite.

S1.1 Proof of Lemma 3.1

We first state a result from [Fang, Kotz, and Ng 1990, Section 3.1], which shows some properties of uniform distribution on the surface of an unit
sphere.

**Lemma 1.** Let \( \mathbf{u}_1 = (u_{11}, ..., u_{1p})^\top \) be a random vector uniformly distributed on the unit sphere in \( \mathbb{R}^p \). Then \( \mathbf{u}_1 \) satisfies \( E(\mathbf{u}_1) = 0, \text{Var}(\mathbf{u}_1) = \frac{1}{p} I_p \). For \( \forall j \neq k \), \( E(u_{1j}^4) = \frac{3}{p(p+2)}, E(u_{1j}^2 u_{1k}^2) = \frac{1}{p(p+2)} \). And for any nonnegative integers \( q_1, ..., q_p \), with \( m = \sum_{j=1}^p q_j \), the mixed moments \( E(\Pi_{j=1}^p u_{1j}^{q_j}) = 0 \) if at least one \( q_j \) is odd.

**Proof of Lemma 3.1.** From the definition of \( r_1, \mathbf{u}_1 \) and Lemma 1, we have

\[
E(\mathbf{z}_1) = E(r_1 \mathbf{u}_1) = E(r_1) E(\mathbf{u}_1) = 0,
\]

\[
\text{Var}(\mathbf{z}_1) = \text{Var}(E(\mathbf{z}_1|r_1)) + E(\text{Var}(\mathbf{z}_1|r_1)) = E(r_1^2 \text{Var}(\mathbf{u}_1)) = I_p.
\]

By definition that \( \mathbf{z}_1 = (z_{11}, ..., z_{1p})^\top = r_1 \mathbf{u}_1 \), we have, for \( \forall i \neq j \),

\[
E(z_{1i}^4) = E(r_1^4 u_{1i}^4) = 3 + O(p^{-1}), \quad E(z_{1i}^2 z_{1j}^2) = E(r_1^2 u_{1i}^2 u_{1j}^2) = 1 + O(p^{-1}).
\]

Hence, we have

\[
\text{Var}\left(\frac{\mathbf{z}_1^\top \mathbf{z}_1}{p}\right) = \sum_{i=1}^p E(z_{1i}^4) + \sum_{i\neq j} E(z_{1i}^2 z_{1j}^2) - E\left(\frac{\mathbf{z}_1^\top \mathbf{z}_1}{p}\right)^2 = O(p^{-1}),
\]

and complete the proof. \( \Box \)

**S1.2 Auxiliary lemmas**

We first present a result of asymptotic normality of quadratic form that was discussed by Bhansali, Giraitis, and Kokoszka (2007).
Lemma 2. Consider a general quadratic form

\[ Q_n = z^\top A_n z = \sum_{i,j=1}^{n} z_i a_{ij} z_j, \]

where \(z_i\) are i.i.d. variables with \(E(z_i) = 0\) and \(\text{Var}(z_i) = 1\), and \(a_{ij}\) are entries of a symmetric matrix \(A_n\).

(1) If \(E(z_i^4) < \infty\) and \(\|A_n\|_{sp} \overset{\|A_n\|_F}{\longrightarrow} 0\), then

\[ \text{Var}(Q_n)^{-1/2}(Q_n - E(Q_n)) \overset{D}{\rightarrow} \mathcal{N}(0, 1). \]

(2) If \(\|A_n\|_{sp} \overset{\|A_n\|_F}{\longrightarrow} 0\), \(E(z_i^{2+\delta}) < \infty\) (for some \(\delta > 0\)), and \(\sum_{i=1}^{n} a_{ii}^2 = o(\|A_n\|_F^2)\), then

\[ \frac{1}{\sqrt{2\|A_n\|_F}}(Q_n - E(Q_n)) \overset{D}{\rightarrow} \mathcal{N}(0, 1). \]

Lemma 3 (Woodbury’s formula). Suppose \(G\) is an \(n \times n\) nonsingular matrix, \(U\) and \(V\) are \(n \times k\) matrices, with \(n > k\). If the matrix \((I_k + V^\top G^{-1} U)\) is invertible, we have

\[ (G + UV^\top)^{-1} = G^{-1} - G^{-1} U(I_k + V^\top G^{-1} U)^{-1} V^\top G^{-1}. \]

Suppose \(u\) and \(v\) are vectors. Define \(H = uv^\top\) and \(g = \text{tr}(HG^{-1})\). If \(g \neq -1\), we have

\[ (G + H)^{-1} = G^{-1} - \frac{1}{1 + g} G^{-1} H G^{-1}. \]
We then depict some results about sample covariance matrix in high dimensions. The first is the celebrated work of Marˇ cenko and Pastur (1967), which is named the M-P law by some authors. The second is concerned with the extreme eigenvalues from Bai and Yin (1993, Theorem 2).

Lemma 4. Let \( X = (x_{ij}) \in \mathbb{R}^{k \times n} \) be a matrix of i.i.d. entries with zero mean and unit variance. Define \( S_n = \frac{1}{n}XX^\top \). Suppose the eigenvalues of \( S_n \) are \( \lambda_j, j = 1, \ldots, k \), the empirical spectral distribution (ESD) of the matrix \( S_n \) is defined as \( F_{S_n} = \frac{1}{k} \sum_{j=1}^{k} \mathbb{1}_{\{\lambda_j \leq x\}} \). If \( E(x_{11}^4) < \infty \), as \((n,k) \to \infty \) with relationship \( k/n \to \rho \in (0,1) \), we have

(1) \( F_{S_n} \) tends to the standard M-P law with probability 1, where the standard M-P law \( F_{\rho}(x) \) has a density function

\[
p_{\rho}(x) = \begin{cases} \frac{1}{2\pi \rho} \sqrt{(b-x)(x-a)}, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise,} \end{cases}
\]

where \( a = (1 - \sqrt{\rho})^2 \) and \( b = (1 + \sqrt{\rho})^2 \).

(2) The extreme eigenvalues of \( S_n \) satisfy

\[
\lambda_{\max}(S_n) \to (1 + \sqrt{\rho})^2 \ a.s.,
\]

and

\[
\lambda_{\min}(S_n) \to (1 - \sqrt{\rho})^2 \ a.s..
\]
Lemma 5. Let $X = (x_1, \ldots, x_n)$ be a random matrix with $x_i$ i.i.d. from $\mathcal{N}(0, I_k)$. As $(k, n) \to \infty$ with relationship $k/n \to \rho \in (0, 1)$, we have

(1) $X(I-P_1)X^\top$ and $\bar{x}$ are independent, where $1 = (1, \ldots, 1)^\top$, $P_1 = \frac{1}{n}11^\top$ and $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.

(2) $E((\frac{1}{n-1}x_n^\top S_{n-1}^{-1}x_n - \frac{\rho}{1-\rho})^2) = o(1)$, where $S_{n-1} = \frac{1}{n-1} \sum_{j=1}^{n-1} x_j x_j^\top$, and $E((x_i^\top (XX^\top)^{-1}x_i - \rho)^2) = o(1)$, $x_i^\top (XX^\top)^{-1}x_i \leq \frac{1}{1+(1-\sqrt{\rho})^2}$, a.s..

Proof. (1) We first define an orthogonal matrix $O$ by

$$O = (o_1, \ldots, o_n) = \begin{bmatrix}
\frac{1}{\sqrt{n}} & 0 & 0 & \cdots & -\frac{\sqrt{n-1}}{\sqrt{n}} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{\sqrt{n}} & 0 & -\frac{\sqrt{2}}{\sqrt{3}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\
\frac{1}{\sqrt{n}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\
\frac{1}{\sqrt{n}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}}
\end{bmatrix}.$$  

Let $V = XO$ with the $i$th column denoted as $v_i$. Then the design of orthogonal matrix $O$ implies $XX^\top = XOO^\top X^\top = \sum_{i=1}^n v_i v_i^\top$, $v_1 = \sqrt{n} \bar{x}$ and $X(I-P_1)X^\top = \sum_{i=2}^n v_i v_i^\top$. To study the properties of $v_i$, the random matrix $X$ is divided by rows and denoted as $(r_1, \ldots, r_k)^\top$ with $k$ independent $\mathcal{N}(0, I_n)$ variables. It follows that $v_i = (r_1, \ldots, r_k)^\top o_i$ and is distributed as $\mathcal{N}(0, I_k)$. Let $C_{i,j}^{(i,j)} = (C_{i,j}^{(i,j)})_{s,l=1} = Cov(v_i, v_j)$, for
i \neq j. \text{ Then we have }

\begin{align*}
C_{s,l}^{i,j} &= E(r_s^T o_i r_l^T o_j) - E(r_s^T o_i)E(r_l^T o_j) = 0, \quad s \neq l, \\
C_{s,s}^{i,j} &= E(r_s^T o_i r_s^T o_j) - E(r_s^T o_i)E(r_s^T o_j) = E(o_i^T r_s r_s^T o_j) = 0, \quad s = 1, \ldots, k,
\end{align*}

which indicates \( v_i \) and \( v_j \) are independent. This is sufficient to show that \( X(I - P_1)X^T \) and \( \bar{x} \) are independent.

(2) From the direct calculation, the standard M-P law \( F_\rho(x) \) in Lemma 4 satisfies

\begin{align*}
\int \frac{1}{x} dF_\rho(x) &= \int_a^b \frac{1}{2\pi x^2 \rho} \sqrt{(b - x)(x - a)} dx \\
&= \frac{1}{2\pi \rho} \int_{-2\sqrt{\rho}}^{2\sqrt{\rho}} \frac{1}{(1 + \rho + z)^2} \sqrt{4\rho - z^2} dz \text{ (with } x = 1 + \rho + z) \\
&= \frac{1}{2\pi \rho} \int_{-\pi/2}^{\pi/2} \frac{4\rho \cos^2 \theta}{(1 + \rho + 2\sqrt{\rho} \sin \theta)^2} d\theta \text{ (with } z = 2\sqrt{\rho} \sin \theta) \\
&= \frac{1}{2\pi \rho} \left( \int_{-\pi/2}^{\pi/2} \frac{-2\sqrt{\rho} \cos \theta}{1 + \rho + 2\sqrt{\rho} \sin \theta} d\theta \right) + \int_{-\pi/2}^{\pi/2} \frac{-2\sqrt{\rho} \sin \theta}{1 + \rho + 2\sqrt{\rho} \sin \theta} d\theta \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi \rho} \int_{-\pi/2}^{\pi/2} \frac{1}{1 + \rho + 2\sqrt{\rho} \sin \theta} d\theta \text{ (with } c = 2\sqrt{\rho}(1 + \rho)^{-1} < 1) \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi \rho} \int_{-\pi/2}^{\pi/2} \frac{1}{\cos^2 \frac{\theta}{2} (1 + \tan^2 \frac{\theta}{2} + 2c \tan \frac{\theta}{2})} d\theta \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi \rho} \int_{-1}^{1} \frac{2}{1 + t^2 + 2ct} dt \text{ (with } t = \tan \frac{\theta}{2}) \\
&= -\frac{1}{2\rho} + \frac{1}{2\pi \rho} \cdot \frac{2}{\sqrt{1 - c^2}} \arctan \left( \frac{t + c}{\sqrt{1 - c^2}} \right)_{-1}^{1} = \frac{1}{1 - \rho}.
\end{align*}

We first study the asymptotic behavior of \( \frac{1}{n-1} x_n^T S_{n-1}^{-1} x_n \). From nor-
mality of \( x_i \), Lemma 4 and the above calculation, we have

\[
E\left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n | S_{n-1} \right) = \frac{k}{n-1} \frac{tr(S_{n-1}^{-1})}{k} \\
= \frac{k}{n-1} \int \frac{1}{x} dF S_{n-1}^{-1} \rightarrow \frac{\rho}{1-\rho}, \text{ a.s.,}
\]

\[
Var\left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n | S_{n-1} \right) = \frac{2}{(n-1)^2} tr((S_{n-1}^{-1})^2) \\
\leq \frac{2k}{(n-1)^2} \left( \frac{1}{\lambda_{\min}(S_{n-1})} \right)^2 \rightarrow 0, \text{ a.s.}
\]

Therefore,

\[
E\left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n \right) = E\left( E\left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n | S_{n-1} \right) \right) \rightarrow \frac{\rho}{1-\rho},
\]

\[
Var\left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n \right) \rightarrow 0.
\]

These lead to the first result,

\[
E \left( \left( \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n - \frac{\rho}{1-\rho} \right)^2 \right) \rightarrow 0.
\]

From Lemma 3, we have

\[
x_n^\top (XX^\top)^{-1} x_n = \frac{x_n^\top (\sum_{j \neq n} x_j x_j^\top)^{-1} x_n}{1 + x_n^\top (\sum_{j \neq n} x_j x_j^\top)^{-1} x_n} \\
= \frac{1}{1 + \frac{n-1}{n} x_n^\top S_{n-1}^{-1} x_n}. 
\]

Let \( f(x) = \frac{x}{1+x} \). Its derivative \( f'(x) = \frac{1}{(1+x)^2} \leq 1 \), for \( x \geq 0 \). From \( x_n^\top S_{n-1}^{-1} x_n \geq 0 \) and the mean value theorem, we get

\[
|x_n^\top (XX^\top)^{-1} x_n - \rho| \leq \left| \frac{1}{n-1} x_n^\top S_{n-1}^{-1} x_n - \frac{\rho}{1-\rho} \right|,
\]
which implies
\[
E \left( (x_n^T (XX^T)^{-1} x_n - \rho)^2 \right) \leq E \left( \left( \frac{1}{n-1} x_n^T S_{n-1}^{-1} x_n - \frac{\rho}{1-\rho} \right)^2 \right) \to 0.
\]

