New Tests for High-Dimensional Linear Regression Based on Random Projection

Changyu Liu, Xingqiu Zhao, and Jian Huang

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

Supplementary Material

S1 Lemmas

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

S1.1 Proof of Lemma 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 S1. *Let $\boldsymbol{u}_1 = (u_{11}, ..., u_{1p})^\top$ be a random vector uniformly dis-*

tributed on the unit sphere in $\mathbb{R}^p$*. Then* $\boldsymbol{u}_1$ *satisfies* $E(\boldsymbol{u}_1) = \boldsymbol{0}, Var(\boldsymbol{u}_1) = \frac{1}{p}\boldsymbol{I}_p$*. 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* $E\big(\Pi_{j=1}^p u_{1j}^{q_j}\big) = 0$ *if at least one* q_j *is odd.*

Proof of Lemma 1. From the definition of r_1, $\mathbf{u}_1$ and Lemma S1, we have

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

$$Var(\mathbf{z}_1) = Var(E(\mathbf{z}_1|r_1)) + E(Var(\mathbf{z}_1|r_1)) = E(r_1^2 Var(\mathbf{u}_1)) = \mathbf{I}_p.$$

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

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

Hence, we have

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

and complete the proof. □

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 S2. *Consider a general quadratic form*

$$Q_n = \boldsymbol{z}^\top \boldsymbol{A}_n \boldsymbol{z} = \sum_{i,j=1}^n z_i a_{ij} z_j,$$

where z_i *are i.i.d. variables with* $E(z_i) = 0$ *and* $Var(z_i) = 1$*, and* a_{ij} *are entries of a symmetric matrix* $\boldsymbol{A}_n$*.*

(1) If $E(z_i^4) < \infty$ *and* $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$*, then*

$$Var(Q_n)^{-1/2}\big(Q_n - E(Q_n)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

(2) If $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$*,* $E(z_i^{2+\delta}) < \infty$ *(for some* $\delta > 0$*), and* $\sum_{i=1}^{n} a_{ii}^2 = o(||\boldsymbol{A}_n||_F^2)$*, then*

$$\frac{1}{\sqrt{2}||\boldsymbol{A}_n||_F}\big(Q_n - E(Q_n)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Lemma S3 (Woodbury's formula)**.** *Suppose* $\boldsymbol{G}$ *is an* $n \times n$ *nonsingular matrix,* $\boldsymbol{U}$ *and* $\boldsymbol{V}$ *are* $n \times k$ *matrices, with* $n > k$*. If the matrix* $(\boldsymbol{I}_k + \boldsymbol{V}^\top \boldsymbol{G}^{-1}\boldsymbol{U})$ *is invertible, we have*

$$(\boldsymbol{G} + \boldsymbol{U}\boldsymbol{V}^\top)^{-1} = \boldsymbol{G}^{-1} - \boldsymbol{G}^{-1}\boldsymbol{U}(\boldsymbol{I}_k + \boldsymbol{V}^\top \boldsymbol{G}^{-1}\boldsymbol{U})^{-1}\boldsymbol{V}^\top \boldsymbol{G}^{-1}.$$

Suppose $\boldsymbol{u}$ *and* $\boldsymbol{v}$ *are vectors. Define* $\boldsymbol{H} = \boldsymbol{u}\boldsymbol{v}^\top$ *and* $g = tr(\boldsymbol{H}\boldsymbol{G}^{-1})$*. If* $g \neq -1$*, we have*

$$(\boldsymbol{G} + \boldsymbol{H})^{-1} = \boldsymbol{G}^{-1} - \frac{1}{1+g}\boldsymbol{G}^{-1}\boldsymbol{H}\boldsymbol{G}^{-1}.$$

We then depict some results about sample covariance matrix in high dimensions. The first is the celebrated work of Marčenko 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 S4. *Let $\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{k\times n}$ be a matrix of i.i.d. entries with zero mean and unit variance. Define $\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{X}\boldsymbol{X}^\top$. Suppose the eigenvalues of $\boldsymbol{S}_n$ are λ_j, $j = 1, \ldots, k$, the empirical spectral distribution (ESD) of the matrix $\boldsymbol{S}_n$ is defined as $F^{\boldsymbol{S}_n} = \frac{1}{k}\sum_{j=1}^k \mathbf{1}_{\{\lambda_j \le x\}}$. If $E(x_{11}^4) < \infty$, as $(n, k) \to \infty$ with relationship $k/n \to \rho \in (0, 1)$, we have*

(1) $F^{\boldsymbol{S}_n}$ tends to the standard M-P law with probability 1, where the standard M-P law $F_\rho(x)$ has a density function

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

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

(2) The extreme eigenvalues of $\boldsymbol{S}_n$ satisfy

$$\lambda_{\max}(\boldsymbol{S}_n) \to (1 + \sqrt{\rho})^2 \text{ a.s.},$$

and

$$\lambda_{\min}(\boldsymbol{S}_n) \to (1 - \sqrt{\rho})^2 \text{ a.s.}.$$

Lemma S5. *Let $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ be a random matrix with $\boldsymbol{x}_i$ i.i.d. from $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$. As $(k, n) \to \infty$ with relationship $k/n \to \rho \in (0, 1)$, we have*

(1) $\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top$ and $\bar{\boldsymbol{x}}$ are independent, where $\boldsymbol{1} = (1, ..., 1)^\top$, $\boldsymbol{P}_1 = \frac{1}{n}\boldsymbol{1}\boldsymbol{1}^\top$ and $\bar{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i$.

(2) $E\Big((\frac{1}{n-1}\boldsymbol{x}_n^\top \boldsymbol{S}_{n-1}^{-1}\boldsymbol{x}_n - \frac{\rho}{1-\rho})^2\Big) = o(1)$, *where* $\boldsymbol{S}_{n-1} = \frac{1}{n-1}\sum_{j=1}^{n-1}\boldsymbol{x}_j\boldsymbol{x}_j^\top$, *and* $E\left((\boldsymbol{x}_i^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{x}_i - \rho)^2\right) = o(1)$, $\boldsymbol{x}_i^\top(\boldsymbol{X}\boldsymbol{X}^\top)^{-1}\boldsymbol{x}_i \le \frac{1}{1+(1-\sqrt{\rho})^2}$, *a.s..*

Proof. (1) We first define an orthogonal matrix $\mathbf{O}$ by

$$\mathbf{O} = (\mathbf{o}_1, ..., \mathbf{o}_n) = \begin{bmatrix} \frac{1}{\sqrt{n}} & 0 & 0 & \cdots & -\frac{\sqrt{n-1}}{\sqrt{n}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\sqrt{n}} & 0 & -\frac{\sqrt{2}}{\sqrt{3}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \cdots & \frac{1}{\sqrt{n(n-1)}} \end{bmatrix}.$$

Let $\mathbf{V} = \mathbf{X}\mathbf{O}$ with the ith column denoted as $\mathbf{v}_i$. Then the design of orthogonal matrix $\mathbf{O}$ implies $\mathbf{X}\mathbf{X}^\top = \mathbf{X}\mathbf{O}\mathbf{O}^\top\mathbf{X}^\top = \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i^\top$, $\mathbf{v}_1 = \sqrt{n}\bar{\mathbf{x}}$ and $\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top = \sum_{i=2}^n \mathbf{v}_i\mathbf{v}_i^\top$. To study the properties of $\mathbf{v}_i$, the random matrix $\mathbf{X}$ is divided by rows and denoted as $(\mathbf{r}_1, ..., \mathbf{r}_k)^\top$ with k independent $\mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ variables. It follows that $\mathbf{v}_i = (\mathbf{r}_1, ..., \mathbf{r}_k)^\top\mathbf{o}_i$ and is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I}_k)$. Let $\mathbf{C}^{i,j} = (C_{s,l}^{i,j})_{s,l=1}^k = Cov(\mathbf{v}_i, \mathbf{v}_j)$, for $i \neq j$. Then we have

$$C_{s,l}^{i,j} = E(\mathbf{r}_s^\top\mathbf{o}_i\mathbf{r}_l^\top\mathbf{o}_j) - E(\mathbf{r}_s^\top\mathbf{o}_i)E(\mathbf{r}_l^\top\mathbf{o}_j) = 0,\ s \neq l,$$

$$C_{s,s}^{i,j} = E(\mathbf{r}_s^\top\mathbf{o}_i\mathbf{r}_s^\top\mathbf{o}_j) - E(\mathbf{r}_s^\top\mathbf{o}_i)E(\mathbf{r}_s^\top\mathbf{o}_j) = E(\mathbf{o}_i^\top\mathbf{r}_s\mathbf{r}_s^\top\mathbf{o}_j) = 0, s = 1, .., k,$$

which indicates $\mathbf{v}_i$ and $\mathbf{v}_j$ are independent. This is sufficient to show that $\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top$ and $\bar{\mathbf{x}}$ are independent.

(2) From the direct calculation, the standard M-P law $F_\rho(x)$ in Lemma S4 satisfies

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

We first study the asymptotic behavior of $\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n$. From normality of $\mathbf{x}_i$, Lemma S4 and the above calculation, we have

$$
\begin{aligned}
E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n | \mathbf{S}_{n-1}) &= \frac{k}{n-1} \frac{tr(\mathbf{S}_{n-1}^{-1})}{k} \\
&= \frac{k}{n-1} \int \frac{1}{x} dF^{\mathbf{S}_{n-1}} \to \frac{\rho}{1-\rho}, \; a.s.,
\end{aligned}
$$

$$Var(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1}) = \frac{2}{(n-1)^2}tr\left((\mathbf{S}_{n-1}^{-1})^2\right)$$
$$\leq \frac{2k}{(n-1)^2}(\frac{1}{\lambda_{\min}(\mathbf{S}_{n-1})})^2 \to 0,\ a.s..$$

Therefore,

$$E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n) = E\big(E(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1})\big) \to \frac{\rho}{1-\rho},$$
$$Var(\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n) \to 0.$$

These lead to the first result,

$$E\left((\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\rho}{1-\rho})^2\right) \to 0.$$

From Lemma S3, we have

$$\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n = \frac{\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}{1+\mathbf{x}_n^\top(\sum_{j\neq n}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_n}$$
$$= \frac{\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n}{1+\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n}.$$

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

$$\left|\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n - \rho\right| \leq \left|\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\rho}{1-\rho}\right|,$$

which implies

$$E\left((\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n - \rho)^2\right) \leq E\big((\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n - \frac{\rho}{1-\rho})^2\big) \to 0.$$

Furthermore, from $\frac{1}{n-1}\mathbf{x}_n^\top \mathbf{S}_{n-1}^{-1}\mathbf{x}_n \le \lambda_{\min}^{-1}(\mathbf{S}_{n-1})\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{x}_n \to \frac{1}{(1-\sqrt{\rho})^2}$ a.s., we obtain

$$\mathbf{x}_n^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_n \le \frac{1}{1+(1-\sqrt{\rho})^2}, \quad a.s.,$$

and complete the proof.

