A Random Projection Approach to Hypothesis Tests in High-Dimensional Single-Index Models

Changyu Liu, Xingqiu Zhao and Jian Huang

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

Supplementary Materials

Appendix A

We first introduce some notation. For a vector $\mathbf{v}$, we write $||\mathbf{v}||_2$ for the Euclidean norm. 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 $||\mathbf{B}||_{sp} = \max\limits_{||\mathbf{x}||_2=1} ||\mathbf{B}\mathbf{x}||_2$.

A.1 Proof of Lemma 2.1

Proof. As $\mathbf{x}^\top \boldsymbol{\beta}$ follows the normal distribution and $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, we could make a division

$$\mathbf{x}^\top \boldsymbol{\beta} = \mathbf{u}^\top \boldsymbol{\eta} + q,$$

where $\boldsymbol{\eta} = E(\mathbf{u}\mathbf{x}^\top \boldsymbol{\beta})$ and q is independent of $\mathbf{u}$.

When $\boldsymbol{\eta} \neq \mathbf{0}$, for any $\mathbf{b} \in \mathbb{R}^k$, it could be expressed as a sum $\mathbf{b} = c\boldsymbol{\eta} + \mathbf{r}$, where $c = \boldsymbol{\eta}^\top \mathbf{b}/||\boldsymbol{\eta}||_2^2$ and $\mathbf{r} = \mathbf{b} - c\boldsymbol{\eta}$. The orthogonality between $\boldsymbol{\eta}$ and $\mathbf{r}$ implies that $\mathbf{u}^\top \boldsymbol{\eta}$ and

$\mathbf{u}^\top\mathbf{r}$ are independent, which further leads to the independence between $\mathbf{u}^\top\mathbf{r}$ and the response y. Consequently, we have

$$
\begin{aligned}
R(\alpha,\mathbf{b}) &= E\left\{L\left(\alpha+\mathbf{u}^\top\mathbf{b},y\right)\right\} \\
&= E\left\{L\left(\alpha+c\mathbf{u}^\top\boldsymbol{\eta}+\mathbf{u}^\top\mathbf{r},y\right)\right\} \\
&= E\left[E\left\{L\left(\alpha+\mathbf{u}^\top\mathbf{r}+c\mathbf{u}^\top\boldsymbol{\eta},y\right)\left|\mathbf{u}^\top\mathbf{r}\right.\right\}\right] \\
&\geq \min_{\tilde{\alpha}} E\left\{L(\tilde{\alpha}+c\mathbf{u}^\top\boldsymbol{\eta},y)\right\} \\
&\geq \min_{\tilde{\alpha},\tilde{c}} E\left\{L(\tilde{\alpha}+\tilde{c}\mathbf{u}^\top\boldsymbol{\eta},y)\right\}.
\end{aligned}
$$

Since there is an unique solution to the minimization problem, we have

$$\mathbf{b}^\star = c^\star\boldsymbol{\eta},$$

where $c^\star$ is a constant and could be calculated through $c^\star = \boldsymbol{\eta}^\top\mathbf{b}^\star/||\boldsymbol{\eta}||_2^2$. □

A.2 Proof of Theorem 3.1

Lemma A.1. *Suppose $(x_i,y_i), i=1,\ldots n$, are i.i.d. from a distribution satisfying $E(x_i) = E(y_i) = 0$, $Var(x_i) = \sigma_x^2$, $Var(y_i) = \sigma_y^2$ and $E(x_iy_i) = \tau$. Let $\boldsymbol{x} = (x_1,...,x_n)^\top$ and $\boldsymbol{y} = (y_1,...,y_n)^\top$. For a symmetric $n\times n$ matrix $\boldsymbol{M} = (m_{ij})$, we have*

$$
\begin{aligned}
&E\left(\boldsymbol{x}^\top\boldsymbol{M}\boldsymbol{y}\right) = \tau tr(\boldsymbol{M}) \quad \text{and} \\
&Var\left(\boldsymbol{x}^\top\boldsymbol{M}\boldsymbol{y}\right) = \left(E(x_1^2y_1^2)-2\tau^2-\sigma_x^2\sigma_y^2\right)\sum_{i=1}^n m_{ii}^2 + \left(\tau^2+\sigma_x^2\sigma_y^2\right)tr(\boldsymbol{M}^2).
\end{aligned}
$$

The property of the diagonal entries of the hat matrix is investigated by the following lemma, whose proof is deferred to Appendix B.

Lemma A.2. *Suppose* $\boldsymbol{X}$ *is an* $n \times p$ *random matrix, where each entry is i.i.d. from* $\mathcal{N}(0,1)$. *Let* $\boldsymbol{H} = (h_{ij}) = \boldsymbol{X}\big(\boldsymbol{X}^\top \boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top$. *As* $(p,n) \to \infty$ *with* $p/n \to \zeta \in (0,1)$, *we have*

$$\max_{i=1,\ldots,n} E\left\{(h_{ii} - \zeta)^2\right\} \to 0.$$

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

Lemma A.3. *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 $\boldsymbol{A}_n = (a_{ij})$ *is a symmetric matrix, and* z_i *are i.i.d. variables satisfying* $E(z_i) = 0$ *and* $Var(z_i) = 1$. *When* $E(z_1^4) < \infty$ *and* $\frac{||\boldsymbol{A}_n||_{sp}}{||\boldsymbol{A}_n||_F} \to 0$, *then*

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

Proof of Theorem 3.1. Under $\mathbf{H}_0$, the response y is independent of $\mathbf{x}$ and satisfies $y = e$, where e is the residual defined in the linear form (2.3). Let $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}}{p} - \frac{\mathbf{I}-\mathbf{H}}{n-p}$. The matrix is independent of $\mathbf{y}$ and satisfies $tr(\mathbf{M}) = 0$ and $\mathbf{M}^2 = \frac{\mathbf{H}}{p^2} + \frac{\mathbf{I}-\mathbf{H}}{(n-p)^2}$. Hence, $||\mathbf{M}||_{sp}^2/||\mathbf{M}||_F^2 \le O(n^{-1})$ and $tr(\mathbf{M}^2) = \frac{1}{p} + \frac{1}{n-p}$. From Lemma A.1, we have

$$\begin{aligned}
&E\left(\mathbf{y}^\top \mathbf{M}\mathbf{y}|\mathbf{M}\right) = 0 \quad \text{and} \\
&Var\left(\mathbf{y}^\top \mathbf{M}\mathbf{y}|\mathbf{M}\right) = \left(E(y^4) - 3E(y^2)^2\right)\sum_{i=1}^{n} m_{ii}^2 + 2E(y^2)^2 tr(\mathbf{M}^2).
\end{aligned}$$

Under the condition $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$, which will be verified subsequently, Lemma A.3 demonstrates that

$$\frac{\mathbf{y}^\top \mathbf{M}\mathbf{y}}{\sigma^2\sqrt{2/n\zeta(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{A.1}$$

where $\sigma^2 = Var(e) = Var(y)$.

Let $G_n = n\sum_{i=1}^{n} m_{ii}^2$. It can be written as

$$G_n = n\sum_{i=1}^{n}\left\{\frac{h_{ii}}{p} - \frac{1-h_{ii}}{n-p}\right\}^2 = \frac{n^3}{p^2(n-p)^2}\sum_{i=1}^{n}\left\{h_{ii} - \frac{p}{n}\right\}^2,$$

where h_{ij} denotes the ij-th entry of $\mathbf{H}$. Based on the assumption $p/n \to \zeta$, Lemma A.2 implies that $E(G_n) = o(1)$. Therefore, we have $G_n = o_p(1)$ by Markov's inequality.

To study $\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}$, the calculation shows

$$E\left(\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}\right) = \sigma^2(n-p),\ Var\left(\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}\right) \le 3E(y^4)(n-p).$$

From Markov's inequality,

$$\frac{\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}}{n-p} = \sigma^2 + o_p(1). \tag{A.2}$$

Consequently, (A.1) and (A.2) lead to

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

which completes the proof. □

A.3 Proof of Theorem 3.2

First, we introduce a decomposition of the projection matrix in the following lemma, whose proof is deferred to Appendix B.

Lemma A.4. *Suppose $\boldsymbol{X}$ is an $n \times p$ random matrix, where each entry is i.i.d. from $\mathcal{N}(0,1)$. Consider a decomposition $\boldsymbol{X} = (\boldsymbol{w}, \boldsymbol{G})$, where $\boldsymbol{w} \in \mathbb{R}^n$ and $\boldsymbol{G} \in \mathbb{R}^{n\times(p-1)}$. Let*

$\boldsymbol{H} = \boldsymbol{X}\big(\boldsymbol{X}^\top \boldsymbol{X}\big)^{-1}\boldsymbol{X}^\top$. *We have*

$$\boldsymbol{H} = \boldsymbol{H_w} + (\boldsymbol{I} - \boldsymbol{H_w})[\boldsymbol{H_G} + \frac{\boldsymbol{H_G H_w H_G}}{1 - tr(\boldsymbol{H_G H_w})}](\boldsymbol{I} - \boldsymbol{H_w}),$$

where $\boldsymbol{H_w} = \boldsymbol{w}(\boldsymbol{w}^\top \boldsymbol{w})^{-1}\boldsymbol{w}^\top$ *and* $\boldsymbol{H_G} = \boldsymbol{G}(\boldsymbol{G}^\top \boldsymbol{G})^{-1}\boldsymbol{G}^\top$. *Suppose* $\boldsymbol{e} = (e_1, \ldots, e_n)^\top$ *is a random vector composed of i.i.d. entries, satisfying* $E(e_1) = 0$ *and* $E(e_1^4) < \infty$. *In addition,* $\boldsymbol{e}$ *is independent of* $\boldsymbol{G}$. *For* $i \neq j$, e_i *is independent of* w_j, *otherwise,* $E(e_i w_i) = 0$, *where* w_i *denotes the i-th entry of* $\boldsymbol{w}$. *As* $(p, n) \to \infty$ *with* $p/n \to \zeta \in (0, 1)$, *for any* $\delta > 0$, *we have*

$$\boldsymbol{e}^\top \boldsymbol{He} = \boldsymbol{e}^\top \boldsymbol{H_G e} + Re,$$

where $Re = o_p(n^\delta)$.

Proof of Theorem 3.2. From the linear form (2.3) and the condition $E(y) = 0$, the model can be written as

$$\mathbf{y} = c_0 \mathbf{X}\boldsymbol{\beta} + \mathbf{e},$$

where $\mathbf{e} = (e_1, \ldots, e_n)^\top$. The direct calculation shows that

$$\mathbf{y}^\top \mathbf{H}\mathbf{y} = c_0^2 \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} + 2c_0 \boldsymbol{\beta}^\top \mathbf{X}^\top \mathbf{e} + \mathbf{e}^\top \mathbf{H}\mathbf{e} \text{ and}$$

$$\mathbf{y}^\top (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{e}^\top (\mathbf{I} - \mathbf{H})\mathbf{e}.$$

First, we investigate $\mathbf{e}^\top \mathbf{H}\mathbf{e}$. Let $\mathbf{w} = \frac{\mathbf{X}\boldsymbol{\beta}}{||\boldsymbol{\beta}||_2}$ and $\mathbf{G} = \mathbf{X}(\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{p-1})$, where $\{\frac{\boldsymbol{\beta}}{||\boldsymbol{\beta}||_2}, \boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{p-1}\}$ forms an orthonormal basis of $\mathbb{R}^p$. The normality of $\mathbf{X}$ implies that $\mathbf{w}$ is independent of $\mathbf{G}$, which further indicates that $\mathbf{e}$ and $\mathbf{G}$ are independent, since e_i is determined by $\mathbf{x}_i^\top \boldsymbol{\beta}$ and ϵ_i. When $\mathbf{X}$ is replaced by $\mathbf{XO}$, where $\mathbf{O}$ is a $p \times p$ orthogonal matrix, $\mathbf{H}$ stays the same. Based on Lemma A.4, we have

$$\mathbf{e}^\top \mathbf{H}\mathbf{e} = \mathbf{e}^\top \mathbf{H_G}\mathbf{e} + Re,$$

where $Re = o_p(n^{\delta})$ with sufficiently small $\delta > 0$. Define $\mathbf{M} = \frac{\mathbf{H_G}}{p-1} - \frac{\mathbf{I}-\mathbf{H_G}}{n-p+1}$. The matrix is independent of $\mathbf{e}$. With a similar method in A.2, we have

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

and

$$\frac{\mathbf{e}^{\top}(\mathbf{I}-\mathbf{H_G})\mathbf{e}}{n-p} = \sigma^2 + o_p(1),$$

where $\sigma^2 = Var(e_1) = Var(y) - c_0^2||\boldsymbol{\beta}||_2^2$.

To investigate $c_0^2\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta} + 2c_0\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{e}$, the calculation shows

$$E(c_0^2\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta}) = nc_0^2||\boldsymbol{\beta}||_2^2, \quad Var(c_0^2\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta}) = 2nc_0^4||\boldsymbol{\beta}||_2^4,$$

$$E(c_0\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{e}) = 0, \quad Var(c_0\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{e}) \leq 4nc_0^2||\boldsymbol{\beta}||_2^2 E(e_1^4)^{1/2}.$$

Under the condition $c_0^2||\boldsymbol{\beta}||_2^2 = o(1)$, Markov's inequality implies that

$$\frac{1}{\sqrt{n}}c_0^2\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta} = \sqrt{n}c_0^2||\boldsymbol{\beta}||_2^2 + o_p(1), \quad \frac{1}{\sqrt{n}}c_0\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{e} = o_p(1).$$

Consequently,

$$\begin{aligned}\frac{T_n - 1}{\sqrt{2/n\zeta(1-\zeta)}} &= \frac{\sqrt{\frac{n\zeta(1-\zeta)}{2}}\left\{\frac{c_0^2\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{X}\boldsymbol{\beta}}{p} + \frac{2c_0\boldsymbol{\beta}^{\top}\mathbf{X}^{\top}\mathbf{e}}{p} + \mathbf{e}^{\top}\left(\frac{\mathbf{H}}{p} - \frac{\mathbf{I}-\mathbf{H}}{n-p}\right)\mathbf{e}\right\}}{\frac{\mathbf{e}^{\top}(\mathbf{I}-\mathbf{H})\mathbf{e}}{n-p}} \\ &= \frac{\sqrt{\frac{n(1-\zeta)}{2\zeta}}c_0^2||\boldsymbol{\beta}||_2^2 + \sqrt{\frac{n\zeta(1-\zeta)}{2}}\mathbf{e}^{\top}\mathbf{M}\mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)}.\end{aligned}$$

It shows that the power function satisfies

$$\begin{aligned}\Psi_n(\boldsymbol{\beta}) &= P(\frac{T_n - 1}{\sqrt{2/n\zeta(1-\zeta)}} > z_{\alpha}) \\ &= \Phi(-z_{\alpha} + \sqrt{\frac{n(1-\zeta)}{2\zeta}}\frac{c_0^2||\boldsymbol{\beta}||_2^2}{\sigma^2}) + o(1),\end{aligned}$$

which completes the proof. $\square$

A.4 Proof of Theorem 3.3

Proof. Under $\mathbf{H}_{0,\mathbf{B}}$, according to the linear form (3.4), the model can be written as

$$\mathbf{y} = c_0 \mathbf{W}\boldsymbol{\gamma} + \mathbf{e},$$

where $\mathbf{e} = (e_1, \ldots, e_n)^\top$. Then,

$$\mathbf{y}^\top(\mathbf{H} - \mathbf{H}_{\mathbf{W}})\mathbf{y} = \mathbf{e}^\top(\mathbf{H} - \mathbf{H}_{\mathbf{W}})\mathbf{e}, \quad \mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{e}^\top(\mathbf{I} - \mathbf{H})\mathbf{e}.$$

Hence, we have

$$T_{n,p2} = \frac{\mathbf{e}^\top(\mathbf{H} - \mathbf{H}_{\mathbf{W}})\mathbf{e}/b}{\mathbf{e}^\top(\mathbf{I} - \mathbf{H})\mathbf{e}/(n-p)}. \tag{A.3}$$

First, we investigate $\mathbf{e}^\top\mathbf{H}\mathbf{e}$ and $\mathbf{e}^\top\mathbf{H}_{\mathbf{W}}\mathbf{e}$. Let $\widetilde{\mathbf{X}} = \mathbf{X}\boldsymbol{\Sigma}^{-1/2}$ and $\widetilde{\mathbf{W}} = \mathbf{W}\boldsymbol{\Sigma}_{\mathbf{w}}^{-1/2}$, where $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_{\mathbf{w}}$ are the covariance matrices of $\mathbf{x}$ and $\mathbf{w}$, respectively. Then,