Furthermore, from
\[
\frac{1}{n-1} x_n^T S_{n-1}^{-1} x_n \leq \lambda_{\min}^{-1}(S_{n-1}) \frac{1}{n-1} x_n^T x_n \to \frac{1}{(1-\sqrt{\rho})^2}
\]
a.s., we obtain
\[
x_n^T (XX^T)^{-1} x_n \leq \frac{1}{1 + (1 - \sqrt{\rho})^2}, \quad \text{a.s.,}
\]
and complete the proof.

\[\square\]

**Lemma 6.** Let \( X = (x_1, \ldots, x_n) \) be a random matrix with \( x_i \) i.i.d. from \( \mathcal{N}(0, I_k) \). The matrix \( H \) is defined as \( H = (I - P_1) X^T (X(I - P_1) X^T)^{-1} X(I - P_1) \) and has its entries denoted by \( H_{ij} \). As \( (k,n) \to \infty \) with \( k/n \to \rho \in (0,1) \), we have
\[
\max_{i=1,\ldots,n} E \left[ (H_{ii} - \rho)^2 \right] \to 0.
\]

**Proof.** From Lemma 5 we get
\[
E \left( (nx^T (X(I - P_1) X^T)^{-1} x - \frac{\rho}{1-\rho})^2 \right) \to 0, \quad (S1.1)
\]
\[
E \left( (nx^T (XX^T)^{-1} x - \rho)^2 \right) \to 0, \quad n x^T (XX^T)^{-1} x \leq \frac{1}{1 + (1 - \sqrt{\rho})^2}, \quad \text{a.s.,} \quad (S1.2)
\]
\[
E \left( (x_1^T (XX^T)^{-1} x_1 - \rho)^2 \right) \to 0. \quad (S1.3)
\]

The proof proceeds in two steps. First, we study \( x_1^T (X(I - P_1) X^T)^{-1} x_1 \) and show that it converges to \( \rho \) in quadratic mean. Second, we divide
H_ii into three parts and investigate them separately. Then we reach the statement in the lemma and complete the proof.

In the first step, we would show x_1^T(X(I - P_1)X)^{-1}x_1 is a well approximation to x_1^T(X(I-P_1)X)^{-1}x_1 and then the convergence is guaranteed by (S1.3). Lemma 3 and (S1.2) imply

$$ (X(I - P_1)X^T)^{-1} = (XX^T)^{-1} + \frac{1}{1 + g} (XX^T)^{-1} n \bar{x} \bar{x}^T (XX^T)^{-1}, $$

where $g = -n \bar{x}^T (XX^T)^{-1} \bar{x} \geq -\frac{1}{1+(1-\sqrt{\rho})^2}$ a.s. is lower-bounded. Then, we have

$$ |x_1^T(X(I - P_1)X^T)^{-1}x_1 - x_1^T(XX^T)^{-1}x_1| $$

$$ = \frac{1}{1 + g} x_1^T(XX^T)^{-1} n \bar{x} \bar{x}^T (XX^T)^{-1} x_1 $$

$$ = \frac{n}{1 + g} (x_1^T(XX^T)^{-1} \bar{x})^2 $$

$$ \leq \frac{2}{1 + g} \left[ \frac{1}{n} (x_1^T(XX^T)^{-1} x_1)^2 + \frac{1}{n} \left( \sum_{j \neq 1} x_1^T(XX^T)^{-1} x_j \right)^2 \right]. $$

Based on (S1.3), the expectation of the first part in the sum goes to 0. Then we show the second part $\frac{1}{n} (\sum_{j \neq 1} x_1^T(XX^T)^{-1} x_j)^2$ would also converge to 0 in the first mean. Define $A_{1,j} = \sum_{k \neq 1,j} x_k x_k^T$ and $S_{1,j} = \frac{1}{n-2} A_{1,j}$. We have $XX^T = A_{1,j} + x_1 x_1^T + x_j x_j^T$. From Lemma 3,

$$ x_1^T(XX^T)^{-1} x_j = \frac{x_1^T A_{1,j}^{-1} x_j}{D_{1,j}}, $$
where $D_{1,j} = (1 + x_i^\top A_{1,j}^{-1} x_i)(1 + x_j^\top A_{1,j}^{-1} x_j) - (x_i^\top A_{1,j}^{-1} x_j)^2 \geq 1$. Then,

$$E\left(\frac{1}{n} \sum_{j \neq 1} x_i^\top (XX^\top)^{-1} x_j\right)^2 = E\left(\frac{1}{n} \sum_{j \neq 1} x_i^\top A_{1,j}^{-1} x_j\right)^2$$

$$= \sum_{j \neq 1} E\left(\frac{x_i^\top A_{1,j}^{-1} x_j}{n D_{1,j}}\right)^2 + \sum_{j \neq \ell \neq 1} E\left(\frac{x_j^\top A_{1,j}^{-1} A_{1,\ell}^{-1} x_\ell}{n D_{1,j} D_{1,\ell}}\right).$$

For any $j \neq \ell \neq 1$, we have

$$E\left(\frac{(x_i^\top A_{1,j}^{-1} x_j)^2}{D_{1,j}^2}\right) = E\left(\frac{(x_i^\top A_{1,2}^{-1} x_2)^2}{D_{1,2}^2}\right) \quad \text{and} \quad E\left(\frac{x_j^\top A_{1,j}^{-1} A_{1,\ell}^{-1} x_\ell}{D_{1,j} D_{1,\ell}}\right) = E\left(\frac{x_2^\top A_{1,2}^{-1} A_{1,3}^{-1} x_3}{D_{1,2} D_{1,3}}\right).$$

Therefore,

$$E\left(\frac{1}{n} \sum_{j \neq 1} x_i^\top (XX^\top)^{-1} x_j\right)^2 = \frac{n-1}{n} E\left(\frac{(x_i^\top A_{1,2}^{-1} x_2)^2}{D_{1,2}^2}\right) + \frac{(n-1)(n-2)}{n} E\left(\frac{x_2^\top A_{1,2}^{-1} A_{1,3}^{-1} x_3}{D_{1,2} D_{1,3}}\right).$$

(S1.4)

Lemma 4 asserts the first part in (S1.4) converges to 0 by

$$E\left(\frac{(x_i^\top A_{1,2}^{-1} x_2)^2}{D_{1,2}^2}\right) \leq E\left((x_i^\top A_{1,2}^{-1} x_2)^2\right) = \frac{k}{(n-2)^2} E\left(\frac{tr(S_{1,2}^{-1})^2}{k}\right) \to 0.$$ 

Next, we study the second part and show it would also go to 0. Let $A_{1,2,3} = \sum s \neq 1,2,3 x_s x_s^\top$, $S_{1,2,3} = \frac{1}{n-3} A_{1,2,3}$. Then, $g_3 = x_3^\top A_{1,2,3}^{-1} x_3 \geq 0$ and Lemma 3 gives the relationship

$$A_{1,2}^{-1} = A_{1,2,3}^{-1} - \frac{1}{1 + g_3} A_{1,2,3}^{-1} x_3 x_3^\top A_{1,2,3}^{-1}.$$
From calculations and Lemma 4 we have

\[ E(x_2^T (A_{1,2,3}^{-1})^2 x_3) = 0, \quad E((x_2^T (A_{1,2,3}^{-1})^2 x_3)^2 | A_{1,2,3}) = \frac{tr((S_{1,2,3}^{-1})^4)}{(n-3)^4} = O(n^{-3}), \]

\[ E(x_2^T A_{1,2,3}^{-1} x_3) = 0, \quad E((x_2^T A_{1,2,3}^{-1} x_3)^2 | A_{1,2,3}) = \frac{1}{(n-3)^2} tr((S_{1,2,3}^{-1})^2) = O(n^{-1}), \]

\[ (x_2^T A_{1,2,3}^{-1} x_3)^2 \leq (x_2^T A_{1,2,3}^{-1} x_2)(x_3^T A_{1,2,3}^{-1} x_3) \leq \frac{\rho^2}{(1 - \sqrt{\rho})^4} \text{ a.s.,} \]

\[ (n-2)x_2^T (A_{1,2,3}^{-1})^2 x_2 \leq \frac{k(n-2)}{(n-3)^2} \lambda_{\min}^2(S_{1,2,3}) \frac{x_2^T x_2}{k} \leq \frac{\rho}{(1 - \sqrt{\rho})^4} \text{ a.s.} \]

These give two upper bounds

\[ \frac{1 + \frac{(x_2^T A_{1,2,3}^{-1} x_3)^2}{(1+g_2)(1+g_3)}}{n-1} D_{1,2} D_{1,3} \leq 1 + \frac{\rho^2}{(1 - \sqrt{\rho})^4} \text{ a.s.,} \]

\[ \frac{(n-2)(x_2^T (A_{1,2,3}^{-1})^2 x_2 + x_3^T (A_{1,2,3}^{-1})^2 x_3)}{n-1} D_{1,2} D_{1,3} \leq \frac{2\rho}{(1 - \sqrt{\rho})^4} \text{ a.s.} \]

Then, we can get

\[ (n-2)^2 E\left((x_2^T (A_{1,2,3}^{-1})^2 x_3 + \frac{n-1}{n-1} D_{1,2} D_{1,3} \right) \rightarrow 0, \]

\[ E\left((x_2^T (A_{1,2,3}^{-1})^2 x_3 + \frac{n-1}{n-1} D_{1,2} D_{1,3} \right) \rightarrow 0. \]

These together show

\[ E\left[ \frac{(n-1)(n-2) x_2^T A_{1,2,3}^{-1} A_{1,2,3}^{-1} x_3}{n} \right] \]

\[ = E\left[ (n-2) x_2^T (A_{1,2,3}^{-1})^2 x_3 + \frac{1}{(1+g_2)(1+g_3)} \right] \]

\[ - E\left[ x_2^T (A_{1,2,3}^{-1})^2 x_2 + \frac{1}{(1+g_2)(1+g_3)} \right] \]

\[ \rightarrow 0. \]
Hence, from (S1.4), we derive \( E\left(\frac{1}{n}(\sum_{j\neq 1} x_1^\top (XX^\top)^{-1}x_j)^2\right) \to 0 \). This together with an upper-bound inferred from (S1.3) and (S1.2) leads to
\[
E\left[(x_1^\top (X(I-P_1)X^\top)^{-1}x_1 - x_1^\top (XX^\top)^{-1}x_1)^2\right] \to 0.
\]

And then (S1.3) further shows
\[
E\left[(x_1^\top (X(I-P_1)X^\top)^{-1}x_1 - \rho)^2\right] \to 0. \tag{S1.5}
\]

For any \( i \in \{1, \ldots, n\} \), we divide \( H_{ii} \) into three parts
\[
H_{ii} = (x_i - \bar{x})^\top (X(I-P_1)X^\top)^{-1}(x_i - \bar{x})
\]
\[
= \bar{x}^\top (X(I-P_1)X^\top)^{-1}x - 2x_i^\top (X(I-P_1)X^\top)^{-1}x + x_i^\top (X(I-P_1)X^\top)^{-1}x_i.
\]

Based on (S1.1) and (S1.5), we obtain
\[
E[(H_{ii} - \rho)^2] = E\left[((x_i - \bar{x})^\top (X(I-P_1)X^\top)^{-1}(x_i - \bar{x}) - \rho)^2\right]
\]
\[
= E\left[((x_1 - \bar{x})^\top (X(I-P_1)X^\top)^{-1}(x_1 - \bar{x}) - \rho)^2\right]
\]
\[
\leq E\left[3(x_1^\top (X(I-P_1)X^\top)^{-1}x_1 - \rho)^2 + 3(\bar{x}^\top (X(I-P_1)X^\top)^{-1}\bar{x})^2 + 12(x_1^\top (X(I-P_1)X^\top)^{-1}x)^2\right]
\]
\[
= o(1).
\]

Therefore,
\[
\max_{i=1,\ldots,n} E[(H_{ii} - \rho)^2] \to 0,
\]
which completes the proof.
Lemma 7. Let $z_1, \ldots, z_n$ be i.i.d. $m$-variate random vectors satisfying $E(z_i) = 0$, $\text{Var}(z_i) = I_m$ and $\text{Var}(\frac{z_i^\top z_i}{m}) = O(m^{-1})$. Suppose matrix $A$ is uniformly distributed on the Stiefel manifold $V_k(\mathbb{R}^m) = \{ A \in \mathbb{R}^{m \times k} : A^\top A = I_k \}$ and is independent of $z_i$. Let $Z = (z_1, \ldots, z_n)^\top$ and

$$H = (I - P_1)ZA \left( A^\top Z^\top (I - P_1)ZA \right)^{-1} A^\top Z^\top (I - P_1).$$

As $n, k, m \to \infty$, with $k/n \to \rho \in (0, 1)$ and $m$ sufficiently larger than $n$, we have

$$\frac{1}{n} \sum_{i=1}^{n} (H_{ii} - \rho)^2 = o_p(1),$$

where $H_{ii}$ denote the $i$th diagonal entries of $H$.