□

Lemma S6. *Let* $\boldsymbol{X} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ *be a random matrix with* $\boldsymbol{x}_i$ *i.i.d. from* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_k)$*. The matrix* $\boldsymbol{H}$ *is defined as* $\boldsymbol{H} = (\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top\left(\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)\boldsymbol{X}^\top\right)^{-1}\boldsymbol{X}(\boldsymbol{I}-\boldsymbol{P}_1)$ *and has its entries denoted by* $\boldsymbol{H}_{ij}$*. As* $(k,n)\to\infty$ *with* $k/n \to \rho \in (0,1)$*, we have*

$$\max_{i=1,\dots,n} E\left[\left(\boldsymbol{H}_{ii}-\rho\right)^2\right] \to 0.$$

Proof. From Lemma S5, we get

$$E\big((n\bar{\mathbf{x}}^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}} - \frac{\rho}{1-\rho})^2\big) \to 0, \tag{S1.1}$$

$$E\big((n\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}} - \rho)^2\big) \to 0, n\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}} \le \frac{1}{1+(1-\sqrt{\rho})^2}, \ a.s., \tag{S1.2}$$

$$E\big((\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1 - \rho)^2\big) \to 0. \tag{S1.3}$$

The proof proceeds in two steps. First, we study $\mathbf{x}_1^\top\left(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\right)^{-1}\mathbf{x}_1$ and show that it converges to ρ in quadratic mean. Second, we divide $\mathbf{H}_{ii}$ 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 $\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1$ is a well approximation to $\mathbf{x}_1^\top\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_1$ and then the convergence is guaranteed by (S1.3). Lemma S3 and (S1.2) imply

$$\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1} = (\mathbf{X}\mathbf{X}^\top)^{-1} + \frac{1}{1+g}(\mathbf{X}\mathbf{X}^\top)^{-1}n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1},$$

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

$$\begin{aligned}
&\big|\mathbf{x}_1^\top\big(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_1 - \mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1\big| \\
&= \frac{1}{1+g}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}n\bar{\mathbf{x}}\bar{\mathbf{x}}^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1 \\
&= \frac{n}{1+g}(\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\bar{\mathbf{x}})^2 \\
&\leq \frac{2}{1+g}\Big[\frac{1}{n}(\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1)^2 + \frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2\Big].
\end{aligned}$$

Based on (S1.3), the expectation of the first part in the sum goes to 0. Then we show the second part $\frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2$ would also converge to 0 in the first mean. Define $\mathbf{A}_{1,j} = \sum_{k\neq 1,j}\mathbf{x}_k\mathbf{x}_k^\top$ and $\mathbf{S}_{1,j} = \frac{1}{n-2}\mathbf{A}_{1,j}$. We have $\mathbf{X}\mathbf{X}^\top = \mathbf{A}_{1,j} + \mathbf{x}_1\mathbf{x}_1^\top + \mathbf{x}_j\mathbf{x}_j^\top$. From Lemma S3,

$$\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j = \frac{\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j}{D_{1,j}},$$

where $D_{1,j} = (1+\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_1)(1+\mathbf{x}_j^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j) - (\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2 \geq 1$. Then,

$$\begin{aligned}
E(\frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2) &= E(\frac{1}{n}(\sum_{j\neq 1}\frac{\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j}{D_{1,j}})^2) \\
&= \sum_{j\neq 1}E(\frac{(\mathbf{x}_1^\top\mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2}{nD_{1,j}^2}) + \sum_{j\neq\ell\neq 1}E(\frac{\mathbf{x}_j^\top\mathbf{A}_{1,j}^{-1}\mathbf{A}_{1,\ell}^{-1}\mathbf{x}_\ell}{nD_{1,j}D_{1,\ell}}).
\end{aligned}$$

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

$$E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,j}^{-1}\mathbf{x}_j)^2}{D_{1,j}^2}) = E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) \text{ and } E(\frac{\mathbf{x}_j^\top \mathbf{A}_{1,j}^{-1}\mathbf{A}_{1,\ell}^{-1}\mathbf{x}_\ell}{D_{1,j}D_{1,\ell}}) = E(\frac{\mathbf{x}_2^\top \mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}).$$

Therefore,

$$E(\frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2) = \frac{n-1}{n}E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) + \frac{(n-1)(n-2)}{n}E(\frac{\mathbf{x}_2^\top \mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}). \tag{S1.4}$$

Lemma S4 asserts the first part in (S1.4) converges to 0 by

$$E(\frac{(\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2}{D_{1,2}^2}) \leq E\big((\mathbf{x}_1^\top \mathbf{A}_{1,2}^{-1}\mathbf{x}_2)^2\big) = \frac{k}{(n-2)^2}E(\frac{tr(\mathbf{S}_{1,2}^{-1})^2}{k}) \to 0.$$

Next, we study the second part and show it would also go to 0. Let $\mathbf{A}_{1,2,3} = \sum_{s\neq 1,2,3}\mathbf{x}_s\mathbf{x}_s^\top$, $\mathbf{S}_{1,2,3} = \frac{1}{n-3}\mathbf{A}_{1,2,3}$. Then, $g_3 = \mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3 \geq 0$ and Lemma S3 gives the relationship

$$\mathbf{A}_{1,2}^{-1} = \mathbf{A}_{1,2,3}^{-1} - \frac{1}{1+g_3}\mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3\mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}.$$

From calculations and Lemma S4, we have

$$E(\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3) = 0,\ E\big((\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3)^2|\mathbf{A}_{1,2,3}\big) = \frac{tr\big((\mathbf{S}_{1,2,3}^{-1})^4\big)}{(n-3)^4} = O(n^{-3}),$$

$$E(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3) = 0,\ E\big((\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2|\mathbf{A}_{1,2,3}\big) = \frac{1}{(n-3)^2}tr\big((\mathbf{S}_{1,2,3}^{-1})^2\big) = O(n^{-1}),$$

$$(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2 \leq (\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_2)(\mathbf{x}_3^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3) \leq \frac{\rho^2}{(1-\sqrt{\rho})^4}\ a.s.,$$

$$(n-2)\mathbf{x}_2^\top(\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2 \leq \frac{k(n-2)}{(n-3)^2}\lambda_{\min}^{-2}(\mathbf{S}_{1,2,3})\frac{\mathbf{x}_2^\top\mathbf{x}_2}{k} \leq \frac{\rho}{(1-\sqrt{\rho})^4}\ a.s..$$

These give two upper bounds

$$\frac{1+\frac{(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}} \le 1+\frac{\rho^2}{(1-\sqrt{\rho})^4} \ a.s.,$$
$$\frac{(n-2)(\frac{\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}} \le \frac{2\rho}{(1-\sqrt{\rho})^4} \ a.s..$$

Then, we can get

$$(n-2)^2 E\Big((\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3 \frac{1+\frac{(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}})^2\Big) \to 0,$$
$$E\Big((\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3 \frac{(n-2)(\frac{\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}})^2\Big) \to 0.$$

These together show

$$\begin{aligned}
&E\Big[\frac{(n-1)(n-2)}{n}\frac{\mathbf{x}_2^\top \mathbf{A}_{1,2}^{-1}\mathbf{A}_{1,3}^{-1}\mathbf{x}_3}{D_{1,2}D_{1,3}}\Big] \\
&= E\Big[(n-2)\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3 \frac{1+\frac{(\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3)^2}{(1+g_2)(1+g_3)}}{\frac{n}{n-1}D_{1,2}D_{1,3}}\Big] \\
&- E\Big[\mathbf{x}_2^\top \mathbf{A}_{1,2,3}^{-1}\mathbf{x}_3 \frac{(n-2)(\frac{\mathbf{x}_2^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_2}{1+g_2}+\frac{\mathbf{x}_3^\top (\mathbf{A}_{1,2,3}^{-1})^2\mathbf{x}_3}{1+g_3})}{\frac{n}{n-1}D_{1,2}D_{1,3}}\Big] \\
&\to 0.
\end{aligned}$$

Hence, from (S1.4), we derive $E(\frac{1}{n}(\sum_{j\neq 1}\mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_j)^2) \to 0$. This together with an upper-bound inferred from (S1.3) and (S1.2) leads to

$$E\big[(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1 - \mathbf{x}_1^\top(\mathbf{X}\mathbf{X}^\top)^{-1}\mathbf{x}_1)^2\big] \to 0.$$

And then (S1.3) further shows

$$E\big[(\mathbf{x}_1^\top(\mathbf{X}(\mathbf{I}-\mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1 - \rho)^2\big] \to 0. \tag{S1.5}$$

For any $i \in \{1, \ldots, n\}$, we divide $\mathbf{H}_{ii}$ into three parts

$$\begin{aligned}\mathbf{H}_{ii} &= (\mathbf{x}_i - \bar{\mathbf{x}})^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) \\ &= \bar{\mathbf{x}}^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\bar{\mathbf{x}} - 2\mathbf{x}_i^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\bar{\mathbf{x}} \\ &+ \mathbf{x}_i^\top \big(\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top\big)^{-1}\mathbf{x}_i.\end{aligned}$$

Based on (S1.1) and (S1.5), we obtain

$$\begin{aligned}E\big[(\mathbf{H}_{ii} - \rho)^2\big] &= E\Big[\big((\mathbf{x}_i - \bar{\mathbf{x}})^\top (\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}(\mathbf{x}_i - \bar{\mathbf{x}}) - \rho\big)^2\Big] \\ &= E\Big[\big((\mathbf{x}_1 - \bar{\mathbf{x}})^\top (\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}(\mathbf{x}_1 - \bar{\mathbf{x}}) - \rho\big)^2\Big] \\ &\leq E\Big[3\big(\mathbf{x}_1^\top (\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\mathbf{x}_1 - \rho\big)^2 + 3\big(\bar{\mathbf{x}}^\top (\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\big)^2 \\ &+ 12\big(\mathbf{x}_1^\top (\mathbf{X}(\mathbf{I} - \mathbf{P}_1)\mathbf{X}^\top)^{-1}\bar{\mathbf{x}}\big)^2\Big] \\ &= o(1).\end{aligned}$$

Therefore,

$$\max_{i=1,\ldots,n} E\big[(\mathbf{H}_{ii} - \rho)^2\big] \to 0,$$

which completes the proof. □

Lemma S7. *Let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ be i.i.d. m-variate random vectors satisfying $E(\boldsymbol{z}_i) = \boldsymbol{0}, Var(\boldsymbol{z}_i) = \boldsymbol{I}_m$ and $Var(\frac{\boldsymbol{z}_i^\top \boldsymbol{z}_i}{m}) = O(m^{-1})$. Suppose matrix $\boldsymbol{A}$ is uniformly distributed on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m) = \{\boldsymbol{A} \in \mathbb{R}^{m\times k} : \boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{I}_k\}$ and is independent of $\boldsymbol{z}_i$. Let $\boldsymbol{Z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n)^\top$ and*

$$\boldsymbol{H} = (\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Z}\boldsymbol{A}\left(\boldsymbol{A}^\top \boldsymbol{Z}^\top (\boldsymbol{I} - \boldsymbol{P}_1)\boldsymbol{Z}\boldsymbol{A}\right)^{-1} \boldsymbol{A}^\top \boldsymbol{Z}^\top (\boldsymbol{I} - \boldsymbol{P}_1).$$

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

$$\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{H}_{ii}-\rho\right)^2 = o_p(1),$$

where $\boldsymbol{H}_{ii}$ denote the ith diagonal entries of $\boldsymbol{H}$.

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

$$\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{O}_n^\top. \tag{S1.6}$$

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

$$\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top = \frac{1}{m}\mathbf{U}\mathbf{D}^2\mathbf{U}^\top.$$

This indicates the diagonal entries of $\frac{1}{m}\mathbf{D}^2$ are the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$, then

$$\max_{i=1,\ldots,n}\left(\frac{d_i^2}{m}-1\right)^2 = \lambda_{\max}\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right\} \le tr\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right\}$$

From the properties of $\mathbf{z}_i$, we have

$$\begin{aligned} E\left\{tr\left[\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}\right)^2\right]\right\} &= \sum_{i=1}^{n} E\left\{\left(\frac{\mathbf{z}_i^\top \mathbf{z}_i}{m} - 1\right)^2\right\} + \sum_{i\neq j}^{n} E\left\{\left(\frac{\mathbf{z}_i^\top \mathbf{z}_j}{m}\right)^2\right\} \\ &= nVar\left(\frac{\mathbf{z}_1^\top \mathbf{z}_1}{m}\right) + \frac{n^2-n}{m} \\ &= O(n^2 m^{-1}). \end{aligned}$$

Therefore, from Markov's inequality, for any $t > 0$,

$$P\left\{\max_{i=1,\dots,n}\left(\frac{d_i}{\sqrt{m}} - 1\right)^2 > t\right\} \leq P\left\{\max_{i=1,\dots,n}\left(\frac{d_i^2}{m} - 1\right)^2 > t\right\} \leq O(n^2 m^{-1} t^{-1}), \tag{S1.7}$$

which shows the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$ 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 $(\mathbf{I} - \mathbf{P}_1)\,\mathbf{Z}\mathbf{A}$ are the same, the hat matrix for $\tilde{\mathbf{Z}}$ and $\mathbf{X}$ are denoted as

$$\mathbf{H} = \tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top, \qquad \mathbf{S} = \mathbf{X}\left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\mathbf{X}^\top,$$

where $\mathbf{H}$ is the target matrix of the lemma. Let $\mathbf{S}_{ii}$ denote the ith diagonal entry of the matrix $\mathbf{S}$. We will show $\mathbf{H}_{ii}$ and $\mathbf{S}_{ii}$ are close. Let $\mathbf{e}_i$ denote the vector with 1 in the ith coordinate and 0's elsewhere. Define $\hat{\boldsymbol{\gamma}}_i^{ls} = \left(\mathbf{X}^\top \mathbf{X}\right)^{-1}\mathbf{X}^\top \mathbf{e}_i$. Based on the least square, then $\hat{\boldsymbol{\gamma}}_i^{ls}$ satisfies

$$\hat{\boldsymbol{\gamma}}_i^{ls} = \underset{\boldsymbol{\gamma}\in\mathbb{R}^k}{\operatorname{argmin}}\left|\left|\,(\mathbf{I} - \mathbf{P}_1)\,\mathbf{e}_i - \mathbf{X}\boldsymbol{\gamma}\right|\right|_2^2. \tag{S1.8}$$