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top = \widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^\top \text{ and}$$
$$\mathbf{H}_{\mathbf{W}} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top = \widetilde{\mathbf{W}}(\widetilde{\mathbf{W}}^\top\widetilde{\mathbf{W}})^{-1}\widetilde{\mathbf{W}}^\top,$$

where the entries of $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}$ are i.i.d. from $\mathcal{N}(0,1)$, respectively. Let $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\mathbf{w}}^{1/2}\boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\xi}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}$, where $\boldsymbol{\xi} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top(\mathbf{I} - \mathbf{H}_{\mathbf{B}}))^\top$. Then, $\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}} = \widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}$, since $\mathbf{X}\boldsymbol{\xi} = \mathbf{W}\boldsymbol{\gamma}$. Based on this relationship, we make decomposition to the projection matrices. For the matrix $\mathbf{H}$, let $\mathbf{v} = \frac{\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}}}{||\widetilde{\boldsymbol{\xi}}||_2}$ and $\mathbf{G} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{p-1})$, where $\{\frac{\widetilde{\boldsymbol{\xi}}}{||\widetilde{\boldsymbol{\xi}}||_2}, \boldsymbol{\xi}_1, ..., \boldsymbol{\xi}_{p-1}\}$ forms an orthonormal basis of $\mathbb{R}^p$. The orthogonality implies that $\mathbf{v}$ and $\mathbf{G}$ are independent, which further shows that $\mathbf{e}$ is independent of $\mathbf{G}$, since e_i is determined by $\mathbf{w}_i^\top\boldsymbol{\gamma}$ and ϵ_i. From Lemma A.4, we obtain

$$\mathbf{e}^\top\mathbf{H}\mathbf{e} = \mathbf{e}^\top\mathbf{H}_{\mathbf{G}}\mathbf{e} + Re_1, \tag{A.4}$$

where $\mathbf{H_G} = \mathbf{G}(\mathbf{G}^\top\mathbf{G})^{-1}\mathbf{G}^\top$, and $Re_1 = o_p(n^{\delta_1})$ with sufficiently small $\delta_1 > 0$. For the matrix $\mathbf{H_W}$, let $\mathbf{F} = \widetilde{\mathbf{W}}(\boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_{p-b-1})$, where $\{\frac{\widetilde{\boldsymbol{\gamma}}}{||\widetilde{\boldsymbol{\gamma}}||_2}, \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_{p-b-1}\}$ forms an orthonormal basis of $\mathbb{R}^{p-b}$. According to $\widetilde{\mathbf{X}}\widetilde{\boldsymbol{\xi}} = \widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}$, there is a constant c such that $\frac{\widetilde{\mathbf{W}}\widetilde{\boldsymbol{\gamma}}}{||\widetilde{\boldsymbol{\gamma}}||_2} = c\mathbf{v}$. The orthogonality implies that $\mathbf{v}$ and $\mathbf{F}$ are independent, which then shows that $\mathbf{e}$ is independent of $\mathbf{F}$. From Lemma A.4, we obtain

$$\mathbf{e}^\top\mathbf{H_W}\mathbf{e} = \mathbf{e}^\top\mathbf{H_F}\mathbf{e} + Re_2, \tag{A.5}$$

where $\mathbf{H_F} = \mathbf{F}(\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top$ and $Re_2 = o_p(n^{\delta_2})$ with sufficiently small $\delta_2 > 0$. Let $\sigma^2 = Var(e_1)$. The calculation yields

$$E\{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}\} = \sigma^2(n+1-p),\ Var\{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}\} \leq 3E(e_1^4)(n+1-p).$$

Markov's inequality implies

$$\frac{\mathbf{e}^\top(\mathbf{I} - \mathbf{H_G})\mathbf{e}}{n+1-p} = \sigma^2 + o_p(1).$$

Combining (A.3), (A.4) and (A.5), we obtain

$$\frac{T_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} = \frac{\sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\zeta_1)}}\mathbf{e}^\top\mathbf{M}\mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)}, \tag{A.6}$$

where $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H_G} - \mathbf{H_F}}{b} - \frac{\mathbf{I} - \mathbf{H_G}}{n+1-p}$.

To study $\mathbf{e}^\top\mathbf{M}\mathbf{e}$, we follow a similar method as that given in A.2. First, note that $Span\{\mathbf{F}\} \subseteq Span\{\mathbf{G}\}$. The property of the projection matrix shows that

$$\mathbf{H_G}\mathbf{H_F} = \mathbf{H_F}\mathbf{H_G} = \mathbf{H_F},$$

and $tr(\mathbf{M}) = 0$. Then, $\mathbf{M}^\top\mathbf{M} = \frac{\mathbf{H_G} - \mathbf{H_F}}{b^2} + \frac{\mathbf{I} - \mathbf{H_G}}{(n+1-p)^2}$, leading to $||\mathbf{M}||_{sp}^2/||\mathbf{M}||_F^2 \leq O(n^{-1})$ and

$$Var(\mathbf{e}^\top\mathbf{M}\mathbf{e}|\mathbf{M}) = (E(e_1^4) - 3\sigma^4)\sum_{i=1}^n m_{ii}^2 + 2\sigma^4(\frac{1}{b} + \frac{1}{n+1-p}).$$

We show that $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$. Let $(\mathbf{H_G})_{ii}$ and $(\mathbf{H_F})_{ii}$ denote the i-th diagonal entries of $\mathbf{H_G}$ and $\mathbf{H_F}$, respectively. Then,

$$n\sum_{i=1}^{n} m_{ii}^2 \le \frac{2h_1^2}{n}\sum_{i=1}^{n}\left\{(\mathbf{H_G})_{ii} - \frac{p-1}{n}\right\}^2 + \frac{2h_2^2}{n}\sum_{i=1}^{n}\left\{(\mathbf{H_F})_{ii} - \frac{p-b-1}{n}\right\}^2,$$

where $h_1 = \frac{1-\frac{p}{n}+\frac{b}{n}+\frac{1}{n}}{\frac{b}{n}(1-\frac{p}{n}+\frac{1}{n})} \to \frac{1+\zeta_1-\zeta}{\zeta_1(1-\zeta)}$ and $h_2 = \frac{n}{b} \to \frac{1}{\zeta_1}$ as $n \to \infty$. From the definitions of $\mathbf{G}$ and $\mathbf{F}$, Lemma A.2 demonstrates that

$$E\left\{(\mathbf{H_G})_{ii} - \frac{p-1}{n}\right\}^2 \to 0,\ E\left\{(\mathbf{H_F})_{ii} - \frac{p-b-1}{n}\right\}^2 \to 0.$$

It leads to $E(n\sum_{i=1}^{n} m_{ii}^2) = o(1)$. Hence, $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$ by Markov's inequality. From Lemma A.3, we obtain

$$\frac{\mathbf{e}^\top\mathbf{M}\mathbf{e}}{\sigma^2\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Taking this into (A.6), we have

$$\frac{T_{n,p_2}-1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

which completes the proof. □

A.5 Proof of Theorem 3.4

Proof. Let

$$\boldsymbol{\psi} = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\mathbf{D}^\top\begin{pmatrix}\boldsymbol{\Sigma}_{12}\\ \boldsymbol{\Sigma}_{22}\end{pmatrix}\mathbf{H_B}\boldsymbol{\beta}_2, \quad \mathbf{D} = \begin{pmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp}\end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\mathbf{w}}$ is the covariance matrix of $\mathbf{w}$. Then, $\mathbf{w} = \mathbf{D}^\top\mathbf{x}$ and $\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{D}^\top\boldsymbol{\Sigma}\mathbf{D}$. From the normality assumption of $\mathbf{x}$, we derive a decomposition

$$\mathbf{x}_2^\top\mathbf{H_B}\boldsymbol{\beta}_2 = \mathbf{w}^\top\boldsymbol{\psi} + q,$$

where q is independent of $\mathbf{w}$. Let $\tau^2 = Var(q)$. It satisfies

$$\tau^2 = \boldsymbol{\beta}_2^\top \mathbf{H}_\mathbf{B} \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \mathbf{H}_\mathbf{B} \boldsymbol{\beta}_2.$$

Based on the linear form (3.4), the model can be written as

$$y = c_0 \mathbf{w}^\top (\boldsymbol{\gamma} + \boldsymbol{\psi}) + c_0 q + e.$$

Then, we obtain

$$\begin{aligned} &\mathbf{y}^\top (\mathbf{H} - \mathbf{H}_\mathbf{W})\mathbf{y} = c_0^2 \mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{q} + 2c_0 \mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{e} + \mathbf{e}^\top (\mathbf{H} - \mathbf{H}_\mathbf{W})\mathbf{e} \ \text{ and} \\ &\mathbf{y}^\top (\mathbf{I} - \mathbf{H})\mathbf{y} = \mathbf{e}^\top (\mathbf{I} - \mathbf{H})\mathbf{e}, \end{aligned}$$

where $\mathbf{q} = (q_1, ..., q_n)^\top$ and $\mathbf{e} = (e_1, \dots, e_n)^\top$.

To investigate $c_0^2 \mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{q}$, Lemma A.1 indicates

$$E\left\{\mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{q}\right\} = \tau^2 (n - p + b),\ Var\left\{\mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{q}\right\} = 2\tau^4 (n - p + b).$$

Under the condition $c_0^2 \tau^2 = o(1)$, Markov's inequality shows

$$\frac{\sqrt{n} c_0^2 \mathbf{q}^\top (\mathbf{I} - \mathbf{H}_\mathbf{W})\mathbf{q}}{b} = \frac{\sqrt{n}(1 - \zeta + \zeta_1)}{\zeta_1} c_0^2 \tau^2 + o_p(1). \tag{A.7}$$

Next, we study $\mathbf{e}^\top (\mathbf{H} - \mathbf{H}_\mathbf{W})\mathbf{e}$ and $\mathbf{e}^\top (\mathbf{I} - \mathbf{H})\mathbf{e}$. Let $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}$ have the same definitions as that in A.4 and define $\widetilde{\boldsymbol{\beta}} = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta}$. We introduce another matrix $\mathbf{R} = (\mathbf{W}, \mathbf{q})$ with its projection matrix defined as $\mathbf{H}_\mathbf{R} = \mathbf{R}(\mathbf{R}^\top \mathbf{R})^{-1}\mathbf{R}^\top$. Let $\mathbf{v} = \mathbf{X}\boldsymbol{\beta}$. Then, $\mathbf{v} \in Span\{\mathbf{R}\} \subseteq Span\{\mathbf{X}\}$. Let $\mathbf{F} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{p-b})$ and $\mathbf{G} = \widetilde{\mathbf{X}}(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{p-1})$, where $\boldsymbol{\xi}_i$ are selected to make $\{\frac{\widetilde{\boldsymbol{\beta}}}{||\widetilde{\boldsymbol{\beta}}||_2}, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{p-1}\}$ form an orthonormal basis for $\mathbb{R}^p$ and $Span\{\mathbf{R}\} = Span\{\mathbf{F}, \mathbf{v}\}$. It indicates that $\mathbf{v}$ is independent of $\mathbf{F}$ and $\mathbf{G}$. Then, $\mathbf{e}$ is independent of $\mathbf{F}$ and $\mathbf{G}$, since e_i is determined by ϵ_i and $\mathbf{x}_i^\top \boldsymbol{\beta}$. From Lemma A.4, we have

$$\mathbf{e}^\top \mathbf{H}_\mathbf{R} \mathbf{e} = \mathbf{e}^\top \mathbf{H}_\mathbf{F} \mathbf{e} + Re_1, \quad \mathbf{e}^\top \mathbf{H} \mathbf{e} = \mathbf{e}^\top \mathbf{H}_\mathbf{G} \mathbf{e} + Re_2, \tag{A.8}$$

where $Re_1 = o_p(n^{\delta_1})$ and $Re_2 = o_p(n^{\delta_2})$ for sufficiently small $\delta_1, \delta_2 > 0$. Since $Span\{\mathbf{F}\} \subseteq Span\{\mathbf{G}\}$,

$$\mathbf{H}_\mathbf{G}\mathbf{H}_\mathbf{F} = \mathbf{H}_\mathbf{F}\mathbf{H}_\mathbf{G} = \mathbf{H}_\mathbf{F}.$$

Let $\mathbf{M} = \frac{\mathbf{H}_\mathbf{G} - \mathbf{H}_\mathbf{F}}{b-1} - \frac{\mathbf{I} - \mathbf{H}_\mathbf{G}}{n+1-p}$. The matrix is independent of $\mathbf{e}$ and satisfies $tr(\mathbf{M}) = 0$ and $\mathbf{M}^\top \mathbf{M} = \frac{\mathbf{H}_\mathbf{G} - \mathbf{H}_\mathbf{F}}{(b-1)^2} + \frac{\mathbf{I} - \mathbf{H}_\mathbf{G}}{(n+1-p)^2}$, leading to

$$\frac{||\mathbf{M}||_{sp}^2}{||\mathbf{M}||_F^2} \le O(n^{-1}), \quad n||\mathbf{M}||_F^2 = \frac{1}{\zeta_1} + \frac{1}{1-\zeta} + o(1).$$

With a similar method in A.4, we have

$$\frac{\mathbf{e}^\top \mathbf{M} \mathbf{e}}{\sigma^2 \sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{A.9}$$

where $\sigma^2 = Var(e) = Var(y) - c_0^2 \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$.

We then study the relationship between $\mathbf{e}^\top \mathbf{H}_\mathbf{R} \mathbf{e}$ and $\mathbf{e}^\top \mathbf{H}_\mathbf{W} \mathbf{e}$, since the test statistic contains $\mathbf{H}_\mathbf{W}$ rather than $\mathbf{H}_\mathbf{R}$. From the definition of $\mathbf{R}$, the matrix $\mathbf{H}_\mathbf{R}$ could be divided based on $\mathbf{q}$ and $\mathbf{H}_\mathbf{W}$. Under the condition that $\tau^{-1}\mathbf{e}^\top \mathbf{H}_\mathbf{W} \mathbf{q} = o_p(n^{\gamma+0.5})$ for any $\gamma > 0$, which will be verified subsequently, the proof of Lemma A.4 shows

$$\mathbf{e}^\top \mathbf{H}_\mathbf{R} \mathbf{e} = \mathbf{e}^\top \mathbf{H}_\mathbf{W} \mathbf{e} + Re_3, \tag{A.10}$$

with $Re_3 = o_p(n^{\delta_3})$ for any $\delta_3 > 0$. Note that $\mathbf{e}$ might not be independent of $\mathbf{W}$. To verify the condition, let $\boldsymbol{\ell} = \boldsymbol{\Sigma}_\mathbf{w}^{1/2}(\boldsymbol{\gamma} + \boldsymbol{\psi})$ and $\mathbf{E} = \widetilde{\mathbf{W}}(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1})$, where $\boldsymbol{\gamma}_i$ are selected to make

$\{\frac{\boldsymbol{\ell}}{||\boldsymbol{\ell}||_2}, \boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_{p-b-1}\}$ form an orthonormal basis for $\mathbb{R}^{p-b}$. Then, $\mathbf{e}$ is independent of $\mathbf{E}$, since $\widetilde{\mathbf{W}}\boldsymbol{\ell}$ and $\mathbf{E}$ are independent. From Lemma A.4, it gives a decomposition, $\mathbf{H}_{\mathbf{W}} = \mathbf{H}_{\mathbf{E}} + \mathbf{K}$, where $\mathbf{K}$ satisfies $\mathbf{e}^\top \mathbf{K}\mathbf{e} = o_p(n^{\delta_4})$ for any $\delta_4 > 0$. The calculation shows

$$E(\mathbf{e}^\top \mathbf{H}_{\mathbf{E}}\mathbf{q}) = 0,\ Var(\mathbf{e}^\top \mathbf{H}_{\mathbf{E}}\mathbf{q}) \leq 4(p-b-1)\tau^2 E(e^4)^{1/2}$$

and

$$|\mathbf{e}^\top \mathbf{K}\mathbf{q}| \leq \sqrt{\mathbf{e}^\top \mathbf{K}\mathbf{e}}\sqrt{\mathbf{q}^\top \mathbf{q}} = \tau o_p(n^{\delta_4+0.5}).$$

Hence, $\tau^{-1}\mathbf{e}^\top \mathbf{H}_{\mathbf{W}}\mathbf{q} = o_p(n^{\gamma+0.5})$ for any $\gamma > 0$. This together with the condition $c_0^2\tau^2 = o(1)$ implies

$$\frac{1}{\sqrt{n}} c_0 \mathbf{q}^\top (\mathbf{I} - \mathbf{H}_{\mathbf{W}})\mathbf{e} = o_p(1). \tag{A.11}$$

Finally, we combine the above results to derive the asymptotic local power function. For any idempotent matrix $\mathbf{A}$, Lemma A.1 shows

$$E(\mathbf{e}^\top \mathbf{A}\mathbf{e}) = \sigma^2 tr(\mathbf{A}),\ Var(\mathbf{e}^\top \mathbf{A}\mathbf{e}) \leq 3E(e^4) tr(\mathbf{A}).$$

This together with Markov's inequality and the derived independence indicates

$$\frac{\mathbf{e}^\top (\mathbf{H}_{\mathbf{G}} - \mathbf{H}_{\mathbf{F}})\mathbf{e}}{b-1} = \sigma^2 + o_p(1) \ \text{ and } \ \frac{\mathbf{e}^\top (\mathbf{I} - \mathbf{H}_{\mathbf{G}})\mathbf{e}}{n-p+1} = \sigma^2 + o_p(1).$$

Combining it with (A.7), (A.8), (A.9), (A.10) and (A.11), we have

$$\begin{aligned}
&\frac{T_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \\
&= \frac{\sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\zeta_1)}}\left\{\frac{c_0^2 \mathbf{q}^\top (\mathbf{I}-\mathbf{H}_{\mathbf{W}})\mathbf{q}}{b} + \frac{2c_0 \mathbf{q}^\top (\mathbf{I}-\mathbf{H}_{\mathbf{W}})\mathbf{e}}{b} + \mathbf{e}^\top (\frac{\mathbf{H}-\mathbf{H}_{\mathbf{W}}}{b} - \frac{\mathbf{I}-\mathbf{H}}{n-p})\mathbf{e}\right\}}{\frac{\mathbf{e}^\top (\mathbf{I}-\mathbf{H})\mathbf{e}}{n-p}} \\
&= \frac{\sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}} c_0^2\tau^2 + \sqrt{\frac{n\zeta_1(1-\zeta)}{2(1-\zeta+\rho_1)}}\mathbf{e}^\top \mathbf{M}\mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)}.
\end{aligned}$$

Therefore, the asymptotic power function satisfies

$$\begin{aligned}\Psi_n(\boldsymbol{\beta}_2; \mathbf{B}) &= P(\frac{T_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} > z_\alpha) \\ &= \Phi(-z_\alpha + \sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}\frac{c_0^2\tau^2}{\sigma^2}}) + o(1),\end{aligned}$$

which completes the proof. □

A.6 Proof of Theorem 4.1

An upper-bound and a lower-bound on the extreme eigenvalues of Wishart matrices were given in Davidson and Szarek (2001, Theorem 2.13) .