Proof. Let $U \Lambda O^\top$ be the singular value decomposition (SVD) of $Z$, where $U$ is an $n \times n$ orthogonal matrix, $O$ is an $m \times m$ orthogonal matrix, and $A = (D, 0)$ with $D = \text{diag}(d_1, \ldots, d_n)$. Let $O_n$ be the matrix consisting of first $n$ columns of $O$, then $Z$ can be denoted as

$$Z = UDO_n^\top. \quad \text{(S1.6)}$$

In the first step, we study the properties of the entries of $D$. Based on (S1.6), we have

$$\frac{1}{m} ZZ^\top = \frac{1}{m} UD^2 U^\top.$$

This indicates the diagonal entries of $\frac{1}{m} D^2$ are the eigenvalues of $\frac{1}{m} ZZ^\top$. 
then
\[
\max_{i=1,...,n} \left( \frac{d_i^2}{m} - 1 \right)^2 = \lambda_{\text{max}} \left\{ \left( \frac{1}{m} \mathbf{ZZ}^\top - \mathbf{I} \right)^2 \right\} \leq \text{tr} \left\{ \left( \frac{1}{m} \mathbf{ZZ}^\top - \mathbf{I} \right)^2 \right\}
\]

From the properties of \( \mathbf{z}_i \), we have
\[
E \left\{ \text{tr} \left\{ \left( \frac{1}{m} \mathbf{ZZ}^\top - \mathbf{I} \right)^2 \right\} \right\} = \sum_{i=1}^{n} E \left\{ \left( \frac{\mathbf{z}_i^\top \mathbf{z}_i}{m} - 1 \right)^2 \right\} + \sum_{i \neq j}^{n} E \left\{ \left( \frac{\mathbf{z}_i^\top \mathbf{z}_j}{m} \right)^2 \right\}
\]
\[
= n \text{Var} \left( \frac{\mathbf{z}_1^\top \mathbf{z}_1}{m} \right) + \frac{n^2 - n}{m}
\]
\[
= O(n^2 m^{-1}).
\]

Therefore, from Markov’s inequality, for any \( t > 0 \),
\[
P \left\{ \max_{i=1,...,n} \left( \frac{d_i}{\sqrt{m}} - 1 \right)^2 > t \right\} \leq P \left\{ \max_{i=1,...,n} \left( \frac{d_i^2}{m} - 1 \right)^2 > t \right\} \leq O(n^2 m^{-1} t^{-1}),
\]
(S1.7)

which shows the eigenvalues of \( \frac{1}{m} \mathbf{ZZ}^\top \) are close to 1 when \( m \) is sufficiently larger than \( n \).

Let \( \mathbf{X} = (\mathbf{I} - \mathbf{P}_1) \mathbf{U} \mathbf{O}_n^\top \mathbf{A} \) and \( \tilde{\mathbf{Z}} = (\mathbf{I} - \mathbf{P}_1) \mathbf{U} \frac{\mathbf{D}}{\sqrt{m}} \mathbf{O}_n^\top \mathbf{A} \). Since the hat matrix for \( \tilde{\mathbf{Z}} \) and \( (\mathbf{I} - \mathbf{P}_1) \mathbf{ZA} \) are the same, the hat matrix for \( \tilde{\mathbf{Z}} \) and \( \mathbf{X} \) are denoted as
\[
\mathbf{H} = \tilde{\mathbf{Z}} \left( \tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}} \right)^{-1} \tilde{\mathbf{Z}}^\top, \quad \mathbf{S} = \mathbf{X} \left( \mathbf{X}^\top \right)^{-1} \mathbf{X}^\top,
\]
where \( \mathbf{H} \) is the target matrix of the lemma. Let \( S_{ii} \) denote the \( i \)th diagonal entry of the matrix \( \mathbf{S} \). We will show \( H_{ii} \) and \( S_{ii} \) are close. Let \( e_i \) denote the vector with 1 in the \( i \)th coordinate and 0’s elsewhere. Define \( \gamma_i^{ls} = \)
Based on the least square, then \( \hat{\gamma}_{ls} \) satisfies
\[
\hat{\gamma}_{ls} = \arg\min_{\gamma \in \mathbb{R}^k} \| (I - P_1) e_i - X \gamma \|_2^2.
\] (S1.8)

Similarly, define \( \hat{\eta}_{ls} = \left( \tilde{Z}^\top \tilde{Z} \right)^{-1} \tilde{Z}^\top e_i \). Then, it satisfies
\[
\hat{\eta}_{ls} = \arg\min_{\eta \in \mathbb{R}^k} \| (I - P_1) e_i - \tilde{Z} \eta \|_2^2.
\] (S1.9)

Based on (S1.8) and (S1.9), we have
\[
\| (I - P_1) e_i - \tilde{Z} \hat{\eta}_{ls} \|_2^2 \leq \| (I - P_1) e_i - \tilde{Z} \hat{\gamma}_{ls} \|_2^2
\]
\[
= \| (I - P_1) e_i - X \hat{\gamma}_{ls} + (X - \tilde{Z}) \hat{\gamma}_{ls} \|_2^2
\]
\[
\leq \left( \| (I - P_1) e_i - X \hat{\gamma}_{ls} \|_2 + \| (X - \tilde{Z}) \hat{\gamma}_{ls} \|_2 \right)^2,
\] (S1.10)

and
\[
\| (I - P_1) e_i - X \hat{\gamma}_{ls} \|_2^2 \leq \| (I - P_1) e_i - \tilde{Z} \hat{\eta}_{ls} \|_2^2
\]
\[
= \| (I - P_1) e_i - \tilde{Z} \hat{\eta}_{ls} + (\tilde{Z} - X) \hat{\eta}_{ls} \|_2^2
\]
\[
\leq \left( \| (I - P_1) e_i - \tilde{Z} \hat{\eta}_{ls} \|_2 + \| (\tilde{Z} - X) \hat{\eta}_{ls} \|_2 \right)^2.
\] (S1.11)

To study (S1.10) and (S1.11), we first investigate the values of \( \| (X - \tilde{Z}) \hat{\gamma}_{ls} \|_2 \) and \( \| (\tilde{Z} - X) \hat{\eta}_{ls} \|_2 \). From Theorem 2.2.1 in Chikuse (2003), matrix \( A \) can be expressed as \( A = G \left( G^\top G \right)^{-1/2} \), where the elements of \( m \times k \) matrix \( G \) are i.i.d. from \( \mathcal{N}(0,1) \). Let \( E = O_n^\top G \). Then
\( O_{n}^\top A = E (G^\top G)^{-1/2} \). From Lemma 11, for any \( h_1 > 0 \) and \( h_2 > 0 \), the independence between \( A \) and \( Z \) leads to

\[
P \left[ \lambda_{\max} \left( \frac{1}{n} E^\top E \right) \geq (1 + \sqrt{k/n + h_1})^2 \right] \leq \exp \left( -nh_1^2/2 \right),
\]

\[
P \left[ \lambda_{\min} \left( \frac{1}{n} E^\top E \right) \leq (1 - \sqrt{k/n - h_2})^2 \right] \leq \exp \left( -nh_2^2/2 \right).
\]

(S1.12)

For any matrix \( M \), SVD shows the nonzero eigenvalues of \( M^\top M \) and \( MM^\top \) are the same. Therefore, with \( k < n \), it indicates \( \lambda_{\min}(E^\top U^\top (I - P_1) U E) = \lambda_{\min}(E^\top E) \) and \( \lambda_{\min}(E^\top D\sqrt{m} U^\top (I - P_1) U D\sqrt{m} E) = \lambda_{\min}(E^\top D^2 E) \). Based on the property \( \lambda_{\max}(M^\top M) = \lambda_{\max}(MM^\top) \) and (S1.12), we have

\[
\lambda_{\max} \left( X(X^\top X)^{-1} A^\top O_{n} O_{n}^\top A (X^\top X)^{-1} X^\top \right) = \lambda_{\max} \left( E \left( E^\top U^\top (I - P_1) U E \right)^{-1} E^\top \right)
\]

\[
\leq \lambda_{\max} \left( \frac{1}{n} E^\top E \right) \frac{1}{\lambda_{\min}(E^\top U^\top (I - P_1) U E)}
\]

\[
\leq \frac{(1 + \sqrt{k/n + h_1})^2}{(1 - \sqrt{k/n - h_2})^2}.
\]

(S1.13)

and

\[
\lambda_{\max} \left( \tilde{Z}(\tilde{Z}^\top \tilde{Z})^{-1} A^\top O_{n} O_{n}^\top A (\tilde{Z}^\top \tilde{Z})^{-1} \tilde{Z}^\top \right) = \lambda_{\max} \left( E \left( E^\top D\sqrt{m} U^\top (I - P_1) U D\sqrt{m} E \right)^{-1} E^\top \right)
\]

\[
\leq \lambda_{\max} \left( \frac{1}{n} E^\top E \right) \frac{1}{\lambda_{\min}(E^\top D\sqrt{m} U^\top (I - P_1) U D\sqrt{m} E)}
\]

\[
\leq \frac{1}{\lambda_{\min}(D^2)} \cdot \frac{(1 + \sqrt{k/n + h_1})^2}{(1 - \sqrt{k/n - h_2})^2}
\]

(S1.14)

with probability at least \( 1 - \exp(-nh_1^2/2) - \exp(-nh_2^2/2) \). Based on (S1.7),
Combining (S1.10), (S1.11), (S1.15) and (S1.16), upper bounds can be derived as follows.

\[ \| (X - \tilde{Z}) \hat{\gamma}_i^{ls} \|^2_F = \| (I - P_1) U (I - D/m) O_n A (X^T X)^{-1} X^T e_i \|^2_F \]

\[ \leq \max_{i=1, \ldots, n} \left( 1 - \frac{d_i}{m} \right)^2 \| O_n^T A (X^T X)^{-1} X^T e_i \|^2_F \]

\[ \leq t \cdot \frac{(1 + \sqrt{k/n + h_1})^2}{(1 - \sqrt{k/n - h_2})^2} \]

and

\[ \| (\tilde{Z} - X) \hat{n}_i^{ls} \|^2_F = \| (I - P_1) U (I - D/m) O_n A (\tilde{Z}^T \tilde{Z})^{-1} \tilde{Z}^T e_i \|^2_F \]

\[ \leq \max_{i=1, \ldots, n} \left( 1 - \frac{d_i}{m} \right)^2 \| O_n^T A (\tilde{Z}^T \tilde{Z})^{-1} \tilde{Z}^T e_i \|^2_F \]

\[ \leq \max_{i=1, \ldots, n} \left( 1 - \frac{d_i}{m} \right)^2 \cdot \frac{1}{\min_{i=1, \ldots, n} \left( \frac{d_i^2}{m} \right)} \cdot \frac{(1 + \sqrt{k/n + h_1})^2}{(1 - \sqrt{k/n - h_2})^2} \]

\[ \leq \frac{t}{(1 - \sqrt{t})^2} \cdot \frac{(1 + \sqrt{k/n + h_1})^2}{(1 - \sqrt{k/n - h_2})^2} \]

(S1.16)

with probability at least \( 1 - O(n^2 m^{-1} t^{-1}) - \exp(-nh_1^2/2) - \exp(-nh_2^2/2) \).

Combining (S1.16), (S1.11), (S1.15) and (S1.16), with \( h_1 = n^{-1/4}, h_2 = n^{-1/4} \) and \( t = n^{-c} \), where \( c \) is a positive constant, we have

\[ \| (I - P_1) e_i - \tilde{Z} n_i^{ls} \|^2_F \leq \| (I - P_1) e_i - X \hat{\gamma}_i^{ls} \|^2_F + 3n^{-c/2} \cdot \frac{1 + \sqrt{k/n + n^{-1/4}}}{1 - \sqrt{k/n - n^{-1/4}}}; \]

\[ \| (I - P_1) e_i - X \hat{\gamma}_i^{ls} \|^2_F \leq \| (I - P_1) e_i - \tilde{Z} n_i^{ls} \|^2_F + \frac{3}{n^{-c/2} - 1} \cdot \frac{1 + \sqrt{k/n + n^{-1/4}}}{1 - \sqrt{k/n - n^{-1/4}}} \]

with probability at least \( 1 - O(n^{2+c} m^{-1}) - 2 \exp(-n^{1/2}/2) \). Since \( \| (I - P_1) e_i - \tilde{Z} n_i^{ls} \|^2_F = e_i^T (I - P_1) e_i - H_i \) and \( \| (I - P_1) e_i - X \hat{\gamma}_i^{ls} \|^2_F = e_i^T (I - P_1) e_i - \).
$S_{ii}$, and the above derivation is valid for any $e_i$, we obtain

$$|H_{ii} - S_{ii}| \leq \frac{3}{n^{c/2} - 1} \cdot \frac{1 + \sqrt{k/n + n^{-1/4}}}{1 - \sqrt{k/n - n^{-1/4}}}, \quad i = 1, \ldots, n,$$

with probability at least $1 - O(n^{2+c}m^{-1}) - 2 \exp(-n^{1/2}/2)$. When $n \to \infty$ and $n^{2+c}m^{-1} = o(1)$, there is a constant $C \geq \frac{12+12\sqrt{\rho}}{1-\sqrt{\rho}}$ such that

$$P\left[ \max_{i=1, \ldots, n} |H_{ii} - S_{ii}| \geq Cn^{-c/2} \right] = o(1). \quad \text{(S1.17)}$$

According to the definitions of $X$ and $A$, the hat matrix $S$ can be denoted as

$$S = (I - P_1)UO_n^\top G \left( G^\top O_n U^\top (I - P_1)UO_n^\top G \right)^{-1} G^\top O_n U^\top (I - P_1),$$

where $UO_n^\top$ is independent of $G$ and satisfies $UO_n^\top O_n U^\top = I_n$. From the definition of $G$, Lemma 5 and the dominated convergence theorem, we obtain

$$E\left[ \frac{1}{n} \sum_{i=1}^n (S_{ii} - \rho)^2 \right] \to 0.$$

Then, $\frac{1}{n} \sum_{i=1}^n (S_{ii} - \rho)^2 = o_p(1)$ can be derived based on Markov’s inequality.