Similarly, define $\hat{\boldsymbol{\eta}}_i^{ls} = \left(\tilde{\mathbf{Z}}^\top \tilde{\mathbf{Z}}\right)^{-1} \tilde{\mathbf{Z}}^\top \mathbf{e}_i$. Then, it satisfies

$$\hat{\boldsymbol{\eta}}_i^{ls} = \underset{\boldsymbol{\eta} \in \mathbb{R}^k}{\operatorname{argmin}} \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \tilde{\mathbf{Z}} \boldsymbol{\eta} \big|\big|_2^2. \tag{S1.9}$$

Based on (S1.8) and (S1.9), we have

$$\begin{aligned} \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \tilde{\mathbf{Z}} \hat{\boldsymbol{\eta}}_i^{ls} \big|\big|_2^2 &\leq \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \tilde{\mathbf{Z}} \hat{\boldsymbol{\gamma}}_i^{ls} \big|\big|_2^2 \\ &= \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \mathbf{X} \hat{\boldsymbol{\gamma}}_i^{ls} + \big(\mathbf{X} - \tilde{\mathbf{Z}}\big) \hat{\boldsymbol{\gamma}}_i^{ls} \big|\big|_2^2 \\ &\leq \left(\big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \mathbf{X} \hat{\boldsymbol{\gamma}}_i^{ls} \big|\big|_2 + \big|\big| \big(\mathbf{X} - \tilde{\mathbf{Z}}\big) \hat{\boldsymbol{\gamma}}_i^{ls} \big|\big|_2 \right)^2, \end{aligned} \tag{S1.10}$$

and

$$\begin{aligned} \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \mathbf{X} \hat{\boldsymbol{\gamma}}_i^{ls} \big|\big|_2^2 &\leq \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \mathbf{X} \hat{\boldsymbol{\eta}}_i^{ls} \big|\big|_2^2 \\ &= \big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \tilde{\mathbf{Z}} \hat{\boldsymbol{\eta}}_i^{ls} + \big(\tilde{\mathbf{Z}} - \mathbf{X}\big) \hat{\boldsymbol{\eta}}_i^{ls} \big|\big|_2^2 \\ &\leq \left(\big|\big| \left(\mathbf{I} - \mathbf{P}_1\right) \mathbf{e}_i - \tilde{\mathbf{Z}} \hat{\boldsymbol{\eta}}_i^{ls} \big|\big|_2 + \big|\big| \big(\tilde{\mathbf{Z}} - \mathbf{X}\big) \hat{\boldsymbol{\eta}}_i^{ls} \big|\big|_2 \right)^2. \end{aligned} \tag{S1.11}$$

To study (S1.10) and (S1.11), we first investigate the values of $\big|\big|\big(\mathbf{X} - \tilde{\mathbf{Z}}\big)\hat{\boldsymbol{\gamma}}_i^{ls}\big|\big|_2$ and $\big|\big|\big(\tilde{\mathbf{Z}} - \mathbf{X}\big)\hat{\boldsymbol{\eta}}_i^{ls}\big|\big|_2$. From Theorem 2.2.1 in Chikuse (2003), matrix $\mathbf{A}$ can be expressed as $\mathbf{A} = \mathbf{G}\left(\mathbf{G}^\top \mathbf{G}\right)^{-1/2}$, where the elements of $m \times k$ matrix $\mathbf{G}$ are i.i.d. from $\mathcal{N}(0, 1)$. Let $\mathbf{E} = \mathbf{O}_n^\top \mathbf{G}$. Then $\mathbf{O}_n^\top \mathbf{A} = \mathbf{E}\left(\mathbf{G}^\top \mathbf{G}\right)^{-1/2}$. From Lemma S11, for any $h_1 > 0$ and $h_2 > 0$, the independence between $\mathbf{A}$ and $\mathbf{Z}$ leads to

$$\begin{aligned} &P\left[\lambda_{\max}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \geq (1+\sqrt{k/n}+h_1)^2\right] \leq \exp\left(-nh_1^2/2\right), \\ &P\left[\lambda_{\min}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \leq (1-\sqrt{k/n}-h_2)^2\right] \leq \exp\left(-nh_2^2/2\right). \end{aligned} \tag{S1.12}$$

For any matrix $\mathbf{M}$, SVD shows the nonzero eigenvalues of $\mathbf{M}^\top\mathbf{M}$ and $\mathbf{M}\mathbf{M}^\top$ are the same. Therefore, with $k < n$, it indicates $\lambda_{\min}\big(\mathbf{E}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\mathbf{E}\big) = \lambda_{\min}\big(\mathbf{E}^\top\mathbf{E}\big)$ and $\lambda_{\min}\big(\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big) = \lambda_{\min}\big(\mathbf{E}^\top\frac{\mathbf{D}^2}{m}\mathbf{E}\big)$. Based on the property $\lambda_{\max}(\mathbf{M}^\top\mathbf{M}) = \lambda_{\max}(\mathbf{M}\mathbf{M}^\top)$ and (S1.12) , we have

$$\begin{aligned} &\lambda_{\max}\left(\mathbf{X}\big(\mathbf{X}^\top\mathbf{X}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{X}^\top\mathbf{X}\big)^{-1}\mathbf{X}^\top\right) \\ &= \lambda_{\max}\left(\mathbf{E}\left(\mathbf{E}^\top\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\mathbf{E}\right)^{-1}\mathbf{E}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\mathbf{E}\big)} \\ &\leq \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}, \end{aligned} \tag{S1.13}$$

and

$$\begin{aligned} &\lambda_{\max}\left(\tilde{\mathbf{Z}}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\tilde{\mathbf{Z}}^\top\right) \\ &= \lambda_{\max}\left(\mathbf{E}\left(\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\right)^{-1}\mathbf{E}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top\big(\mathbf{I}-\mathbf{P}_1\big)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big)} \\ &\leq \frac{1}{\lambda_{\min}\big(\frac{\mathbf{D}^2}{m}\big)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{S1.14}$$

with probability at least $1-\exp\left(-nh_1^2/2\right)-\exp\left(-nh_2^2/2\right)$. Based on (S1.7), (S1.13) and (S1.14), upper bounds can be derived as follows.

$$\begin{aligned} ||(\mathbf{X}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2 &= ||(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i||_2^2 \\ &\leq \max_{i=1,\ldots,n}(1-\frac{d_i}{\sqrt{m}})^2||\mathbf{O}_n^\top\mathbf{A}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{e}_i||_2^2 \\ &\leq t\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{S1.15}$$

and

$$\begin{aligned} ||(\tilde{\mathbf{Z}}-\mathbf{X})\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 &= ||(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i||_2^2 \\ &\leq \max_{i=1,\ldots,n}(1-\frac{d_i}{\sqrt{m}})^2||\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i||_2^2 \\ &\leq \max_{i=1,\ldots,n}(1-\frac{d_i}{\sqrt{m}})^2\cdot\frac{1}{\min_{i=1,\ldots,n}(\frac{d_i^2}{m})}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \\ &\leq \frac{t}{(1-\sqrt{t})^2}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}, \end{aligned} \tag{S1.16}$$

with probability at least $1-O(n^2m^{-1}t^{-1})-\exp(-nh_1^2/2)-\exp(-nh_2^2/2)$. Combining (S1.10), (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

$$\begin{aligned} ||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 &\leq ||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2+3n^{-c/2}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}}, \\ ||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2 &\leq ||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2+\frac{3}{n^{c/2}-1}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}} \end{aligned}$$

with probability at least $1-O(n^{2+c}m^{-1})-2\exp\left(-n^{1/2}/2\right)$. Since $||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2=\mathbf{e}_i^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\mathbf{H}_{ii}$ and $||(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-\mathbf{X}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2=\mathbf{e}_i^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}_i-$

$\mathbf{S}_{ii}$, and the above derivation is valid for any $\mathbf{e}_i$, we obtain

$$|\mathbf{H}_{ii}-\mathbf{S}_{ii}| \leq \frac{3}{n^{c/2}-1}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}}, \qquad i=1,\ldots,n,$$

with probability at least $1-O(n^{2+c}m^{-1})-2\exp\left(-n^{1/2}/2\right)$. 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

$$P\left[\max_{i=1\ldots,n}|\mathbf{H}_{ii}-\mathbf{S}_{ii}|\geq Cn^{-c/2}\right]=o(1). \tag{S1.17}$$

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

$$\mathbf{S}=\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{U}\mathbf{O}_n^\top\mathbf{G}\left(\mathbf{G}^\top\mathbf{O}_n\mathbf{U}^\top\left(\mathbf{I}-\mathbf{P}_1\right)\mathbf{U}\mathbf{O}_n^\top\mathbf{G}\right)^{-1}\mathbf{G}^\top\mathbf{O}_n\mathbf{U}^\top\left(\mathbf{I}-\mathbf{P}_1\right),$$

where $\mathbf{U}\mathbf{O}_n^\top$ is independent of $\mathbf{G}$ and satisfies $\mathbf{U}\mathbf{O}_n^\top\mathbf{O}_n\mathbf{U}^\top=\mathbf{I}_n$. From the definition of $\mathbf{G}$, Lemma S6 and the dominated convergence theorem, we obtain

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

Then, $\frac{1}{n}\sum_{i=1}^n(\mathbf{S}_{ii}-\rho)^2=o_p(1)$ can be derived based on Markov's inequality. Combining this with (S1.17) and Slutsky's theorem, it shows

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\left(\mathbf{H}_{ii}-\rho\right)^2 &= \frac{1}{n}\sum_{i=1}^n\left(\mathbf{H}_{ii}-\mathbf{S}_{ii}+\mathbf{S}_{ii}-\rho\right)^2\\ &\leq \frac{2}{n}\sum_{i=1}^n\left(\mathbf{H}_{ii}-\mathbf{S}_{ii}\right)^2+\frac{2}{n}\sum_{i=1}^n\left(\mathbf{S}_{ii}-\rho\right)^2\\ &\leq \max_{i=1\ldots,n}2\left(\mathbf{H}_{ii}-\mathbf{S}_{ii}\right)^2+\frac{2}{n}\sum_{i=1}^n\left(\mathbf{S}_{ii}-\rho\right)^2\\ &= o_p(1),\end{aligned}$$

which completes the proof. □

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

Lemma S8. *Suppose m-variate random vector* $\boldsymbol{z} = (z_1, \ldots, z_m)^\top$ *has a Lebesgue density* $f_{\boldsymbol{z}}$ *and satisfies* $E(\boldsymbol{z}) = \boldsymbol{0}$ *and* $E(\boldsymbol{z}\boldsymbol{z}^\top) = \boldsymbol{I}_m$. *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* $\boldsymbol{z}$ *are bounded by a constant* $D \geq 1$. *Suppose matrix* $\boldsymbol{A}$ *is uniformly distributed on the Stiefel manifold* $\mathcal{V}_k(\mathbb{R}^m) = \{\boldsymbol{A} \in \mathbb{R}^{m\times k} : \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}_k\}$. *Let* $\nu_{m,k}$ *denote the uniform distribution on* $\mathcal{V}_k(\mathbb{R}^m)$. *Let* $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_n$ *be the i.i.d. copies of* $\boldsymbol{z}$ *and* $\boldsymbol{A}$ *be independent of* $\boldsymbol{z}_i$. *For any nonzero vector* $\boldsymbol{b} \in \mathbb{R}^m$, *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* $F_n \subseteq \mathcal{V}_k(\mathbb{R}^m)$ *such that*

$$\sup_{\boldsymbol{A}\in F_n} P\left(\sum_{i=1}^n \left(E(\boldsymbol{b}^\top\boldsymbol{z}_i|\boldsymbol{A}^\top\boldsymbol{z}_i) - \boldsymbol{b}^\top\boldsymbol{A}\boldsymbol{A}^\top\boldsymbol{z}_i\right)^2 > ||\boldsymbol{b}||_2^2\right) = o(1),$$

$$\sup_{\boldsymbol{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left|Var\big(\boldsymbol{b}^\top\boldsymbol{z}_i|\boldsymbol{A}^\top\boldsymbol{z}_i\big) - \boldsymbol{b}^\top\left(\boldsymbol{I}_m - \boldsymbol{A}\boldsymbol{A}^\top\right)\boldsymbol{b}\right| > 5||\boldsymbol{b}||_2^2\right) = o(1),$$

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

$$\sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}|\mathbf{A}^\top\mathbf{z}\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}\right|\right|_2 > t\right) \le \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1-\tau}\frac{2k}{\log m},$$

$$\sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}\mathbf{z}^\top|\mathbf{A}^\top\mathbf{z}\big) - \big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top + \mathbf{A}\mathbf{A}^\top\mathbf{z}\mathbf{z}^\top\mathbf{A}\mathbf{A}^\top\big)\right|\right|_{sp} > t\right) \le \frac{m^{-\tau/10}}{t} + \frac{\gamma_2}{1-\tau}\frac{2k}{\log m},$$

for each $t > 0$, and such that $\nu_{m,k}(F_n^{\mathsf{c}}) \le \kappa_2 m^{-(\tau/10)\cdot(1-\frac{\gamma_2}{\tau}\frac{10k}{\log m})}$, where κ_2 and γ_2 are constants. Therefore, when $t = n^{-1/2}$, we have

$$\begin{aligned}
&\sup_{\mathbf{A}\in F_n} P\left(\sum_{i=1}^n \left|\left|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right|\right|_2^2 > 1\right)\\
&\le \sum_{i=1}^n \sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right|\right|_2 > t\right)\\
&\le n^{3/2}m^{-\tau/10} + \frac{\gamma_2}{1-\tau}\frac{2nk}{\log m},
\end{aligned} \tag{S1.18}$$

$$\begin{aligned}
&\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\big|\mathbf{A}^\top\mathbf{z}_i\right] - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top - \big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\right|\right|_{sp} > 1\right)\\
&\le \sum_{i=1}^n \sup_{\mathbf{A}\in F_n} P\left(\left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\big|\mathbf{A}^\top\mathbf{z}_i\right] - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top - \big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\right|\right|_{sp} > t\right)\\
&\le n^{3/2}m^{-\tau/10} + \frac{\gamma_2}{1-\tau}\frac{2nk}{\log m},
\end{aligned} \tag{S1.19}$$

and $\nu_{m,k}(F_n^{\mathsf{c}}) \le \kappa_2 m^{-(\tau/10)\cdot(1-\frac{\gamma_2}{\tau}\frac{10k}{\log m})}$.