Lemma A.5. *For $k \le p$, 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 $t \ge 0$, we have*

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

Proof of Theorem 4.1. Under $\mathbf{H}_0$, the model becomes

$$y = f(\epsilon). \tag{A.12}$$

Consider a linear regression model

$$y = \alpha + \mathbf{u}^\top \boldsymbol{\eta} + e, \tag{A.13}$$

where α is an intercept, $\mathbf{u} = \mathbf{P}_k^\top \mathbf{x}$, $\boldsymbol{\eta}$ is a vector of coefficients and e is a error term independent of $\mathbf{x}$. When $\alpha = E(f(\epsilon))$, $\boldsymbol{\eta} = \mathbf{0}$, and $e = f(\epsilon) - E(f(\epsilon))$, $(\mathbf{x}, y)$ from model

(A.12) has the same distribution as that of $(\mathbf{x}, y)$ from model (A.13). Hence, it suffices to study the test statistic under (A.13). Let $\mathbf{M} = (m_{ij}) = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-k-1}$. Under the condition $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$, which will be verified subsequently, with a similar proof method in A.2, we can derive

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

Therefore, the asymptotic normality of the test statistic is demonstrated.

To verify the condition $n\sum_{i=1}^n m_{ii}^2 = o_p(1)$, it is sufficient to prove

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

where h_{ii} is the i-th diagonal entry of $\mathbf{H}_k$. For simplicity of notation, we denote $\mathbf{H}_k$ by $\mathbf{H}$ in the following. Based on Assumption H1, the matrix $\mathbf{H}$ can be denoted as

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

where $\mathbf{Z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$ and $\mathbf{A}$ belongs to the Stiefel manifold $\mathcal{V}_k(\mathbb{R}^m) = \{\mathbf{A} \in \mathbb{R}^{m\times k} : \mathbf{A}^\top\mathbf{A} = \mathbf{I}\}$. Considering the randomness of $\mathbf{P}_k$, to cover general cases, we assume that the matrix $\mathbf{A}$ is uniformly distributed on $\mathcal{V}_k(\mathbb{R}^m)$ and is independent of $\mathbf{Z}$. Let $\mathbf{U}\mathbf{\Lambda}\mathbf{O}^\top$ 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 $\mathbf{\Lambda} = (\mathbf{D}, \mathbf{0})$ 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 $\mathbf{Z} = \mathbf{U}\mathbf{D}\mathbf{O}_n^\top$. The diagonal entries of $\frac{1}{m}\mathbf{D}^2$ are the eigenvalues of $\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top$. The calculation shows

$$E\left[\max_{i=1,\ldots,n}\left(\frac{d_i^2}{m}-1\right)^2\right] \le E\left[tr\left\{\left(\frac{1}{m}\mathbf{Z}\mathbf{Z}^\top-\mathbf{I}\right)^2\right\}\right] = O(n^2m^{-1}).$$

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

$$P\left\{\max_{i=1,\dots,n}\left(\frac{d_i}{\sqrt{m}}-1\right)^2 > t\right\} \le P\left\{\max_{i=1,\dots,n}\left(\frac{d_i^2}{m}-1\right)^2 > t\right\} \le O(n^2m^{-1}t^{-1}). \tag{A.14}$$

Let $\mathbf{V} = (\mathbf{I}-\mathbf{P}_1)\,\mathbf{U}\mathbf{O}_n^\top\mathbf{A}$ and $\tilde{\mathbf{Z}} = (\mathbf{I}-\mathbf{P}_1)\,\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{O}_n^\top\mathbf{A}$. The hat matrix for $\tilde{\mathbf{Z}}$ and $\mathbf{V}$ can be denoted as

$$\mathbf{H} = (h_{ij}) = \tilde{\mathbf{Z}}\left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top \quad\text{and}\quad \mathbf{S} = (s_{ij}) = \mathbf{V}\left(\mathbf{V}^\top\mathbf{V}\right)^{-1}\mathbf{V}^\top,$$

where $\mathbf{H}$ is the target matrix. We will show that h_{ii} and s_{ii} are close. Let $\mathbf{e}_i$ denote the unit vector with 1 in the i-th coordinate. Define $\hat{\boldsymbol{\gamma}}_i^{ls} = \left(\mathbf{V}^\top\mathbf{V}\right)^{-1}\mathbf{V}^\top\mathbf{e}_i$ and $\hat{\boldsymbol{\eta}}_i^{ls} = \left(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\right)^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i$. Then, $\hat{\boldsymbol{\gamma}}_i^{ls}$ and $\hat{\boldsymbol{\eta}}_i^{ls}$ are the solutions to the least square problems $\min\limits_{\boldsymbol{\gamma}\in\mathbb{R}^k}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\boldsymbol{\gamma}\right|\right|_2^2$ and $\min\limits_{\boldsymbol{\eta}\in\mathbb{R}^k}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\boldsymbol{\eta}\right|\right|_2^2$, respectively. Therefore,

$$\begin{aligned}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2^2 &\le \left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2^2 \\ &\le \left(\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2 + \left|\left|(\mathbf{V}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2\right)^2\end{aligned} \tag{A.15}$$

and

$$\begin{aligned}\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2^2 &\le \left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \mathbf{V}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2^2 \\ &\le \left(\left|\left|\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i - \tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2 + \left|\left|(\tilde{\mathbf{Z}}-\mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2\right)^2.\end{aligned} \tag{A.16}$$

To study (A.15) and (A.16), we first investigate the value of $\left|\left|(\mathbf{V}-\tilde{\mathbf{Z}})\hat{\boldsymbol{\gamma}}_i^{ls}\right|\right|_2$ and $\left|\left|(\tilde{\mathbf{Z}}-\mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}\right|\right|_2$. 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 each element of $m\times k$ matrix $\mathbf{G}$ is 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 A.5 and the independence between $\mathbf{A}$ and $\mathbf{Z}$, for any $h_1 > 0$ and $h_2 > 0$, we have

$$\begin{aligned} &P\left[\lambda_{\max}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \geq (1+\sqrt{k/n}+h_1)^2\right] \leq \exp\left(-nh_1^2/2\right) \quad \text{and} \\ &P\left[\lambda_{\min}\left(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\right) \leq (1-\sqrt{k/n}-h_2)^2\right] \leq \exp\left(-nh_2^2/2\right). \end{aligned} \tag{A.17}$$

Based on SVD, for any matrix $\mathbf{B}$, the nonzero eigenvalues of $\mathbf{B}^\top\mathbf{B}$ and $\mathbf{B}\mathbf{B}^\top$ are the same. Hence, we have

$$\begin{aligned} &\lambda_{\max}\left(\mathbf{V}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\mathbf{E}\big)} \\ &\leq \frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{A.18}$$

and

$$\begin{aligned} &\lambda_{\max}\left(\tilde{\mathbf{Z}}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\mathbf{A}^\top\mathbf{O}_n\mathbf{O}_n^\top\mathbf{A}\big(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}}\big)^{-1}\tilde{\mathbf{Z}}^\top\right) \\ &\leq \lambda_{\max}\big(\frac{1}{n}\mathbf{E}^\top\mathbf{E}\big)\frac{1}{\lambda_{\min}\big(\frac{1}{n}\mathbf{E}^\top\frac{\mathbf{D}}{\sqrt{m}}\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\frac{\mathbf{D}}{\sqrt{m}}\mathbf{E}\big)} \\ &\leq \frac{1}{\lambda_{\min}\big(\frac{\mathbf{D}^2}{m}\big)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{A.19}$$

with probability at least $1-\exp(-nh_1^2/2)-\exp(-nh_2^2/2)$. Based on (A.14), (A.18) and (A.19), upper bounds can be derived as follows:

$$\begin{aligned} ||(\mathbf{V}-\tilde{\mathbf{Z}})\hat{\gamma}_i^{ls}||_2^2 &= ||(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\mathbf{e}_i||_2^2 \\ &\leq \max_{i=1,\ldots,n}(1-\frac{d_i}{\sqrt{m}})^2||\mathbf{O}_n^\top\mathbf{A}\big(\mathbf{V}^\top\mathbf{V}\big)^{-1}\mathbf{V}^\top\mathbf{e}_i||_2^2 \\ &\leq t\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \end{aligned} \tag{A.20}$$

and

$$\begin{aligned}
||(\tilde{\mathbf{Z}}-\mathbf{V})\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 &= ||(\mathbf{I}-\mathbf{P}_1)\mathbf{U}(\mathbf{I}-\frac{\mathbf{D}}{\sqrt{m}})\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i||_2^2 \\
&\le \max_{i=1,\dots,n}(1-\frac{d_i}{\sqrt{m}})^2||\mathbf{O}_n^\top\mathbf{A}(\tilde{\mathbf{Z}}^\top\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}^\top\mathbf{e}_i||_2^2 \\
&\le \max_{i=1,\dots,n}(1-\frac{d_i}{\sqrt{m}})^2\cdot\frac{1}{\min\limits_{i=1,\dots,n}\left(\frac{d_i^2}{m}\right)}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2} \\
&\le \frac{t}{(1-\sqrt{t})^2}\cdot\frac{(1+\sqrt{k/n}+h_1)^2}{(1-\sqrt{k/n}-h_2)^2}
\end{aligned} \tag{A.21}$$

with probability at least $1-O(n^2m^{-1}t^{-1})-\exp(-nh_1^2/2)-\exp(-nh_2^2/2)$. Combining (A.15), (A.16), (A.20) and (A.21), with $h_1=n^{-1/4}$, $h_2=n^{-1/4}$ and $t=n^{-c}$, where c is a positive constant, we have

$$||\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i-\tilde{\mathbf{Z}}\hat{\boldsymbol{\eta}}_i^{ls}||_2^2 \le ||\,(\mathbf{I}-\mathbf{P}_1)\,\mathbf{e}_i-\mathbf{V}\hat{\boldsymbol{\gamma}}_i^{ls}||_2^2+3n^{-c/2}\cdot\frac{1+\sqrt{k/n}+n^{-1/4}}{1-\sqrt{k/n}-n^{-1/4}} \quad \text{and}$$

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

with probability at least $1-O(n^{2+c}m^{-1})-2\exp\left(-n^{1/2}/2\right)$. As the above derivation is valid for any $\mathbf{e}_i$, when $n\to\infty$ and m is sufficiently large, we obtain

$$\max_{i=1\dots,n}|h_{ii}-s_{ii}|^2=o_p(1). \tag{A.22}$$

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

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

where $\mathbf{U}\mathbf{O}_n^\top$ is independent of $\mathbf{G}$ and satisfies $\mathbf{U}\mathbf{O}_n^\top\mathbf{O}_n\mathbf{U}^\top=\mathbf{I}$. From the definition of $\mathbf{G}$, Lemma A.2 and the dominated convergence theorem, we obtain

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

Combining this with (A.22) and Markov's inequality, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\big(h_{ii}-\rho\big)^2 &= \frac{1}{n}\sum_{i=1}^{n}\big(h_{ii}-s_{ii}+s_{ii}-\rho\big)^2 \\ &\le \max_{i=1\ldots,n} 2\big(h_{ii}-s_{ii}\big)^2+\frac{2}{n}\sum_{i=1}^{n}\big(s_{ii}-\rho\big)^2 \\ &= o_p(1),\end{aligned}$$

which completes the proof. □

Remark A.1. *In Section A.2, the proof of Theorem 3.1 proceeds in the following steps.*

1. *Suppose the condition $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$ hold. By Lemma A.3, we show*

$$\frac{\boldsymbol{y}^\top \boldsymbol{M}\boldsymbol{y}}{\sigma^2\sqrt{2/n\zeta(1-\zeta)}} = \frac{\boldsymbol{y}^\top(\frac{\boldsymbol{H}}{p}-\frac{\boldsymbol{I}-\boldsymbol{H}}{n-p})\boldsymbol{y}}{\sigma^2\sqrt{2/n\zeta(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

2. *Based on the normality of $\boldsymbol{x}$ and Lemma A.2, we prove $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$.*

3. *By the Markov's inequality, we show*

$$\frac{\boldsymbol{y}^\top(\boldsymbol{I}-\boldsymbol{H})\boldsymbol{y}}{n-p} = \sigma^2 + o_p(1).$$

4. *The result of Theorem 3.1 follows from the Slutsky's lemma.*

As shown above, the normality assumption of the covariate is only applied in the step 2. Suppose we have $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$, then the proof of Theorem 4.1 proceeds in the same steps as above. In addition, Assumption H1 is shown to be adequate enough to verify the condition $n\sum_{i=1}^{n} m_{ii}^2 = o_p(1)$ in the proof of Theorem 4.1. Consequently, the result of Theorem 4.1 is derived under the weaker Assumption H1.

A.7 Proof of Theorem 4.2

Proof. Let $\mathbf{U} = \mathbf{X}\mathbf{P}_k\boldsymbol{\Sigma}_1^{-1/2}$, with the i-th row denoted by $\mathbf{u}_i$ and $\boldsymbol{\Sigma}_1 = \mathbf{P}_k^\top\boldsymbol{\Sigma}\mathbf{P}_k$. Define $\boldsymbol{\eta} = \boldsymbol{\Sigma}_1^{-1/2}\mathbf{P}_k^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$ and $\omega^2 = ||\boldsymbol{\eta}||_2^2$. Based on the normality assumption, we derive a decomposition, $\mathbf{x}_i^\top\boldsymbol{\beta} = \mathbf{u}_i^\top\boldsymbol{\eta} + q_i$, where q_i is independent of $\mathbf{u}_i$. According to Lemma 2.1, the model can be written as

$$y_i = c_{0,k}\mathbf{u}_i^\top\boldsymbol{\eta} + e_i, \tag{A.23}$$

where $c_{0,k} = E(\boldsymbol{\eta}^\top\mathbf{u}_1 y_1)/\omega^2$ when $\boldsymbol{\eta} \neq \mathbf{0}$, and e_i satisfies $E(e_i) = 0$ and $E(\mathbf{u}_i e_i) = \mathbf{0}$. Let σ^2 denote the variance of e_i. It satisfies $\sigma^2 = Var(y) - c_{0,k}^2\omega^2$. The calculation shows that

$$\frac{T_{n,k}-1}{\sqrt{2/n\rho(1-\rho)}} = \frac{\sqrt{\frac{n\rho(1-\rho)}{2}}\left(\frac{c_{0,k}^2\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}}{k} + \frac{2c_{0,k}\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}}{k} + \mathbf{e}^\top\mathbf{F}\mathbf{e}\right)}{\frac{\mathbf{e}^\top(\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k)\mathbf{e}}{n-1-k}}, \tag{A.24}$$

where $\mathbf{F} = \frac{\mathbf{H}_k}{k} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H}_k}{n-1-k}$ and $\mathbf{e} = (e_1, ..., e_n)^\top$.

According to Assumption H4, $c_{0,k}^2\omega^2 = o(1)$. The calculation shows

$$E(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}) = (n-1)\omega^2,\ Var(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta}) \leq 10(n-1)\omega^4,$$

$$E(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}) = 0,\ Var(\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e}) \leq 16\sqrt{2+E(y^4)}(n-1)\omega^2.$$

By Markov's inequality, we have

$$\frac{\sqrt{n}}{k}c_{0,k}^2\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{U}\boldsymbol{\eta} = \frac{\sqrt{n}}{\rho}c_{0,k}^2\omega^2 + o_p(1),\ \frac{\sqrt{n}}{k}c_{0,k}\boldsymbol{\eta}^\top\mathbf{U}^\top(\mathbf{I}-\mathbf{P}_1)\mathbf{e} = o_p(1). \tag{A.25}$$

For the study of $\mathbf{e}^\top\mathbf{F}\mathbf{e}$, we follow a similar method in A.3 and derive

$$\frac{T_{n,k}-1}{\sqrt{2/n\rho(1-\rho)}} = \frac{\sqrt{\frac{n(1-\rho)}{2\rho}}c_{0,k}^2\omega^2 + \sqrt{\frac{n\rho(1-\rho)}{2}}\mathbf{e}^\top\mathbf{M}\mathbf{e} + o_p(1)}{\sigma^2 + o_p(1)},$$

where $\mathbf{M} = \frac{\mathbf{H_G}}{k-1} - \frac{\mathbf{I}-\mathbf{P}_1-\mathbf{H_G}}{n-k}$, with the projection matrix $\mathbf{H_G}$ independent of $\mathbf{e}$. Based on a similar method in A.6, we can derive

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

Consequently, the asymptotic power function satisfies

$$\begin{aligned}\Psi_n^{RP}(\boldsymbol{\beta};\mathbf{P}_k) &= P\Big(\frac{T_{n,k}-1}{\sqrt{2/n\rho(1-\rho)}} > z_\alpha\Big)\\ &= \Phi\Big(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}\frac{c_{0,k}^2\omega^2}{\sigma^2}}\Big) + o(1),\end{aligned}$$

which completes the proof. □

A.8 Proof of Lemma 4.3

Proof. Let $\mathbf{U}^\top\mathbf{D}\mathbf{U}$ be a spectral decomposition of $\boldsymbol{\Sigma}$, where $\mathbf{D} = \mathrm{diag}(d_1, ..., d_p)$ with $d_1 \geq d_2 \geq \cdots \geq d_p \geq 0$. To cover general cases, we consider that the direction of $\boldsymbol{\beta}$ is uniformly distributed on the unit sphere. Since the standard normal distribution in high dimensions is close to the uniform distribution on the sphere of radius $\sqrt{p}$, the direction of $\boldsymbol{\beta}$ is distributed as $\frac{1}{\sqrt{p}}\boldsymbol{\delta}$, where $\boldsymbol{\delta}$ follows $\mathcal{N}(\mathbf{0},\mathbf{I})$. Based on the rotation invariance of the normal distribution, with a slight abuse of notation, we denote $\frac{1}{||\boldsymbol{\beta}||_2}\mathbf{U}\boldsymbol{\beta}$ and $\mathbf{U}\mathbf{P}_k$ by $\frac{1}{\sqrt{p}}\boldsymbol{\delta}$ and $\mathbf{P}_k$, respectively. Then, we have

$$\min_{\boldsymbol{\zeta}\in\mathbb{R}^k} \frac{\sqrt{n}||\boldsymbol{\Gamma}^\top\boldsymbol{\beta} - \boldsymbol{\Gamma}^\top\mathbf{P}_k\boldsymbol{\zeta}||_2^2}{||\boldsymbol{\beta}||_2^2} = \min_{\boldsymbol{\zeta}\in\mathbb{R}^k}\sqrt{n}||\sqrt{\mathbf{D}}\frac{\boldsymbol{\delta}}{\sqrt{p}} - \sqrt{\mathbf{D}}\mathbf{P}_k\boldsymbol{\zeta}||_2^2. \tag{A.26}$$