Combining this with (S1.17) and Slutsky’s theorem, it shows

$$\frac{1}{n} \sum_{i=1}^n (H_{ii} - \rho)^2 = \frac{1}{n} \sum_{i=1}^n (H_{ii} - S_{ii} + S_{ii} - \rho)^2$$

$$\leq \frac{2}{n} \sum_{i=1}^n (H_{ii} - S_{ii})^2 + \frac{2}{n} \sum_{i=1}^n (S_{ii} - \rho)^2$$

$$\leq \max_{i=1, \ldots, n} 2(H_{ii} - S_{ii})^2 + \frac{2}{n} \sum_{i=1}^n (S_{ii} - \rho)^2$$

$$= o_p(1),$$
which completes the proof.

Conditional on \( A^\top z \), Theorem 2.1 in Steinberger and Leeb (2018) showed that the mean of \( z \) is approximately linear in \( A^\top z \) under certain conditions. Based on this result, we derived the following lemma.

**Lemma 8.** Suppose \( m \)-variate random vector \( z = (z_1, \ldots, z_m)^\top \) has a Lebesgue density \( f_z \) and satisfies \( E(z) = 0 \) and \( E(zz^\top) = I_m \). For all \( i = 1, \ldots, m \), the components \( z_i \) are independent and the moments satisfy \( E(z_i^{20}) \leq C \) for some constant \( C \). And all the marginal densities of the components of \( z \) are bounded by a constant \( D \geq 1 \). Suppose matrix \( A \) is uniformly distributed on the Stiefel manifold \( V_k(\mathbb{R}^m) = \{ A \in \mathbb{R}^{m \times k} : A^\top A = I_k \} \). Let \( \nu_{m,k} \) denote the uniform distribution on \( V_k(\mathbb{R}^m) \). Let \( z_1, \ldots, z_n \) be the i.i.d. copies of \( z \) and \( A \) be independent of \( z_i \). For any nonzero vector \( b \in \mathbb{R}^m \), as \( n \to \infty \), with \( k/n \to \rho \in (0,1) \) and \( m \) sufficiently larger than \( n \), there is a series of Borel set \( F_n \subseteq V_k(\mathbb{R}^m) \) such that

\[
\sup_{A \in F_n} P \left( \sum_{i=1}^{n} \left( E(b^\top z_i|A^\top z_i) - b^\top AA^\top z_i \right)^2 > ||b||_2^2 \right) = o(1),
\]

\[
\sup_{A \in F_n} P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| Var(b^\top z_i|A^\top z_i) - b^\top (I_m - AA^\top) b \right| > 5||b||_2^2 \right) = o(1),
\]

and \( \nu_{m,k}(F_n) \to 1 \).

**Proof.** Based on Example 3.1 and Theorem 2.1 given in Steinberger and
Leeb (2018), for each $\tau \in (0, 1)$, there is a Borel set $F_n \subseteq \mathcal{V}_k(\mathbb{R}^m)$ such that

$$
\sup_{A \in F_n} P \left( \left\| E(z|A^\top z) - AA^\top z \right\|_2 > t \right) \leq \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1 - \tau \log m},
$$

$$
\sup_{A \in F_n} P \left( \left\| E(zz^\top|A^\top z) - (I_m - AA^\top + AA^\top zz^\top AA^\top) \right\|_{sp} > t \right) \leq \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1 - \tau \log m},
$$

for each $t > 0$, and such that $\nu_{m,k}(F_n^c) \leq 2\kappa_1 m^{-(\tau/10) \cdot (1 - \gamma_2 \log m)}$, where $\kappa_2$ and $\gamma_2$ are constants. Therefore, when $t = n^{-1/2}$, we have

$$
\sup_{A \in F_n} P \left( \sum_{i=1}^n \left\| E(z_i|A^\top z_i) - AA^\top z_i \right\|_2 > 1 \right) \leq \sum_{i=1}^n \sup_{A \in F_n} P \left( \left\| E(z_i|A^\top z_i) - AA^\top z_i \right\|_2 > 1 \right)
\leq \frac{n^{3/2} m^{-\tau/10}}{1 - \tau \log m} + \frac{\gamma_2}{1 - \tau \log m},
$$

(S1.18)

$$
\sup_{A \in F_n} P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\| E[z_i z_i^\top|A^\top z_i] - AA^\top z_i z_i^\top AA^\top - (I_m - AA^\top) \right\|_{sp} > 1 \right)
\leq \sum_{i=1}^n \sup_{A \in F_n} P \left( \left\| E[z_i z_i^\top|A^\top z_i] - AA^\top z_i z_i^\top AA^\top - (I_m - AA^\top) \right\|_{sp} > 1 \right)
\leq \frac{n^{3/2} m^{-\tau/10}}{1 - \tau \log m} + \frac{\gamma_2}{1 - \tau \log m},
$$

(S1.19)

and $\nu_{m,k}(F_n^c) \leq 2\kappa_1 m^{-(\tau/10) \cdot (1 - \gamma_2 \log m)}$.

For each $i$, define $r_i = E(b^\top z_i|A^\top z_i) - b^\top AA^\top z_i$ and $q_i = b^\top z_i - E(b^\top z_i|A^\top z_i)$. Based on the definition of the conditional variance, we
could derive

$$Var(q_i|A^\top z_i) = b^\top E[z_i z_i^\top |A^\top z_i] b - E[b^\top z_i |A^\top z_i]^2,$$

then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} |Var(q_i|A^\top z_i) - b^\top (I_m - AA^\top) b|$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |b^\top \{E[z_i z_i^\top |A^\top z_i] - AA^\top z_i z_i^\top AA^\top - (I_m - AA^\top)\} b - 2b^\top AA^\top z_i r_i - r_i^2|$$

$$\leq ||b||_2^2 \frac{1}{\sqrt{n}} \sum_{i=1}^{n} ||E[z_i z_i^\top |A^\top z_i] - AA^\top z_i z_i^\top AA^\top - (I_m - AA^\top)||_{sp}$$

$$+ 2\sqrt{\sum_{i=1}^{n} \left(\frac{b^\top AA^\top z_i}{n}\right)^2} \sqrt{\sum_{i=1}^{n} r_i^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} r_i^2}.$$  

(S1.20)

From the calculation,

$$Var \left\{ (b^\top AA^\top z_i)^2 \right\} \leq (C^{1/5} + 1)(b^\top AA^\top b)^2,$$

Markov’s inequality leads to

$$P \left( \sum_{i=1}^{n} \frac{(b^\top AA^\top z_i)^2}{n} > 2b^\top AA^\top b \right) \leq \frac{C^{1/5} + 1}{n}. \quad \text{(S1.21)}$$

According to Cauchy–Schwarz inequality,

$$r_i^2 = \{E(b^\top z_i |A^\top z_i) - b^\top AA^\top z_i\}^2 \leq ||b||_2^2 \cdot ||E(z_i |A^\top z_i) - AA^\top z_i||_2^2.$$
Therefore, combining (S1.18), (S1.19), (S1.20) and (S1.21), we can derive

\[
\sup_{A \in F_n} P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \text{Var}(q_i | A^\top z_i) - b^\top (I_m - AA^\top) b \right| > 5 ||b||_2^2 \right)
\]

\[
\leq \sup_{A \in F_n} P \left( \sum_{i=1}^{n} r_i^2 > ||b||_2^2 \right) + \sup_{A \in F_n} P \left( \sum_{i=1}^{n} \frac{(b^\top AA^\top z_i)^2}{n} > 2b^\top AA^\top b \right)
\]

\[
+ \sup_{A \in F_n} P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \left| E[z_i z_i^\top | A^\top z_i] - AA^\top z_i z_i^\top AA^\top - (I_m - AA^\top) \right| \right|_sp > 1 \right)
\]

\[
\leq 2n^{3/2}m^{-\tau/10} + \frac{\gamma_2}{1 - \tau \log m} + \frac{2C}{n}.
\]

When \(m\) is sufficiently large such that \(n^2 = o(\log m)\), as \(n \to \infty\), we have

\[
\sup_{A \in F_n} P \left( \sum_{i=1}^{n} r_i^2 > ||b||_2^2 \right) = o(1).
\]

and

\[
\sup_{A \in F_n} P \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| \text{Var}(q_i | A^\top z_i) - b^\top (I_m - AA^\top) b \right| > 5 ||b||_2^2 \right) = o(1),
\]

where \(\nu_{m,k}(F_n) \to 1\). The proof is completed.

### S1.3 Proof of Lemma 3.2

First we present a trace inequality (Lopes, Jacob, and Wainwright, 2011, Lemma 2).

**Lemma 9.** If \(A\) and \(B\) are square matrices of the same size with \(A \succeq 0\) and \(B = B^\top\), then

\[
\lambda_{\min}(B) \text{tr}(A) \leq \text{tr}(AB) \leq \lambda_{\max}(B) \text{tr}(A).
\]
Some results for Gaussian concentration inequalities will be introduced. The following concentration bounds for Gaussian quadratic forms are given in Bechar (2009).

**Lemma 10.** Let $A \in \mathbb{R}^{p \times p}$ with $A \succeq 0$ and $z \sim \mathcal{N}(0, I_p)$. For any $t > 0$, we have

$$P \left[ z^\top A z \geq \text{tr}(A) + 2\|A\|_F \sqrt{t} + 2\|A\|_{sp} t \right] \leq \exp(-t),$$  
and

$$P \left[ z^\top A z \leq \text{tr}(A) - 2\|A\|_F \sqrt{t} \right] \leq \exp(-t).$$

Davidson and Szarek (2001, Theorem 2.13) gave an upper-bound and a lower-bound on the extreme eigenvalues of Wishart matrices.

**Lemma 11.** For $k \leq p$, let $P_k \in \mathbb{R}^{p \times k}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then, for all $t \geq 0$, we have

$$P \left[ \lambda_{\max}(\frac{1}{p} P_k^\top P_k) \geq (1 + \sqrt{kp} + t)^2 \right] \leq \exp(-pt^2/2),$$  
and

$$P \left[ \lambda_{\min}(\frac{1}{p} P_k^\top P_k) \leq (1 - \sqrt{kp} - t)^2 \right] \leq \exp(-pt^2/2).$$

As a restatement of partial proof in Lopes, Jacob, and Wainwright (2011, Lemma 5), we obtain an upper bound for $\text{tr}(P_k^\top \Sigma P_k)$.

**Lemma 12.** For $k \leq p$, let $P_k \in \mathbb{R}^{p \times k}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Suppose matrix $\Sigma \in \mathbb{R}^{p \times p}$ satisfies $\Sigma \succeq 0$. Then, as
(k, p) \rightarrow \infty$, for any constant $C > 1$, we have

$$P \left[ \text{tr}(P_k^\top \Sigma P_k) \leq C \text{tr}(\Sigma) \right] \rightarrow 1.$$  

**Proof.** Let $U^\top DU$ be a spectral decomposition of $\Sigma$. Then $P_k^\top \Sigma P_k$ can be written as $(UP_k)^\top D(UP_k)$. As $UP_k$ has the same distribution as $P_k$, $P_k^\top \Sigma P_k$ is distributed as $P_k^\top DP_k$. In the following, we work under $P_k^\top DP_k$.

Let $\xi_i$ be the $i$th column of $P_k$ and $Z^\top = (\xi_1^\top, ..., \xi_k^\top)$. Then $Z \in \mathbb{R}^{pk \times 1}$ and is distributed as $\mathcal{N}(0, I_{pk})$. Likewise, let $\tilde{D} \in \mathbb{R}^{pk \times pk}$ be a diagonal matrix obtained by arranging $k$ copies of $D$ along the diagonal, i.e.

$$\tilde{D} := \begin{pmatrix} D & \cdots & \cdots & \cdots \\ \cdots & D & \cdots & \cdots \\ \cdots & \cdots & \cdots & D \\ \cdots & \cdots & \cdots & \cdots \end{pmatrix}.$$  

Consider the diagonal entries of $P_k^\top DP_k$

$$\text{tr}(P_k^\top DP_k) = \sum_{i=1}^k \xi_i^\top D \xi_i = Z^\top \tilde{D} Z.$$  

Applying Lemma 10 to the quadratic form $Z^\top \tilde{D} Z$, and noting that $\frac{\|D\|_F}{\text{tr}(D)}$ and $\frac{\|D\|_{sp}}{\text{tr}(D)}$ are at most 1, we get

$$\text{tr}(P_k^\top DP_k) \leq \text{tr}(\tilde{D}) + 2\|\tilde{D}\|_F \sqrt{t_1} + 2\|\tilde{D}\|_{sp} t_1$$  

$$= k \text{tr}(D) + 2\|D\|_F \sqrt{t_1 k} + 2\|D\|_{sp} t_1$$  

$$\leq k \text{tr}(\Sigma) \left( 1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k} \right)$$
with probability at least $1 - \exp(-t_1)$.

Choose $t_1 = \sqrt{k}$. The probability of the event tends to 1 as $(k, p) \to \infty$ with

$$(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) \to 1.$$ 

Hence, for large $k$ and any constant $C > 1$, we can obtain $(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) < C$ and complete the proof. \hfill $\square$

**Proof of Lemma 3.2.** Let $U^T D U$ be a spectral decomposition of $\Sigma$, where $D = \text{diag}(d_1, ..., d_p)$ and $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. From this decomposition,

$$\sqrt{n}||\Gamma^T \beta - \Gamma^T P_k \eta||^2_2 = \sqrt{n}||\sqrt{D}U \beta - \sqrt{D}UP_k \eta||^2_2.$$ 

(S1.22)

To cover general cases, we assume $\beta/||\beta||_2$ distributed uniformly on the unit sphere. Then, we work under the assumption $\beta/||\beta||_2 = \delta/\sqrt{p}$, where $\delta$ follows $\mathcal{N}(0, I_p)$. In light of this, $U \beta/||\beta||_2$ and $\beta/||\beta||_2$ have the same distributions and then $U \beta/||\beta||_2$ is denoted by $\delta/\sqrt{p}$ for simplicity. For the same reason, we denote $U P_k$ as $P_k$.