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

$$Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big) = \mathbf{b}^\top E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right]\mathbf{b} - E\left[\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\right]^2,$$

then

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{b}^\top\big(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\big)\mathbf{b}\right| \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\left|\mathbf{b}^\top\left\{E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right] - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top - \left(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\right)\right\}\mathbf{b} - 2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i r_i - r_i^2\right| \\
&\le ||\mathbf{b}||_2^2\frac{1}{\sqrt{n}}\sum_{i=1}^n\left|\left|E\left[\mathbf{z}_i\mathbf{z}_i^\top\middle|\mathbf{A}^\top\mathbf{z}_i\right] - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top - \left(\mathbf{I}_m - \mathbf{A}\mathbf{A}^\top\right)\right|\right|_{sp} \\
&+ 2\sqrt{\sum_{i=1}^n\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n}}\sqrt{\sum_{i=1}^n r_i^2} + \frac{1}{\sqrt{n}}\sum_{i=1}^n r_i^2.
\end{aligned} \tag{S1.20}$$

From the calculation,

$$Var\left\{\big(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\big)^2\right\} \le (C^{1/5}+1)\big(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\big)^2,$$

Markov's inequality leads to

$$P\left(\sum_{i=1}^n\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n} > 2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\right) \le \frac{C^{1/5}+1}{n}. \tag{S1.21}$$

According to Cauchy–Schwarz inequality,

$$r_i^2 = \left\{E(\mathbf{b}^\top\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i) - \mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\right\}^2 \le \big|\big|\mathbf{b}\big|\big|_2^2\cdot\big|\big|E\big(\mathbf{z}_i|\mathbf{A}^\top\mathbf{z}_i\big) - \mathbf{A}\mathbf{A}^\top\mathbf{z}_i\big|\big|_2^2.$$

Therefore, combining (S1.18), (S1.19), (S1.20) and (S1.21), we can derive

$$\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big)-\mathbf{b}^\top\big(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\big)\mathbf{b}\right|>5||\mathbf{b}||_2^2\right)$$
$$\le \sup_{\mathbf{A}\in F_n} P\big(\sum_{i=1}^{n} r_i^2>||\mathbf{b}||_2^2\big)+\sup_{\mathbf{A}\in F_n} P\big(\sum_{i=1}^{n}\frac{(\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{z}_i)^2}{n}>2\mathbf{b}^\top\mathbf{A}\mathbf{A}^\top\mathbf{b}\big)$$
$$+\sup_{\mathbf{A}\in F_n} P\big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big|\big|E\big[\mathbf{z}_i\mathbf{z}_i^\top|\mathbf{A}^\top\mathbf{z}_i\big]-\mathbf{A}\mathbf{A}^\top\mathbf{z}_i\mathbf{z}_i^\top\mathbf{A}\mathbf{A}^\top-\big(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\big)\big|\big|_{sp}>1\big)$$
$$\le 2n^{3/2}m^{-\tau/10}+\frac{\gamma_2}{1-\tau}\frac{4nk}{\log m}+\frac{2C}{n}.$$

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

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

and

$$\sup_{\mathbf{A}\in F_n} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|Var\big(q_i|\mathbf{A}^\top\mathbf{z}_i\big)-\mathbf{b}^\top\big(\mathbf{I}_m-\mathbf{A}\mathbf{A}^\top\big)\mathbf{b}\right|>5||\mathbf{b}||_2^2\right)=o(1),$$

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

S1.3 Proof of Lemma 2

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

Lemma S9. *If $\boldsymbol{A}$ and $\boldsymbol{B}$ are square matrices of the same size with $\boldsymbol{A}\succeq\boldsymbol{0}$ and $\boldsymbol{B}=\boldsymbol{B}^\top$, then*

$$\lambda_{\min}(\boldsymbol{B})tr(\boldsymbol{A})\le tr(\boldsymbol{A}\boldsymbol{B})\le\lambda_{\max}(\boldsymbol{B})tr(\boldsymbol{A}).$$

Some results for Gaussian concentration inequalities will be introduced. The following concentration bounds for Gaussian quadratic forms are given in Bechar (2009).

Lemma S10. *Let* $\boldsymbol{A} \in \mathbb{R}^{p\times p}$ *with* $\boldsymbol{A} \succeq \boldsymbol{0}$ *and* $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_p)$. *For any* $t > 0$, *we have*

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

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

Lemma S11. *For* $k \leq p$, *let* $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ *be a random matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries. Then, for all* $t \geq 0$, *we have*

$$P\left[\lambda_{\max}(\frac{1}{p}\boldsymbol{P}_k^\top \boldsymbol{P}_k) \geq (1 + \sqrt{k/p} + t)^2\right] \leq \exp(-pt^2/2), \quad and$$
$$P\left[\lambda_{\min}(\frac{1}{p}\boldsymbol{P}_k^\top \boldsymbol{P}_k) \leq (1 - \sqrt{k/p} - t)^2\right] \leq \exp(-pt^2/2).$$

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

Lemma S12. *For* $k \leq p$, *let* $\boldsymbol{P}_k \in \mathbb{R}^{p\times k}$ *be a random matrix with i.i.d.* $\mathcal{N}(0,1)$ *entries. Suppose matrix* $\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ *satisfies* $\boldsymbol{\Sigma} \succeq \boldsymbol{0}$. *Then, as*

$(k,p) \to \infty$, for any constant $C > 1$, we have

$$P\left[tr(\boldsymbol{P}_k^\top \boldsymbol{\Sigma} \boldsymbol{P}_k) \le Cktr(\boldsymbol{\Sigma})\right] \to 1.$$

Proof. Let $\mathbf{U}^\top \mathbf{D} \mathbf{U}$ be a spectral decomposition of $\boldsymbol{\Sigma}$. Then $\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k$ can be written as $(\mathbf{U}\mathbf{P}_k)^\top \mathbf{D}(\mathbf{U}\mathbf{P}_k)$. As $\mathbf{U}\mathbf{P}_k$ has the same distribution as $\mathbf{P}_k$, $\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k$ is distributed as $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$. In the following, we work under $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$.

Let $\boldsymbol{\xi}_i$ be the ith column of $\mathbf{P}_k$ and $\mathbf{Z}^\top = (\boldsymbol{\xi}_1^\top, ..., \boldsymbol{\xi}_k^\top)$. Then $\mathbf{Z} \in \mathbb{R}^{pk \times 1}$ and is distributed as $\mathcal{N}(\mathbf{0}, \mathbf{I}_{pk})$. Likewise, let $\tilde{\boldsymbol{D}} \in \mathbb{R}^{pk \times pk}$ be a diagonal matrix obtained by arranging k copies of $\mathbf{D}$ along the diagonal, i.e.

$$\tilde{\boldsymbol{D}} := \begin{pmatrix} \mathbf{D} & & \\ & \ddots & \\ & & \mathbf{D} \end{pmatrix}.$$

Consider the diagonal entries of $\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k$

$$tr(\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k) = \sum_{i=1}^k \boldsymbol{\xi}_i^\top \mathbf{D} \boldsymbol{\xi}_i = \mathbf{Z}^\top \tilde{\boldsymbol{D}} \mathbf{Z}.$$

Applying Lemma S10 to the quadratic form $\mathbf{Z}^\top \tilde{\boldsymbol{D}} \mathbf{Z}$, and noting that $\frac{||\mathbf{D}||_F}{tr(\mathbf{D})}$ and $\frac{||\mathbf{D}||_{sp}}{tr(\mathbf{D})}$ are at most 1, we get

$$\begin{aligned} tr(\mathbf{P}_k^\top \mathbf{D} \mathbf{P}_k) &\le tr(\tilde{\boldsymbol{D}}) + 2||\tilde{\boldsymbol{D}}||_F \sqrt{t_1} + 2||\tilde{\boldsymbol{D}}||_{sp} t_1 \\ &= ktr(\mathbf{D}) + 2||\mathbf{D}||_F \sqrt{t_1 k} + 2||\mathbf{D}||_{sp} t_1 \\ &\le ktr(\boldsymbol{\Sigma})(1 + \frac{2\sqrt{t_1}}{\sqrt{k}} + \frac{2t_1}{k}) \end{aligned}$$

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. □

Proof of Lemma 2. Let $\mathbf{U}^\top \mathbf{D} \mathbf{U}$ be a spectral decomposition of $\boldsymbol{\Sigma}$, where $\mathbf{D} = \text{diag}(d_1, ..., d_p)$ and $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. From this decomposition,

$$\sqrt{n}||\boldsymbol{\Gamma}^\top \boldsymbol{\beta} - \boldsymbol{\Gamma}^\top \mathbf{P}_k \boldsymbol{\eta}||_2^2 = \sqrt{n}||\sqrt{\mathbf{D}}\mathbf{U}\boldsymbol{\beta} - \sqrt{\mathbf{D}}\mathbf{U}\mathbf{P}_k\boldsymbol{\eta}||_2^2. \tag{S1.22}$$

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

For the s given in Assumption 6, we let $\boldsymbol{\delta} = (\boldsymbol{\delta}_s^\top, \boldsymbol{\delta}_{p-s}^\top)^\top$, where $\boldsymbol{\delta}_s \in \mathbb{R}^s$ and $\boldsymbol{\delta}_{p-s} \in \mathbb{R}^{p-s}$. Correspondingly, $\mathbf{D}$ is divided into $\mathbf{D}_s$ and $\mathbf{D}_{p-s}$, where $\mathbf{D}_s = \text{diag}(d_1, ..., d_s)$ and $\mathbf{D}_{p-s} = \text{diag}(d_{s+1}, ..., d_p)$. Let $\mathbf{P}_k = (\mathbf{P}_{s,k}^\top, \mathbf{P}_{p-s,k}^\top)^\top$ with $\mathbf{P}_{s,k} \in \mathbb{R}^{s\times k}$ and $\mathbf{P}_{p-s,k} \in \mathbb{R}^{(p-s)\times k}$. We define $\boldsymbol{\eta}_0 \in \mathbb{R}^k$

as

$$\boldsymbol{\eta}_0 = \mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}.$$

Plugging $\boldsymbol{\eta}_0$ into (S1.22), we have

$$\begin{aligned}
&\min_{\boldsymbol{\eta}\in\mathbb{R}^k} \frac{\sqrt{n}||\boldsymbol{\Gamma}^\top\boldsymbol{\beta} - \boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\eta}||_2^2}{||\boldsymbol{\beta}||_2^2} \\
&= \min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}||\sqrt{\mathbf{D}}\frac{\boldsymbol{\delta}}{\sqrt{p}} - \sqrt{\mathbf{D}}\mathbf{P}_k\boldsymbol{\eta}||_2^2 \\
&= \min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}\left(||\sqrt{\mathbf{D}_s}(\frac{\boldsymbol{\delta}_s}{\sqrt{p}} - \mathbf{P}_{s,k}\boldsymbol{\eta})||_2^2 + ||\sqrt{\mathbf{D}_{p-s}}(\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k}\boldsymbol{\eta})||_2^2\right) \\
&\le \sqrt{n}||\sqrt{\mathbf{D}_s}(\frac{\boldsymbol{\delta}_s}{\sqrt{p}} - \mathbf{P}_{s,k}\boldsymbol{\eta}_0)||_2^2 + \sqrt{n}||\sqrt{\mathbf{D}_{p-s}}(\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k}\boldsymbol{\eta}_0)||_2^2 \\
&= \sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \sqrt{\mathbf{D}_{p-s}}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}||_2^2 \\
&\le 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}}||_2^2 + 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}||_2^2 \\
&= T_1 + T_2. \qquad \text{(S1.23)}
\end{aligned}$$