Let $\boldsymbol{\delta} = (\boldsymbol{\delta}_s^\top, \boldsymbol{\delta}_{p-s}^\top)^\top$ with $\boldsymbol{\delta}_s \in \mathbb{R}^s$ and $\boldsymbol{\delta}_{p-s} \in \mathbb{R}^{p-s}$, where s is given in Assumption H5. Correspondingly, $\mathbf{D}$ is divided into $\mathbf{D}_s$ and $\mathbf{D}_{p-s}$, with $\mathbf{D}_s = \mathrm{diag}(d_1, ..., d_s)$ and $\mathbf{D}_{p-s} =$

$\text{diag}(d_{s+1}, ..., d_p)$, and $\mathbf{P}_k = (\mathbf{P}_{s,k}^\top, \mathbf{P}_{p-s,k}^\top)^\top$ with $\mathbf{P}_{s,k} \in \mathbb{R}^{s\times k}$ and $\mathbf{P}_{p-s,k} \in \mathbb{R}^{(p-s)\times k}$. Define $\boldsymbol{\zeta}_0 = \mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}$. The calculation shows that

$$\begin{aligned}
&\sqrt{n}||\sqrt{\mathbf{D}}\frac{\boldsymbol{\delta}}{\sqrt{p}} - \sqrt{\mathbf{D}}\mathbf{P}_k\boldsymbol{\zeta}_0||_2^2, \\
&\leq \sqrt{n}||\sqrt{\mathbf{D}_s}(\frac{\boldsymbol{\delta}_s}{\sqrt{p}} - \mathbf{P}_{s,k}\boldsymbol{\zeta}_0)||_2^2 + \sqrt{n}||\sqrt{\mathbf{D}_{p-s}}(\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}} - \mathbf{P}_{p-s,k}\boldsymbol{\zeta}_0)||_2^2 \\
&\leq 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\frac{\boldsymbol{\delta}_{p-s}}{\sqrt{p}}||_2^2 + 2\sqrt{n}||\sqrt{\mathbf{D}_{p-s}}\mathbf{P}_{p-s,k}\mathbf{P}_{s,k}^\top(\mathbf{P}_{s,k}\mathbf{P}_{s,k}^\top)^{-1}\frac{\boldsymbol{\delta}_s}{\sqrt{p}}||_2^2 \\
&= T_1 + T_2.
\end{aligned} \tag{A.27}$$

For T_1, based on the concentration inequality in Bechar (2009), an upper bound can be derived as follows.

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

where h_1 is a positive real number that may vary with n. Let $h_1 = n^\gamma$. According to Assumption H5, we get

$$\begin{aligned}
||\boldsymbol{\beta}||_2^2 T_1 &\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p}\left(tr(\mathbf{D}_{p-s}) + 2\sqrt{h_1}||\mathbf{D}_{p-s}||_F + 2h_1||\mathbf{D}_{p-s}||_{sp}\right) \\
&\leq \frac{2\sqrt{n}||\boldsymbol{\beta}||_2^2}{p}tr(\mathbf{D}_{p-s})\left(1 + 2\sqrt{h_1} + 2h_1\right) \\
&\leq \frac{10n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2 tr(\mathbf{D}_{p-s})}{p} = o(1)
\end{aligned} \tag{A.28}$$

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

For T_2, an upper bound for $tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k})$ can be derived, based on the proof procedure in Lopes, Jacob, and Wainwright (2011, Lemma 5). Specifically,

$$P\left[tr(\mathbf{P}_{p-s,k}^\top\mathbf{D}_{p-s}\mathbf{P}_{p-s,k}) \leq 2k\,tr(\mathbf{D}_{p-s})\right] \to 1.$$

This together with Lemma A.5 leads to

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

with probability tending to 1. Note that the first inequality is from Lemma 2 of Lopes, Jacob, and Wainwright (2011), which demonstrates that $tr(\mathbf{AB}) \leq \lambda_{\max}(\mathbf{B})tr(\mathbf{A})$, for any positive semi-definite matrices $\mathbf{A}$ and $\mathbf{B}$. With a similar method in the study of $||\boldsymbol{\beta}||_2^2T_1$, we have

$$
||\boldsymbol{\beta}||_2^2T_2 \leq \frac{20n^{0.5+\gamma}||\boldsymbol{\beta}||_2^2tr(\mathbf{D}_{p-s})}{p(1-\sqrt{s/k}-k^{-1/4})^2} = o(1) \tag{A.29}
$$

with probability tending to 1.

Therefore, combining (A.26), (A.27), (A.28) and (A.29), we have

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

with probability tending to 1. □

A.9 Proof of Theorem 4.3

Proof. Under $\mathbf{H}_{0,\mathbf{B}}$, the model becomes

$$
y = f(\mathbf{w}^T\boldsymbol{\gamma}, \epsilon). \tag{A.30}
$$

Define

$$
\mathbf{D} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp} \end{pmatrix}, \quad \mathbf{R}_{k_2} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{\mathbf{B}^\perp} & \mathbf{S}_{\mathbf{B}}\mathbf{P}_{k_2} \end{pmatrix}.
$$

Then, $\mathbf{w} = \mathbf{D}^\top \mathbf{x}$ and $\mathbf{u}_{k2} = \mathbf{R}_{k_2}^\top \mathbf{x}$. Consider a working model

$$y = f(\mathbf{u}_{k2}^\top \boldsymbol{\xi}, \epsilon), \tag{A.31}$$

where $\mathbf{u}_{k2}^\top \boldsymbol{\xi} = \mathbf{w}^\top \boldsymbol{\gamma} + \mathbf{x}_2^\top \mathbf{S}_{\mathbf{B}} \mathbf{P}_{k_2} \boldsymbol{\xi}_2$ with $\boldsymbol{\xi}_2$ being a k_2-dimensional vector. When $\boldsymbol{\xi}_2 = \mathbf{0}$, the distribution of y under model (A.31) is the same as that in model (A.30). Therefore, it is sufficient to study the test statistic under model (A.31) with $\boldsymbol{\xi}_2 = \mathbf{0}$. With a similar analysis in A.4, we can derive

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \to \mathcal{N}(0,1),$$

which completes the proof. $\square$

A.10 Proof of Theorem 4.4

Proof. Let the definition of $\mathbf{D}$, $\mathbf{R}_{k_2}$ are the same as that in A.9. Let $\boldsymbol{\Sigma}_{\mathbf{w}} = \mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D}$ and $\boldsymbol{\Sigma}_1 = \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2}$, which are the covariance matrix of $\mathbf{w}$ and $\mathbf{u}_{k2}$, respectively. Define $\mathbf{U} = \mathbf{X}\mathbf{R}_{k_2}\boldsymbol{\Sigma}_1^{-1/2}$ and denote its i-th row by $\mathbf{u}_i$. The assumption of normality leads to a decomposition, $\mathbf{x}_i^\top \boldsymbol{\beta} = \mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} + r_i$, where $\boldsymbol{\xi} = \boldsymbol{\Sigma}_1^{-1} \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$ and r_i is independent of $\mathbf{u}_i$. Therefore, the model can be written as

$$y_i = f(\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} + r_i, \epsilon_i), \tag{A.32}$$

where r_i and ϵ_i are independent of $\mathbf{u}_i$. Consider $\mathbf{w}_i$. It can be denoted as $\mathbf{w}_i = \mathbf{D}^\top \mathbf{x}_i$ and satisfies $\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} = (\mathbf{w}_i^\top, \mathbf{x}_{2i}^\top \mathbf{S}_{\mathbf{B}} \mathbf{P}_{k_2})$. Let $\boldsymbol{\delta} = \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2} \boldsymbol{\Sigma}_1^{-1} \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$. Then,

$$\mathbf{u}_i^\top \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{\xi} = \mathbf{w}_i^\top \boldsymbol{\delta} + q_i,$$

where q_i is independent of $\mathbf{w}_i$. Let $\boldsymbol{\eta} = \mathbf{R}_{k_2}\boldsymbol{\xi} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{p_2}$. Define $\nu^2 = Var(q_1)$. The calculation yields

$$\nu^2 = \boldsymbol{\eta}_2^\top \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top\boldsymbol{\Sigma}\mathbf{D})^{-1}\mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \boldsymbol{\eta}_2.$$

Let $c_{0,k_2} = \boldsymbol{\xi}^\top\boldsymbol{\Sigma}_1^{1/2}E(\mathbf{u}_i y_i)/\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_1\boldsymbol{\xi}$, where $\mathbf{P}_{k_2}$ is treated as fixed in the expectation. Under the condition $c_{0,k_2}^2\nu^2 = o(1)$, the method of A.5 is applicable. Therefore, we obtain

$$\begin{aligned} \Psi_n^{RP}(\boldsymbol{\beta}_2; \mathbf{B}, \mathbf{P}_{k_2}) &= P\Big(\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha\Big) \\ &= \Phi\Big(-z_\alpha + \sqrt{\frac{n(1-\rho_1)(1-\rho_1-\rho_2)}{2\rho_2}}\frac{c_{0,k_2}^2\nu^2}{\sigma^2}\Big) + o(1), \end{aligned}$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2\boldsymbol{\xi}^\top\boldsymbol{\Sigma}_1\boldsymbol{\xi}$. The proof is completed. $\square$

Appendix B

Proof of Lemma A.2. Let $\mathbf{S}_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1}\mathbf{x}_i\mathbf{x}_i^\top$, where $\mathbf{x}_i$ denotes the i-th row of $\mathbf{X}$. According to Bai and Yin (1993, Theorem 2), it shows that the extreme eigenvalues of $\mathbf{S}_{n-1}$ satisfy $\lambda_{\max}(\mathbf{S}_{n-1}) \to (1+\sqrt{\zeta})^2$ and $\lambda_{\min}(\mathbf{S}_{n-1}) \to (1-\sqrt{\zeta})^2, a.s.$. Based on the result in Marčenko and Pastur (1967), $\frac{1}{p}tr(\mathbf{S}_{n-1}^{-1}) \to \frac{1}{1-\zeta}, a.s.$. Then, we obtain

$$E\Big(\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n|\mathbf{S}_{n-1}\Big) = \frac{p}{n-1}\frac{tr(\mathbf{S}_{n-1}^{-1})}{p} \to \frac{\zeta}{1-\zeta},\ a.s.\ \text{ and}$$

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

Therefore,

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

From Woodbury formula,

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

Let $f(x)=\frac{x}{1+x}$. It satisfies $f^{'}(x)\leq 1$, for $x\geq 0$. Based on the mean value theorem, we get

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

which implies

$$E\left\{(h_{nn}-\zeta)^2\right\} \leq E\left\{\left(\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n-\frac{\zeta}{1-\zeta}\right)^2\right\} \to 0. \tag{B.1}$$

For $i=1,\dots,n-1$, we follow a similar method and derive

$$\begin{aligned}E\left\{(h_{ii}-\zeta)^2\right\} &\leq E\left\{\left(\frac{1}{n-1}\mathbf{x}_i^\top(\frac{1}{n-1}\sum_{j\neq i}\mathbf{x}_j\mathbf{x}_j^\top)^{-1}\mathbf{x}_i-\frac{\zeta}{1-\zeta}\right)^2\right\} \\ &= E\left\{\left(\frac{1}{n-1}\mathbf{x}_n^\top\mathbf{S}_{n-1}^{-1}\mathbf{x}_n-\frac{\zeta}{1-\zeta}\right)^2\right\} \to 0.\end{aligned} \tag{B.2}$$

Therefore,

$$\max_{i=1,\dots,n} E\left\{(h_{ii}-\zeta)^2\right\} \to 0,$$

which completes the proof.

□

Proof of Lemma A.4. According to Woodbury formula, we have

$$\begin{aligned}\mathbf{H} &= \mathbf{H_w}+(\mathbf{I}-\mathbf{H_w})\mathbf{G}[\mathbf{G}^\top(\mathbf{I}-\mathbf{H_w})\mathbf{G}]^{-1}\mathbf{G}^\top(\mathbf{I}-\mathbf{H_w}) \\ &= \mathbf{H_w}+(\mathbf{I}-\mathbf{H_w})[\mathbf{H_G}+\frac{\mathbf{H_G}\mathbf{H_w}\mathbf{H_G}}{1-tr(\mathbf{H_G}\mathbf{H_w})}](\mathbf{I}-\mathbf{H_w}).\end{aligned}$$

This leads to

$$\mathbf{e}^\top \mathbf{H}\mathbf{e} = \mathbf{e}^\top [\mathbf{H}_\mathbf{w} + (\mathbf{I} - \mathbf{H}_\mathbf{w})(\mathbf{H}_\mathbf{G} + \frac{\mathbf{H}_\mathbf{G}\mathbf{H}_\mathbf{w}\mathbf{H}_\mathbf{G}}{1 - tr(\mathbf{H}_\mathbf{G}\mathbf{H}_\mathbf{w})})(\mathbf{I} - \mathbf{H}_\mathbf{w})]\mathbf{e}$$

$$= \mathbf{e}^\top \mathbf{H}_\mathbf{G}\mathbf{e} + Re,$$

where Re is defined as

$$Re = \frac{r_1^2}{r_4} + \frac{r_3^2}{r_4(1-g)} - \frac{2r_1r_3}{r_4} - \frac{2r_1r_2r_3}{r_4^2(1-g)} + \frac{r_1^2r_2}{r_4^2} + \frac{r_1^2r_2^2}{r_4^3(1-g)},$$

with $r_1 = \mathbf{e}^\top\mathbf{w}$, $r_2 = \mathbf{w}^\top\mathbf{H}_\mathbf{G}\mathbf{w}$, $r_3 = \mathbf{e}^\top\mathbf{H}_\mathbf{G}\mathbf{w}$, $r_4 = \mathbf{w}^\top\mathbf{w}$, and $g = tr(\mathbf{H}_\mathbf{G}\mathbf{H}_\mathbf{w})$.

For r_1, Lemma A.1 shows that $E(r_1) = 0, Var(r_1) = nE(e_1^2w_1^2)$. Then, $r_1 = o_p(n^{\gamma_1+0.5})$, for any $\gamma_1 > 0$, based on Markov's inequality. For r_2, the independence between $\mathbf{w}$ and $\mathbf{G}$ implies that $E(r_2) = p-1$ and $Var(r_2) = 2(p-1)$. Then, $\frac{r_2}{n} = \zeta + o_p(1)$ by Markov's inequality. For r_3, Lemmas A.1 and A.2 show that

$$E(r_3) = 0 \text{ and } E(r_3^2) \leq Cn(1+o(1)) + (p-1)E(e_1^2)$$

for a constant C. Then, $r_3 = o_p(n^{\gamma_2+0.5})$, for any $\gamma_2 > 0$, based on Markov's inequality. For r_4, the strong law of large numbers shows that $\frac{r_4}{n} \to 1$, a.s.. From $g = r_2/r_4$, we obtain $g = \zeta + o_p(1)$. Consequently, for any $\delta > 0$, $Re = o_p(n^\delta)$, which completes the proof. □

Appendix C

C.1 Partial Test in Relatively High Dimensions

In this subsection, we investigate the problem of testing partial coefficients in a relatively high-dimensional regime. Specifically, the classical F-statistic designed for testing partial coefficients is studied. We derive the asymptotic normality and the asymptotic local power function of the proposed test.

Let $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$, where $\mathbf{x}_1$ is a p_1-dimensional nuisance covariate and $\mathbf{x}_2$ is a p_2-dimensional covariate of interest. SIM (1.1) is then denoted as

$$y = f(\mathbf{x}_1^\top \boldsymbol{\beta}_1 + \mathbf{x}_2^\top \boldsymbol{\beta}_2, \epsilon). \tag{C.1}$$

Suppose that $\mathbf{B}^\top$ is a $b \times p_2$ matrix of full row rank. We are interested in testing the following linear hypothesis problem

$$\mathbf{H}_{0,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{C.2}$$

When $\mathbf{B}^\top = \mathbf{I}$, the problem is converted to detecting the significance of $\mathbf{x}_2$. Therefore, hypothesis testing problem (C.2) is general. We rewrite the model to highlight the part of testing interest. Let $\mathbf{H}_\mathbf{B} = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top$, $\mathbf{S}_\mathbf{B}$ be a $p_2 \times b$ matrix such that $\mathbf{H}_\mathbf{B} = \mathbf{S}_\mathbf{B}\mathbf{S}_\mathbf{B}^\top$ and $\mathbf{S}_\mathbf{B}^\top\mathbf{S}_\mathbf{B} = \mathbf{I}$, and $\mathbf{S}_{\mathbf{B}^\perp}$ be a $p_2 \times (p_2 - b)$ matrix such that $\mathbf{I} - \mathbf{H}_\mathbf{B} = \mathbf{S}_{\mathbf{B}^\perp}\mathbf{S}_{\mathbf{B}^\perp}^\top$ and $\mathbf{S}_{\mathbf{B}^\perp}^\top\mathbf{S}_{\mathbf{B}^\perp} = \mathbf{I}$. The existences of $\mathbf{S}_\mathbf{B}$ and $\mathbf{S}_{\mathbf{B}^\perp}$ follow from the singular value decomposition of $\mathbf{B}$. Let $\mathbf{w} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top\mathbf{S}_{\mathbf{B}^\perp})^\top$ and $\boldsymbol{\gamma} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top\mathbf{S}_{\mathbf{B}^\perp})^\top$. Based on linear form (2.3), model (C.1) can be denoted as

$$y = c_0\mathbf{w}^\top\boldsymbol{\gamma} + c_0\mathbf{x}_2^\top\mathbf{S}_\mathbf{B}\mathbf{S}_\mathbf{B}^\top\boldsymbol{\beta}_2 + e, \tag{C.3}$$

where $\mathbf{w}$ is the nuisance covariate, $\mathbf{x}_2^\top\mathbf{S}_\mathbf{B}$ is the b-dimensional covariate of testing interest, and e is the residual term.