For the $s$ given in Assumption A6, we let $\delta = (\delta_s^T, \delta_{p-s}^T)^T$, where $\delta_s \in \mathbb{R}^s$ and $\delta_{p-s} \in \mathbb{R}^{p-s}$. Correspondingly, $D$ is divided into $D_s$ and $D_{p-s}$, where $D_s = \text{diag}(d_1, ..., d_s)$ and $D_{p-s} = \text{diag}(d_{s+1}, ..., d_p)$. Let $P_k = (P_{s,k}^T, P_{p-s,k}^T)^T$ with $P_{s,k} \in \mathbb{R}^{s \times k}$ and $P_{p-s,k} \in \mathbb{R}^{(p-s) \times k}$. We define $\eta_0 \in \mathbb{R}^k$
as

$$\eta_0 = \mathbf{P}_{s,k}^\top (\mathbf{P}_{s,k} \mathbf{P}_{s,k}^\top)^{-1} \frac{\delta_s}{\sqrt{p}}$$

Plugging $\eta_0$ into (S1.22), we have

$$\min_{\eta \in \mathbb{R}^k} \sqrt{n} \frac{||\Gamma^\top \beta - \Gamma^\top \mathbf{P}_k \eta||_2^2}{||\beta||_2^2} = \min_{\eta \in \mathbb{R}^k} \sqrt{n} \frac{\|\sqrt{D} \delta - \sqrt{D} \mathbf{P}_k \eta\|_2^2}{\|\beta\|_2^2}$$

$$= \min_{\eta \in \mathbb{R}^k} \sqrt{n} \left( \|\sqrt{D_s} (\frac{\delta_s}{\sqrt{p}} - \mathbf{P}_{s,k} \eta_0)\|_2^2 + \|\sqrt{D_{p-s}} (\frac{\delta_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k} \eta_0)\|_2^2 \right)$$

$$\leq \sqrt{n} ||\sqrt{D_s} (\frac{\delta_s}{\sqrt{p}} - \mathbf{P}_{s,k} \eta_0)\|_2^2 + \sqrt{n} ||\sqrt{D_{p-s}} (\frac{\delta_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k} \eta_0)\|_2^2$$

$$= \sqrt{n} ||\sqrt{D_{p-s}} \frac{\delta_{p-s}}{\sqrt{p}} - \sqrt{D_{p-s}} \mathbf{P}_{p-s,k} \mathbf{P}_{s,k}^\top (\mathbf{P}_{s,k} \mathbf{P}_{s,k}^\top)^{-1} \frac{\delta_s}{\sqrt{p}}\|_2^2$$

$$\leq 2\sqrt{n} ||\sqrt{D_{p-s}} \frac{\delta_{p-s}}{\sqrt{p}}\|_2^2 + 2\sqrt{n} ||\sqrt{D_{p-s}} \mathbf{P}_{p-s,k} \mathbf{P}_{s,k}^\top (\mathbf{P}_{s,k} \mathbf{P}_{s,k}^\top)^{-1} \frac{\delta_s}{\sqrt{p}}\|_2^2$$

$$= T_1 + T_2.$$  \hspace{1cm} \text{(S1.23)}$$

Next we show that $||\beta||_2^2 T_1$ and $||\beta||_2^2 T_2$ both converge to 0 with probability tending to 1.

In the first step, the concentration inequality for quadratic forms in Lemma \[10\] gives an upper bound on $T_1$

$$P \left[ T_1 \leq 2\sqrt{n} \left( \text{tr}(D_{p-s}) + 2\sqrt{h_1} ||D_{p-s}||_F + 2h_1 ||D_{p-s}||_{sp} \right) \right] \geq 1 - \exp(-h_1),$$

where $h_1$ is a positive real number that may vary with $n$. From Assumption
A6 and the properties of $|| \cdot ||_F$ and $|| \cdot ||_{sp}$, we select $h_1 = n^\gamma$ and get

$$||\beta||^2_2 T_1 \leq \frac{2 \sqrt{n} ||\beta||^2}{p} \left( tr(D_{p-s}) + 2 \sqrt{h_1}||D_{p-s}||_F + 2h_1||D_{p-s}||_{sp} \right)$$

$$\leq \frac{2 \sqrt{n} ||\beta||^2}{p} tr(D_{p-s}) \left( 1 + 2 \sqrt{h_1} + 2h_1 \right)$$

$$\leq \frac{10n^{0.5+\gamma} ||\beta||^2}{p} tr(D_{p-s}) = o(1)$$

(S1.24)

with probability at least $1 - \exp(-n^\gamma)$.

In the next step, Lemmas 12 and 11 give upper bounds by

$$k \lambda_{\max} \left( \left( P_{s,k} P_{s,k}^T \right)^{-1} \right) = \frac{1}{\lambda_{\min} \left( \frac{P_{s,k} P_{s,k}^T}{k} \right)} \leq \frac{1}{(1 - \sqrt{s/k - k^{-1/4}})^2},$$

$$\frac{tr(P_{p-k}^T D_{p-s} P_{p-s,k})}{k} \leq 2tr(D_{p-s})$$

with probability converging to 1. These inequalities together with Lemma 9 lead to

$$tr \left( \left( P_{s,k} P_{s,k}^T \right)^{-1} P_{s,k} P_{p,s-k} D_{p-s} P_{p-s,k} \left( P_{s,k} P_{s,k}^T \right)^{-1} \right)$$

$$\leq k \lambda_{\max} \left( P_{s,k} P_{s,k}^T \right)^{-1} tr \left( P_{p-s,k}^T D_{p-s} P_{p-s,k} \right)$$

$$= k \lambda_{\max} \left( P_{s,k} P_{s,k}^T \right)^{-1} tr \left( P_{p-s,k}^T D_{p-s} P_{p-s,k} \right)$$

(S1.25)

with probability converging to 1. To study the randomness from $\delta_s$, we apply the same method in the first step of investigating $||\beta||^2_2 T_1$ with the
help from upper bound in (S1.25) and get

\[ \|\beta\|^2 T^2_2 \leq \frac{20n^{0.5+\gamma}\|\beta\|^2 tr(D_{p-s})}{p(1 - \sqrt{s/k - k^{-1/4}})^2} = o(1) \]  

(S1.26)

with probability tending to 1.

Combining (S1.23), (S1.24) and (S1.26), we have

\[ \min_{\eta \in \mathbb{R}^k} \sqrt{n}|\Gamma^\top \beta - \Gamma^\top P_k \eta|^2 = o(1) \]

with probability tending to 1 and complete the proof. \(\square\)

S2  Proof of theorems

S2.1  Proof of Theorem 3.1

Under \(H_0\), we have

\[ T_n - 1 = \frac{\epsilon^\top M \epsilon}{\epsilon^\top (I - P_1 - H_k) \epsilon / (n - 1 - k)}, \]

where \(M = (m_{ij}) = \frac{H_k}{k} - \frac{I - P_1 - H_k}{n - k - 1}\). The property that \(H_k\) is idempotent with rank \(k\) leads to \(tr(M) = 0\) and \(M^\top M = \frac{H_k^+}{k^2} + \frac{I - P_1 - H_k}{(n - k - 1)^2}\). Therefore,

\[ \frac{\|M\|_{sp}^2}{\|M\|_{tr}^2} = \frac{\lambda_{\text{max}}(M^\top M)}{tr(M^\top M)} \leq \frac{\lambda_{\text{max}}(H_k^+)}{k} + \frac{\lambda_{\text{max}}(I - P_1 - H_k)}{(n - k - 1)^2} = O(n^{-1}). \]

And we have

\[ E(\epsilon^\top M \epsilon | M) = \sigma^2 tr(M) = 0, \]

\[ Var(\epsilon^\top M \epsilon | M) = (\mu_4 - 3\sigma^4) \sum_{i=1}^n m_{ii}^2 + 2\sigma^4 \left( \frac{1}{k} + \frac{1}{n - k - 1} \right), \]
where the error term \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)^\top \) has \( E(\epsilon_i) = 0, \) \( Var(\epsilon_i) = \sigma^2, \) and \( E(\epsilon_i^4) = \mu_4. \) When \( M \) is given, these together with Lemma 2 imply

\[
\frac{\epsilon^\top M \epsilon}{\sqrt{Var(\epsilon^\top M \epsilon | M)}} \xrightarrow{D} \mathcal{N}(0, 1).
\]

The randomness brought from \( M \) in fact does not influence the asymptotic normality. From the law of total expectation, we have, for \( \forall \alpha \in \mathbb{R}, \)

\[
P\left( \frac{\epsilon^\top M \epsilon}{\sqrt{Var(\epsilon^\top M \epsilon | M)}} \leq \alpha \right) = E\left( P\left( \frac{\epsilon^\top M \epsilon}{\sqrt{Var(\epsilon^\top M \epsilon | M)}} \leq \alpha | M \right) \right).
\]

And the aforementioned result shows

\[
P\left( \frac{\epsilon^\top M \epsilon}{\sqrt{Var(\epsilon^\top M \epsilon | M)}} \leq \alpha | M \right) \rightarrow \Phi(\alpha).
\]

Based on the dominated convergence theorem, we get

\[
\frac{\epsilon^\top M \epsilon}{\sigma^2 \sqrt{(\frac{\mu_4}{\sigma^4} - 3) \sum_{i=1}^n m_{ii}^2 + 2(\frac{1}{k} + \frac{1}{n-k-1})}} \xrightarrow{D} \mathcal{N}(0, 1). \tag{S2.1}
\]

Let \( G_n = \sum_{i=1}^n m_{ii}^2. \) Next, we will show \( nG_n = \text{op}(1). \) From the definition,

\[
nG_n = n \sum_{i=1}^n m_{ii}^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \left( \frac{\mathcal{H}_k \sum_{i=1} m_{ii}^2}{k} \right)^2 + \left\{ \frac{\rho - \frac{k}{n-1}(1 - \frac{1}{n})}{\frac{k}{n-1}(1 - \frac{k+1}{n})} \right\}^2 \right\} \leq 2 \sum_{i=1}^n \left\{ \frac{\mathcal{H}_k \sum_{i=1} m_{ii}^2}{k} + \left\{ \frac{\rho - \frac{k}{n-1}(1 - \frac{1}{n})}{\frac{k}{n-1}(1 - \frac{k+1}{n})} \right\}^2 \right\}. \tag{S2.2}
\]

Let \( \Sigma_1 = P_k^\top \Sigma P_k. \) From Lemma 4, we find the smallest eigenvalue of

\[
\frac{1}{p} P_k^\top P_k
\]

is bounded away from 0 a.s., showing \( P_k \) is of full column rank with probability 1. Therefore, \( \Sigma_1 \) is of full rank with probability 1. Define
\( \tilde{U}_k = XP_k \Sigma_i^{-1/2} \). Since \( H_k \) is invariant to the full rank linear transform of \( U_k \), the hat matrix can be expressed as

\[
H_k = U_k (U_k^T U_k)^{-1} U_k^T = (I - P_1) \tilde{U}_k (U_k^T (I - P_1) \tilde{U}_k)^{-1} \tilde{U}_k^T (I - P_1).
\]

From Assumption A1, \( \tilde{U}_k \) can be denoted by \( ZA \), where \( A = \Gamma^T P_k \Sigma_i^{-1/2} \) is an \( m \times k \) matrix. From Section 2.4.2 in [Chikuse (2003)], matrix \( A \) is on the Stiefel manifold \( \mathcal{V}_k(\mathbb{R}^m) \) with probability 1, which demonstrates \( U_k^T U_k \) is of full rank with probability 1. From Lemma 7 and (S2.2), we obtain \( nG_n = op(1) \).

Assumption A3 implies \( \frac{n}{k} + \frac{n}{n-k-1} \rightarrow \frac{1}{\rho(1-\rho)} \), as \( n \rightarrow \infty \). Therefore, (S2.1) leads to

\[
\frac{\epsilon^T M \epsilon}{\sigma^2 \sqrt{2/n \rho(1-\rho)}} \overset{D}{\rightarrow} \mathcal{N}(0,1).
\]

In addition, from \( E\left( \frac{\epsilon^T (I - P_1 - H_k) \epsilon}{n-k-1} \right) = \sigma^2 \), \( Var\left( \frac{\epsilon^T (I - P_1 - H_k) \epsilon}{n-k-1} \right) \leq \frac{\mu_4 - \sigma^4}{n-k-1} \rightarrow 0 \) and Markov’s inequality, we have

\[
\frac{\epsilon^T (I - P_1 - H_k) \epsilon}{n-k-1} = \sigma^2 + o_p(1).
\]

Hence, under \( H_0 \),

\[
\frac{T_n - 1}{\sqrt{2/n \rho(1-\rho)}} \overset{D}{\rightarrow} \mathcal{N}(0,1),
\]

which completes the proof.
S2. PROOF OF THEOREMS

S2.2 Proof of Theorem 3.2

First, we derive a decomposition of \( \mathbf{x}_i^\top \beta \). Let \( \xi = (\mathbf{P}_k^\top \Sigma \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \Sigma \beta \). For each \( i \), define

\[
    r_i = E(\mathbf{x}_i^\top \beta | \mathbf{P}_k^\top \mathbf{x}_i) - \mathbf{x}_i^\top P_k \xi, \quad q_i = \mathbf{x}_i^\top \beta - E(\mathbf{x}_i^\top \beta | \mathbf{P}_k^\top \mathbf{x}_i).
\]

Then, we have \( \mathbf{x}_i^\top \beta = \mathbf{x}_i^\top \mathbf{P}_k \xi + r_i + q_i \), where \( q_i \) satisfies \( E(q_i | \mathbf{P}_k^\top \mathbf{x}_i) = 0 \). Let \( \omega^2 = \beta^\top \Sigma \beta - \xi^\top \mathbf{P}_k^\top \Sigma \mathbf{P}_k \xi \) and \( \tau_i = Var(q_i | \mathbf{P}_k^\top \mathbf{x}_i) - \omega^2 \). According to Lemma 8 and the condition \( \beta^\top \Sigma \beta = o(1) \), it shows

\[
    \sum_{i=1}^{n} r_i^2 = o_p(1) \text{ and } \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\tau_i| = o_p(1), \quad (S2.3)
\]

when the event \( \mathbf{A} \in F_n \) is satisfied, where \( F_n \) is a series of sets that satisfy \( v_{m,k}(F_n) \to 1 \), as \( n \to \infty \), and \( \mathbf{A} = \Gamma^\top \mathbf{P}_k (\mathbf{P}_k^\top \Sigma \mathbf{P}_k)^{-1/2} \). The probability of the event tends to 1, based on the randomness of \( \mathbf{P}_k \).