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

In the first step, the concentration inequality for quadratic forms in Lemma S10 gives an upper bound on T_1

$$P\left[T_1 \le \frac{2\sqrt{n}}{p}\left(tr(\mathbf{D}_{p-s}) + 2\sqrt{h_1}||\mathbf{D}_{p-s}||_F + 2h_1||\mathbf{D}_{p-s}||_{sp}\right)\right] \ge 1 - \exp(-h_1),$$

where h_1 is a positive real number that may vary with n. From Assumption

6 and the properties of $||\cdot||_F$ and $||\cdot||_{sp}$, we select $h_1 = n^\gamma$ and get

$$\begin{aligned}
||\boldsymbol{\beta}||_2^2 T_1 &\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p}\left(tr(\mathbf{D}_{p-s}) + 2\sqrt{h_1}||\mathbf{D}_{p-s}||_F + 2h_1||\mathbf{D}_{p-s}||_{sp}\right) \\
&\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p} tr(\mathbf{D}_{p-s})\left(1 + 2\sqrt{h_1} + 2h_1\right) \\
&\leq \frac{10n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2 tr(\mathbf{D}_{p-s})}{p} = o(1)
\end{aligned} \tag{S1.24}$$

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

In the next step, Lemmas S12 and S11 give upper bounds by

$$\begin{aligned}
&k\lambda_{\max}\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right) = \frac{1}{\lambda_{\min}(\frac{\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top}{k})} \leq \frac{1}{(1 - \sqrt{s/k} - k^{-1/4})^2}, \\
&\frac{tr(\mathbf{P}_{p-s,k}^\top \mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k} \leq 2tr(\mathbf{D}_{p-s})
\end{aligned}$$

with probability converging to 1. These inequalities together with Lemma S9 lead to

$$\begin{aligned}
&tr\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\mathbf{P}_{s,k}\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right) \\
&\leq k\lambda_{\max}\left(\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-2}\mathbf{P}_{s,k}\right)\frac{tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k} \\
&= k\lambda_{\max}\left((\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\right)\frac{tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})}{k} \\
&\leq \frac{2tr(\mathbf{D}_{p-s})}{(1 - \sqrt{s/k} - k^{-1/4})^2}
\end{aligned} \tag{S1.25}$$

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

$$||\boldsymbol{\beta}||_2^2 T_2 \leq \frac{20n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2 tr(\mathbf{D}_{p-s})}{p(1-\sqrt{s/k}-k^{-1/4})^2} = o(1) \tag{S1.26}$$

with probability tending to 1.

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

$$\min_{\boldsymbol{\eta}\in\mathbb{R}^k} \sqrt{n}||\mathbf{\Gamma}^\top\boldsymbol{\beta} - \mathbf{\Gamma}^\top\mathbf{P}_k\boldsymbol{\eta}||_2^2 = o(1)$$

with probability tending to 1 and complete the proof. □

S2 Proof of theorems

S2.1 Proof of Theorem 1

Under $\mathbf{H}_0$, we have

$$T_n - 1 = \frac{\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}}{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}/(n-1-k)},$$

where $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-k-1}$. The property that $\mathbf{H}_k$ is idempotent with rank k leads to $tr(\mathbf{M}) = 0$ and $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H}_k}{k^2} + \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{(n-k-1)^2}$. Therefore,

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} = \frac{\lambda_{\max}(\mathbf{M}^\top\mathbf{M})}{tr(\mathbf{M}^\top\mathbf{M})} \leq \frac{\lambda_{\max}(\frac{\mathbf{H}_k}{k^2}) + \lambda_{\max}(\frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{(n-k-1)^2})}{\frac{1}{k}+\frac{1}{n-k-1}} = O(n^{-1}).$$

And we have

$$E(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = \sigma^2 tr(\mathbf{M}) = 0,$$

$$Var(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = (\mu_4 - 3\sigma^4)\sum_{i=1}^n m_{ii}^2 + 2\sigma^4(\frac{1}{k} + \frac{1}{n-k-1}),$$

where the error term $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^\top$ has $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$, and $E(\epsilon_i^4) = \mu_4$. When $\mathbf{M}$ is given, these together with Lemma S2 imply

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var\left(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M}\right)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

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

$$P\big(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha\big) = E\big(P(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha | \mathbf{M})\big).$$

And the aforementioned result shows

$$P(\frac{\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon}}{\sqrt{Var(\boldsymbol{\epsilon}^\top \mathbf{M} \boldsymbol{\epsilon} | \mathbf{M})}} \leq \alpha | \mathbf{M}) \to \Phi(\alpha).$$

Based on the dominated convergence theorem, we get

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

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

$$\begin{aligned} nG_n = n \sum_{i=1}^n m_{ii}^2 &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(\mathbf{H}_k)_{ii} - \frac{k}{n-1}(1 - \frac{1}{n})}{\frac{k}{n-1}(1 - \frac{k+1}{n})} \right\}^2 \\ &\leq \frac{2}{n} \sum_{i=1}^n \frac{\left\{(\mathbf{H}_k)_{ii} - \rho\right\}^2 + \left\{\rho - \frac{k}{n-1}(1 - \frac{1}{n})\right\}^2}{\left\{\frac{k}{n-1}(1 - \frac{k+1}{n})\right\}^2}. \end{aligned} \tag{S2.2}$$

Let $\boldsymbol{\Sigma}_1 = \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k$. From Lemma S4, we find the smallest eigenvalue of $\frac{1}{p}\mathbf{P}_k^\top \mathbf{P}_k$ is bounded away from 0 a.s., showing $\mathbf{P}_k$ is of full column rank with probability 1. Therefore, $\boldsymbol{\Sigma}_1$ is of full rank with probability 1. Define

$\widetilde{\mathbf{U}}_k = \mathbf{X}\mathbf{P}_k\mathbf{\Sigma}_1^{-1/2}$. Since $\mathbf{H}_k$ is invariant to the full rank linear transform of $\mathbf{U}_k$, the hat matrix can be expressed as

$$\mathbf{H}_k = \mathbf{U}_k(\mathbf{U}_k^\top \mathbf{U}_k)^{-1}\mathbf{U}_k^\top = (\mathbf{I}-\mathbf{P}_1)\widetilde{\mathbf{U}}_k(\widetilde{\mathbf{U}}_k^\top(\mathbf{I}-\mathbf{P}_1)\widetilde{\mathbf{U}}_k)^{-1}\widetilde{\mathbf{U}}_k^\top(\mathbf{I}-\mathbf{P}_1).$$

From Assumption 1, $\widetilde{\mathbf{U}}_k$ can be denoted by $\mathbf{Z}\mathbf{A}$, where $\mathbf{A} = \mathbf{\Gamma}^\top\mathbf{P}_k\mathbf{\Sigma}_1^{-1/2}$ is an $m \times k$ matrix. From Section 2.4.2 in Chikuse (2003), matrix $\mathbf{A}$ is on the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m)$ with probability 1, which demonstrates $\mathbf{U}_k^\top\mathbf{U}_k$ is of full rank with probability 1. From Lemma S7 and (S2.2), we obtain $nG_n = op(1)$.

Assumption 3 implies $\frac{n}{k} + \frac{n}{n-k-1} \to \frac{1}{\rho(1-\rho)}$, as $n \to \infty$. Therefore, (S2.1) leads to

$$\frac{\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}}{\sigma^2\sqrt{2/n\rho(1-\rho)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

In addition, from $E(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1}) = \sigma^2$, $Var(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1}) \leq \frac{\mu_4-\sigma^4}{n-k-1} \to 0$ and Markov's inequality, we have

$$\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\boldsymbol{\epsilon}}{n-k-1} = \sigma^2 + o_p(1).$$

Hence, under $\mathbf{H}_0$,

$$\frac{T_n - 1}{\sqrt{2/n\rho(1-\rho)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

which completes the proof.

S2.2 Proof of Theorem 2

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

$$r_i = E(\mathbf{x}_i^\top \boldsymbol{\beta} | \mathbf{P}_k^\top \mathbf{x}_i) - \mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi}, \qquad q_i = \mathbf{x}_i^\top \boldsymbol{\beta} - E(\mathbf{x}_i^\top \boldsymbol{\beta} | \mathbf{P}_k^\top \mathbf{x}_i).$$

Then, we have $\mathbf{x}_i^\top \boldsymbol{\beta} = \mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi} + r_i + q_i$, where q_i satisfies $E(q_i | \mathbf{P}_k^\top \mathbf{x}_i) = 0$. Let $\omega^2 = \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} - \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi}$ and $\tau_i = Var(q_i | \mathbf{P}_k^\top \mathbf{x}_i) - \omega^2$. According to Lemma S8 and the condition $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} = o(1)$, it shows

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

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 $\mathbf{A} = \boldsymbol{\Gamma}^\top \mathbf{P}_k (\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k)^{-1/2}$. 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

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X} \mathbf{P}_k \boldsymbol{\xi} + \mathbf{r} + \mathbf{e}, \tag{S2.4}$$

where $\mathbf{r} = (r_1, ..., r_n)^\top$, and $\mathbf{e} = (e_1, ..., e_n)^\top$ with each elements of $\mathbf{e}$ satisfying $E(e_i) = 0$, $E(e_i | \mathbf{P}_k^\top \mathbf{x}_i) = 0$, $Var(e_i | \mathbf{P}_k^\top \mathbf{x}_i) = \sigma^2 + \omega^2 + \tau_i$, and $E(e_i^4 | \mathbf{P}_k^\top \mathbf{x}_i) = \mu_4 + 6\sigma^2 Var(q_i | \mathbf{P}_k^\top \mathbf{x}_i) + E(q_i^4 | \mathbf{P}_k^\top \mathbf{x}_i)$. For matrix $\mathbf{M} =$

$(m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-k-1}$, calculation shows

$$E(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = \sum_{i=1}^n m_{ii}\tau_i,$$

$$Var(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = \sum_{i=1}^n m_{ii}^2 \left\{E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2\right\}$$
$$+ 2\sum_{i,j} m_{ij}^2 E(e_i^2|\mathbf{X}\mathbf{P}_k)E(e_j^2|\mathbf{X}\mathbf{P}_k)$$
$$= 2(\sigma^2+\omega^2)^2 tr(\mathbf{M}^\top\mathbf{M}) + g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k),$$

where $g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k) = \sum_{i=1}^n m_{ii}^2 \left\{E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2\right\} + 2\sum_{i,j} m_{ij}^2\{(\sigma^2 + \omega^2)(\tau_i+\tau_j) + \tau_i\tau_j\}$. For a constant $a \le 2/\rho(1-\rho)$ and large n, $\mathbf{M}$ satisfies $||\mathbf{M}||_{sp} \le a/n$ and $|m_{ii}| = |\mathbf{e}_i^\top \mathbf{M}\mathbf{e}_i| \le ||\mathbf{M}||_{sp}$. Then, (S2.3) leads to

$$\sqrt{n}E(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = o_p(1). \tag{S2.5}$$

To investigate the conditional variance, based on (S2.2) and Lemma S7, we can derive

$$\sum_{i=1}^n m_{ii}^2 \left\{E(e_i^4|\mathbf{X}\mathbf{P}_k) - 3E(e_i^2|\mathbf{X}\mathbf{P}_k)^2\right\} \le \sum_{i=1}^n m_{ii}^2 \left\{\mu_4 - 3\sigma^4 + E(q_i^4|\mathbf{X}\mathbf{P}_k)\right\} = o_p(n^{-1}).$$

In addition, $\sum_{j=1}^n m_{ij}^2 = \mathbf{e}_i^\top \mathbf{M}\mathbf{M}^\top \mathbf{e}_i \le ||\mathbf{M}||_{sp}^2 \le a^2/n^2$ and (S2.3) lead to