Suppose that observations $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$ are i.i.d. from SIM (C.1). Let $\mathbf{x}_{\ell i}$ be the ith row of the matrix $\mathbf{X}_\ell$, for $\ell = 1, 2$, and $\mathbf{y} = (y_1, ..., y_n)^\top$. Define $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2\mathbf{S}_{\mathbf{B}^\perp})$. The

F-statistic for testing problem (C.2) is defined as

$$T_{n,p_2} = \frac{\mathbf{y}^\top(\mathbf{H} - \mathbf{H}_{\mathbf{W}})\mathbf{y}/b}{\mathbf{y}^\top(\mathbf{I} - \mathbf{H})\mathbf{y}/(n-p)},$$

where $\mathbf{H}_{\mathbf{W}} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. The test statistic T_{n,p_2} is well-defined, since the matrix $\mathbf{W}^\top\mathbf{W}$ is invertible with the probability 1 when $p < n$ and $\mathbf{w}$ has the normal distribution.

C.1.1 Asymptotic Normality

First, we derive the asymptotic normality of T_{n,p_2} under $\mathbf{H}_{0,\mathbf{B}}$. The following assumptions are needed.

Assumption L5. *$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{x}$ is independent of ϵ.*

Assumption L6. *There are constants $\zeta, \zeta_1 \in (0,1)$, with $\zeta_1 < \zeta$, such that $\frac{p}{n} \to \zeta$ and $\frac{b}{n} \to \zeta_1$.*

From Assumption L6, the orders of b and p are asymptotically the same, indicating that linear hypothesis testing (C.2) is a high-dimensional testing problem. The asymptotic normality of the test statistic is shown as follows.

Theorem 3.3. *Under $\boldsymbol{H}_{0,\boldsymbol{B}}$ and Assumptions L2, L5 and L6, as $n \to \infty$, we have*

$$\frac{T_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This asymptotic normality result justifies the following test procedure. Given an α-level of significance, the test rejects $\mathbf{H}_{0,\mathbf{B}}$ if

$$\frac{T_{n,p_2} - 1}{\sqrt{2(1-\zeta+\zeta_1)/n\zeta_1(1-\zeta)}} > z_\alpha,$$

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

C.1.2 Asymptotic Local Power Function

Then, we move to investigate the asymptotic local power function of T_{n,p_2}. Let $\mathbf{D} = \text{diag}(\mathbf{I}, \mathbf{S}_{\mathbf{B}^\perp})$. Then $\mathbf{w} = \mathbf{D}^\top \mathbf{x}$. Let $\mathbf{\Sigma}_{22} = Var(\mathbf{x}_2)$, $\mathbf{\Sigma}_{12} = Cov(\mathbf{x}_1, \mathbf{x}_2)$ and $\mathbf{\Sigma}_{21} = Cov(\mathbf{x}_2, \mathbf{x}_1)$. Define

$$\tau^2 = \boldsymbol{\beta}_2^\top \mathbf{H}_{\mathbf{B}} \left[\mathbf{\Sigma}_{22} - (\mathbf{\Sigma}_{21}\ \mathbf{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \mathbf{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{22} \end{pmatrix} \right] \mathbf{H}_{\mathbf{B}} \boldsymbol{\beta}_2.$$

The scalar $c_0 = E(\boldsymbol{\beta}^\top \mathbf{x} y)/||\boldsymbol{\beta}||_2^2$. Additional assumption is needed for the study.

Assumption L7. $c_0^2 \tau^2 = o(1)$.

Theorem 3.4. *Suppose Assumptions L2, L5–L7 hold. Let $\Psi_n(\boldsymbol{\beta}_2; \boldsymbol{B})$ denote the power function of the test statistic T_{n,p_2}. As $n \to \infty$, we have*

$$\Psi_n(\boldsymbol{\beta}_2; \boldsymbol{B}) - \Phi(-z_\alpha + \sqrt{\frac{n(1-\zeta)(1-\zeta+\zeta_1)}{2\zeta_1}} \frac{c_0^2 \tau^2}{\sigma^2}) \to 0,$$

where $\sigma^2 = Var(y) - c_0^2 \boldsymbol{\beta}^\top \mathbf{\Sigma} \boldsymbol{\beta}$, Φ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

It indicates that $\Psi_n(\boldsymbol{\beta}_2; \mathbf{B})$ is an increasing function of $c_0^2 \tau^2$. In addition, when ζ increases, the loss of testing power demonstrates that the test is influenced by the effect of high dimensionality, and the test becomes powerless when ζ is close to 1.

C.2 New Partial Test in Ultrahigh Dimensions

In this subsection, we focus on the problem of testing partial regression coefficients in an ultrahigh-dimensional regime. Specifically, we propose a new testing procedure and derive its asymptotic null distribution and asymptotic local power function. While our theoretical analysis focuses on the normal distribution, the proposed testing procedure can also be applicable for general cases, based on the property of randomly projected data, which can be demonstrated by the simulation in Appendix D.5.

To emphasize the target model, SIM is restated as follows.

$$y = f(\mathbf{x}_1^\top \boldsymbol{\beta}_1 + \mathbf{x}_2^\top \boldsymbol{\beta}_2, \epsilon), \tag{C.4}$$

where $\mathbf{x}_1$ is a p_1-dimensional nuisance covariate and $\mathbf{x}_2$ is a p_2-dimensional covariate of interest. Suppose that $\mathbf{B}^\top$ is a $b \times p_2$ matrix of full row rank, we are interested in testing the linear hypothesis

$$\mathbf{H}_{0,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,\mathbf{B}} : \mathbf{B}^\top \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{C.5}$$

With the same definitions of $\mathbf{H}_\mathbf{B}$, $\mathbf{S}_\mathbf{B}$, $\mathbf{S}_{\mathbf{B}^\perp}$, $\mathbf{w}$ and $\boldsymbol{\gamma}$ in Section C.1, we obtain

$$\mathbf{x}_1^\top \boldsymbol{\beta}_1 + \mathbf{x}_2^\top \boldsymbol{\beta}_2 = \mathbf{w}^T \boldsymbol{\gamma} + \mathbf{x}_2^\top \mathbf{S}_\mathbf{B} \mathbf{S}_\mathbf{B}^\top \boldsymbol{\beta}_2.$$

For an integer $1 \leq k_2 < \min\{n + b - p, b\}$, let $\mathbf{P}_{k_2} \in \mathbb{R}^{b \times k_2}$ denote a random projection matrix with random entries, drawn independently of the data. Define $\mathbf{W} = (\mathbf{X}_1, \mathbf{X}_2 \mathbf{S}_{\mathbf{B}^\perp})$ and $\mathbf{U}_{k2} = (\mathbf{W}, \mathbf{X}_2 \mathbf{S}_\mathbf{B} \mathbf{P}_{k_2})$. Their projection matrices are given as

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

$$\mathbf{H}_{\mathbf{U}_{k_2}} = (\mathbf{I} - \mathbf{P}_1)\mathbf{U}_{k2}(\mathbf{U}_{k2}^\top (\mathbf{I} - \mathbf{P}_1)\mathbf{U}_{k2})^{-1}\mathbf{U}_{k2}^\top (\mathbf{I} - \mathbf{P}_1).$$

We propose a test statistic

$$T_{n,k_2} = \frac{\mathbf{y}^\top (\mathbf{H}_{\mathbf{U}_{k_2}} - \mathbf{H}_\mathbf{W})\mathbf{y}/k_2}{\mathbf{y}^\top (\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_{\mathbf{U}_{k_2}})\mathbf{y}/(n + b - p - k_2 - 1)}.$$

C.2.1 Asymptotic Normality

We first derive the asymptotic normality of the test statistic under $\mathbf{H}_{0,\mathbf{B}}$. The following assumption is made to facilitate our analysis.

Assumption H6. *$p = p_1 + p_2 \gg n$, $b \gg p_1$, and there are constants $\rho_1, \rho_2 \in (0,1)$, with $\rho_1 + \rho_2 < 1$, such that $\frac{p-b}{n} \to \rho_1$ and $\frac{k_2}{n} \to \rho_2$.*

Since T_{n,k_2} is invariant to the location shift of $\mathbf{y}$, $\mathbf{X}_1$ and $\mathbf{X}_2$, we assume $E(y) = 0$ and $E(\mathbf{x}) = \mathbf{0}$ in the following.

Theorem 4.5. *Suppose that Assumptions H1, H2 and H6 hold and $\boldsymbol{z}$ follows the standard normal distribution. Under $\boldsymbol{H}_{0,\boldsymbol{B}}$, as $n \to \infty$, we have*

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This asymptotic normality result justifies the following test procedure. Given a α-level of significance, $\mathbf{H}_{0,\mathbf{B}}$ is rejected when

$$\frac{T_{n,k_2} - 1}{\sqrt{2(1-\rho_1)/n\rho_2(1-\rho_1-\rho_2)}} > z_\alpha,$$

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

C.2.2 Asymptotic Local Power Function

We now investigate the asymptotic local power function of the test statistic. With the same definitions of $\mathbf{D}$, $\mathbf{\Sigma}_{22}$, $\mathbf{\Sigma}_{12}$, and $\mathbf{\Sigma}_{21}$ in Section C.1, we define $\mathbf{R}_{k_2} = (\mathbf{D}, (\mathbf{0}, \mathbf{P}_{k_2}^\top \mathbf{S}_{\mathbf{B}}^\top)^\top)$

and $\boldsymbol{\xi} = (\mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2})^{-1} \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$. Then $\mathbf{U}_{k2} = \mathbf{X}\mathbf{R}_{k_2}$. We divide the p-dimensional vector $\mathbf{R}_{k_2}\boldsymbol{\xi} = (\boldsymbol{\eta}_1^\top, \boldsymbol{\eta}_2^\top)^\top$, where $\boldsymbol{\eta}_1 \in \mathbb{R}^{p_1}$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^{p_2}$. Let

$$\nu^2 = \boldsymbol{\eta}_2^\top \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \boldsymbol{\eta}_2,$$

and $\tau^2 = \boldsymbol{\beta}_2^\top \mathbf{H}_\mathbf{B} \left[\boldsymbol{\Sigma}_{22} - (\boldsymbol{\Sigma}_{21}\ \boldsymbol{\Sigma}_{22})\, \mathbf{D}(\mathbf{D}^\top \boldsymbol{\Sigma} \mathbf{D})^{-1} \mathbf{D}^\top \begin{pmatrix} \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{22} \end{pmatrix} \right] \mathbf{H}_\mathbf{B} \boldsymbol{\beta}_2$. Conditional on $\mathbf{P}_{k_2}$, define $c_{0,k_2} = E(\boldsymbol{\xi}^\top \mathbf{R}_{k_2}^\top \mathbf{x} y)/\boldsymbol{\xi}^\top \mathbf{R}_{k_2}^\top \boldsymbol{\Sigma} \mathbf{R}_{k_2} \boldsymbol{\xi}$. Additional assumption is needed to facilitate the study.

Assumption H7. $(\boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta})^{-1} E(y\boldsymbol{x})^\top \boldsymbol{\Sigma}^{-1} E(y\boldsymbol{x}) \tau^2 = o(1)$.

This is known as a local alternative. In the linear model, Assumption H7 is converted to $\tau^2 = o(1)$. Considering a family of models where $E(y|\mathbf{x}) = g(\mathbf{x}^\top \boldsymbol{\beta})$ for a differentiable function g and $\mathbf{x}$ follows the normal distribution, Assumption H7 is converted to $E\{g'(\mathbf{x}^\top \boldsymbol{\beta})\}^2 \tau^2 = o(1)$.

Theorem 4.6. *Suppose that Assumptions H1, H2, H6 and H7 hold and $\boldsymbol{z}$ follows the standard normal distribution. Let $\Psi_n^{RP}(\boldsymbol{\beta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2})$ denote the power function of the proposed test T_{n,k_2}. As $n \to \infty$, we have*

$$\Psi_n^{RP}(\boldsymbol{\beta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2}) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2} \frac{c_{0,k_2}^2 \nu^2}{\sigma^2}} \right) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2 \boldsymbol{\xi}^\top \boldsymbol{R}_{k_2}^\top \boldsymbol{\Sigma} \boldsymbol{R}_{k_2} \boldsymbol{\xi}$, Φ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

From the result, the asymptotic local power function is an increasing function of $c_{0,k_2}^2\nu^2$. It is found that $\nu^2 \leq \tau^2$. We give a sufficient condition when the upper bound is asymptotically reached.

Assumption H8. *There is an integer $s_2 < k_2$ and a real number $\gamma_2 > 0$, such that $\frac{\sqrt{n}}{b}||\boldsymbol{S}_{\boldsymbol{B}}^\top\boldsymbol{\beta}_2||_2^2\sum_{i=s_2+1}^{b} d_i = o(n^{-\gamma_2})$, where d_i are the eigenvalues of $\boldsymbol{S}_{\boldsymbol{B}}^\top\boldsymbol{\Sigma}_{22}\boldsymbol{S}_{\boldsymbol{B}}$ satisfying $d_1 \geq d_2 \geq \cdots \geq d_b \geq 0$.*

This assumption ensures that Lemma 4.3 is valid for $\mathbf{S}_{\mathbf{B}}^\top\boldsymbol{\beta}_2$ and $\mathbf{S}_{\mathbf{B}}^\top\boldsymbol{\Sigma}_{22}\mathbf{S}_{\mathbf{B}}$. In this case, the asymptotic local power function of the proposed test statistic is shown as below.

Corollary 4.2. *Suppose that Assumptions H1, H2, H6, H7 and H8 hold and $\boldsymbol{z}$ follows the standard normal distribution. As $n \to \infty$, we have*

$$\Psi_n^{RP}(\boldsymbol{\beta}_2; \boldsymbol{B}, \boldsymbol{P}_{k_2}) - \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho_1-\rho_2)(1-\rho_1)}{2\rho_2}\frac{c_{0,k_2}^2\tau^2}{\sigma^2}}\right) \to 0,$$

where $\sigma^2 = Var(y) - c_{0,k_2}^2\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$, Φ is the cumulative distribution function of the standard normal distribution, and z_α is the upper α-quantile of Φ.

When $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\beta})$ for a differentiable function g, the scalar $c_{0,k_2} = E\{g^{'}(\mathbf{x}^\top\boldsymbol{\beta})\}$. Hence, it is proved that $\Psi_n^{RP}(\boldsymbol{\beta}_2; \mathbf{B}, \mathbf{P}_{k_2})$ can be a nonrandom function in some certain conditions, and it is a decreasing function of ρ_1.

C.3 Illustrative Examples

We give some examples to illustrate the forms of $c_{0,k}$ and σ^2. Let $\lambda^2 = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$ and $\mathbf{x}$ follow the normal distribution.

Example 1 (Linear model)**.** *Suppose the observation* $(\boldsymbol{x}, y)$ *is generated from*

$$y = \boldsymbol{x}^\top \boldsymbol{\beta} + \epsilon,\ \epsilon \textit{ is independent of } \boldsymbol{x}. \tag{C.6}$$

Then $c_{0,k} = 1$ *and* $\sigma^2 = Var(\epsilon)$.

Example 2 (Logistic model)**.** *Suppose the observation* $(\boldsymbol{x}, y)$ *is generated from*

$$y|\boldsymbol{x} \sim \textit{Bernoulli}\big(g(\boldsymbol{x}^\top \boldsymbol{\beta})\big),\ g(t) = \frac{\exp(t)}{1+\exp(t)}. \tag{C.7}$$

The scalar $c_{0,k}$ *and the variance* σ^2 *can be derived by*

$$c_{0,k} = E\{g^{'}(\boldsymbol{x}^\top \boldsymbol{\beta})\} \quad \textit{and}$$

$$\sigma^2 = Var(y) - c_{0,k}^2 \omega^2 = E\{g(\boldsymbol{x}^\top \boldsymbol{\beta})\}\left(1 - E\{g(\boldsymbol{x}^\top \boldsymbol{\beta})\}\right) - c_{0,k}^2 \omega^2.$$

This result is also available to the probit model, where the link function satisfies $g(t) = \Phi(t)$.

Example 3 (Poisson model)**.** *Suppose the observation* $(\boldsymbol{x}, y)$ *is generated from*

$$y|\boldsymbol{x} \sim \textit{Poisson}\big(g(\boldsymbol{x}^\top \boldsymbol{\beta})\big),\ g(t) = \exp(t).$$

The scalar $c_{0,k}$ *and the variance* σ^2 *can be derived by*

$$c_{0,k} = E\{g^{'}(\boldsymbol{x}^\top \boldsymbol{\beta})\} = \exp(0.5\lambda^2) \quad \textit{and}$$

$$\sigma^2 = Var(y) - c_{0,k}^2 \omega^2 = \exp(\lambda^2)\left(\exp(\lambda^2) + \exp(-0.5\lambda^2) - 1 - \omega^2\right),$$

where $c_{0,k}$ *is an increasing function of* λ^2.