Define a new error term \( e_i = q_i + \epsilon_i \). Let \( \sigma^2 = Var(\epsilon_i) \). The model can be denoted as

\[
    y = \alpha \mathbf{1} + \mathbf{X} \mathbf{P}_k \xi + \mathbf{r} + \mathbf{e}, \quad (S2.4)
\]

where \( \mathbf{r} = (r_1, ..., r_n)^\top \), and \( \mathbf{e} = (\epsilon_1, ..., \epsilon_n)^\top \) with each elements of \( \mathbf{e} \) satisfying \( E(\epsilon_i) = 0 \), \( E(\epsilon_i | \mathbf{P}_k^\top \mathbf{x}_i) = 0 \), \( Var(\epsilon_i | \mathbf{P}_k^\top \mathbf{x}_i) = \sigma^2 + \omega^2 + \tau_i \), and \( E(\epsilon_i^4 | \mathbf{P}_k^\top \mathbf{x}_i) = \mu_4 + 6\sigma^2 Var(q_i | \mathbf{P}_k^\top \mathbf{x}_i) + E(q_i^4 | \mathbf{P}_k^\top \mathbf{x}_i) \). For matrix \( \mathbf{M} = \)
\( (m_{ij}) = \frac{H_j}{k} - \frac{1-P_1-H_n}{n-k-1} \), calculation shows

\[
E(e^\top Me|XP_k) = \sum_{i=1}^{n} m_{ii}\tau_i,
\]

\[
Var(e^\top Me|XP_k) = \sum_{i=1}^{n} m_{ii}^2 \left\{ E(e_i^4|XP_k) - 3E(e_i^2|XP_k)^2 \right\}
+ 2 \sum_{i,j} m_{ij}^2 E(e_i^2|XP_k)E(e_j^2|XP_k)
= 2(\sigma^2 + \omega^2) tr(M^\top M) + g(M, X, \epsilon, P_k),
\]

where \( g(M, X, \epsilon, P_k) = \sum_{i=1}^{n} m_{ii}^2 \left\{ \mu_4 - 3\sigma^4 + E(q_i^4|XP_k) \right\} \).

For a constant \( a \leq 2/\rho(1-\rho) \) and large \( n \), \( M \) satisfies

\[
||M||_{sp} \leq a/n \quad \text{and} \quad |m_{ii}| = |e_i^\top Me_i| \leq ||M||_{sp}.
\]

Then, (S2.3) leads to

\[
\sqrt{n}E(e^\top Me|XP_k) = o_p(1). \quad \text{(S2.5)}
\]

To investigate the conditional variance, based on (S2.2) and Lemma 7, we can derive

\[
\sum_{i=1}^{n} m_{ii}^2 \left\{ E(e_i^4|XP_k) - 3E(e_i^2|XP_k)^2 \right\} \leq \sum_{i=1}^{n} m_{ii}^2 \left\{ \mu_4 - 3\sigma^4 + E(q_i^4|XP_k) \right\} = o_p(n^{-1}).
\]

In addition, \( \sum_{j=1}^{n} m_{ij}^2 = e_i^\top MM^\top e_i \leq ||M||_{sp}^2 \leq a^2/n^2 \) and (S2.3) lead to

\[
\sum_{i,j} m_{ij}^2 \left\{ (\sigma^2 + \omega^2)(\tau_i + \tau_j) + \tau_i\tau_j \right\} \leq 2(\sigma^2 + \omega^2)a^2 \sum_{i=1}^{n} \frac{\tau_i}{n^2} + a^2 \left( \sum_{i=1}^{n} |\tau_i| \right)^2 = o_p(n^{-1}).
\]

Therefore, \( g(M, X, \epsilon, P_k) = o_p(n^{-1}) \), from which we obtain

\[
Var(e^\top Me|XP_k) = 2(\sigma^2 + \omega^2) tr(M^\top M) + o_p(n^{-1}). \quad \text{(S2.6)}
\]
According to $\text{tr}(M^T M) = \frac{1}{k} + \frac{1}{n-1-k}$, (S2.5), (S2.6) and the condition $k/n \to \rho$, Lemma 2 shows

$$\sqrt{n\rho(1-\rho)}e^T M e - o_p(1) \quad \mathcal{D} \to \mathcal{N}(0, 1).$$

(S2.7)

To investigate the numerator of the test statistic, (S2.3) shows that $r$ satisfies

$$\frac{1}{\sqrt{n}} r^T E r \leq \frac{1}{\sqrt{n}} r^T r = o_p(n^{-1/2}),$$

(S2.8)

for any $n \times n$ idempotent matrix $E$. Based on Jensen’s inequality, the fourth moment of $q_i$ satisfies $E(q_i^4) \leq 16 E \{(x_i^\top \beta)^4\}$. According to

$$E\{(x_i^\top \beta)^4\} = \sum_{i=1}^{m} (\Gamma^\top \beta)^4_1 E(z_{i1}^4) + 3 \sum_{i \neq j} (\Gamma^\top \beta)^2_i (\Gamma^\top \beta)^2_j E(z_{i2} z_{j2}),$$

and $\text{Var}(q_i) \leq \omega^2 \leq \beta^\top \Sigma \beta$, the condition $\beta^\top \Sigma \beta = o(1)$ leads to $E(q_i^4) = o(1)$ and

$$\left|E(e_i^4) - \mu_4\right| \leq c_1 \beta^\top \Sigma \beta = o(1), \quad E\{r_i^2\} \leq E(q_i^4) + \omega^4 \leq c_1 (\beta^\top \Sigma \beta)^2 = o(1),$$

(S2.9)
Changyu Liu, Xingqiu Zhao and Jian Huang

for a constant $c_1$. In addition, the calculation shows

$$
\left| E \left( \frac{e^\top (I - P_1 - H_k) e}{n - 1 - k} \right) - (\sigma^2 + \omega^2) \right| = \left| E \left\{ \sum_{i=1}^{n} \frac{(I - P_1 - H_k) \tau_i}{n - 1 - k} \right\} \right|
$$

$$
\leq \sum_{i=1}^{n} \frac{1}{n - 1 - k} \sqrt{E\{\tau_i^2\}} = o(1),
$$

$$
E \left\{ Var \left( \frac{e^\top (I - P_1 - H_k) e}{n - 1 - k} | XP_k \right) \right\}
$$

$$
\leq \sum_{i=1}^{n} E(c_i^4) \frac{2(\sigma^2 + \omega^2)^2}{(n - 1 - k)^2} + \frac{4n(\sigma^2 + \omega^2) \sum_{i=1}^{n} \sqrt{E(\tau_i^2)} + 2 \sum_{i,j} \sqrt{E(\tau_i^2)E(\tau_j^2)}}{(n - 1 - k)^2} = o(1),
$$

$$
Var \left\{ E \left( \frac{e^\top (I - P_1 - H_k) e}{n - 1 - k} | XP_k \right) \right\} = Var \left( \sum_{i=1}^{n} \frac{(I - P_1 - H_k) \tau_i}{n - 1 - k} \right)
$$

$$
\leq E \left\{ \left( \sum_{i=1}^{n} \frac{(I - P_1 - H_k)^2 \tau_i}{(n - 1 - k)^2} \right) \left( \sum_{i=1}^{n} \tau_i^2 \right) \right\}
$$

$$
\leq \frac{n}{(n - 1 - k)^2} \sum_{i=1}^{n} E(\tau_i^2) = o(1).
$$

Consequently, Markov’s inequality leads to

$$
\frac{e^\top (I - P_1 - H_k) e}{n - 1 - k} = \sigma^2 + \omega^2 + o_p(1).
$$

This combines with (S2.8) shows

$$
\frac{(e + r)^\top (I - P_1 - H_k)(e + r)}{n - 1 - k} = \sigma^2 + \omega^2 + o_p(1). \tag{S2.10}
$$

Next, we study $\frac{\sqrt{n}}{k} \xi^\top P_k^\top X^\top (I - P_1)XP_k \xi$. From Assumption A1, we have

$$
E \left\{ \frac{1}{\sqrt{n}} \xi^\top P_k^\top X^\top (I - P_1)XP_k \xi \right\} = \frac{n - 1}{\sqrt{n}} \xi^\top P_k^\top \Sigma P_k \xi \tag{S2.11}
$$
and the fourth moment of $x_i^T P_k \xi$ satisfies

$$E\{ (x_i^T P_k \xi)^4 \} = \sum_{i=1}^{m} (\Gamma^T P_k \xi)^4_i E(z_i^4) + 3 \sum_{i \neq j}^{m} (\Gamma^T P_k \xi)^2_i (\Gamma^T P_k \xi)^2_j E(z_i^2 z_j^2).$$

Based on $\xi^T P_k^T \Sigma P_k \xi = \beta^T \Sigma P_k (P_k^T \Sigma P_k)^{-1} P_k^T \Sigma \beta \leq \beta^T \Sigma \beta = o(1)$, we have

$$Var\left( \frac{1}{\sqrt{n}} \xi^T P_k^T X^T (I - P_1) X P_k \xi \right) \leq E\{ (x_i^T P_k \xi)^4 \} + 2(\xi^T P_k^T \Sigma P_k \xi) = o(1).$$

From Markov’s inequality and $k/n \to \rho$, we have

$$\sqrt{n} \xi^T P_k^T X^T (I - P_1) X P_k \xi = \frac{\sqrt{n}}{\rho} \xi^T P_k^T \Sigma P_k \xi + o_p(1). \quad (S2.12)$$

To investigate $\sqrt{n} \xi^T P_k^T X^T (I - P_1) e$, the condition $\beta^T \Sigma \beta = o(1)$, (S2.9) and (S2.11) lead to

$$E\left\{ \left( \frac{1}{\sqrt{n}} \xi^T P_k^T X^T (I - P_1) e \right)^2 \right\}$$

$$= E \left\{ E\left\{ \left( \frac{1}{\sqrt{n}} \xi^T P_k^T X^T (I - P_1) e \right)^2 \mid XP_k \right\} \right\}$$

$$= E \left\{ \frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \omega^2 + \tau_i) (x_i^T P_k \xi - \frac{1}{n} \sum_{j=1}^{n} x_j^T P_k \xi)^2 \right\}$$

$$\leq (\sigma^2 + \omega^2) E \left\{ \frac{1}{n} \xi^T P_k^T X^T (I - P_1) XP_k \xi \right\}$$

$$+ \sqrt{E \left\{ \frac{1}{n} \sum_{i=1}^{n} \tau_i^2 \right\}} \sqrt{E \left\{ \frac{1}{n} \sum_{i=1}^{n} (x_i^T P_k \xi - \sum_{j=1}^{n} x_j^T P_k \xi)^4 \right\}}$$

$$\leq (\sigma^2 + \omega^2) \xi^T P_k^T \Sigma P_k \xi + \sqrt{c_1} \beta^T \Sigma \beta \sqrt{E \left\{ \frac{16}{n} \sum_{i=1}^{n} (x_i^T P_k \xi)^4 \right\}}$$

$$= o(1).$$
Therefore, Markov’s inequality and $k/n \to \rho$ demonstrate

$$\frac{\sqrt{n}}{k} \xi^\top P_k^\top X^\top (I - P_1)e = o_p(1).$$

This combines with (S2.8) and (S2.12) implies

$$\frac{\sqrt{n}}{k} \xi^\top P_k^\top X^\top (I - P_1)(e + r) = o_p(1).$$

(S2.13)

Based on the new expression (S2.4), together with (S2.8), (S2.10), (S2.12) and (S2.13), we have

$$T_n - 1 = \frac{\sqrt{n(1-\rho)}}{\sqrt{2/[n\rho(1-\rho)]}} \left\{ \frac{\xi^\top P_k^\top X^\top (I - P_1)X\xi_k}{k} + \frac{2\xi^\top P_k^\top X^\top (I - P_1)(e + r)}{n-k-1} \right\} + \frac{(e + r)^\top M(e + r)}{\sigma^2 + \omega^2 + o_p(1)}.$$