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

Therefore, $g(\mathbf{M}, \mathbf{X}, \boldsymbol{\epsilon}, \mathbf{P}_k) = o_p(n^{-1})$, from which we obtain

$$Var(\mathbf{e}^\top \mathbf{M}\mathbf{e}|\mathbf{X}\mathbf{P}_k) = 2(\sigma^2+\omega^2)^2 tr(\mathbf{M}^\top\mathbf{M}) + o_p(n^{-1}). \tag{S2.6}$$

According to $tr(\mathbf{M}^\top\mathbf{M}) = \frac{1}{k} + \frac{1}{n-1-k}$, (S2.5), (S2.6) and the condition $k/n \to \rho$, Lemma S2 shows

$$\frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\mathbf{e}^\top\mathbf{M}\mathbf{e} - o_p(1)}{(\sigma^2+\omega^2)\sqrt{1+o_p(1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{S2.7}$$

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

$$\frac{1}{\sqrt{n}}\mathbf{r}^\top\mathbf{E}\mathbf{r} \leq \frac{1}{\sqrt{n}}\mathbf{r}^\top\mathbf{r} = o_p(n^{-1/2}), \tag{S2.8}$$

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

$$E\big\{(\mathbf{x}_1^\top\boldsymbol{\beta})^4\big\} = \sum_{i=1}^m(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_i^4E(z_{1i}^4) + 3\sum_{i\neq j}^m(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_i^2(\boldsymbol{\Gamma}^\top\boldsymbol{\beta})_j^2E(z_{1i}^2z_{1j}^2),$$

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

$$|E(e_i^4)-\mu_4| \leq c_1\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} = o(1), \quad E\{\tau_i^2\} \leq E(q_i^4)+\omega^4 \leq c_1(\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta})^2 = o(1), \tag{S2.9}$$

for a constant c_1. In addition, the calculation shows

$$\left|E\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}\right)-(\sigma^2+\omega^2)\right| = \left|E\left\{\sum_{i=1}^n\left(\frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-1-k}\right)_{ii}\tau_i\right\}\right|$$
$$\leq \sum_{i=1}^n \frac{1}{n-1-k}\sqrt{E\{\tau_i^2\}} = o(1),$$

$$E\left\{Var\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}|\mathbf{X}\mathbf{P}_k\right)\right\}$$
$$\leq \frac{\sum_{i=1}^n E\left(e_i^4\right)}{(n-1-k)^2}+\frac{2(\sigma^2+\omega^2)^2}{n-1-k}+\frac{4n(\sigma^2+\omega^2)\sum_{i=1}^n\sqrt{E(\tau_i^2)}+2\sum_{i,j}\sqrt{E(\tau_i^2)E(\tau_j^2)}}{(n-1-k)^2} = o(1),$$

$$Var\left\{E\left(\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}|\mathbf{X}\mathbf{P}_k\right)\right\} = Var\left(\sum_{i=1}^n\frac{(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)_{ii}}{n-1-k}\tau_i\right)$$
$$\leq E\left\{\left(\sum_{i=1}^n\frac{(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)_{ii}^2}{(n-1-k)^2}\right)\left(\sum_{i=1}^n\tau_i^2\right)\right\}$$
$$\leq \frac{n}{(n-1-k)^2}\sum_{i=1}^n E(\tau_i^2) = o(1).$$

Consequently, Markov's inequality leads to

$$\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k} = \sigma^2+\omega^2+o_p(1).$$

This combines with (S2.8) shows

$$\frac{(\mathbf{e}+\mathbf{r})^\top\big(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k\big)(\mathbf{e}+\mathbf{r})}{n-1-k} = \sigma^2+\omega^2+o_p(1). \tag{S2.10}$$

Next, we study $\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}$. From Assumption 1, we have

$$E\left\{\frac{1}{\sqrt{n}}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\mathbf{X}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}\right\} = \frac{n-1}{\sqrt{n}}\boldsymbol{\xi}^\top\mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} \tag{S2.11}$$

and the fourth moment of $\mathbf{x}_1^\top \mathbf{P}_k \boldsymbol{\xi}$ satisfies

$$E\big\{(\mathbf{x}_1^\top \mathbf{P}_k \boldsymbol{\xi})^4\big\} = \sum_{i=1}^{m} (\boldsymbol{\Gamma}^\top \mathbf{P}_k \boldsymbol{\xi})_i^4 E(z_{1i}^4) + 3\sum_{i\neq j}^{m} (\boldsymbol{\Gamma}^\top \mathbf{P}_k \boldsymbol{\xi})_i^2 (\boldsymbol{\Gamma}^\top \mathbf{P}_k \boldsymbol{\xi})_j^2 E(z_{1i}^2 z_{1j}^2).$$

Based on $\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi} = \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \mathbf{P}_k (\mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k)^{-1} \mathbf{P}_k^\top \boldsymbol{\Sigma} \boldsymbol{\beta} \leq \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} = o(1)$, we have

$$Var\left(\frac{1}{\sqrt{n}} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{X} \mathbf{P}_k \boldsymbol{\xi}\right) \leq E\left\{(\mathbf{x}_1^\top \mathbf{P}_k \boldsymbol{\xi})^4\right\} + 2(\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi})^2 = o(1).$$

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

$$\frac{\sqrt{n}}{k} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{X} \mathbf{P}_k \boldsymbol{\xi} = \frac{\sqrt{n}}{\rho} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi} + o_p(1). \tag{S2.12}$$

To investigate $\frac{\sqrt{n}}{k} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{e}$, the condition $\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} = o(1)$, (S2.9) and (S2.11) lead to

$$\begin{aligned}
& E\left\{\left(\frac{1}{\sqrt{n}} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top \big(\mathbf{I} - \mathbf{P}_1\big) \mathbf{e}\right)^2\right\} \\
&= E\left[E\left\{\left(\frac{1}{\sqrt{n}} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top \big(\mathbf{I} - \mathbf{P}_1\big) \mathbf{e}\right)^2 \big| \mathbf{X} \mathbf{P}_k\right\}\right] \\
&= E\left\{\frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \omega^2 + \tau_i) \big(\mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi} - \frac{1}{n} \sum_{j=1}^{n} \mathbf{x}_j^\top \mathbf{P}_k \boldsymbol{\xi}\big)^2\right\} \\
&\leq (\sigma^2 + \omega^2) E\left\{\frac{1}{n} \boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1) \mathbf{X} \mathbf{P}_k \boldsymbol{\xi}\right\} \\
&\quad + \sqrt{E\left(\frac{1}{n} \sum_{i=1}^{n} \tau_i^2\right)} \sqrt{E\left\{\frac{1}{n} \sum_{i=1}^{n} \big(\mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi} - \sum_{j=1}^{n} \mathbf{x}_j^\top \mathbf{P}_k \boldsymbol{\xi}\big)^4\right\}} \\
&\leq (\sigma^2 + \omega^2) \boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma} \mathbf{P}_k \boldsymbol{\xi} + \sqrt{c_1} \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} \sqrt{E\left\{\frac{16}{n} \sum_{i=1}^{n} \big(\mathbf{x}_i^\top \mathbf{P}_k \boldsymbol{\xi}\big)^4\right\}} \\
&= o(1).
\end{aligned}$$

Therefore, Markov's inequality and $k/n \to \rho$ demonstrate

$$\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{e} = o_p(1).$$

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

$$\frac{\sqrt{n}}{k}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I} - \mathbf{P}_1)(\mathbf{e}+\mathbf{r}) = o_p(1). \tag{S2.13}$$

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

$$\begin{aligned}
\frac{T_n - 1}{\sqrt{2/[n\rho(1-\rho)]}} &= \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\left\{\frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I}-\mathbf{P}_1)\mathbf{X}\mathbf{P}_k\boldsymbol{\xi}}{k} + \frac{2\boldsymbol{\xi}^\top \mathbf{P}_k^\top \mathbf{X}^\top (\mathbf{I}-\mathbf{P}_1)(\mathbf{e}+\mathbf{r})}{k} + (\mathbf{e}+\mathbf{r})^\top \mathbf{M}(\mathbf{e}+\mathbf{r})\right\}}{\frac{(\mathbf{e}+\mathbf{r})^\top (\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)(\mathbf{e}+\mathbf{r})}{n-k-1}} \\
&= \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}(\frac{1}{\rho}\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi} + \mathbf{e}^\top \mathbf{M}\mathbf{e}) + o_p(1)}{\sigma^2 + \omega^2 + o_p(1)}.
\end{aligned}$$

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

$$\begin{aligned}
\Psi_n^{RP}(\boldsymbol{\beta};\mathbf{P}_k) &= P(\frac{T_n - 1}{\sqrt{2/[n\rho(1-\rho)]}} > z_\alpha) \\
&= \Phi(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}}\frac{\boldsymbol{\xi}^\top \mathbf{P}_k^\top \boldsymbol{\Sigma}\mathbf{P}_k\boldsymbol{\xi}}{\delta_k^2}) + o(1),
\end{aligned}$$

which completes the proof.

S2.3 Proof of Theorem 3

Recall the definitions of projection matrices.

$$\mathbf{P}_1 = \frac{1}{n}\mathbf{1}\mathbf{1}^\top,$$

$$\mathbf{P}_{\mathbf{X}_1} = (\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1(\mathbf{X}_1^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1)^{-1}\mathbf{X}_1^\top(\mathbf{I} - \mathbf{P}_1),$$

$$\mathbf{H}_{k_2} = (\mathbf{I} - \mathbf{P}_1)\mathbf{W}(\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{I} - \mathbf{P}_1),$$

where $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{P}_{k_2})$. Under $\mathbf{H}_{part,0}$, we have

$$T_{n,p_2} = \frac{\boldsymbol{\epsilon}^\top(\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1})\boldsymbol{\epsilon}/k_2}{\boldsymbol{\epsilon}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2})\boldsymbol{\epsilon}/(n-1-p_1-k_2)}.$$

Define $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2} - \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{n-1-p_1-k_2}$. From $Span\{(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_1\} \subseteq Span\{(\mathbf{I} - \mathbf{P}_1)\mathbf{W}\}$ and properties of projection matrices, we have

$$\mathbf{P}_{\mathbf{X}_1}\mathbf{H}_{k_2} = \mathbf{H}_{k_2}\mathbf{P}_{\mathbf{X}_1} = \mathbf{P}_{\mathbf{X}_1}.$$

Hence, $tr(\mathbf{M}) = 0$, $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2^2} + \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{(n-1-p_1-k_2)^2}$, and

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} = \frac{\lambda_{\max}(\mathbf{M}^\top\mathbf{M})}{tr(\mathbf{M}^\top\mathbf{M})} \leq \frac{\lambda_{\max}(\frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2^2}) + \lambda_{\max}(\frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{(n-1-p_1-k_2)^2})}{\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}} = O(n^{-1}).$$

For given $\mathbf{M}$, we have

$$E(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = \sigma^2 tr(\mathbf{M}) = 0,$$

$$Var(\boldsymbol{\epsilon}^\top\mathbf{M}\boldsymbol{\epsilon}|\mathbf{M}) = (\mu_4 - 3\sigma^4)\sum_{i=1}^n m_{ii}^2 + 2\sigma^4\left(\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}\right).$$

Then, Lemma S2 leads to

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

This together with the law of total expectation and the dominated convergence theorem shows

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

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

$$\frac{\boldsymbol{\epsilon}^\top \mathbf{M}\boldsymbol{\epsilon}}{\sigma^2\sqrt{[E\{(\frac{\epsilon_1}{\sigma})^4\} - 3]\sum_{i=1}^n m_{ii}^2 + 2(\frac{1}{k_2} + \frac{1}{n-1-p_1-k_2})}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

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

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

Let $G_n = \sum_{i=1}^n m_{ii}^2$. Next, we will verify $nG_n = o_p(1)$. From the definition, $m_{ii} = \frac{(\mathbf{H}_{k_2})_{ii} - (\mathbf{P}_{\mathbf{X}_1})_{ii}}{k_2} - \frac{1-\frac{1}{n}-(\mathbf{H}_{k_2})_{ii}}{n-1-p_1-k_2}$. Then

$$\begin{aligned} nG_n = n\sum_{i=1}^n m_{ii}^2 &= \frac{1}{n}\sum_{i=1}^n\left\{\frac{(1-\frac{1}{n}-\frac{p_1}{n})\big((\mathbf{H}_{k_2})_{ii} - \frac{p_1+k_2}{n}\big)}{\frac{k_2}{n}(1-\frac{1}{n}-\frac{p_1}{n}-\frac{k_2}{n})} - \frac{(\mathbf{P}_{\mathbf{X}_1})_{ii} - \frac{p_1}{n}}{\frac{k_2}{n}}\right\}^2 \\ &\le \frac{2h_1}{n}\sum_{i=1}^n\left\{(\mathbf{H}_{k_2})_{ii} - \frac{p_1+k_2}{n}\right\}^2 + \frac{2h_2}{n}\sum_{i=1}^n\left\{(\mathbf{P}_{\mathbf{X}_1})_{ii} - \frac{p_1}{n}\right\}^2, \end{aligned} \tag{S2.15}$$

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

$$h_1 \to \frac{(1-\rho_1)^2}{\rho_2^2(1-\rho_1-\rho_2)^2},\ h_2 \to \frac{1}{\rho_2^2}.$$

Consequently, we only need to consider the sum parts in (S2.15). From the definition,

$$\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{P}_{k_2}) = \mathbf{Z}\mathbf{\Gamma}^\top \begin{pmatrix} \mathbf{I}_{p_1} & 0 \\ 0 & \mathbf{P}_{k_2} \end{pmatrix} \triangleq \mathbf{Z}\mathbf{\Gamma}^\top\mathbf{V},$$

where $\mathbf{Z} = (\mathbf{z}_1, ..., \mathbf{z}_n)^\top$ and $\mathbf{V}$ is a full column rank matrix with probability 1. Define $\mathbf{\Sigma}_2 = \mathbf{V}^\top\mathbf{\Sigma}\mathbf{V}$. The matrix $\mathbf{\Sigma}_2$ is of full rank with probability 1, then $\mathbf{\Gamma}^\top\mathbf{V}\mathbf{\Sigma}_2^{-1/2}$ is well defined on the Stiefel manifold $\mathcal{V}_{p_1+k_2}(\mathbb{R}^m)$. Let $\mathbf{W}_1 = \mathbf{W}\mathbf{\Sigma}_2^{-1/2} = \mathbf{Z}\mathbf{\Gamma}^\top\mathbf{V}\mathbf{\Sigma}_2^{-1/2}$. The hat matrix $\mathbf{H}_{k_2}$ can be denoted as

$$\mathbf{H}_{k_2} = (\mathbf{I} - \mathbf{P}_1)\mathbf{W}_1(\mathbf{W}_1^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{W}_1)^{-1}\mathbf{W}_1^\top(\mathbf{I} - \mathbf{P}_1).$$

According to Lemma S7 and the condition $(p_1 + k_2)/n \to \rho_1 + \rho_2$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left\{(\mathbf{H}_{k_2})_{ii} - \frac{p_1 + k_2}{n}\right\}^2 = o_p(1).$$

Let $\mathbf{R}_1 = \mathbf{Z}\mathbf{\Gamma}_1\mathbf{\Sigma}_{11}^{-1/2}$. The hat matrix $\mathbf{P}_{\mathbf{X}_1}$ can be denoted as

$$\mathbf{P}_{\mathbf{X}_1} = (\mathbf{I} - \mathbf{P}_1)\mathbf{R}_1(\mathbf{R}_1^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{R}_1)^{-1}\mathbf{R}_1^\top(\mathbf{I} - \mathbf{P}_1).$$

Based on Lemma S7 and the condition $p_1/n \to \rho_1$, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left\{(\mathbf{P}_{\mathbf{X}_1})_{ii} - \frac{p_1}{n}\right\}^2 = o_p(1).$$

Therefore, $nG_n = o_p(1)$ is verified, and then (S2.14) is demonstrated.

To study the denominator of T_{n,p_2}, calculation shows

$$E\left\{\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\right\} = E\left[E\left\{\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\big|\mathbf{H}_{k_2}\right\}\right] = \sigma^2,$$
$$Var\left(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\right) = E\left\{Var\left(\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2}\big|\mathbf{H}_{k_2}\right)\right\} = o(1).$$

From Markov's inequality, we have

$$\frac{\boldsymbol{\epsilon}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})\boldsymbol{\epsilon}}{n-1-p_1-k_2} = \sigma^2 + o_p(1).$$

Combining this with (S2.14), we obtain

$$\frac{T_{n,p_2}-1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

which completes the proof.

S2.4 Proof of Theorem 4

Define $\mathbf{V} = \mathrm{diag}(\mathbf{I}_{p_1}, \mathbf{P}_{k_2})$. The matrix is a full column rank matrix with probability 1, and $\mathbf{W} = \mathbf{X}\mathbf{V}$, with the ith row $\mathbf{w}_i = \mathbf{V}^\top\mathbf{x}_i$. Let $\boldsymbol{\gamma} = (\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma}\boldsymbol{\Gamma}_2^\top\boldsymbol{\beta}_2$. For each i, define

$$r_i = E\left(\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2\big|\mathbf{V}^\top\mathbf{x}_i\right) - \mathbf{x}_i^\top\mathbf{V}\boldsymbol{\gamma}, \quad q_i = \mathbf{x}_{2i}^\top\boldsymbol{\beta}_2 - E\left(\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2\big|\mathbf{V}^\top\mathbf{x}_i\right).$$

Then, a decomposition of $\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2$ can be derived, given as $\mathbf{x}_{2i}^\top\boldsymbol{\beta}_2 = \mathbf{w}_i^\top\boldsymbol{\gamma} + r_i + q_i$. Let $\omega^2 = \boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2 - \boldsymbol{\gamma}^\top\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V}\boldsymbol{\gamma}$ and $\tau_i = Var(q_i|\mathbf{V}^\top\mathbf{x}_i) - \omega^2$. According to Lemma S8 and the condition $\boldsymbol{\beta}_2^\top\boldsymbol{\Sigma}_{22}\boldsymbol{\beta}_2 = o(1)$, we have

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

when the event $\mathbf{A} \in F_n$ is satisfied, where $\mathbf{A} = \mathbf{\Gamma}^\top \mathbf{V}(\mathbf{V}^\top \mathbf{\Sigma} \mathbf{V})^{-1/2}$ 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 $\mathbf{P}_{k_2}$.

Define a new error term $e_i = q_i + \epsilon_i$. Let σ^2 denote the variance of ϵ_i. The model can be expressed as

$$\mathbf{y} = \alpha \mathbf{1} + \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{W}\boldsymbol{\gamma} + \mathbf{r} + \mathbf{e}, \tag{S2.17}$$

where $\mathbf{r} = (r_1, \ldots, r_n)^\top$, and $\mathbf{e} = (e_1, ..., e_n)^\top$ with each elements of $\mathbf{e}$ satisfying $E(e_i) = 0$, $E(e_i|\mathbf{V}^\top \mathbf{x}_i) = 0$, $Var(e_i|\mathbf{V}^\top \mathbf{x}_i) = \sigma^2 + \omega^2 + \tau_i$, and $E(e_i^4|\mathbf{V}^\top \mathbf{x}_i) = \mu_4 + 6\sigma^2 Var(q_i|\mathbf{V}^\top \mathbf{x}_i) + E(q_i^4|\mathbf{V}^\top \mathbf{x}_i)$. Define $\mathbf{M} = \frac{\mathbf{H}_{k_2} - \mathbf{P}_{\mathbf{X}_1}}{k_2} - \frac{\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{k_2}}{n-1-p_1-k_2}$. The matrix satisfies $tr(\mathbf{M}) = 0$, $tr(\mathbf{M}\mathbf{M}^\top) = \frac{1}{k_2} + \frac{1}{n-1-p_1-k_2}$, and $||\mathbf{M}||_{sp}^2 \le \frac{1}{k_2^2} + \frac{1}{(n-1-p_1-k_2)^2}$. 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 $a \le 2/\rho_2(1-\rho_1-\rho_2)$ such that $||\mathbf{M}||_{sp} \le a/n$. With a similar proof method in Appendix S2.2, we can derive

$$\frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}}\mathbf{e}^\top \mathbf{M}\mathbf{e} - o_p(1)}{(\sigma^2+\omega^2)\sqrt{1+o_p(1)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1). \tag{S2.18}$$

The condition $\boldsymbol{\beta}_2^\top \mathbf{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1)$ leads to $E(q_i^4) = o(1)$ as well as

$$|E(e_i^4) - \mu_4| \le c_1 \boldsymbol{\beta}_2^\top \mathbf{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1), \quad E\{\tau_i^2\} \le E(q_i^4) + \omega^4 \le c_1(\boldsymbol{\beta}_2^\top \mathbf{\Sigma}_{22} \boldsymbol{\beta}_2)^2 = o(1), \tag{S2.19}$$

for a constant c_1, from which we could obtain

$$\frac{(\mathbf{e}+\mathbf{r})^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})(\mathbf{r}+\mathbf{e})}{n-1-p_1-k_2} = \sigma^2+\omega^2+o_p(1). \tag{S2.20}$$

Let $\mathbf{V}\boldsymbol{\gamma} = (\boldsymbol{\xi}_1^\top, \boldsymbol{\xi}_2^\top)^\top$ with $\boldsymbol{\xi}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\xi}_2 \in \mathbb{R}^{p_2}$. Define $\nu^2 = \boldsymbol{\xi}_2^\top(\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\xi}_2$. Then

$$\nu^2 = \boldsymbol{\beta}_2^\top\boldsymbol{\Gamma}_2\left(\boldsymbol{\Gamma}^\top\mathbf{V}(\mathbf{V}^\top\boldsymbol{\Sigma}\mathbf{V})^{-1}\mathbf{V}^\top\boldsymbol{\Gamma} - \boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1\right)\boldsymbol{\Gamma}_2^\top\boldsymbol{\beta}_2 \le \boldsymbol{\beta}_2^\top(\boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})\boldsymbol{\beta}_2 = o(1).$$

To investigate $\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma}$, the term could be denoted as

$$\begin{aligned}&\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma}\\ &= \boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} - \boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\left(\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\right)^{-1}\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\end{aligned}$$

where $\boldsymbol{\phi} = (\mathbf{I}-\boldsymbol{\Gamma}_1^\top\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Gamma}_1)\boldsymbol{\Gamma}_2^\top\boldsymbol{\xi}_2$ and $\boldsymbol{\phi}^\top\boldsymbol{\phi} = \nu^2 = o(1)$. From the calculation

$$\begin{aligned}E\left\{\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\right\} &= \frac{n-1}{\sqrt{n}}\nu^2,\\ Var\left\{\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi}\right\} &\le 6\nu^4 = o(1)\end{aligned}$$

Markov's inequality implies,

$$\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} = \sqrt{n}\nu^2 + o_p(1).$$

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

$$\frac{1}{\sqrt{n}}\boldsymbol{\phi}^\top\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\left(\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\Gamma}_1^\top\right)^{-1}\boldsymbol{\Gamma}_1\mathbf{Z}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{Z}\boldsymbol{\phi} = \frac{p_1}{\sqrt{n}}\nu^2 + o_p(1).$$

Therefore,

$$\frac{1}{\sqrt{n}}\boldsymbol{\gamma}^\top\mathbf{W}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})\mathbf{W}\boldsymbol{\gamma} = \frac{n-p_1}{\sqrt{n}}\nu^2 + o_p(1). \tag{S2.21}$$

To study $\boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})(\mathbf{e} + \mathbf{r})$, (S2.16) and (S2.21) lead to

$$\left| \frac{1}{\sqrt{n}} \boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1}) \mathbf{r} \right| \le \sqrt{\mathbf{r}^\top \mathbf{r}} \cdot \sqrt{\frac{1}{n} \boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1}) \mathbf{W} \boldsymbol{\gamma}} = o_p(1).$$

the condition $\boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 = o(1)$, (S2.19) and (S2.21) lead to

$$\begin{aligned} E\left\{ \left(\frac{1}{\sqrt{n}} \boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1}) \mathbf{e} \right)^2 \right\} &= E\left[E\left\{ \left(\frac{1}{\sqrt{n}} \boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1}) \mathbf{e} \right)^2 \Big| \mathbf{W} \right\} \right] \\ &\le (c_3 + \sigma^2 + \omega^2) \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 \\ &= o(1), \end{aligned}$$

where c_3 is a constant. Therefore, we obtain

$$\frac{1}{\sqrt{n}} \boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{P}_{\mathbf{X}_1})(\mathbf{e} + \mathbf{r}) = o_p(1). \tag{S2.22}$$

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

$$\begin{aligned} \frac{T_{n,p_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} &= \frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}} \left\{ \frac{\boldsymbol{\gamma}^\top \mathbf{W}^\top (\mathbf{I}-\mathbf{P}_1-\mathbf{P}_{\mathbf{X}_1})(\mathbf{W}\boldsymbol{\gamma}+2\mathbf{e}+2\mathbf{r})}{k_2} + (\mathbf{r}+\mathbf{e})^\top \mathbf{M} (\mathbf{r}+\mathbf{e}) \right\}}{\frac{(\mathbf{r}+\mathbf{e})^\top (\mathbf{I}-\mathbf{P}_1-\mathbf{H}_{k_2})(\mathbf{r}+\mathbf{e})}{n-1-p_1-k_2}} \\ &= \frac{\sqrt{\frac{n\rho_2(1-\rho_1-\rho_2)}{2(1-\rho_1)}} \left\{ \frac{(1-\rho_1)}{\rho_2} \nu^2 + \mathbf{e}^\top \mathbf{M} \mathbf{e} \right\} + o_p(1)}{\sigma^2 + \omega^2 + o_p(1)}. \end{aligned}$$

Define $\tau_k^2 = \sigma^2 + \omega^2$. Then, ν^2 and τ_k^2 can also be calculated as follows. Let $\widetilde{\boldsymbol{\gamma}} = (\mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V})^{-1} \mathbf{V}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$ and $\mathbf{V} \widetilde{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{\xi}}_1^\top, \widetilde{\boldsymbol{\xi}}_2^\top)^\top$, where $\widetilde{\boldsymbol{\xi}}_1 \in \mathbb{R}^{p_1}$ and $\widetilde{\boldsymbol{\xi}}_2 \in \mathbb{R}^{p_2}$.

Then,

$$\begin{aligned}\nu^2 &= \boldsymbol{\beta}_2^\top \boldsymbol{\Gamma}_2 \left(\boldsymbol{\Gamma}^\top \mathbf{V}(\mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V})^{-1} \mathbf{V}^\top \boldsymbol{\Gamma} - \boldsymbol{\Gamma}_1^\top \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Gamma}_1\right) \boldsymbol{\Gamma}_2^\top \boldsymbol{\beta}_2 \\ &= \boldsymbol{\beta}^\top \boldsymbol{\Gamma} \left(\boldsymbol{\Gamma}^\top \mathbf{V}(\mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V})^{-1} \mathbf{V}^\top \boldsymbol{\Gamma} - \boldsymbol{\Gamma}_1^\top \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Gamma}_1\right) \boldsymbol{\Gamma}^\top \boldsymbol{\beta} \\ &= \widetilde{\boldsymbol{\xi}}_2^\top (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}) \widetilde{\boldsymbol{\xi}}_2.\end{aligned}$$

and

$$\tau_k^2 = \sigma^2 + \boldsymbol{\beta}_2^\top \boldsymbol{\Sigma}_{22} \boldsymbol{\beta}_2 - \boldsymbol{\gamma}^\top \mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V} \boldsymbol{\gamma} = \sigma^2 + \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta} - \widetilde{\boldsymbol{\gamma}}^\top \mathbf{V}^\top \boldsymbol{\Sigma} \mathbf{V} \widetilde{\boldsymbol{\gamma}}.$$

From (S2.18), the asymptotic power function of the proposed test T_{n,p_2} is

$$\begin{aligned}\Psi_{n,p_2}^{RP}(\boldsymbol{\beta}_2; \mathbf{P}_{k_2}) &= P\left(\frac{T_{n,p_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha\right) \\ &= \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}}\frac{\nu^2}{\tau_k^2}\right) + o(1),\end{aligned}$$

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

$$y_i = \alpha + \mathbf{x}_{1i}^\top \boldsymbol{\beta}_1 + \mathbf{x}_{2i}^\top \boldsymbol{\beta}_2 + \epsilon_i.$$

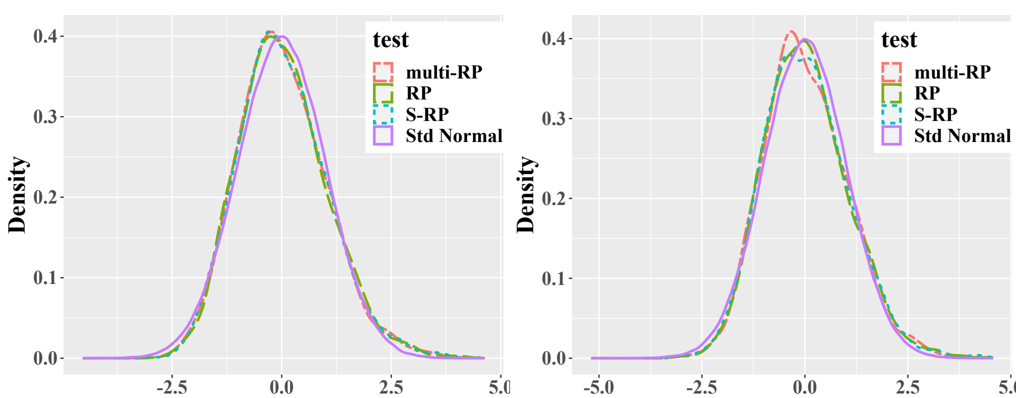
The covariate $(\mathbf{x}_{1i}^\top, \mathbf{x}_{2i}^\top)^\top$ is generated from $\boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \mathbf{z}_i$. The setup is almost the same as the first simulation study with differences lying in the design of $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ and $\boldsymbol{\Sigma}^{1/2}$. Specifically, we generated $\boldsymbol{\Sigma}^{1/2}$ by

$$\begin{pmatrix} c_1\mathbf{U}_1\sqrt{\mathbf{D}_1}\mathbf{U}_1^\top & c_2\mathbf{U}_1(\sqrt{\mathbf{D}_1},\mathbf{0})\mathbf{U}_2^\top \\ \mathbf{0} & \mathbf{U}_2\sqrt{\mathbf{D}_2}\mathbf{U}_2^\top \end{pmatrix},$$

where $\mathbf{U}_1$ ($\mathbf{U}_2$) 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 $\mathbf{D}_1$ are from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{p_1})$ with absolute values taken and the entries of diagonal matrix $\mathbf{D}_2$ 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 $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ keep unchanged in the two cases. The regression coefficient $\boldsymbol{\beta}_1$ is generated from $\mathcal{N}(\mathbf{0}, \mathbf{I}_{p_1})$ and $\boldsymbol{\beta}_2$ is randomly selected from the space generated by the first s columns of $\mathbf{U}_2$ with $||\boldsymbol{\beta}_2||_2^2$ 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 $\mathbf{H}_{part,0}$, indicating that the asymptotic null distribution of the proposed tests can be well approximated by the standard normal distribution. Here, ρ 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 3.

Table 1 reports the empirical power and type-I error of the proposed tests for the error term ϵ 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 $||\boldsymbol{\beta}_2||_2^2$. Compared with the correlated case, the tests show large power when there is no correlation between $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$, which is consistent with the feature in the asymptotic power in Theorem 4. Moreover, we find the empirical power is close to the asymptotic power, which further confirms the result in Theorem 4.



(a) Norm $\mathbf{z}$, norm ϵ and R=0.

(b) Norm $\mathbf{z}$, norm ϵ and R=1.

Figure 1: The kernel density estimation of RP,multi-RP, and S-RP tests under $\mathbf{H}_{part,0}$.

Table 1: Empirical power and type-I error of the RP, multi-RP, S-RP at the significance level 0.05 when $(n, p_1, p_2) = (400, 40, 3960)$ and $\rho = 0.2$.

Z	R	$\|\|\boldsymbol{\beta}_2\|\|_2^2$	$\epsilon \sim \sqrt{3/5}t(5)$			$\epsilon \sim \mathcal{N}(0,1)$		
			RP	multi-RP	S-RP	RP	multi-RP	S-RP
$\mathcal{N}(0,1)$	0	0	0.056	0.059	0.053	0.063	0.057	0.060
		0.1	0.715	0.729	0.717	0.704	0.707	0.715
		0.2	0.981	0.978	0.979	0.980	0.982	0.980
		0.3	0.999	0.998	0.999	1.000	0.999	0.999
	1	0	0.063	0.060	0.064	0.063	0.060	0.062
		0.1	0.533	0.548	0.532	0.544	0.532	0.545
		0.2	0.903	0.897	0.904	0.898	0.900	0.904
		0.3	0.988	0.983	0.985	0.992	0.991	0.990
$U(-\sqrt{3}, \sqrt{3})$	0	0	0.064	0.058	0.060	0.063	0.066	0.065
		0.1	0.716	0.716	0.720	0.717	0.711	0.722
		0.2	0.983	0.981	0.984	0.981	0.981	0.986
		0.3	1.000	1.000	1.000	1.000	1.000	1.000
	1	0	0.058	0.057	0.056	0.059	0.062	0.060
		0.1	0.533	0.537	0.539	0.533	0.542	0.542
		0.2	0.901	0.895	0.901	0.905	0.911	0.916
		0.3	0.991	0.992	0.991	0.991	0.992	0.993

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 $y_i = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i$, where $\alpha = 0$ and ϵ_i is generated from $\mathcal{N}(0, 1)$. The covariate $\mathbf{x}_i$ follows a latent factor structure in Lan et al. (2016). Specifically, $\mathbf{x}_i = \boldsymbol{\gamma}\mathbf{z}_i + \sqrt{\mathbf{D}}\widetilde{\mathbf{x}}_i$, where $\mathbf{z}_i$ is a d-dimensional latent factor, $\boldsymbol{\gamma} \in \mathbb{R}^{p\times d}$ is an associated factor loadings, $\widetilde{\mathbf{x}}_i$ is a p-dimensional factor profiled predictor that is independent of $\mathbf{z}_i$, and $\mathbf{D}$ is a diagonal matrix. From Lan et al. (2016), the factor profiled predictor $\widetilde{\mathbf{x}}_i$ represents the information that is contained in $\mathbf{x}_i$ but cannot be fully explained by the low-dimensional latent factor $\mathbf{z}_i$. In the simulation, each element of $\mathbf{z}_i$ and $\widetilde{\mathbf{x}}_i$ is independently generated from $\mathcal{N}(0,1)$, and each entry of $\boldsymbol{\gamma} \in \mathbb{R}^{p\times d}$ is independently generated from $\mathcal{N}(0, d^{-1})$. The elements of $\sqrt{\mathbf{D}}$ 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 $\boldsymbol{\beta} = ||\boldsymbol{\beta}||_2\boldsymbol{\delta}$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top$ 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 $||\boldsymbol{\beta}||_2^2 = 0.04$ 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	$\boldsymbol{\beta}$	$\|\|\boldsymbol{\beta}\|\|_2^2$	multi-RP	LWT	LDFF	RCV
d=3		0	0.062	0.052	0.050	0.458
	$s=5$	0.04	0.249	0.087	0.086	0.502
		0.08	0.532	0.116	0.118	0.544
	$s=50$	0.04	0.735	0.218	0.216	0.787
		0.08	0.984	0.409	0.388	0.951
d=5		0	0.052	0.071	0.069	0.843
	$s=5$	0.04	0.295	0.183	0.181	0.917
		0.08	0.605	0.312	0.308	0.959
	$s=50$	0.04	0.764	0.387	0.384	0.999
		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 covariates 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 $||\boldsymbol{\beta}||_2$ 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

Bai, Z. D. and Yin, Y. Q. (1993). Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix. *The Annals of Probability* **21**, 1275–1294.

Bechar, I. (2009). A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables. *arXiv preprint arXiv:0909.3595*.

Bhansali, R. J., Giraitis, L. and Kokoszka, P. S. (2007). Convergence of quadratic forms with nonvanishing diagonal. *Statistics & Probability Letters* **77**, 726–734.

Chikuse, Y. (2003). *Statistics on special manifolds. Lecture Notes in Statistics* **174**. Springer-Verlag, New York.

Davidson, K. R. and Szarek, S. J. (2001). Local operator theory, random matrices and Banach spaces. In *Handbook of the geometry of Banach spaces, Vol. I*, 317–366. North-Holland, Amsterdam.

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *The Annals of Statistics* **12**, 793–815.

Durrett, R. (2019). *Probability—theory and examples. Cambridge Series in Statistical and Probabilistic Mathematics* **49**. Cambridge University Press, Cambridge.

Fang, K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric multivariate and related distributions. Monographs on Statistics and Applied Probability* **36**. Chapman and Hall, Ltd., London.

Lan, W., Ding, Y., Fang, Z. and Fang, K. (2016). Testing covariates in high dimension linear regression with latent factors. *Journal of Multivariate Analysis* **144**, 25-37.

Lopes, M., Jacob, L. and Wainwright, M. J. (2011). A more powerful two-sample test in high dimensions using random projection. In *Advances in Neural Information Processing Systems*, 1206–1214.

Lan, W., Wang, H. and Tsai, C.L. (2014). Testing covariates in high-dimensional regression. *Annals of the Institute of Statistical Mathematics* **66**, 279-301.

Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik* **1**, 457.

Steinberger, L. and Leeb, H. (2018). On conditional moments of high-dimensional random vectors given lower-dimensional projections. *Bernoulli* **24**, 565-591.