Example 4 (Sin model)**.** *Suppose the response* y *is generated from*

$$y = \sin(\boldsymbol{x}^\top \boldsymbol{\beta}) + \epsilon,\ \epsilon \sim \mathcal{N}(0,1) \textit{ and is independent of } \boldsymbol{x}.$$

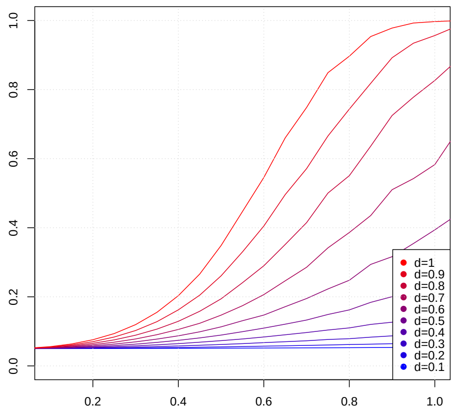
Then, we have

$$c_{0,k} = E\{\cos(\boldsymbol{x}^\top \boldsymbol{\beta})\} = \exp(-0.5\lambda^2) \quad \textit{and}$$

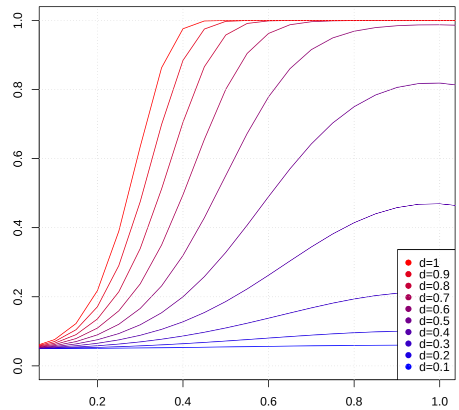
$$\sigma^2 = Var(y) - c_{0,k}^2\omega^2 = 1.5 - 0.5\exp(-2\lambda^2) - \omega^2\exp(-\lambda^2).$$

This indicates that $c_{0,k}$ is a decreasing function of λ^2.

Therefore, explicit expressions of the asymptotic local power functions can be derived in these models. In addition, it is found that the ratio ω^2/λ^2 has a significant influence on the power. The ratio is determined by $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and $\mathbf{P}_k$ and is in the range $[0, 1]$. As shown in Lemma 4.3, the ratio is close to 1 when Assumption H5 is satisfied by $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. We consider a simple situation where $\omega^2 = d^2\lambda^2$ for a specific value d, and we study the influence of the ratio ω^2/λ^2 on the asymptotic local power function. When $n = 600$ and $\rho = 0.4$, Figure 1 illustrates the asymptotic local power functions of the logistic and Poisson models for different values of d. It is found that, for a given λ^2, the largest value of the asymptotic local power can be derived when d is close to 1.



(a) Logistic model.



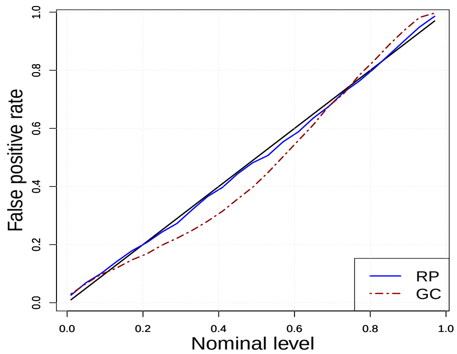
(b) Poisson model.

Figure 1: The asymptotic local power functions of logistic and Poisson models as λ^2 increases when $\omega^2 = d^2\lambda^2$ for a given d.

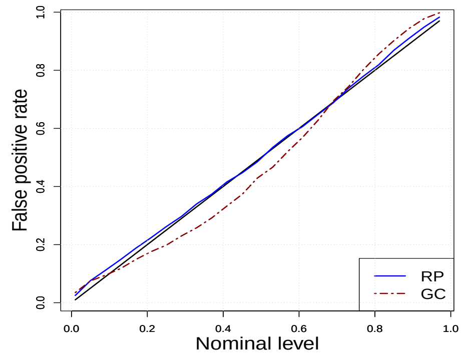
Appendix D

D.1 Numerical Comparison under Poisson Model

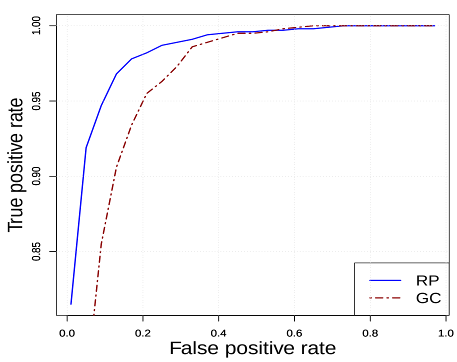
Figure 2 illustrates the numerical comparison between the RP and GC tests under the Poisson model when the nominal significance level varies from 0.1 to 0.9.



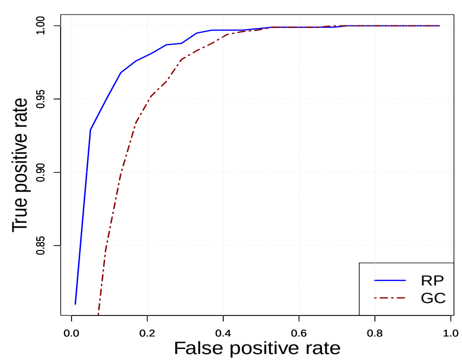
(a) The null hypothesis with $\mathbf{\Sigma}_1$.



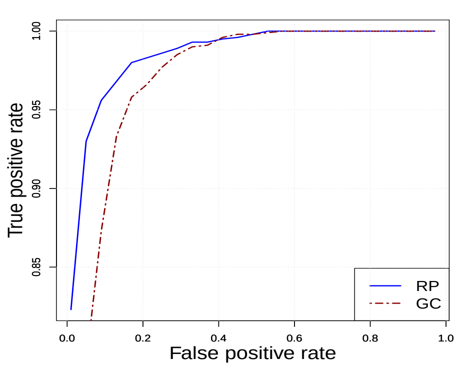
(b) The null hypothesis with $\mathbf{\Sigma}_2$.



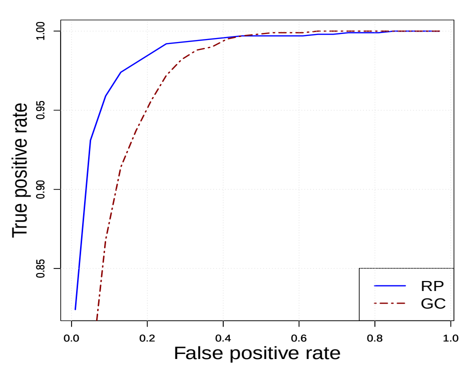
(c) Alternative $\boldsymbol{\delta}_1$ with $\mathbf{\Sigma}_1$.



(d) Alternative $\boldsymbol{\delta}_1$ with $\mathbf{\Sigma}_2$.



(e) Alternative $\boldsymbol{\delta}_2$ with $\mathbf{\Sigma}_1$.



(f) Alternative $\boldsymbol{\delta}_2$ with $\mathbf{\Sigma}_2$.

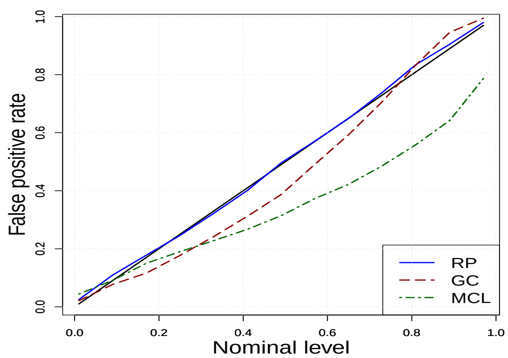
Figure 2: Type I errors and empirical powers of RP and GC tests for the Poisson model when $\rho = 0.4$ and $(n, p) = (400, 1000)$.

D.2 Additional Simulation Studies

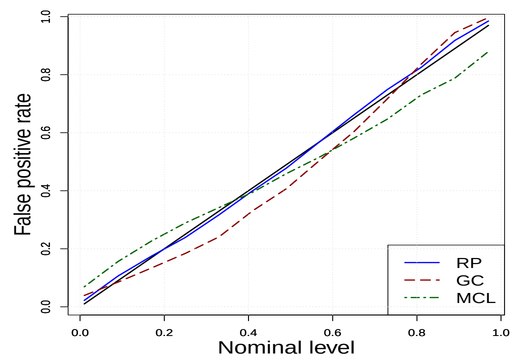
We conducted additional simulation studies to give more illustration of the performance of the proposed test. First, comparison of the RP test and the competing tests was conducted for other types of covariate. The covariate $\mathbf{x}$ was generated from $\mathbf{\Sigma}^{1/2}\mathbf{z}$. A mixture type of $\mathbf{z} = (z_1, \ldots, z_p)^\top$ was considered, with $z_i \overset{i.i.d.}{\sim} \sqrt{3/5}t(5)$ for $i = 1, \ldots, p/2$ and $z_i \overset{i.i.d.}{\sim}$ Rademacher distribution for $i = p/2+1, \ldots, p$. Following the design given in Section 5.1, three different settings of $\mathbf{\Sigma}$ were analyzed: (1)$\mathbf{\Sigma}_1$: $B = 1$. (2)$\mathbf{\Sigma}_2$: $B = 100$. (3)$\mathbf{\Sigma}_3$: $B = 5$. When $\mathbf{\Sigma} = \mathbf{\Sigma}_2$ or $\mathbf{\Sigma}_3$, the covariate $\mathbf{x}$ was a mixture of continuous and discrete variables. The alternative $\boldsymbol{\beta}$ followed the same setting in Section 5.1.

Table 1: Type I errors and empirical powers of RP, GC and MCL tests at the significance level 0.05 in logistic models with mixture type $\mathbf{z}$.

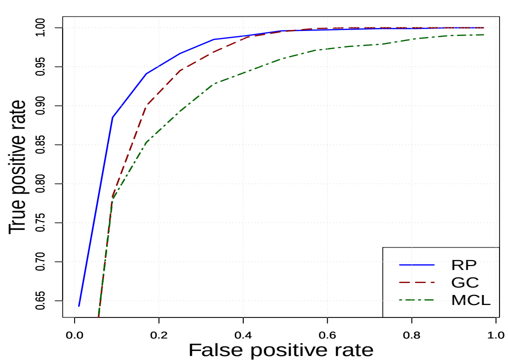
Type of $\mathbf{\Sigma}$	Type of $\boldsymbol{\beta}$	b^2	RP $\rho = 0.2$	RP $\rho = 0.4$	GC	MCL
Logistic model when $(n,p) = (400, 1000)$.						
$\mathbf{\Sigma}_1$	0	0	0.052	0.060	0.064	0.103
	$\boldsymbol{\delta}^1$	0.4	0.455	0.477	0.357	0.289
		0.8	0.774	0.838	0.731	0.581
	$\boldsymbol{\delta}^2$	0.4	0.439	0.461	0.275	0.137
		0.8	0.779	0.822	0.630	0.236
$\mathbf{\Sigma}_2$	0	0	0.053	0.059	0.054	0.072
	$\boldsymbol{\delta}^1$	0.4	0.440	0.469	0.324	0.281
		0.8	0.765	0.831	0.673	0.693
	$\boldsymbol{\delta}^2$	0.4	0.453	0.480	0.292	0.104
		0.8	0.798	0.830	0.616	0.170
$\mathbf{\Sigma}_3$	0	0	0.051	0.056	0.071	0.119
	$\boldsymbol{\delta}^1$	0.4	0.432	0.474	0.320	0.234
		0.8	0.757	0.830	0.669	0.427
	$\boldsymbol{\delta}^2$	0.4	0.467	0.482	0.272	0.161
		0.8	0.789	0.832	0.636	0.291



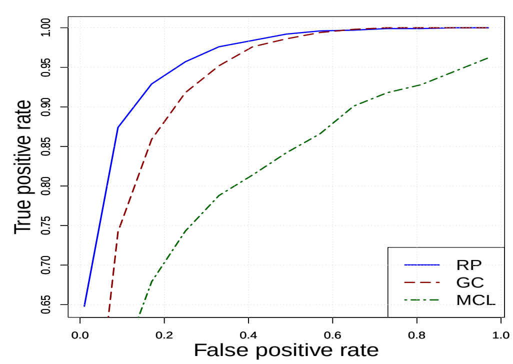
(a) The null hypothesis with $\mathbf{\Sigma}_2$.



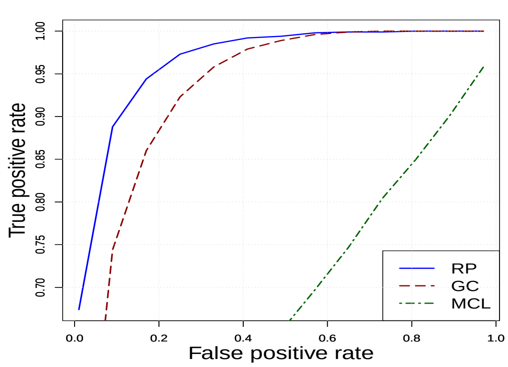
(b) The null hypothesis with $\mathbf{\Sigma}_3$.



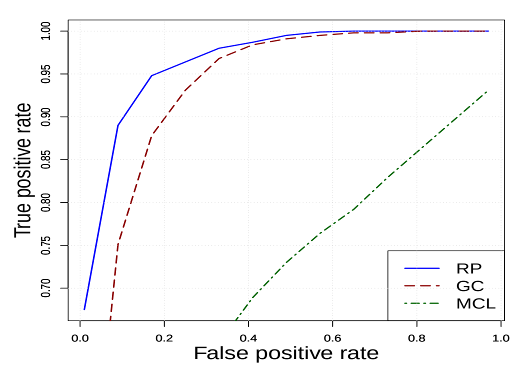
(c) Alternative $\boldsymbol{\delta}_1$ with $\mathbf{\Sigma}_2$.



(d) Alternative $\boldsymbol{\delta}_1$ with $\mathbf{\Sigma}_3$.



(e) Alternative $\boldsymbol{\delta}_2$ with $\mathbf{\Sigma}_2$.



(f) Alternative $\boldsymbol{\delta}_2$ with $\mathbf{\Sigma}_3$.

Figure 3: Type I errors and empirical powers of RP, GC, and MCL tests for the logistic model with mixture type $\mathbf{z}$ when $\rho = 0.4$ and $(n, p) = (400, 1000)$.

Table 1 reports the type I errors and the empirical powers of the RP, GC, and MCL tests in the logistic model based on 1000 simulations. It shows that the type I errors for the RP and GC tests are close to 0.05. And for the MCL test, when the simulation covariance matrix is set to $\mathbf{\Sigma}_1$ or $\mathbf{\Sigma}_3$, the type I errors exceed 0.1. For empirical powers, the proposed RP test show larger values in all the experimented settings. Therefore, the advantages of the proposed test is illustrated.

Figure 3 illustrates the numerical comparison between the RP test and the two competing tests when the nominal significance level varies from 0.1 to 0.9. As shown in the figures, the RP test has well a control of type I error and performs stronger power.

To illustrate the feasibility of the RP test in other nonlinear models, additional SIMs were analyzed:

(1) Model 1: $y = \sin(\mathbf{x}^\top \boldsymbol{\beta}) + \varepsilon$,

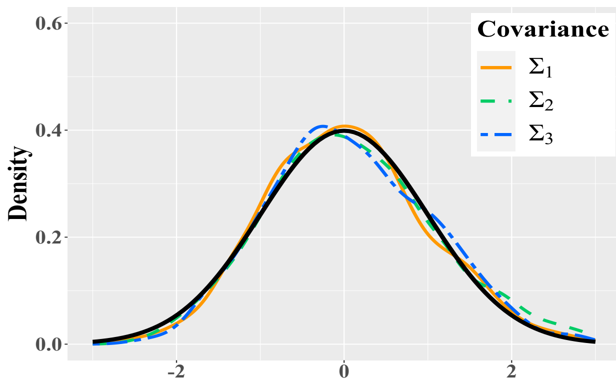
(2) Model 2: $y = \mathbf{x}^\top \beta + (\mathbf{x}^\top \beta)^2 + \varepsilon$,

(3) Model 3: $y = \sin(\mathbf{x}^\top \boldsymbol{\beta}) + \cos(\mathbf{x}^\top \boldsymbol{\beta}) + (\mathbf{x}^\top \beta)^2 + \varepsilon$.

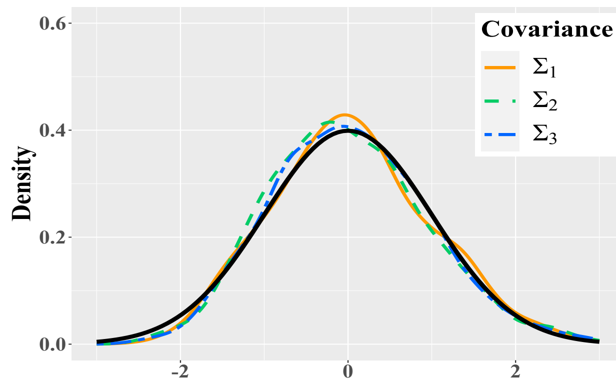
(4) Model 4: $y = \cos(\mathbf{x}^\top \boldsymbol{\beta}) + \varepsilon$.

(5) Model 5: $y = (\mathbf{x}^\top \boldsymbol{\beta})^2 + \varepsilon$.

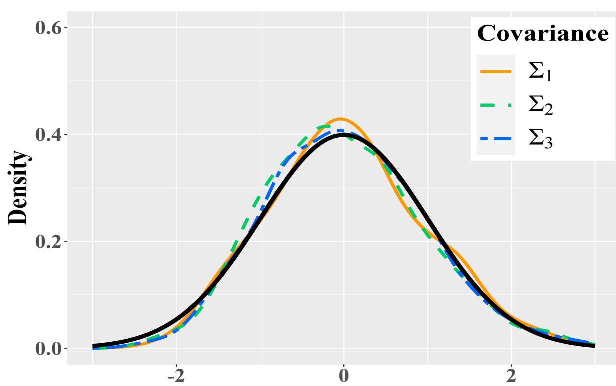
The error ε was independently generated from $\sqrt{3/5}t(5)$. The covariate $\mathbf{x}$ and alternative $\boldsymbol{\beta}$ were generated from the same setting as in the last simulation with different b^2 taken. The dimension (n, p) was set as $(400, 1000)$.