Define $\delta_k^2 = \sigma^2 + \beta^\top \Sigma \beta - \xi^\top P_k^\top \Sigma P_k \xi$. From (S2.7), the asymptotic power function of the proposed test $T_n$ is

$$\Psi_{n, RP}^R(\beta; P_k) = P\left( \frac{T_n - 1}{\sqrt{2/[n\rho(1-\rho)]}} > z_\alpha \right)$$

$$= \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho} \frac{\xi^\top P_k^\top \Sigma P_k \xi}{\delta_k^2}}) + o(1),$$

which completes the proof.
S2.3 Proof of Theorem 4.1

Recall the definitions of projection matrices.

\[ P_1 = \frac{1}{n} \mathbf{1} \mathbf{1}^\top, \]
\[ P_{X_1} = (I - P_1)X_1(I - P_1)X_1^\top(I - P_1), \]
\[ H_{k_2} = (I - P_1)W(W^\top(I - P_1)W)^{-1}W^\top(I - P_1), \]

where \( W = (X_1, X_2P_{k_2}) \). Under \( H_{part,0} \), we have

\[ T_{n,p_2} = \frac{\mathbf{e}^\top(H_{k_2} - P_{X_1})\mathbf{e}/k_2}{\mathbf{e}^\top(I - P_1 - H_{k_2})\mathbf{e}/(n - 1 - p_1 - k_2)}. \]

Define \( M = (m_{ij}) = \frac{H_{k_2} - P_{X_1}}{k_2} - \frac{I - P_1 - H_{k_2}}{n - 1 - p_1 - k_2} \). From \( \text{Span}\{(I - P_1)X_1\} \subseteq \text{Span}\{(I - P_1)W\} \) and properties of projection matrices, we have

\[ P_{X_1}H_{k_2} = H_{k_2}P_{X_1} = P_{X_1}. \]

Hence, \( tr(M) = 0 \), \( M^\top M = \frac{H_{k_2} - P_{X_1}}{k_2^2} + \frac{I - P_1 - H_{k_2}}{(n - 1 - p_1 - k_2)^2} \), and

\[ \frac{||M||^2_{sp}}{||M||^2_{F}} = \frac{\lambda_{\max}(M^\top M)}{tr(M^\top M)} \leq \frac{\lambda_{\max}(\frac{H_{k_2} - P_{X_1}}{k_2^2}) + \lambda_{\max}(\frac{I - P_1 - H_{k_2}}{(n - 1 - p_1 - k_2)^2})}{\frac{1}{k_2^2} + \frac{1}{n - 1 - p_1 - k_2}} = O(n^{-1}). \]

For given \( M \), we have

\[ E(\mathbf{e}^\top M\mathbf{e}|M) = \sigma^2 tr(M) = 0, \]
\[ Var(\mathbf{e}^\top M\mathbf{e}|M) = (\mu_4 - 3\sigma^4) \sum_{i=1}^{n} m_{ii}^2 + 2\sigma^4 \left( \frac{1}{k_2} + \frac{1}{n - 1 - p_1 - k_2} \right). \]

Then, Lemma 2 leads to

\[ \frac{\mathbf{e}^\top M\mathbf{e}}{\sqrt{Var(\mathbf{e}^\top M\mathbf{e}|M)}} \xrightarrow{D} \mathcal{N}(0, 1). \]
This together with the law of total expectation and the dominated convergence theorem shows

\[
P\left( \frac{\epsilon^\top M\epsilon}{\sqrt{\text{Var}(\epsilon^\top M\epsilon|M)}} \leq \alpha \right) = E\left[ P\left( \frac{\epsilon^\top M\epsilon}{\sqrt{\text{Var}(\epsilon^\top M\epsilon|M)}} \leq \alpha|M \right) \right] \to \Phi(\alpha),
\]

for \( \forall \alpha \in \mathbb{R} \). Therefore,

\[
\frac{\epsilon^\top M\epsilon}{\sigma^2 \sqrt{[E\{(\frac{\epsilon}{\sigma})^4\} - 3]\sum_{i=1}^{n} m_{ii}^2 + 2(\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2})}} \xrightarrow{D} \mathcal{N}(0, 1).
\]

When \( n \sum_{i=1}^{n} m_{ii}^2 = o_p(1) \), Assumption S3 and Slutsky’s lemma demonstrate

\[
\frac{\epsilon^\top M\epsilon}{\sigma^2 \sqrt{2(1 - \rho_1)/n\rho_2(1 - \rho_1 - \rho_2)}} \xrightarrow{D} \mathcal{N}(0, 1). \quad (S2.14)
\]

Let \( G_n = \sum_{i=1}^{n} m_{ii}^2 \). Next, we will verify \( nG_n = o_p(1) \). From the definition,

\[
m_{ii} = \frac{(H_{k_2})_{ii} - (P_{X_1})_{ii}}{k_2} - \frac{1 - \frac{n}{n-1-p_1-k_2}}{n-1-p_1-k_2}. \]

Then

\[
nG_n = n \sum_{i=1}^{n} m_{ii}^2 = \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{(1 - \frac{1}{n} - \frac{p_1}{n})((H_{k_2})_{ii} - \frac{p_1+k_2}{n}) - \frac{(P_{X_1})_{ii} - \frac{p_1}{n}}{k_2}}{n} \right\}^2 \leq \frac{2h_1}{n} \sum_{i=1}^{n} \left\{ (H_{k_2})_{ii} - \frac{p_1+k_2}{n} \right\}^2 + \frac{2h_2}{n} \sum_{i=1}^{n} \left\{ (P_{X_1})_{ii} - \frac{p_1}{n} \right\}^2
\]

(S2.15)

where \( h_1 = (1 - \frac{1}{n} - \frac{p_1}{n})^2/((k_2(1 - \frac{1}{n} - \frac{p_1}{n} - k_2))^2 \) and \( h_2 = n^2/k_2^2 \). Based on Assumption S3, as \( n \to \infty \),

\[
h_1 \to \frac{(1 - \rho_1)^2}{\rho_2^2(1 - \rho_1 - \rho_2)^2}, \quad h_2 \to \frac{1}{\rho_2^2},
\]
Consequently, we only need to consider the sum parts in \[(S2.15)\]. From the definition,

\[
W = (X_1, X_2 P_{k_2}) = Z \Gamma^\top \begin{pmatrix} I_{p_1} & 0 \\ 0 & P_{k_2} \end{pmatrix} = Z \Gamma^\top V,
\]

where \(Z = (z_1, ..., z_n)^\top\) and \(V\) is a full column rank matrix with probability 1. Define \(\Sigma_2 = V^\top \Sigma V\). The matrix \(\Sigma_2\) is of full rank with probability 1, then \(\Gamma^\top V \Sigma_2^{-1/2}\) is well defined on the Stiefel manifold \(V_{p_1+k_2}(\mathbb{R}^m)\). Let \(W_1 = W \Sigma_2^{-1/2} = Z \Gamma^\top V \Sigma_2^{-1/2}\). The hat matrix \(H_{k_2}\) can be denoted as

\[
H_{k_2} = (I - P_1) W_1 (W_1^\top (I - P_1) W_1)^{-1} W_1^\top (I - P_1).
\]

According to Lemma \ref{lemma7} and the condition \((p_1 + k_2)/n \to \rho_1 + \rho_2\), we obtain

\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ (H_{k_2})_{ii} - \frac{p_1 + k_2}{n} \right\}^2 = o_p(1).
\]

Let \(R_1 = Z \Gamma_1 \Sigma_{11}^{-1/2}\). The hat matrix \(P_{X_1}\) can be denoted as

\[
P_{X_1} = (I - P_1) R_1 (R_1^\top (I - P_1) R_1)^{-1} R_1^\top (I - P_1).
\]

Based on Lemma \ref{lemma7} and the condition \(p_1/n \to \rho_1\), we obtain

\[
\frac{1}{n} \sum_{i=1}^{n} \left\{ (P_{X_1})_{ii} - \frac{p_1}{n} \right\}^2 = o_p(1).
\]

Therefore, \(nG_n = o_p(1)\) is verified, and then \((S2.14)\) is demonstrated.
To study the denominator of $T_{n,p}$, calculation shows

$$E\left\{ \frac{\epsilon^\top (I - P_1 - H_{k_2}) \epsilon}{n - 1 - p_1 - k_2} \right\} = E\left[ E\left\{ \frac{\epsilon^\top (I - P_1 - H_{k_2}) \epsilon}{n - 1 - p_1 - k_2} | H_{k_2} \right\} \right] = \sigma^2,$$

$$Var\left( \frac{\epsilon^\top (I - P_1 - H_{k_2}) \epsilon}{n - 1 - p_1 - k_2} \right) = E\left\{ Var\left( \frac{\epsilon^\top (I - P_1 - H_{k_2}) \epsilon}{n - 1 - p_1 - k_2} | H_{k_2} \right) \right\} = o(1).$$

From Markov’s inequality, we have

$$\frac{\epsilon^\top (I - P_1 - H_{k_2}) \epsilon}{n - 1 - p_1 - k_2} = \sigma^2 + o_p(1).$$

Combining this with (S2.14), we obtain

$$\frac{T_{n,p} - 1}{\sqrt{2(1 - \rho_1)/n \rho_2 (1 - \rho_1 - \rho_2)}} \xrightarrow{D} N(0, 1),$$

which completes the proof.

### S2.4 Proof of Theorem 4.2

Define $V = \text{diag}(I_{p_1}, P_{k_2})$. The matrix is a full column rank matrix with probability 1, and $W = XV$, with the ith row $w_i = V^\top x_i$. Let $\gamma = (V^\top \Sigma V)^{-1} V^\top \Gamma \Sigma^2 y_2$. For each i, define

$$r_i = E\left( x_{2i}^\top \beta_2 | V^\top x_i \right) - x_{1i}^\top V \gamma, \quad q_i = x_{2i}^\top \beta_2 - E\left( x_{2i}^\top \beta_2 | V^\top x_i \right).$$

Then, a decomposition of $x_{2i}^\top \beta_2$ can be derived, given as $x_{2i}^\top \beta_2 = w_i^\top \gamma + r_i + q_i$. Let $\omega^2 = \beta_2^\top \Sigma_{22} \beta_2 - \gamma^\top V \Sigma V \gamma$ and $\tau_i = Var(q_i | V^\top x_i) - \omega^2$. According to Lemma 8 and the condition $\beta_2^\top \Sigma_{22} \beta_2 = o(1)$, we have

$$\sum_{i=1}^n r_i^2 = o_p(1) \text{ and } \frac{1}{\sqrt{n}} |\tau_i| = o_p(1), \quad (S2.16)$$
S2. PROOF OF THEOREMS

when the event $A \in F_n$ is satisfied, where $A = \Gamma^T V (V^T \Sigma V)^{-1/2}$ and $F_n$ is a series of sets that satisfy $\nu_{m,(p_1+k_2)}(F_n) \to 1$, as $n \to \infty$. The probability of the event tends to 1, based on the randomness of $P_{k_2}$.

Define a new error term $e_i = q_i + \epsilon_i$. Let $\sigma^2$ denote the variance of $\epsilon_i$.

The model can be expressed as

$$y = \alpha \mathbf{1} + X_1 \beta_1 + W \gamma + r + e,$$  \hspace{1cm} (S2.17)

where $r = (r_1, \ldots, r_n)^T$, and $e = (e_1, ..., e_n)^T$ with each elements of $e$ satisfying $E(e_i) = 0$, $E(e_i|V^T x_i) = 0$, $Var(e_i|V^T x_i) = \sigma^2 + \omega^2 + \tau_i$, and $E(e_i|V^T x_i) = \mu_4 + 6\sigma^2 Var(q_i|V^T x_i) + E(q_i|V^T x_i)$. Define $M = \frac{H_{k_2} - P X_{1}}{k_2} - \frac{I - P_{1} - H_{k_2}}{n-1-p_1-k_2}$. The matrix satisfies $tr(M) = 0$, $tr(MM^T) = \frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}$, and $\|M\|_{sp}^2 \leq \frac{1}{k_2^2} + \frac{1}{(n-1-p_1-k_2)^2}$. Based on the condition $p_1/n \to \rho_1$ and $k_2/n \to \rho_2$, then for large $n$, there is a constant $a \leq 2/\rho_2(1-\rho_1-\rho_2)$ such that $\|M\|_{sp} \leq a/n$. With a similar proof method in Appendix S2.2, we can derive

$$\sqrt{n \sigma^2 + \omega^2} \sqrt{1 + o_p(1)} \frac{E(e_i)}{M^T Me - o_p(1)} \xrightarrow{d} \mathcal{N}(0, 1).$$  \hspace{1cm} (S2.18)

The condition $\beta_2^T \Sigma_{22} \beta_2 = o(1)$ leads to $E(q_i^4) = o(1)$ as well as

$$|E(e_i^4) - \mu_4| \leq c_1 \beta_2^T \Sigma_{22} \beta_2 = o(1), \quad E\{\tau_i^2\} \leq E(q_i^4) + \omega^4 \leq c_1 (\beta_2^T \Sigma_{22} \beta_2)^2 = o(1),$$  \hspace{1cm} (S2.19)
for a constant $c_1$, from which we could obtain

$$
\frac{(e + r)^\top (I - P_1 - H_{k_2})(r + e)}{n - 1 - p_1 - k_2} = \sigma^2 + \omega^2 + o_p(1).
$$

(S2.20)

Let $V \gamma = (\xi_1^\top, \xi_2^\top)^\top$ with $\xi_1 \in \mathbb{R}^{p_1}$ and $\xi_2 \in \mathbb{R}^{p_2}$. Define $\nu^2 = \xi_2^\top (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}) \xi_2$. Then

$$
\nu^2 = \beta_2^\top \Gamma_2 (\Gamma_1^\top V (V^\top \Sigma V)^{-1} V^\top \Gamma_1)^{-1} \Gamma_1 \xi_2.
$$

To investigate $\gamma^\top W^\top (I - P_1 - P_{X_1}) W \gamma$, the term could be denoted as

$$
\gamma^\top W^\top (I - P_1 - P_{X_1}) W \gamma
$$

$$
= \phi^\top Z^\top (I - P_1) Z \phi - \phi^\top Z^\top (I - P_1) Z \Gamma_1^\top (\Gamma_1 Z^\top (I - P_1) Z \Gamma_1)^{-1} \Gamma_1 Z^\top (I - P_1) Z \phi
$$

where $\phi = (I - \Gamma_1^\top \Sigma_{11}^{-1} \Sigma_{12}) \Gamma_2^\top \xi_2$ and $\phi^\top \phi = \nu^2 = o(1)$. From the calculation

$$
E \left\{ \frac{1}{\sqrt{n}} \phi^\top Z^\top (I - P_1) Z \phi \right\} = \frac{n - 1}{\sqrt{n}} \nu^2,
$$

$$
Var \left\{ \frac{1}{\sqrt{n}} \phi^\top Z^\top (I - P_1) Z \phi \right\} \leq 6\nu^4 = o(1).
$$

Markov’s inequality implies,

$$
\frac{1}{\sqrt{n}} \phi^\top Z^\top (I - P_1) Z \phi = \sqrt{n} \nu^2 + o_p(1).
$$

From a similar derivation method for (S2.20), we obtain

$$
\frac{1}{\sqrt{n}} \phi^\top Z^\top (I - P_1) Z \Gamma_1^\top (\Gamma_1 Z^\top (I - P_1) Z \Gamma_1)^{-1} \Gamma_1 Z^\top (I - P_1) Z \phi = \frac{p_1}{\sqrt{n}} \nu^2 + o_p(1).
$$

Therefore,

$$
\frac{1}{\sqrt{n}} \gamma^\top W^\top (I - P_1 - P_{X_1}) W \gamma = \frac{n - p_1}{\sqrt{n}} \nu^2 + o_p(1).
$$

(S2.21)
S2. PROOF OF THEOREMS

To study $\gamma^T W^T (I - P_1 - P_{X_1})(e + r)$, (S2.16) and (S2.21) lead to

$$\left| \frac{1}{\sqrt{n}} \gamma^T W^T (I - P_1 - P_{X_1}) r \right| \leq \sqrt{r^T r} \sqrt{\frac{1}{n} \gamma^T W^T (I - P_1 - P_{X_1}) W \gamma} = o_p(1).$$

the condition $\beta_2^T \Sigma_{22} \beta_2 = o(1)$, (S2.19) and (S2.21) lead to

$$E \left\{ \left( \frac{1}{\sqrt{n}} \gamma^T W^T (I - P_1 - P_{X_1}) e \right)^2 \right\} = E \left[ E \left\{ \left( \frac{1}{\sqrt{n}} \gamma^T W^T (I - P_1 - P_{X_1}) e \right)^2 | W \right\} \right]$$

$$\leq (c_3 + \sigma^2 + \omega^2) \beta_2^T \Sigma_{22} \beta_2$$

$$= o(1),$$

where $c_3$ is a constant. Therefore, we obtain

$$\frac{1}{\sqrt{n}} \gamma^T W^T (I - P_1 - P_{X_1})(e + r) = o_p(1). \quad (S2.22)$$

From the new expression (S2.17), together with (S2.16), (S2.20), (S2.21) and (S2.22), we have

$$T_{n,p_2} - 1 \frac{\sqrt{2(n - 1 - \rho_1) / n \rho_2}}{\sqrt{2 \rho_1 (1 - \rho_1) / n \rho_2}} = \sqrt{\frac{n \rho_2 (1 - \rho_1 - \rho_2)}{2(1 - \rho_1)}} \left\{ \frac{\gamma^T W^T (I - P_1 - P_{X_1})(W \gamma + 2e + 2r)}{k_2} + (r + e)^T M (r + e) \right\}$$

$$= \sqrt{\frac{n \rho_2 (1 - \rho_1 - \rho_2)}{2(1 - \rho_1)}} \left\{ \frac{\left( \frac{1}{\rho_2} \right) \nu^2 + e^T M e}{\sigma^2 + \omega^2 + o_p(1)} \right\} + o_p(1).$$

Define $\tau_k^2 = \sigma^2 + \omega^2$. Then, $\nu^2$ and $\tau_k^2$ can also be calculated as follows. Let

$$\bar{\gamma} = (V^T \Sigma V)^{-1} V^T \Sigma \beta$$

and

$$V \bar{\gamma} = (\bar{\xi}_1, \bar{\xi}_2)^T,$$

where $\bar{\xi}_1 \in \mathbb{R}^{p_1}$ and $\bar{\xi}_2 \in \mathbb{R}^{p_2}$. 

Then,
\[
\nu^2 = \beta_2^\top \Gamma_2 \left( \Gamma^\top V (V^\top \Sigma V)^{-1} V^\top \Gamma - \Gamma_1^\top \Sigma_{11}^{-1} \Gamma_1 \right) \Gamma_2^\top \beta_2
\]
\[
= \beta^\top \Gamma \left( \Gamma^\top V (V^\top \Sigma V)^{-1} V^\top \Gamma - \Gamma_1^\top \Sigma_{11}^{-1} \Gamma_1 \right) \Gamma^\top \beta
\]
\[
= \tilde{\xi}_2^\top (\Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}) \tilde{\xi}_2.
\]
and
\[
\tau_k^2 = \sigma^2 + \beta_2^\top \Sigma_{22} \beta_2 - \gamma^\top V^\top \Sigma V \gamma = \sigma^2 + \beta^\top \Sigma \beta - \tilde{\gamma}^\top V^\top \Sigma V \tilde{\gamma}.
\]

From (S2.18), the asymptotic power function of the proposed test \( T_{n,p_2} \) is
\[
\Psi_{n,p_2}^{RP} (\beta_2; \mathbf{P}_{k_2}) = P\left( \frac{T_{n,p_2} - 1}{\sqrt{2(1 - \rho_1)/n \rho_2 (1 - \rho_1 - \rho_2)}} > z_\alpha \right)
\]
\[
= \Phi(-z_\alpha + \sqrt{\frac{n(1 - \rho_1 - \rho_2)(1 - \rho_1) \nu^2}{2 \rho_2 \tau_k^2}}) + o(1),
\]
which completes the proof.

S3 Simulations

In the second simulation study, we consider the problem of testing the partial regression coefficient in the linear model
\[
y_i = \alpha + x_{1i}^\top \beta_1 + x_{2i}^\top \beta_2 + \epsilon_i.
\]

The covariate \((x_{1i}^\top, x_{2i}^\top)^\top\) is generated from \( \mu + \Sigma^{1/2} z_i \). The setup is almost the same as the first simulation study with differences lying in the
S3. SIMULATIONS

design of $\beta_1$, $\beta_2$ and $\Sigma^{1/2}$. Specifically, we generated $\Sigma^{1/2}$ by

$$
\begin{pmatrix}
    c_1 U_1 \sqrt{D_1} U_1^\top & c_2 U_1 (\sqrt{D_1}, 0) U_2^\top \\
    0 & U_2 \sqrt{D_2} U_2^\top
\end{pmatrix},
$$

where $U_1$ ($U_2$) is an orthogonal matrix generated from the uniform distribution on the $p_1 \times p_1$ ($p_2 \times p_2$) orthogonal group, the entries of diagonal matrix $D_1$ are from $\mathcal{N}(0, I_{p_1})$ with absolute values taken and the entries of diagonal matrix $D_2$ are generated in the same way as the first simulation study for the small tail eigenvalue requirement. We used an indicator $R$ for the different cases: (i) uncorrelated case ($R = 0$): $c_1 = 1, c_2 = 0$; (ii) correlated case ($R = 1$): $c_1 = c_2 = 1/\sqrt{2}$. Here, the values of $c_1$ and $c_2$ are selected to ensure the variances of $x_{1i}$ and $x_{2i}$ keep unchanged in the two cases. The regression coefficient $\beta_1$ is generated from $\mathcal{N}(0, I_{p_1})$ and $\beta_2$ is randomly selected from the space generated by the first $s$ columns of $U_2$ with $||\beta_2||_2^2$ taking 0.1, 0.2, and 0.3. This selection is aimed for a better display of the impact from the correlation on the power of the tests. For a high-dimensional design, we chose $(n, p_1, p_2)$ to be (400, 40, 3960).

Figures 1a and 1b display the kernel density estimation of the proposed test statistics under $H_{\text{part}, 0}$, indicating that the asymptotic null distribution of the proposed tests can be well approximated by the standard normal distribution. Here, $\rho$ takes the value 0.2. We show both the correlated
and uncorrelated cases. The good resemblance to the normal distribution confirms the theoretical results in Theorem 4.1.

Table 1 reports the empirical power and type-I error of the proposed tests for the error term $\epsilon$ distributed from $\mathcal{N}(0, 1)$ and $\sqrt{3/5}t(5)$, based on 2000 simulations. It can be observed that the performances of the three proposed tests have negligible differences. The type-I errors of the proposed tests are close to 0.05 and the power of the tests are increasing functions of the norm $||\beta_2||_2^2$. Compared with the correlated case, the tests show large power when there is no correlation between $x_{1i}$ and $x_{2i}$, which is consistent with the feature in the asymptotic power in Theorem 4.2. Moreover, we find the empirical power is close to the asymptotic power, which further confirms the result in Theorem 4.2.

Figure 1: The kernel density estimation of RP, multi-RP, and S-RP tests under $H_{part,0}$.
In the third simulation, we conducted numerical comparison with the LWT test and LDFF test proposed in Lan, Wang, and Tsai (2014) and Lan et al. (2016), respectively. The data are generated from $y_i = \alpha + x_i^\top \beta + \epsilon_i$, where $\alpha = 0$ and $\epsilon_i$ is generated from $\mathcal{N}(0, 1)$. The covariate $x_i$ follows a latent factor structure in Lan et al. (2016). Specifically, $x_i = \gamma z_i + \sqrt{D} \tilde{x}_i$, where $z_i$ is a d-dimensional latent factor, $\gamma \in \mathbb{R}^{p \times d}$ is an associated factor loadings, $\tilde{x}_i$ is a p-dimensional factor profiled predictor that is independent of $z_i$, and $D$ is a diagonal matrix. From Lan et al. (2016), the factor profiled predictor $\tilde{x}_i$ represents the information that is contained in $x_i$ but cannot be fully explained by the low-dimensional latent factor $z_i$. In the simulation,
each element of \( z_i \) and \( \tilde{x}_i \) is independently generated from \( \mathcal{N}(0, 1) \), and each entry of \( \gamma \in \mathbb{R}^{p \times d} \) is independently generated from \( \mathcal{N}(0, d^{-1}) \). The elements of \( \sqrt{D} \) are generated in the same way as that in the first set of simulation, when \( s = [n^{0.5}] \) and \( L = [n^{1.5}] \). For the alternative hypothesis, we considered \( \beta = ||\beta||_2 \delta \), where \( \delta = (\delta_1, \ldots, \delta_p)^\top \) with \( \delta_j = s^{-1/2} \), for \( j \leq s \), and otherwise, \( \delta_j = 0 \). The integer \( s \) takes values 5 and 50 to denote different levels of sparsity, and the norm \( ||\beta||_2^2 = 0.04 \) and 0.08. In the simulation, \( (n, p) = (300, 3000) \).

Table 2: Empirical power and type-I error of the multi-RP, RCV, LWT, and LDFF tests at the significance level 0.05.

| d  | \( \beta \) | \( ||\beta||_2^2 \) | multi-RP | LWT | LDFF | RCV |
|----|-------------|-----------------|----------|-----|------|-----|
| d=3 |             |                 |          |     |      |     |
|     | 0           | 0.062           | 0.052    | 0.050 | 0.458 |
| s=5 | 0.04        | 0.249           | 0.087    | 0.116 | 0.544 |
|     | 0.08        | 0.532           | 0.116    | 0.502 | 0.454 |
| s=50| 0.04        | 0.735           | 0.218    | 0.787 | 0.951 |
|     | 0.08        | 0.984           | 0.388    | 0.951 | 0.987 |
| d=5 |             |                 |          |     |      |     |
| s=5 | 0           | 0.052           | 0.071    | 0.069 | 0.843 |
|     | 0.04        | 0.295           | 0.183    | 0.181 | 0.917 |
| s=50| 0.08        | 0.605           | 0.312    | 0.308 | 0.959 |
|     | 0.04        | 0.764           | 0.387    | 0.384 | 0.999 |
| s=50| 0.08        | 0.987           | 0.698    | 0.681 | 1.000 |

As shown in Table 2, the type-I errors of the multi-RP, LWT and LDFF tests are around 0.05, which indicates that the type-I error can be well controlled at the nominal level by the tests. But for the RCV test, the type-I errors are alarmingly larger than the given significance level, which indicates
the test might not be applicable in this experimented setting, where the co-
variates have high correlations based on the latent factor structure. Therefore, the comparison for the empirical powers is only considered among the
multi-RP test, LWT test and LDFF test. Table 2 indicates that empirical
powers grow when $\|\mathcal{A}\|_2$ increases and the performances of the LWT and
LDFF tests are similar. The large empirical powers demonstrate that our
proposed test has superior performances in all the experimented alternatives. Therefore, the simulation results demonstrate that our proposed test
is applicable in the highly correlated setting and has higher testing power
than the competing tests in some cases.

Bibliography

large-dimensional sample covariance matrix. The Annals of Probability
21, 1275–1294.

Bechar, I. (2009). A Bernstein-type inequality for stochastic processes of

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007). Convergence of