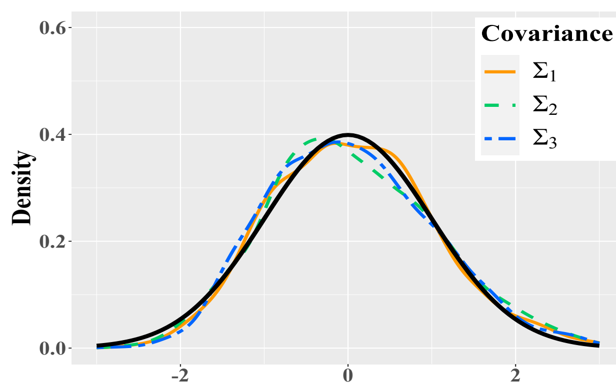
(a) Model 1.



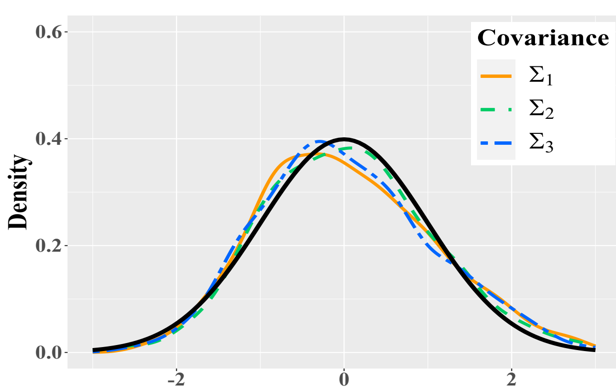
(b) Model 2.



(c) Model 3.



(d) Model 4.

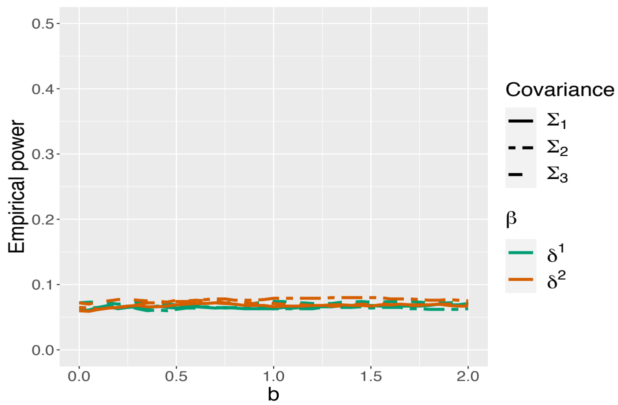


(e) Model 5.

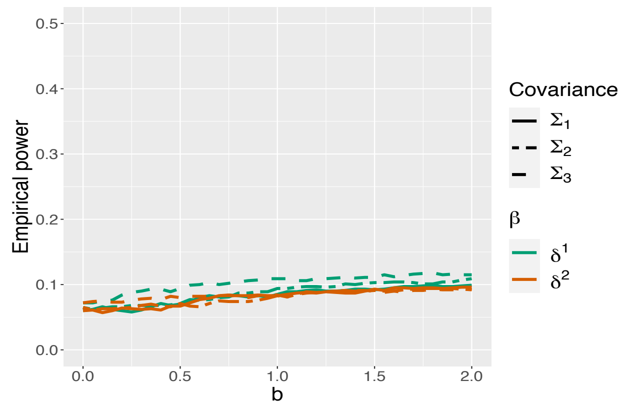
Figure 4: The kernel density estimation of the RP test with $\rho = 0.4$ under $\mathbf{H}_0$ for different nonlinear models.

Table 2: Type I errors and empirical powers of the RP test at the significance level 0.05 for models 1–3.

Model	Type of $\boldsymbol{\Sigma}$	Type of $\boldsymbol{\beta}$	b^2	RP $\rho = 0.2$	RP $\rho = 0.4$
Model 1	$\boldsymbol{\Sigma}_1$	0	0	0.050	0.059
		$\boldsymbol{\delta}^1$	0.1	0.501	0.451
			0.2	0.832	0.805
		$\boldsymbol{\delta}^2$	0.1	0.435	0.484
			0.2	0.762	0.794
	$\boldsymbol{\Sigma}_2$	0	0	0.053	0.057
		$\boldsymbol{\delta}^1$	0.1	0.443	0.460
			0.2	0.757	0.823
		$\boldsymbol{\delta}^2$	0.1	0.439	0.468
			0.2	0.756	0.813
	$\boldsymbol{\Sigma}_3$	0	0	0.060	0.059
		$\boldsymbol{\delta}^1$	0.1	0.516	0.454
			0.2	0.808	0.802
		$\boldsymbol{\delta}^2$	0.1	0.460	0.460
			0.2	0.749	0.829
Model 2	$\boldsymbol{\Sigma}_1$	0	0	0.050	0.057
		$\boldsymbol{\delta}^1$	0.1	0.536	0.507
			0.2	0.866	0.857
		$\boldsymbol{\delta}^2$	0.1	0.482	0.528
			0.2	0.824	0.855
	$\boldsymbol{\Sigma}_2$	0	0	0.053	0.057
		$\boldsymbol{\delta}^1$	0.1	0.480	0.504
			0.2	0.802	0.878
		$\boldsymbol{\delta}^2$	0.1	0.482	0.517
			0.2	0.814	0.855
	$\boldsymbol{\Sigma}_3$	0	0	0.058	0.057
		$\boldsymbol{\delta}^1$	0.1	0.546	0.503
			0.2	0.867	0.849
		$\boldsymbol{\delta}^2$	0.1	0.493	0.504
			0.2	0.821	0.881
Model 3	$\boldsymbol{\Sigma}_1$	0	0	0.050	0.057
		$\boldsymbol{\delta}^1$	0.1	0.495	0.462
			0.2	0.815	0.788
		$\boldsymbol{\delta}^2$	0.1	0.439	0.482
			0.2	0.749	0.785
	$\boldsymbol{\Sigma}_2$	0	0	0.053	0.057
		$\boldsymbol{\delta}^1$	0.1	0.439	0.458
			0.2	0.740	0.816
		$\boldsymbol{\delta}^2$	0.1	0.439	0.471
			0.2	0.733	0.787
	$\boldsymbol{\Sigma}_3$	0	0	0.058	0.057
		$\boldsymbol{\delta}^1$	0.1	0.509	0.463
			0.2	0.807	0.786
		$\boldsymbol{\delta}^2$	0.1	0.452	0.460
			0.2	0.746	0.814



(a) Model 4.



(b) Model 5.

Figure 5: The empirical power of the RP test for models 4 and 5.

Figure 4 reports the kernel density estimation of the RP test statistic under $\mathbf{H}_0$ for different nonlinear models. In each figure, the lines in orange, green and blue represent the kernel density estimations in the simulation setting with $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$ and $\mathbf{\Sigma}_3$, respectively. The black line represents the density of the standard normal distribution. The lines in orange, green and blue are shown to be close to the black line. Therefore, the asymptotic null distribution of the RP test statistic can be well approximated by the standard normal distribution, which confirms the result of Theorem 4.1.

Table 2 reports the type I errors and the empirical powers of the RP test under models 1–3 based on 1000 simulations. It shows that the type I errors are around 0.05 and the empirical powers increase as b^2 grows.

Figure 5 displays the empirical power of the proposed RP test under models 4 and 5. For a specific model, when the types of $\mathbf{\Sigma}$ and $\boldsymbol{\beta}$ vary, the performances of the empirical power have negligible differences. However, as b increases, it approximates a horizon line, which indicates that the proposed RP test has imperceptible power under the alternative.

To explore the reason, we recall the asymptotic local power function $\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k)$ shown in Theorem 4.2. It satisfies

$$\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k) = \Phi\left(-z_\alpha + \sqrt{\frac{n(1-\rho)}{2\rho}\frac{c_{0,k}^2\omega^2}{\sigma^2}}\right) + o(1), \tag{D.1}$$

where $\sigma^2 = Var(y) - c_{0,k}^2\omega^2$. We consider a specific case where the model satisfies $E(y|\mathbf{x}) = g(\mathbf{x}^\top\boldsymbol{\beta})$ for a differentiable function g while $\mathbf{x}$ follows the normal distribution with mean zero. By Stein's lemma, the coefficient $c_{0,k}$ satisfies $c_{0,k} = E[g'(\mathbf{x}^\top\boldsymbol{\beta})]$. Note that $c_{0,k} = 0$ when g' is an odd function. Plugging this into equation (D.1), we have $\Psi_n^{RP}(\boldsymbol{\beta}; \mathbf{P}_k) = \Phi(-z_\alpha) + o(1)$, which means the proposed test will show a negligible power under the alternative. The analysis above offers certain explanation for the loss of power of the proposed test under models 4 and 5, since the derivatives of both cosine and square link functions are odd. Such a problem is also met in Neykov et al. (2016).

Following the suggestion in Neykov et al. (2016), a possible strategy is provided to avoid such problem. Suppose $\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta} \neq 0$, by (2.3) on page 8, we have

$$E[y\mathbf{x}^\top\boldsymbol{\beta}] = c_0\boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}.$$

Therefore, $c_0 = 0$ if and only if $E[y\mathbf{x}^\top\boldsymbol{\beta}] = 0$. When $c_0 = 0$, we consider a transformed response $\tilde{y} = \psi(y)$ by a function ψ such that $E[\tilde{y}\mathbf{x}^\top\boldsymbol{\beta}] \neq 0$. The transformation exists if and only if $E[\mathbf{x}^\top\boldsymbol{\beta}|y] \neq 0$. The verification of the statement is given as below. For a measurable function ψ,

$$E[\psi(y)\mathbf{x}^\top\boldsymbol{\beta}] = E\big\{\psi(y)E[\mathbf{x}^\top\boldsymbol{\beta}|y]\big\}.$$

When $E[\mathbf{x}^\top \boldsymbol{\beta}|y] = 0$, it follows that $E[\psi(y)\mathbf{x}^\top \boldsymbol{\beta}] = 0$ for any measure function ψ. When $E[\mathbf{x}^\top \boldsymbol{\beta}|y] \neq 0$, by taking $\psi(y) = E[\mathbf{x}^\top \boldsymbol{\beta}|y]$, the existence of the transformation is demonstrated. Therefore, when $E[\mathbf{x}^\top \boldsymbol{\beta}|y] \neq 0$, it is possible for us to avoid the problem by transforming the response. However, how to search for such transformation in practice remains a problem, which needs to be explored in the further research work.

Another possible method to overcome the problem is to use the strategy proposed in Cannings and Samworth (2017) and Shi, Lu, and Song (2020) as suggested by an anonymous reviewer. Specifically, multiple sparse random projections could be generated independently and the sparse random projection that maximizes certain criterion may be adaptively selected to construct the test statistic.

D.3 Multiple-Random-Projection-Based Tests

Tests based on different types of multiple random projection were considered:

- $\widetilde{T}^{D}_{n,k}$: independently generate the normal random projection D times and calculate the hat matrix based on each random projection. The value of the test is the value of $T_{n,k}$ based on the average of the hat matrix.

- $\bar{T}^{D}_{n,k}$: independently generate the normal random projection D times and calculate $T_{n,k}$ for each random projection. The value of the test is the average of these $T_{n,k}$.

Letting half of the random projection matrix be generated from the normal random projection and the other half be generated from the sparse random projection as defined in (4.4) with $\ell = 400$, we defined $\widetilde{T}^{\mathrm{mix},D}_{n,k}$ and $\bar{T}^{\mathrm{mix},D}_{n,k}$ following a similar structure of $\widetilde{T}^{D}_{n,k}$ and $\bar{T}^{D}_{n,k}$, respectively. We also experimented with the quantile aggregation method introduced in Meinshausen et al. (2009). Specifically, we used the following procedure.

1. For $d = 1, \dots, D$:
   (a) Randomly generate the random projection matrix $P_k^{(d)}$;
   (b) Calculate the proposed test statistic $T_{n,k}^{(d)}$ based on $P_k^{(d)}$;
   (c) Calculate the p-value $P^{(d)}$ using the normal approximation.
2. Aggregate over the p-values $\{P^{(d)}\}_{d=1}^{D}$. For $\gamma \in (0, 1)$, define

$$Q(\gamma) = \min\{1, q_\gamma(\{P^{(d)}/\gamma : d = 1, \dots, D\})\},$$

   where $q_\gamma(\cdot)$ is the (empirical) γ-quantile function. The proposed p-value is determined by the selection of γ. Specifically, the p-value P_{fixed}^D determined by fixed γ and P_{adap}^D determined by adaptive γ are introduced as below.
   (a) Fixed γ:

$$P_{\text{fixed}}^D = Q(0.5),$$

   where $\gamma = 0.5$ is considered.
   (b) Adaptive γ :

$$P_{\text{adap}}^D = \min\{1, (1 - \log \gamma_{\min}) \inf_{\gamma \in (\gamma_{\min}, 1)} Q(\gamma)\},$$

   where the recommended value $\gamma_{\min} = 0.05$ is taken in the experiment.
3. For a given α-level of significance, reject the null hypothesis if p-value is smaller than α.

We investigated the performances of the tests in the setting of Appendix D.2, where the logistic model with $\mathbf{\Sigma} = \mathbf{\Sigma}_1$ was considered. For a positive integer D, the results derived by the quantile aggregation method with fixed and adaptive γ were denoted by A^D_{fixed} and A^D_{adap}, respectively. Different numbers of random projections were experimented, such as $D = 10, 20, 50$ and 100.

Table 3: Type I errors and empirical powers of the multiple-random-projection-based tests at the significance level 0.05 for $D = 10, 20, 50, 100$.

Type of $\boldsymbol{\beta}$	b^2	$\widetilde{T}^D_{n,k}$				$\bar{T}^D_{n,k}$			
		10	20	50	100	10	20	50	100
0	0	0.055	0.054	0.050	0.051	0.055	0.055	0.053	0.053
	0.4	0.445	0.463	0.456	0.458	0.452	0.466	0.463	0.462
$\boldsymbol{\delta}^1$	0.8	0.864	0.861	0.869	0.867	0.866	0.865	0.870	0.870
	0.4	0.464	0.463	0.462	0.459	0.467	0.467	0.467	0.465
$\boldsymbol{\delta}^2$	0.8	0.847	0.853	0.854	0.850	0.850	0.856	0.856	0.854
Type of $\boldsymbol{\beta}$	b^2	$\widetilde{T}^{\text{mix},D}_{n,k}$				$\bar{T}^{\text{mix},D}_{n,k}$			
		10	20	50	100	10	20	50	100
0	0	0.051	0.049	0.054	0.054	0.052	0.052	0.055	0.056
	0.4	0.477	0.475	0.474	0.476	0.479	0.480	0.482	0.484
$\boldsymbol{\delta}^1$	0.8	0.846	0.853	0.848	0.850	0.847	0.856	0.852	0.853
	0.4	0.462	0.459	0.462	0.465	0.466	0.464	0.469	0.469
$\boldsymbol{\delta}^2$	0.8	0.853	0.850	0.843	0.844	0.857	0.855	0.846	0.849
Type of $\boldsymbol{\beta}$	b^2	A^D_{fixed}				A^D_{adap}			
		10	20	50	100	10	20	50	100
0	0	0.035	0.029	0.031	0.030	0.027	0.029	0.021	0.021
	0.4	0.369	0.346	0.334	0.330	0.305	0.319	0.264	0.264
$\boldsymbol{\delta}^1$	0.8	0.791	0.784	0.784	0.784	0.740	0.747	0.709	0.719
	0.4	0.380	0.363	0.358	0.358	0.324	0.339	0.293	0.289
$\boldsymbol{\delta}^2$	0.8	0.787	0.781	0.780	0.778	0.742	0.758	0.715	0.722

The simulation result is reported in Table 3. The tests $\widetilde{T}^D_{n,k}$, $\bar{T}^D_{n,k}$, $\widetilde{T}^{\text{mix},D}_{n,k}$, and $\bar{T}^{\text{mix},D}_{n,k}$ are based on certain average aggregation methods. As D increases, there are negligible differences between the average based tests in terms of type I error and empirical power.

Therefore, for the random projections varying in type, number and combination in these experiments, the performances of the average based tests are similar. For the quantile aggregation based methods, as D increases, there is negligible difference between A^D_{fixed} and A^D_{adap}. The type I errors of A^D_{fixed} and A^D_{adap} are smaller than the given significance level 0.05. Compare the two aggregation methods, the average based tests show a slightly stronger empirical power.

D.4 Classical Tests Based on Random Projection

To study classical tests based on random projection data, we consider the logistic regression model, where

$$y|\mathbf{x} \sim \text{Bernoulli}(g(\mathbf{x}^\top \boldsymbol{\beta})) \text{ with } g(t) = \frac{\exp(t)}{1+\exp(t)}.$$

The log-likelihood function is

$$\ell_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \mathbf{x}_i^\top \boldsymbol{\beta} - \log(1 + \exp(\mathbf{x}_i^\top \boldsymbol{\beta})),$$

the score function is

$$\dot{\ell}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ y_i - \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{1+\exp(\mathbf{x}_i^\top \boldsymbol{\beta})} \right\} \mathbf{x}_i,$$

and the estimator for the information matrix is

$$\hat{\mathbf{I}}_{\boldsymbol{\beta}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\exp(\mathbf{x}_i^\top \boldsymbol{\beta})}{(1+\exp(\mathbf{x}_i^\top \boldsymbol{\beta}))^2} \mathbf{x}_i \mathbf{x}_i^\top = \frac{1}{n} \mathbf{X}^\top \mathbf{W} \mathbf{X},$$

where $\mathbf{W} = \text{diag}\left(\frac{\exp(\mathbf{x}_1^\top \boldsymbol{\beta})}{(1+\exp(\mathbf{x}_1^\top \boldsymbol{\beta}))^2}, \ldots, \frac{\exp(\mathbf{x}_n^\top \boldsymbol{\beta})}{(1+\exp(\mathbf{x}_n^\top \boldsymbol{\beta}))^2} \right)$. For the global testing problem (1.2), the likelihood ratio test statistic (LRT_n), Wald test statistic (W_n), and score test statistic (S_n)

are defined as

$$LRT_n = -2\{\ell_n(\mathbf{0}) - \ell_n(\hat{\boldsymbol{\beta}})\},$$

$$W_n = n\hat{\boldsymbol{\beta}}^\top \hat{\mathbf{I}}_{\hat{\boldsymbol{\beta}}}\hat{\boldsymbol{\beta}},$$

$$S_n = \frac{1}{n}\dot{\ell}_n(\mathbf{0})^\top \hat{\mathbf{I}}_{\mathbf{0}}^{-1}\dot{\ell}_n(\mathbf{0}).$$

From Wilk's theorem, the likelihood ratio test statistic satisfies $LRT_n \xrightarrow{\mathcal{D}} \chi_p^2$. Here χ_p^2 is the chi-squared distribution with p degrees of freedom. In addition, LRT_n is asymptotically equivalent to both W_n and S_n under the null hypothesis $\boldsymbol{\beta} = \mathbf{0}$ (van der Vaart , 1998, Chapter 16.2).

However, when p grows comparably with n, Wilk's theorem might fail, which is confirmed by Sur, Chen, and Candès (2019). Here we conducted a simple experiment to study the accuracy of the chi-squared approximation. Specifically, we used LRT_n, W_n and S_n to test $\boldsymbol{\beta} = \mathbf{0}$ versus $\boldsymbol{\beta} \neq \mathbf{0}$. In each trail, $n = 1000$ observations were generated with $y_i \overset{i.i.d.}{\sim}$ Bernoulli$(1/2)$ and $\mathbf{x}_i \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$. The dimension of the covariate $\mathbf{x}$ was set as $p = 1, 5, 10, 50, 100, 200$. Figure 6 reports the kernel density estimations of p-values of the three tests under different covariate dimension p based on 40000 simulations. If the chi-squared approximation is true, it is expected to observe uniformly distributed p-values, which is the case when $p = 1, 5, 10$. However, when $p = 50, 100, 200$, the distribution of p-value is far from the uniform distribution. This indicates the inadequacy of Wilks' theorem in high-dimensional setting. Let $LRT_{n,k}$, $W_{n,k}$, and $S_{n,k}$ denote the likelihood ratio test statistic, Wald test statistic, and score test statistic that are constructed by the randomly projected data of dimension k. Note $k = \rho n$ is comparable with n. The null distribution of $S_{n,k}$,

$LRT_{n,k}$, and $W_{n,k}$ will be far from χ^2_k.

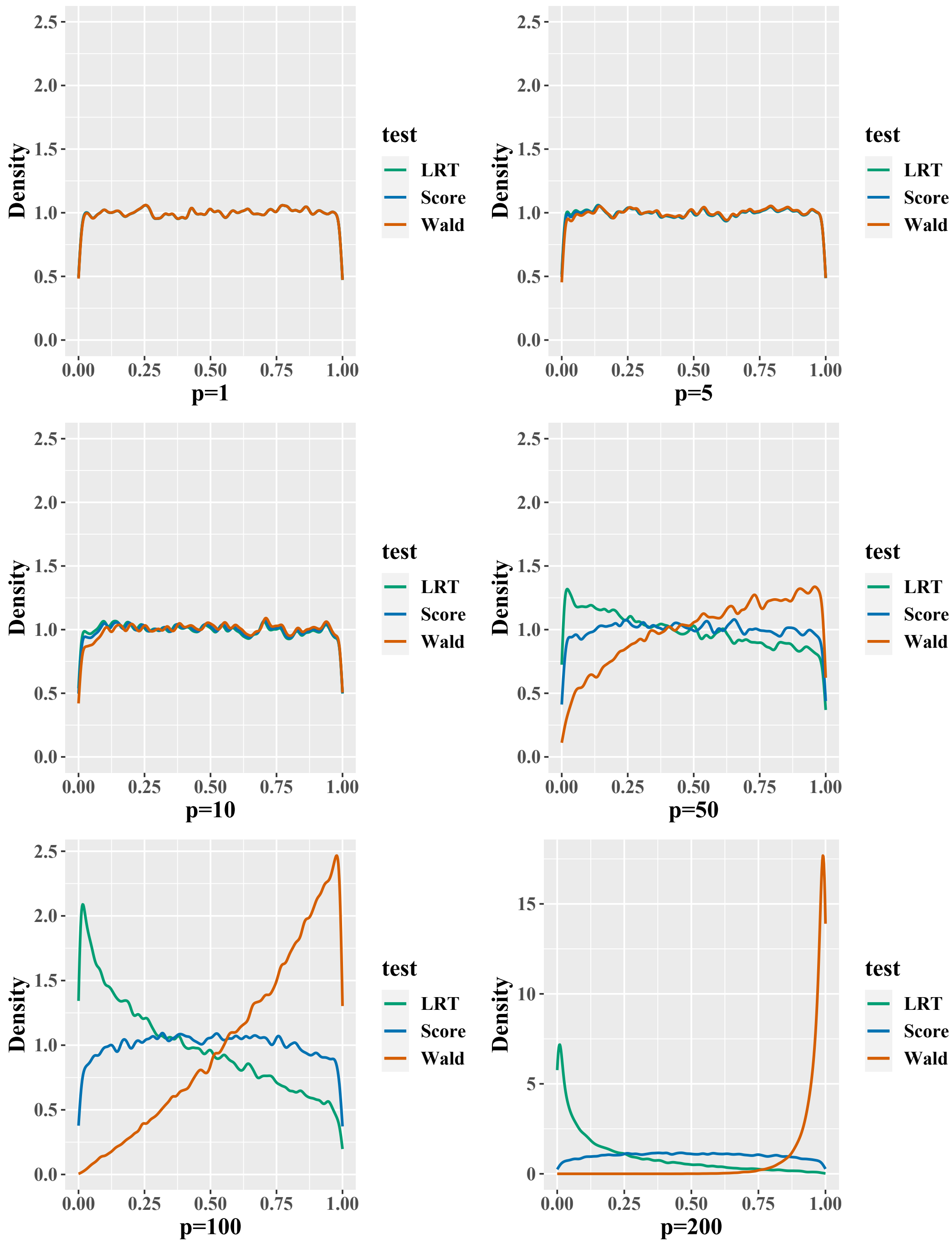


Figure 6: Kernel density estimations of p-values for LRT_n, W_n and S_n tests under the logistic regression with p =1,5,10,50,100 or 200.

We investigate the tests under the setting in Appendix D.2 with $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_2$ and $\rho = 0.4$. As shown in Table 4, the type I error of $LRT_{n,k}$ exceeds 0.8 from the nominal level of 0.05. Therefore, $LRT_{n,k}$ is not a valid test in this setting. On the other hand, the type I errors of $W_{n,k}$ and $S_{n,k}$ are much smaller than 0.05. Moreover, both $W_{n,k}$ and $S_{n,k}$ have lower power than the proposed test in all the settings.

Table 4: Type I errors and empirical powers of RP, $LRT_{n,k}$, $W_{n,k}$ and $S_{n,k}$ tests at the significance level 0.05 when $(n, p) = (400, 1000)$.

Type of $\boldsymbol{\beta}$	b^2	RP	$LRT_{n,k}$	$W_{n,k}$	$S_{n,k}$
0	0	0.059	0.830	0.000	0.021
$\boldsymbol{\delta}^1$	0.4	0.469	0.993	0.000	0.227
	0.8	0.831	0.973	0.026	0.581
$\boldsymbol{\delta}^2$	0.4	0.480	0.987	0.003	0.214
	0.8	0.830	0.981	0.019	0.613

D.5 Simulation Studies for Partial Test

This simulation study was designed for testing the hypothesis

$$\mathbf{H}_{0,p_2} : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad \mathbf{H}_{1,p_2} : \boldsymbol{\beta}_2 \neq \mathbf{0}.$$

The logistic model was analyzed and the covariate $\mathbf{x}$ was generated from $\boldsymbol{\Sigma}^{1/2}\mathbf{z}$. Each entry of $\mathbf{z}$ was i.i.d. from $\mathcal{N}(0,1)$, $U(-\sqrt{3}, \sqrt{3})$, or Rademacher distribution. A specific type of $\boldsymbol{\Sigma}^{1/2}$ was considered. We generated the matrix from $\boldsymbol{\Sigma}^{1/2} = \operatorname{diag}\big(\mathbf{O}_1\sqrt{\mathbf{D}_1}\mathbf{O}_1^\top, \mathbf{O}_2\sqrt{\mathbf{D}_2}\mathbf{O}_2^\top\big)$, where $\mathbf{O}_1$ ($\mathbf{O}_2$) was an orthogonal matrix that was generated from the same way as the matrix $\mathbf{O}$ in Section 5.1, when p took p_1 (p_2) and $B = 1$ ($B = 100$). The entries of diagonal matrix $\mathbf{D}_1$ were from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ with absolute values taken and the entries of diagonal matrix

$\mathbf{D}_2$ were generated in the same way as that in Section 5.1, when $s = [n^{0.72}]$ and $L = [n^{0.8}]$. In the simulation, we considered $(n, p_1, p_2) = (400, 40, 1000)$.

For the alternative, we generated $\boldsymbol{\beta}_1$ from $\mathcal{N}(\mathbf{0}, \mathbf{I})$ and then scaled it to have $\|\boldsymbol{\beta}_1\|_2 = 1$. The vector of coefficients $\boldsymbol{\beta}_2$ was generated from $\boldsymbol{\beta}_2 = b_2\boldsymbol{\delta}_2/\sqrt{\boldsymbol{\delta}_2^\top \boldsymbol{\Sigma}_{22}\boldsymbol{\delta}_2}$, where b_2^2 took values $0.4, 0.8$ in the simulation and $\boldsymbol{\delta}_2$ was a p_2-dimensional vector determining the sparsity level of $\boldsymbol{\beta}_2$. Two types of $\boldsymbol{\delta}_2$ were analyzed as follows: (i) $\boldsymbol{\delta}_2^1$: $\boldsymbol{\delta}_{2j}^1 = 1$, for $j \in S$, where the set S was randomly selected over $\{1, \ldots, p_2\}$ and had size $|S| = 10$, otherwise, $\boldsymbol{\delta}_{2j}^1 = 0$. (ii) $\boldsymbol{\delta}_2^2$: randomly selected from the space that was spanned by the eigenvectors corresponding to the largest 100 eigenvalues of $\boldsymbol{\Sigma}_{22}$.

To demonstrate the feasibility of implementing other random projection, we investigated two other random-projection-based tests: (i) multi-RP test: independently generating normal random projection for 10 times and utilizing their mean; (ii) S-RP test: applying sparse random projection defined in (4.4) with $l = 400$.

Table 5 reports the type I errors and empirical powers of the random-projection-based tests based on 1000 simulations. The type I errors are close to 0.05, and the performances of the three tests have negligible differences, indicating the feasible implementing of other random projection. For the both alternatives, the empirical power increases as b_2^2 grows. And larger empirical powers are shown with smaller ρ_2, which is consistent with the result in Corollary 4.2. When $\mathbf{z}$ belongs to other types of distribution rather than the normal distribution, the results are similar to the one when $\mathbf{z}$ is normal. Therefore, it is possible for the proposed method to be applied in general situations while controlling the type I error well. And for alternatives, the results indicate that the derived asymptotic power function might also be valid for non-normal distribution.

Table 5: Type I errors and empirical powers of RP, multi-RP and S-RP tests at the significance level 0.05 when $(n, p_1, p_2) = (400, 40, 1000)$.

Type of $\mathbf{z}$	Type of $\boldsymbol{\beta}_2$	b_2^2	$\rho_2 = 0.2$			$\rho_2 = 0.4$		
			RP	multi-RP	S-RP	RP	multi-RP	S-RP
$\mathcal{N}(0,1)$	-	0	0.053	0.057	0.054	0.062	0.064	0.050
	$\boldsymbol{\delta}_2^1$	0.4	0.575	0.558	0.518	0.357	0.389	0.366
		0.8	0.906	0.905	0.870	0.716	0.700	0.686
	$\boldsymbol{\delta}_2^2$	0.4	0.559	0.553	0.548	0.351	0.349	0.354
		0.8	0.905	0.893	0.887	0.683	0.668	0.687
$U(-\sqrt{3}, \sqrt{3})$	-	0	0.061	0.056	0.050	0.060	0.053	0.052
	$\boldsymbol{\delta}_2^1$	0.4	0.554	0.554	0.539	0.387	0.368	0.371
		0.8	0.892	0.894	0.857	0.683	0.694	0.691
	$\boldsymbol{\delta}_2^2$	0.4	0.527	0.532	0.517	0.336	0.366	0.336
		0.8	0.904	0.901	0.890	0.701	0.710	0.689
Rademacher	-	0	0.052	0.050	0.058	0.060	0.056	0.057
	$\boldsymbol{\delta}_2^1$	0.4	0.555	0.548	0.544	0.400	0.362	0.365
		0.8	0.901	0.897	0.875	0.706	0.716	0.711
	$\boldsymbol{\delta}_2^2$	0.4	0.556	0.558	0.541	0.328	0.358	0.342
		0.8	0.909	0.903	0.885	0.692	0.700	0.713

D.6 Testing Individual Coefficients under the Global Hypothesis

In this section, we consider testing the global null hypothesis by simultaneously testing each element of $\boldsymbol{\beta}$. Without loss of generality, we consider the first element of $\boldsymbol{\beta}$, which is denoted by β_1. Correspondingly, the design matrix is divided as $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$, where $\mathbf{X}_1 \in \mathbb{R}^{n\times 1}$ and $\mathbf{X}_2 \in \mathbb{R}^{n\times(p-1)}$. We propose a random-projection-based test to test the significance of β_1. For an integer $1 \leq k \leq \min\{n-3, p-1\}$, let $\mathbf{P}_k \in \mathbb{R}^{(p-1)\times k}$ denote a random projection matrix with random entries, drawn independently of the data. Denote the projected data by $\mathbf{X}_{2,k} = \mathbf{X}_2\mathbf{P}_k$ and $\mathbf{U}_k = (\mathbf{X}_1, \mathbf{X}_{2,k})$, and let the projection matrices be denoted by

$$\mathbf{H}_{2,k} = (\mathbf{I} - \mathbf{P}_1)\mathbf{X}_{2,k}(\mathbf{X}_{2,k}^\top(\mathbf{I} - \mathbf{P}_1)\mathbf{X}_{2,k})^{-1}\mathbf{X}_{2,k}^\top(\mathbf{I} - \mathbf{P}_1),$$

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

The proposed test statistic is defined by

$$T_{1,n,k} = \frac{\mathbf{y}^\top(\mathbf{H}_k - \mathbf{H}_{2,k})\mathbf{y}}{\mathbf{y}^\top(\mathbf{I} - \mathbf{P}_1 - \mathbf{H}_k)\mathbf{y}/(n-2-k)}.$$

Based on the observed $T_{1,n,k}$, the p-value P_F is introduced as below. Let $F(d_1, d_2)$ denote the variable following the F-distribution with degrees of freedom (d_1, d_2). Let $d_1 = 1$ and $d_2 = n - 2 - k$. The p-value P_F based on the F-approximation is defined by

$$P_F = P(F(d_1, d_2) > T_{1,n,k}).$$

For a given α-level of significance, β_1 is reported to be significant if P_F is smaller than α.

To investigate the feasibility of P_F, we conducted the simulation study. The logistic model, Poisson model, and model 1 were considered. The covariate $\mathbf{x}$ was generated in the same way as the method in Appendix D.2. The normal random projection matrix was used and the projection dimension k was determined by $k = \rho_2 n$. Different values of ρ_2 were considered, such as $\rho_2 = 0.2, 0.4$, and 0.6. The dimension was set as $(n, p) = (400, 1000)$.

In Table 6, we report the type I error of the proposed test based on 1000 simulations. For all the experimented settings, the proposed test shows a well control of type I error. Therefore, the feasibility of the proposed test is demonstrated.

Table 6: Type I error of the proposed test at the significance level 0.05 for different models.

ρ_2	Logistic model			Poisson model			Model 1		
	$\mathbf{\Sigma}_1$	$\mathbf{\Sigma}_2$	$\mathbf{\Sigma}_3$	$\mathbf{\Sigma}_1$	$\mathbf{\Sigma}_2$	$\mathbf{\Sigma}_3$	$\mathbf{\Sigma}_1$	$\mathbf{\Sigma}_2$	$\mathbf{\Sigma}_3$
0.2	0.047	0.045	0.039	0.051	0.053	0.060	0.042	0.047	0.043
0.4	0.044	0.058	0.056	0.037	0.049	0.044	0.043	0.050	0.052
0.6	0.048	0.041	0.053	0.041	0.049	0.050	0.049	0.048	0.043

Based on the proposed testing procedure, the p-value for testing every single element of $\boldsymbol{\beta}$ can be derived. Therefore, the simultaneous inference via Bonferroni adjustment can be applied to test the global null hypothesis.

References

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.

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.

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

Cannings, T. I. and Samworth, R. J. (2017). Random-projection ensemble classification. *Journal of the Royal Statistical Society. Series B. Statistical Methodology* **79**, 959–1035.

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.

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.

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

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

Neykov, M., Liu, J. S. and Cai, T. (2016). L_1-regularized least squares for support recovery of high dimensional single index models with Gaussian designs. *The Journal of Machine Learning Research* **17**, 2976–3012.

Shi, C., Lu, W. and Song, R. (2020). A sparse random projection-based test for overall qualitative treatment effects. *Journal of the American Statistical Association* **115**, 1201–1213.

Sur, P., Chen, Y. and Candès, E. J. (2019). The likelihood ratio test in high-dimensional logistic regression is asymptotically a *rescaled* chi-square. *Probability Theory and Related Fields* **175**, 487–558.

van der Vaart, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge.