Hypothesis Testing in High-Dimensional Instrumental Variables Regression with an Application to Genomics Data

Jiarui Lu and Hongzhe Li

University of Pennsylvania

Supplementary Material

S1 Detailed conditions for estimation bounds

Before stating the results, we first introduce some assumptions. For any matrix $\mathbf{X}$, we say it satisfies the restricted eigenvalue (RE) condition if its restricted eigenvalue is strictly bounded away from 0. That is, for some $1 \leq s \leq p$, the following condition holds:

$$\kappa(s, \mathbf{X}) \triangleq \min_{\substack{J \subseteq \{1,\dots,p\} \\ |J| \leq s}} \min_{\substack{\boldsymbol{\delta} \neq 0 \\ \|\boldsymbol{\delta}_{J^c}\|_1 \leq 3\|\boldsymbol{\delta}_J\|_1}} \frac{\|\mathbf{X}\boldsymbol{\delta}\|_2}{\sqrt{n}\|\boldsymbol{\delta}_J\|_2} > 0.$$

Denote $s_1 = \|\boldsymbol{\beta}_0\|_0$, $s_2 = \max_j \|\boldsymbol{\Gamma}_{\cdot,j}\|_0$, $r = \max_j \|\boldsymbol{\theta}_j\|_0$ and κ is the restricted eigenvalue defined above. The following assumptions are needed:

(A1) The instrumental variable matrix $\mathbf{Z}$ and matrix $\mathbf{D} = \mathbf{Z}\boldsymbol{\Gamma}_0$ satisfies the restricted eigenvalue condition with some constants $\kappa(s_2, \mathbf{Z}), \kappa(s_1, \mathbf{D}) >$

0, respectively.

(A2) There exists a positive constant C such that $\max\{\|\boldsymbol{\beta}_0\|_1, \|\boldsymbol{\Gamma}_0\|_1, \{\|\boldsymbol{\theta}_i\|_1\}_{i=1,\dots,p}\} \leq C$.

(A3) There exists a positive constant C such that $\max_{1\leq j\leq p}\left(\boldsymbol{\Sigma}^e_{j,j}\right) \leq C^2$.

Since our test statistics rely on the estimation of the parameters in models (**??**) and (**??**), we first provide a lemma on the estimation errors of $\boldsymbol{\Gamma}_{\cdot,j}$ and $\boldsymbol{\beta}$.

Lemma 1 (Estimation error bounds of $\boldsymbol{\Gamma}_{\cdot,j}$ and $\boldsymbol{\beta}_0$ (**?**))**.** *Under assumptions (A1)-(A3), for each $j = 1, 2, \dots, p$, if the tuning parameter λ_{2j} is chosen as*

$$\lambda_{2j} = \widetilde{C}\sqrt{\frac{\boldsymbol{\Sigma}^e_{j,j}\left(\log p + \log q\right)}{n}},$$

*for some $\widetilde{C} \geq 2\sqrt{2}$, then with probability at least $1 - (pq)^{1-\widetilde{C}^2/8}$, $\widehat{\boldsymbol{\Gamma}}$ defined in (**??**) satisfies*

$$\|\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0\|_1 \leq \frac{16\widetilde{C}C}{\kappa^2(s_2, \mathbf{Z})} s_2 \sqrt{\frac{\log p + \log q}{n}},$$

and

$$\|\mathbf{Z}\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}_0\right)\|_F^2 \leq \frac{16\widetilde{C}^2C^2}{\kappa^2(s_2, \mathbf{Z})} s_2 p \left(\log p + \log q\right).$$

Furthermore, if the set of tuning parameters $\{\lambda_{2j} : j = 1, \dots, p\}$ satisfy

$$\lambda_{\max}(2C + \lambda_{\max}) \leq \frac{\kappa^2(s_2, \mathbf{Z})\kappa^2(s_1, \mathbf{D})}{1024 s_1 s_2},$$

where $\lambda_{\max} = \max_{1\le j\le p}\lambda_{2j}$*, if* λ_1 *is chosen as:*

$$\lambda_1 = C_0\sqrt{\frac{s_2\left(\log p + \log q\right)}{n}},$$

then with probability at least $1 - C_1(pq)^{-C_2}$*,* $\widehat{\boldsymbol{\beta}}$ *defined in* (**??**) *satisfies*

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \le C_3 s_1\sqrt{\frac{s_2\left(\log p + \log q\right)}{n}},$$

for some positive constants $C_0 - C_3$*.*

S2 Detailed Proofs

In the appendix we provided the proofs for the lemmas and theorems. We refer the proof of Lemma 1 to **?**. Before proving lemma **??**, we first state a useful proposition.

Proposition 1. Denote $\widehat{\mathbf{D}}_i$ as defined previously, $\widehat{\mathbf{M}}_i$ as $\left(Y, \widehat{\mathbf{D}}_{\cdot,-i}\right)$ for $i = 1, 2.\dots p$. Further for each $l = 1, 2, \dots p$, we use $\mathbf{M}_{i,l}$ and $\widehat{\mathbf{M}}_{i,l}$ to be the l-th column of the matrix $\mathbf{M}_i$ and $\widehat{\mathbf{M}}_i$ respectively (Notice that $\mathbf{M}_i$ is a matrix so $\mathbf{M}_{i,l}$ is a column vector, not the $(i,l)-$th element of matrix $\mathbf{M}$). Then under the assumptions stated in lemma 1, with the same choice of the tuning parameters λ_{2i}, with probability at least $1-(pq)^{1-C^2/8}$ for some $C \ge 2\sqrt{2}$, we have:

$$\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2 \le \frac{4\sqrt{n}\sqrt{s_2}\lambda_{2i}}{\kappa^2(s_2, \mathbf{Z})}, \quad i = 1, 2, \dots, p,$$

$$\|\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l}\|_2 \le \frac{4\sqrt{n}\sqrt{s_2}\lambda_{2i}}{\kappa^2(s_2, \mathbf{Z})}, \quad i = 1, 2, \ldots, p,$$

In addition, the estimated $\widehat{\mathbf{M}}_i$ satisfies the RE condition with some constant $\kappa(r_i, \widehat{\mathbf{M}}_i)$ which satisfies $\kappa(r_i, \widehat{\mathbf{M}}_i) \ge \frac{1}{2}\kappa(r_i, \mathbf{M}_i)$.

Proof of proposition 1. Notice that for $i = 1, 2, \ldots, p$,

$$\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2 = \left\|\mathbf{Z}\left(\widehat{\boldsymbol{\Gamma}}_{0,i} - \boldsymbol{\Gamma}_{0,i}\right)\right\|_2 \le \frac{4\sqrt{n}\sqrt{s_2}\lambda_{2i}}{\kappa(s_2, \mathbf{Z})},$$

where the last inequality follows from **?**.

For the second inequality, notice that $\widehat{\mathbf{M}}_i = \left(Y, \widehat{\mathbf{D}}_{\cdot,-i}\right)$, so there exists some i_0 such that:

$$\|\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l}\|_2 = \begin{cases} 0 \quad \text{if } l = 1, \\ \left\|\mathbf{Z}\left(\widehat{\boldsymbol{\Gamma}}_{0,i_0} - \boldsymbol{\Gamma}_{0,i_0}\right)\right\|_2. \end{cases}$$

So similarly we have:

$$\|\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l}\|_2 \le \frac{4\sqrt{n}\sqrt{s_2}\lambda_{2i}}{\kappa(s_2, \mathbf{Z})}.$$

Furthermore, according to **?**, they proved that $Z\widehat{\boldsymbol{\Gamma}}$ satisfies the RE condition with $\kappa(s_1, Z\widehat{\boldsymbol{\Gamma}}) \ge \frac{1}{2}\kappa(s_1, \mathbf{D})$. Using the relationship between $\mathbf{M}_i$ and $\mathbf{D}$, it is straightforward that $\kappa(r_i, \widehat{\mathbf{M}}_i) \ge \frac{1}{2}\kappa(r_i, \mathbf{M}_i)$. $\square$

Then we provide the proof of lemma **??**.

Proof of lemma **??**. Without lose of generality, we assume $a_i = 0$. For each

$i = 1, 2, \ldots, p$, by the definition of $\widehat{\boldsymbol{\theta}}_i$ in (**??**), we have:

$$\frac{1}{2n}\|\widehat{\mathbf{D}}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\|_2^2 + \mu_i\|\widehat{\boldsymbol{\theta}}_i\|_1 \leq \frac{1}{2n}\|\widehat{\mathbf{D}}_i - \widehat{\mathbf{M}}_i\boldsymbol{\theta}_i\|_2^2 + \mu_i\|\boldsymbol{\theta}_i\|_1. \tag{S2.1}$$

For the left hand side (LHS), notice that:

$$\begin{aligned}
\frac{1}{2n}\|\widehat{\mathbf{D}}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\|_2^2 &= \frac{1}{2n}\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2^2 + \frac{1}{2n}\|\mathbf{D}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\|_2^2 - \frac{1}{n}\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)^\top\left(\mathbf{D}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\right),\\
&= \frac{1}{2n}\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2^2 + \frac{1}{2n}\|\boldsymbol{\zeta}_i\|_2^2 + \frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 + \frac{1}{2n}\|(\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i\|_2^2,\\
&- \frac{1}{n}\boldsymbol{\zeta}^\top\left(\widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i - \mathbf{M}_i\boldsymbol{\theta}\right) + \frac{1}{n}\boldsymbol{\theta}_i\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)^\top\widehat{\mathbf{M}}_i\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right),\\
&- \frac{1}{n}\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)^\top\left(\mathbf{D}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\right).
\end{aligned} \tag{S2.2}$$

While for the right hand side(RHS), similarly,

$$\begin{aligned}
\frac{1}{2n}\|\widehat{\mathbf{D}}_i - \widehat{\mathbf{M}}_i\boldsymbol{\theta}_i\|_2^2 &= \frac{1}{2n}\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2^2 + \frac{1}{2n}\|\mathbf{D}_i - \widehat{\mathbf{M}}_i\boldsymbol{\theta}_i\|_2^2 - \frac{1}{n}\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)^\top\left(\mathbf{D}_i - \widehat{\mathbf{M}}_i\boldsymbol{\theta}_i\right),\\
&= \frac{1}{2n}\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2^2 + \frac{1}{2n}\|\boldsymbol{\zeta}_i\|_2^2 + \frac{1}{2n}\|(\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i\|_2^2 - \frac{1}{n}\boldsymbol{\zeta}_i^\top\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)\boldsymbol{\theta}_i,\\
&- \frac{1}{n}\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)^\top\left(\mathbf{D}_i - \widehat{\mathbf{M}}_i\widehat{\boldsymbol{\theta}}_i\right).
\end{aligned} \tag{S2.3}$$

Combining (S2.2), (S2.3) and (S2.1) we have:

$$\begin{aligned}
\frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 &\leq \frac{1}{n}\boldsymbol{\zeta}_i^\top\widehat{\mathbf{M}}_i\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\right) - \frac{1}{n}\boldsymbol{\theta}_i\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)^\top\widehat{\mathbf{M}}_i\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right),\\
&+ \frac{1}{n}\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)^\top\widehat{\mathbf{M}}_i\left(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\right) + \mu_i\left(\|\boldsymbol{\theta}_i\|_1 - \|\widehat{\boldsymbol{\theta}}_i\|_1\right),\\
&\leq \left\|\frac{1}{n}\widehat{\mathbf{M}}_i^\top\boldsymbol{\zeta}_i - \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)\boldsymbol{\theta}_i + \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)\right\|_\infty \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1,\\
&+ \mu_i\left(\|\boldsymbol{\theta}_i\|_1 - \|\widehat{\boldsymbol{\theta}}_i\|_1\right).
\end{aligned}$$

We first show that the event $\left\|\frac{1}{n}\widehat{\mathbf{M}}_i^\top\boldsymbol{\zeta}_i - \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)\boldsymbol{\theta}_i + \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)\right\|_\infty \leq$

$\dfrac{\mu_i}{2}$ happens with large probability.

As

$$\frac{1}{n}\widehat{\mathbf{M}}_i^\top \boldsymbol{\zeta}_i - \frac{1}{n}\widehat{\mathbf{M}}_i^\top \left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)\boldsymbol{\theta}_i + \frac{1}{n}\widehat{\mathbf{M}}_i^\top \left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right),$$
$$= \underbrace{\frac{1}{n}\mathbf{M}_i^\top \boldsymbol{\zeta}_i}_{T_1} + \underbrace{\frac{1}{n}(\widehat{\mathbf{M}}_i - \mathbf{M}_i)^\top \boldsymbol{\zeta}_i}_{T_2} - \underbrace{\frac{1}{n}\mathbf{M}_i^\top (\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i}_{T_3} - \underbrace{\frac{1}{n}(\widehat{\mathbf{M}}_i - \mathbf{M}_i)^\top (\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i}_{T_4},$$
$$+ \underbrace{\frac{1}{n}\mathbf{M}_i^\top (\widehat{\mathbf{D}}_i - \mathbf{D}_i)}_{T_5} + \underbrace{\frac{1}{n}(\widehat{\mathbf{M}}_i - \mathbf{M}_i)^\top (\widehat{\mathbf{D}}_i - \mathbf{D}_i)}_{T_6},$$

we label these terms from T_1 to T_6. To bound term T_1, it follows from the union bound and the Gaussian tail bound:

$$\mathbb{P}(\|T_1\|_\infty \geq \frac{\mu_i}{12}) = \mathbb{P}\left(\left\|\frac{1}{n}\mathbf{M}_i^\top \boldsymbol{\zeta}\right\|_\infty \geq \frac{\mu_i}{12}\right) \leq p\exp\left\{-\frac{n}{2\sigma_{\boldsymbol{\zeta}_i}^2}\cdot\left(\frac{\mu_i}{12}\right)^2\right\}. \tag{S2.4}$$

To bound term T_2, noticing that $\|\widehat{\boldsymbol{\Gamma}}_{0,i} - \widehat{\boldsymbol{\Gamma}}_{0,i}\|_1 \leq \dfrac{16 s_2 \lambda_{2i}}{\kappa^2(s_2, \mathbf{Z})}$,

$$\begin{aligned}\mathbb{P}(\|T_2\|_\infty \geq \frac{\mu_i}{12}) &\leq \mathbb{P}\left(\left\|\frac{1}{n}\mathbf{Z}^\top \boldsymbol{\zeta}_i\right\|_\infty \geq \frac{\mu_i}{12}\cdot\frac{\kappa^2(s_2,\mathbf{Z})}{16 s_2\lambda_{2i}}\right),\\ &\leq qC^*\exp\left\{-\frac{n}{2\sigma_{\boldsymbol{\zeta}_i}^2}\cdot\left(\frac{\mu_i}{12}\cdot\frac{\kappa^2(s_2,\mathbf{Z})}{16 s_2\lambda_{\max}}\right)^2\right\},\end{aligned} \tag{S2.5}$$

for some positive constant C^*. As for term T_3, as $\|\boldsymbol{\theta}_i\|_\infty \leq C$ and by proposition 1,

$$\begin{aligned}\|T_3\|_\infty = \left\|\frac{1}{n}\mathbf{M}_i^\top(\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i\right\|_\infty &\leq C\max_{1\leq l,k\leq p}\left|\frac{1}{n}\mathbf{M}_{i,l}^\top(\widehat{\mathbf{M}}_{i,k} - \mathbf{M}_{i,k})\right|,\\ &\leq C\max_{1\leq k\leq p}\frac{1}{\sqrt{n}}\left\|\widehat{\mathbf{M}}_{i,k} - \mathbf{M}_{i,k}\right\|_2 \leq \frac{4C\sqrt{s_2}\lambda_{\max}}{\kappa(s_2,\mathbf{Z})}.\end{aligned} \tag{S2.6}$$

For T_4, using the result in proposition 1, we have:

$$\begin{aligned}\|T_4\|_\infty &= \left\|\frac{1}{n}(\widehat{\mathbf{M}}_i - \mathbf{M}_i)^\top(\widehat{\mathbf{M}}_i - \mathbf{M}_i)\boldsymbol{\theta}_i\right\|_\infty, \\ &\leq C \max_{1\leq l,k\leq p} \frac{1}{n}\left\|(\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l})\right\|_2 \cdot \left\|(\widehat{\mathbf{M}}_{i,k} - \mathbf{M}_{i,k})\right\|_2, \\ &\leq \frac{16Cs_2\lambda_{\max}^2}{\kappa^2(s_2,\mathbf{Z})}. \end{aligned} \tag{S2.7}$$

For T_5, similar to T_3, we have

$$\begin{aligned}\|T_5\|_\infty &= \frac{1}{n}\left\|\mathbf{M}_i^\top(\widehat{\mathbf{D}}_i - \mathbf{D}_i)\right\|_\infty \leq \max_{1\leq l\leq p}\frac{1}{n}\left|\mathbf{M}_{i,l}^\top(\widehat{\mathbf{D}}_i - \mathbf{D}_i)\right|, \\ &\leq \frac{1}{\sqrt{n}}\|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2 \leq \frac{4C\sqrt{s_2}\lambda_{\max}}{\kappa(s_2,\mathbf{Z})}. \end{aligned} \tag{S2.8}$$

Finally for T_6,

$$\begin{aligned}\|T_6\|_\infty &= \frac{1}{n}\left\|(\widehat{\mathbf{M}}_i - \mathbf{M}_i)^\top(\widehat{\mathbf{D}}_i - \mathbf{D}_i)\right\|_\infty \leq \max_{1\leq l\leq p}\frac{1}{n}\left|(\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l})^\top(\widehat{\mathbf{D}}_i - \mathbf{D}_i)\right|, \\ &\leq \max_{1\leq l\leq p}\frac{1}{n}\|\widehat{\mathbf{M}}_{i,l} - \mathbf{M}_{i,l}\|_2 \cdot \|\widehat{\mathbf{D}}_i - \mathbf{D}_i\|_2 \leq \frac{16Cs_2\lambda_{\max}^2}{\kappa^2(s_2,\mathbf{Z})}. \end{aligned} \tag{S2.9}$$

Combining the results from (S2.4) to (S2.9), there exists some positive constant C_4, C_5, C_5^*, such that with the tuning parameter μ_i chosen as:

$$\mu_i = \frac{C_4^*}{\kappa(s_2,\mathbf{Z})}\sqrt{\frac{s_2(\log p + \log q)}{n}},$$

with $C_4^* = C_5^* \max(C, \sigma_{\zeta_i})$, then with probability at least $1 - C_4\,(pq)^{-C_5}$,

$$\left\|\frac{1}{n}\widehat{\mathbf{M}}_i^\top\boldsymbol{\zeta}_i - \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{M}}_i - \mathbf{M}_i\right)\boldsymbol{\theta}_i + \frac{1}{n}\widehat{\mathbf{M}}_i^\top\left(\widehat{\mathbf{D}}_i - \mathbf{D}_i\right)\right\|_\infty \leq \frac{\mu_i}{2} \tag{S2.10}$$

Then under (S2.10), we have:

$$\frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 \leq \frac{\mu_i}{2}\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 + \mu_i\left(\|\boldsymbol{\theta}_i\|_1 - \|\widehat{\boldsymbol{\theta}}_i\|_1\right). \tag{S2.11}$$

Let R_i be the support of the true parameter $\boldsymbol{\theta}_i$ and without any abuse of using notations, we use $\boldsymbol{\theta}_{i,R_i}$ and $\widehat{\boldsymbol{\theta}}_{i,R_i}$ to represent the subvector of $\boldsymbol{\theta}_i$ and $\widehat{\boldsymbol{\theta}}_i$ restricted on the set R_i. Also let $|R_i| = r_i$. Adding $\frac{\mu_i}{2}\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1$ to both sides of (S2.11) yields:

$$\begin{aligned}\frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 + \frac{\mu_i}{2}\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 &\le \mu_i\left(\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 + \|\boldsymbol{\theta}_i\|_1 - \|\widehat{\boldsymbol{\theta}}_i\|_1\right),\\ &= \mu_i\left(\|\boldsymbol{\theta}_{i,R_i}\|_1 - \|\widehat{\boldsymbol{\theta}}_{i,R_i}\|_1 + \|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_1\right),\\ &\le 2\mu_i\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_1 \le 2\mu_i\sqrt{r_i}\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2.\end{aligned} \tag{S2.12}$$

The last two inequalities in (S2.12) imply:

$$\frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 \le 2\mu_i\sqrt{r_i}\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2, \tag{S2.13}$$

$$\frac{\mu_i}{2}\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 \le 2\mu_i\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_1, \tag{S2.14}$$

and (S2.14) is equivalent to

$$\|\widehat{\boldsymbol{\theta}}_{i,R_i^c} - \boldsymbol{\theta}_{i,R_i^c}\|_1 \le 3\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_1. \tag{S2.15}$$

As stated in proposition 1, $\widehat{\mathbf{M}}_i$ satisfies the RE condition with some constant $\kappa(r_i, \widehat{\mathbf{M}}_i) \ge \frac{1}{2}\kappa(r_i, \mathbf{M}_i)$, together with (S2.15) we have:

$$\frac{1}{2n}\|\widehat{\mathbf{M}}_i(\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i)\|_2^2 \ge \frac{1}{2}\kappa^2(r_i, \widehat{\mathbf{M}}_i)\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2^2 \ge \frac{1}{8}\kappa^2(r_i, \mathbf{M}_i)\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2^2.$$

Combining with (S2.13),

$$\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2 \le \frac{16\mu_i\sqrt{r_i}}{\kappa^2(r_i, \mathbf{M}_i)}, \tag{S2.16}$$

Plugging in the tuning parameter μ_i gives the final result in lemma **??**:

$$\begin{aligned}
\|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 \leq 4\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_1 &\leq 4\sqrt{r_i}\|\widehat{\boldsymbol{\theta}}_{i,R_i} - \boldsymbol{\theta}_{i,R_i}\|_2,\\
&\leq \frac{64C_4^*}{\kappa^2(r_i, \mathbf{M}_i)\kappa(s_2, \mathbf{Z})} r_i \sqrt{\frac{s_2(\log p + \log q)}{n}},\\
&\leq \frac{64C_4^*}{\kappa^2(Y, \mathbf{D})\kappa(s_2, \mathbf{Z})} r_i \sqrt{\frac{s_2(\log p + \log q)}{n}}.
\end{aligned}$$

□

Based on the previous results, we provide the proof for the main theorems. First we prove the asymptotic distribution of the test statistics for a single hypothesis.

Proof of Theorem **??**. The form of the test statistic T_i is a de-biased version of the sample correlation. To show it follows a standard normal distribution, we list the following notation. Denote:

$$\widetilde{\xi}_k = \xi_k - \overline{\xi},\ \widetilde{\zeta}_{k,i} = \zeta_{k,i} - \overline{\zeta_i},$$

where $\overline{\xi} = \sum_{k=1}^n \xi_k$ and $\overline{\zeta_i} = \sum_{k=1}^n \zeta_{k,i}$. Recall that by the previous definition, we have:

$$\begin{aligned}
&\xi_k = y_k - \mu - \mathbf{D}_k^\top \beta,\\
&\zeta_{k,i} = \mathbf{D}_{k,i} - a_i - (y_k, D_{k,-i}^\top)^\top \boldsymbol{\theta}_i,\\
&\widehat{\xi}_k = y_k - \overline{Y} - (\widehat{\mathbf{D}}_k - \overline{\widehat{\mathbf{D}}})^\top \widehat{\boldsymbol{\beta}},
\end{aligned}$$

$$\widehat{\zeta}_{k,i} = \widehat{\mathbf{D}}_{k,i} - \overline{\widehat{\mathbf{D}}_i} - \left(y_k - \overline{Y}, \left(\widehat{\mathbf{D}}_{k,-i} - \overline{\widehat{\mathbf{D}}}_{-i}\right)^\top\right)\widehat{\boldsymbol{\theta}}_i.$$

Based on these notations, we have the following decomposition:

$$\frac{1}{n}\sum_{k=1}^n \widehat{\xi}_k\widehat{\zeta}_{k,i} = \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k\widetilde{\zeta}_{k,i} - \underbrace{\frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k(\widetilde{\zeta}_{k,i} - \widehat{\zeta}_{k,i})}_{A_1} - \underbrace{\frac{1}{n}\sum_{k=1}^n \widetilde{\zeta}_{k,i}(\widetilde{\xi}_k - \widehat{\xi}_k)}_{A_2} + \underbrace{\frac{1}{n}\sum_{k=1}^n (\widetilde{\xi}_k - \widehat{\xi}_k)(\widetilde{\zeta}_{k,i} - \widehat{\zeta}_{k,i})}_{A_3}.$$

For simplicity, denote the second, third and forth term as A_1, A_2 and A_3.

Then for A_1, we have:

$$\begin{aligned}
A_1 &= \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k^2(\widehat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i}) + \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k(\mathbf{D}_k - \overline{\mathbf{D}})^\top\beta(\widehat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i}),\\
&+ \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k\left\{(\mathbf{D}_{k,i} - \overline{\mathbf{D}_i}) - (\widehat{\mathbf{D}}_{k,i} - \overline{\widehat{\mathbf{D}}_i})\right\} + \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k(\mathbf{D}_{k,-i} - \overline{\mathbf{D}}_{-i})^\top(\widehat{\boldsymbol{\theta}}_{-1,i} - \boldsymbol{\theta}_{-1,i}),\\
&+ \frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k\left\{(\mathbf{D}_{k,-i} - \overline{\mathbf{D}_{-i}}) - (\widehat{\mathbf{D}}_{k,-i} - \overline{\widehat{\mathbf{D}}_{-i}})\right\}^\top\widehat{\boldsymbol{\theta}}_{-1,i}.
\end{aligned}$$

We denote these five terms as $A_{1.1}$ to $A_{1.5}$. For $A_{1.2}$, combining the result in lemma **??** and the fact that ξ and $\mathbf{D}$ are independent, we know that for some positive constant C, there exists some $C' > 0$ such that:

$$\begin{aligned}
&|\widehat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i}| \lesssim_p r\sqrt{\frac{s_2(\log p + \log q)}{n}},\\
&\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k(\mathbf{D}_k - \overline{\mathbf{D}})^\top\boldsymbol{\beta}\right| \geq C\sqrt{\frac{\log p}{n}}\right) = \mathcal{O}(p^{-C'}).
\end{aligned}$$

Hence,

$$\begin{aligned}
A_{1,2} &\leq \left|\frac{1}{n}\sum_{k=1}^n \widetilde{\xi}_k(\mathbf{D}_k - \overline{\mathbf{D}})^\top\boldsymbol{\beta}\right| \cdot |\widehat{\boldsymbol{\theta}}_{1,i} - \boldsymbol{\theta}_{1,i}|,\\
&\lesssim \mathcal{O}_p\left(\sqrt{\frac{\log p}{n}} \cdot r\sqrt{\frac{s_2(\log p + \log q)}{n}}\right). \qquad \text{(S2.17)}
\end{aligned}$$

Similarly for $A_{1.4}$, as

$$\mathbb{P}\left(\max_{1\le j\le p}\left|\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k(\mathbf{D}_{k,j}-\overline{\mathbf{D}}_j)\right|\le C\sqrt{\frac{\log p}{n}}\right)=\mathcal{O}(p^{-C'}),$$

so we have:

$$\begin{aligned}A_{1.4}&\le\left\|\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k(\mathbf{D}_{k,-i}-\overline{\mathbf{D}}_{-i})\right\|_\infty\cdot\|\widehat{\boldsymbol{\theta}}_{-1,i}-\boldsymbol{\theta}_{-1,i}\|_1,\\&\lesssim\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot r\sqrt{\frac{s_2(\log p+\log q)}{n}}\right).\end{aligned}\tag{S2.18}$$

Then for $A_{1.3}$, by the estimation error for $\mathbf{D}_i$ as we used in proposition 1, we have:

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k(\mathbf{D}_{k,i}-\widehat{\mathbf{D}}_{k,i})\right|&=\left|\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k\left(Z_k(\widehat{\boldsymbol{\Gamma}}_{0,i}-\boldsymbol{\Gamma}_{0,i})\right)\right|,\\&\le\left\|\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_kZ_k\right\|_\infty\cdot\|\widehat{\boldsymbol{\Gamma}}_{0,i}-\boldsymbol{\Gamma}_{0,i}\|_1,\\&\lesssim\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_2\sqrt{\frac{\log p+\log q}{n}}\right).\end{aligned}\tag{S2.19}$$

For the last term $A_{1.5}$, similar to $A_{1.3}$,

$$A_{1.5}\lesssim\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_2\sqrt{\frac{\log p+\log q}{n}}\right).\tag{S2.20}$$

Combining the result from (S2.17) to (S2.20) we know that uniformly for $1\le i\le p$:

$$\begin{aligned}A_1=\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k^2(\widehat{\boldsymbol{\theta}}_{1,i}-\boldsymbol{\theta}_{1,i})+\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot r\sqrt{\frac{s_2(\log p+\log q)}{n}}\right)\\+\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_2\sqrt{\frac{\log p+\log q}{n}}\right).\end{aligned}\tag{S2.21}$$

And as for term A_2, we have a similar decomposition given by:

$$
\begin{aligned}
A_2 &= \frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}^2(\widehat{\beta}_i-\beta_i)+\frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}\left(y_k-\overline{Y},\left(\mathbf{D}_{k,-i}-\overline{\mathbf{D}}_{-i}\right)^\top\right)\boldsymbol{\theta}_i(\widehat{\beta}_i-\beta_i),\\
&+\frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}\left(\mathbf{D}_{k,-i}-\overline{\mathbf{D}}_{-i}\right)^\top(\widehat{\boldsymbol{\beta}}_{-i}-\boldsymbol{\beta}_{-i})+\frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}\left[(\widehat{\mathbf{D}}_k-\mathbf{D}_k)+(\overline{\widehat{\mathbf{D}}}-\overline{\mathbf{D}})\right]\widehat{\boldsymbol{\beta}}.
\end{aligned}
$$

By using the similar techniques as in A_1, we know that uniformly over $1\le i\le p$,

$$
\begin{aligned}
A_2 = \frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}^2(\widehat{\beta}_i-\beta_i)&+\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_1\sqrt{\frac{s_2(\log p+\log q)}{n}}\right)\\
&+\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_2\sqrt{\frac{\log p+\log q}{n}}\right). \qquad \text{(S2.22)}
\end{aligned}
$$

For the last term A_3, the decomposition is given as following:

$$
\begin{aligned}
A_3 = \frac{1}{n}\sum_{k=1}^{n}&\left\{\underbrace{(\mathbf{D}_k-\overline{\mathbf{D}})^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{B_1}+\underbrace{\left((\widehat{\mathbf{D}}_k-\mathbf{D}_k)-(\overline{\widehat{\mathbf{D}}}-\overline{\mathbf{D}})\right)^\top(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})}_{B_2}\right\},\\
&\cdot\left\{\underbrace{\left((\mathbf{D}_{k,i}-\overline{\mathbf{D}_i})-(\widehat{\mathbf{D}}_{k,i}-\overline{\widehat{\mathbf{D}}_i})\right)}_{B_3}+\underbrace{\left(y_k-\overline{Y},\left(\mathbf{D}_{k,-i}-\overline{\mathbf{D}}_{-i}\right)^\top\right)(\widehat{\boldsymbol{\theta}}_i-\boldsymbol{\theta}_i)}_{B_4}\right.\\
&\left.+\underbrace{\left((\widehat{\mathbf{D}}_{k,-i}-\overline{\widehat{\mathbf{D}}}_{-i})-(\mathbf{D}_{k,-i}-\overline{\mathbf{D}}_{-i})\right)\widehat{\boldsymbol{\theta}}_i}_{B_5}\right\}.
\end{aligned}
$$

For simplicity we denote the five terms above as B_1 to B_5. Then,

$$
A_3 = B_1(B_3+B_4+B_5)+B_2(B_3+B_4+B_5).
$$

For B_1B_3, based on the bounds in proposition 1, we have:

$$
\begin{aligned}
B_1B_3 &\leq \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \cdot \left\| \frac{1}{n}\sum_{k=1}^{n}(\mathbf{D}_k - \overline{\mathbf{D}})^\top \left((\mathbf{D}_{k,i} - \overline{\mathbf{D}_i}) - (\widehat{\mathbf{D}}_{k,i} - \overline{\widehat{\mathbf{D}}_i})\right)\right\|_\infty, \\
&\lesssim \mathcal{O}_p\left(s_1\sqrt{\frac{s_2(\log p + \log q)}{n}} \cdot \sqrt{\frac{s_2(\log p + \log q)}{n}}\right). \qquad \text{(S2.23)}
\end{aligned}
$$

For B_1B_4, it follows from the proof in **?** that:

$$
\begin{aligned}
B_1B_4 \lesssim\, &\mathcal{O}_p\left(\sqrt{\frac{\log p}{n}} \cdot s_1\sqrt{\frac{s_2(\log p + \log q)}{n}} \cdot r\sqrt{\frac{s_2(\log p + \log q)}{n}}\right) \\
&+ \mathcal{O}_p\left(r\sqrt{\frac{s_2(\log p + \log q)}{n}} \cdot a_n\right), \\
&+ \mathcal{O}_p\left(\lambda_{\max}(\Sigma_{\mathbf{D}}) \cdot a_n^2\right), \qquad \text{(S2.24)}
\end{aligned}
$$

where $\max(\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2, \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}\|_2) = \mathcal{O}_p(a_n)$. For term B_2B_3, we have:

$$
\begin{aligned}
B_2B_3 &\leq \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \cdot \left\| \frac{1}{n}\sum_{k=1}^{n}\left((\widehat{\mathbf{D}}_k - \mathbf{D}_k) - (\overline{\widehat{\mathbf{D}}} - \overline{\mathbf{D}})\right)\left((\mathbf{D}_{k,i} - \overline{\mathbf{D}_i}) - (\widehat{\mathbf{D}}_{k,i} - \overline{\widehat{\mathbf{D}}_i})\right)\right\|_\infty, \\
&\leq \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_1 \cdot \max_{1\leq j\leq p}\left| \frac{1}{n}\sum_{k=1}^{n}\left((\mathbf{D}_{k,j} - \overline{\mathbf{D}_j}) - (\widehat{\mathbf{D}}_{k,j} - \overline{\widehat{\mathbf{D}}_j})\right) \cdot \left((\mathbf{D}_{k,i} - \overline{\mathbf{D}_i}) - (\widehat{\mathbf{D}}_{k,i} - \overline{\widehat{\mathbf{D}}_i})\right)\right|, \\
&\lesssim \mathcal{O}_p\left(s_1\sqrt{\frac{s_2(\log p + \log q)}{n}} \cdot \frac{s_2(\log p + \log q)}{n}\right). \qquad \text{(S2.25)}
\end{aligned}
$$

Then for B_2B_4, it follows from the estimator bounds of $\widehat{\boldsymbol{\theta}}$,

$$
B_2B_4 \leq \|\widehat{\boldsymbol{\theta}}_i - \boldsymbol{\theta}_i\|_1 \cdot \left\| \frac{1}{n}\sum_{k=1}^{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top\left((\widehat{\mathbf{D}}_k - \mathbf{D}_k) - (\overline{\widehat{\mathbf{D}}} - \overline{\mathbf{D}})\right)\left(y_k - \overline{Y}, \left(\mathbf{D}_{k,-i} - \overline{\mathbf{D}}_{-i}\right)^\top\right)\right\|. \qquad \text{(S2.26)}
$$

Notice that the second term is in the same order as the term B_1B_3, so the order of the whole term B_2B_4 is actually dominated by the term B_1B_3.

And terms B_1B_5 and B_2B_5 are in the same order as B_1B_3 and B_2B_3. So together with the result in (S2.23) to (S2.26) and summing up the previous results in (S2.21), (S2.22) and the form of test statistic T_i, we have:

$$
\begin{aligned}
T_i &= \sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n}\widehat{\xi}_k\widehat{\zeta}_{k,i} + \frac{1}{n}\sum_{k=1}^{n}\widehat{\xi}_k^2\widehat{\boldsymbol{\theta}}_{1,i} + \frac{1}{n}\sum_{k=1}^{n}\widehat{\zeta}_{k,i}^2\widehat{\boldsymbol{\beta}}_i\right)/\widehat{\sigma}_\xi\widehat{\sigma}_{\zeta_i}, \\
&= \frac{\sqrt{n}}{\widehat{\sigma}_\xi\widehat{\sigma}_{\zeta_i}}\left\{\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k\widetilde{\zeta}_{k,i} - A_1 - A_2 + A_3 + \frac{1}{n}\sum_{k=1}^{n}\widehat{\xi}_k^2\widehat{\boldsymbol{\theta}}_{1,i} + \frac{1}{n}\sum_{k=1}^{n}\widehat{\zeta}_{k,i}^2\widehat{\boldsymbol{\beta}}_i\right\}, \\
&= \frac{\sqrt{n}}{\widehat{\sigma}_\xi\widehat{\sigma}_{\zeta_i}}\left\{\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k\widetilde{\zeta}_{k,i} + \frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k^2\boldsymbol{\theta}_{1,i} + \frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}^2\boldsymbol{\beta}_i + \frac{1}{n}\sum_{k=1}^{n}\widehat{\boldsymbol{\theta}}_{1,i}(\widehat{\xi}_k^2 - \widetilde{\xi}_k^2)\right. \\
&\quad \left. + \frac{1}{n}\sum_{k=1}^{n}\widehat{\boldsymbol{\beta}}_i(\widehat{\zeta}_{k,i}^2 - \widetilde{\zeta}_{k,i}^2) + order\right\},
\end{aligned}
\tag{S2.27}
$$

where *order* is the sum of all the reminder terms, which is given by:

$$
\begin{aligned}
order &= \mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot r\sqrt{\frac{s_2(\log p + \log q)}{n}}\right) + \mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_2\sqrt{\frac{\log p + \log q}{n}}\right) \\
&\quad + \mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_1\sqrt{\frac{s_2(\log p + \log q)}{n}}\right) + \mathcal{O}_p\left(s_1\sqrt{\frac{s_2(\log p + \log q)}{n}}\cdot\sqrt{\frac{s_2(\log p + \log q)}{n}}\right) \\
&\quad + \mathcal{O}_p\left(\sqrt{\frac{\log p}{n}}\cdot s_1\sqrt{\frac{s_2(\log p + \log q)}{n}}\cdot r\sqrt{\frac{s_2(\log p + \log q)}{n}}\right) + \mathcal{O}_p\left(r\sqrt{\frac{s_2(\log p + \log q)}{n}}\cdot a_n\right) \\
&\quad + \mathcal{O}_p\left(\lambda_{\max}(\Sigma_D)\cdot a_n^2\right) + \mathcal{O}_p\left(s_1\sqrt{\frac{s_2(\log p + \log q)}{n}}\cdot\frac{s_2(\log p + \log q)}{n}\right).
\end{aligned}
\tag{S2.28}
$$

Define $\widetilde{\omega}_{ii} = \boldsymbol{\Omega}_{i,i}^D + \dfrac{\boldsymbol{\beta}_i^2}{\sigma_\xi^2}$. Notice that:

$$
\frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k\widetilde{\zeta}_{k,i} + \frac{1}{n}\sum_{k=1}^{n}\widetilde{\xi}_k^2\boldsymbol{\theta}_{1,i} + \frac{1}{n}\sum_{k=1}^{n}\widetilde{\zeta}_{k,i}^2\boldsymbol{\beta}_i,
$$

$$= \left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\zeta_{k,i} - \mathbb{E}\xi\zeta_i\right) + \left(\mathbb{E}\xi\mathbb{E}\zeta_i - \bar{\xi}\bar{\zeta}_i\right) - \frac{\boldsymbol{\beta}_i}{\widetilde{\omega}_{ii}}\left(1 - \frac{\tilde{\sigma}_\xi^2}{\sigma_\xi^2} - \frac{\tilde{\sigma}_{\zeta_i}^2}{\sigma_{\zeta_i}^2}\right),$$

$$= \left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\zeta_{k,i} - \mathbb{E}\xi\zeta_i\right) + \mathcal{O}_p\left(\frac{\log p}{n}\right) - \frac{\boldsymbol{\beta}_i}{\widetilde{\omega}_{ii}}\left(1 - \frac{\tilde{\sigma}_\xi^2}{\sigma_\xi^2} - \frac{\tilde{\sigma}_{\zeta_i}^2}{\sigma_{\zeta_i}^2}\right). \quad \text{(S2.29)}$$

Denote order $A_n = order + \mathcal{O}_p\left(\frac{\log p}{n}\right)$, followed by the argument in **?** we know that:

$$\widehat{\sigma}_\xi^2 = \sigma_\xi^2 + \mathcal{O}_p(A_n + \sqrt{\frac{\log p}{n}}),\ \widehat{\sigma}_{\zeta_i}^2 = \sigma_{\zeta_i}^2 + \mathcal{O}_p(A_n + \sqrt{\frac{\log p}{n}}). \quad \text{(S2.30)}$$

In addition, notice that the required assumption $\lambda_{\max}(\Sigma_{\mathbf{D}})a_n^2 = o(n^{-\frac{1}{2}})$ and $r\sqrt{s_2(\log p + \log q)} \cdot a_n = o(1)$ are naturally hold for the estimators we are using and under assumptions C1-2. Hence, based on assumptions C1-2, together with (S2.27), (S2.28), (S2.29) and (S2.30) we know that

$$T_i \rightsquigarrow N(0,1).$$

And further recalls the relation between T_i and $\widehat{T}_i$ given by:

$$\widehat{T}_i = \frac{T_i}{1 - \frac{T_i^2}{n}\mathbf{1}\left(\frac{T_i^2}{n} < 1\right)}.$$

So finally Slutsky's theorem, we know that

$$T_i \rightsquigarrow N(0,1).$$

So we have finished the proof of theorem **??**. □

Once we were able to prove that our test statistic follows a standard normal distribution,the proofs for theorem **??** and **??** become rather straight forward. We refer the details of the proofs to section 5.2 of **?**. The proofs for ours differ from theirs in the error terms which have already been shown to be controlled in preferred orders in the proofs of theorem **??**.

S3 Results of simulation studies

S3.1 Evaluation of testing single hypothesis

Figures S1 and S2 show additional simulation results for testing single hypothesis.

S3.2 Obtaining first stage regression via Ridge regression method

In our method, the first stage regression $\mathbf{X} = \mathbf{Z}\boldsymbol{\Gamma}_0 + \mathbf{E}$, the regression efficients are obtained via Lasso and are used to predict $\mathbf{X}$. In genetic studies, the gene expression could also be predicted by imposing a Gaussian prior, which leads to the Ridge regression method. Although we proof replies on the theoretical results that are built upon the first stage regression using Lasso, it is still of interest to compare the numerical performance when using Ridge regression instead. As a result, we present in this supplementary some additional simulation results. For the Ridge regression, the tuning parameter is selected by a 10-fold cross-validation method.

Table S1 shows the empirical FDR and FDV for the proposed procedure

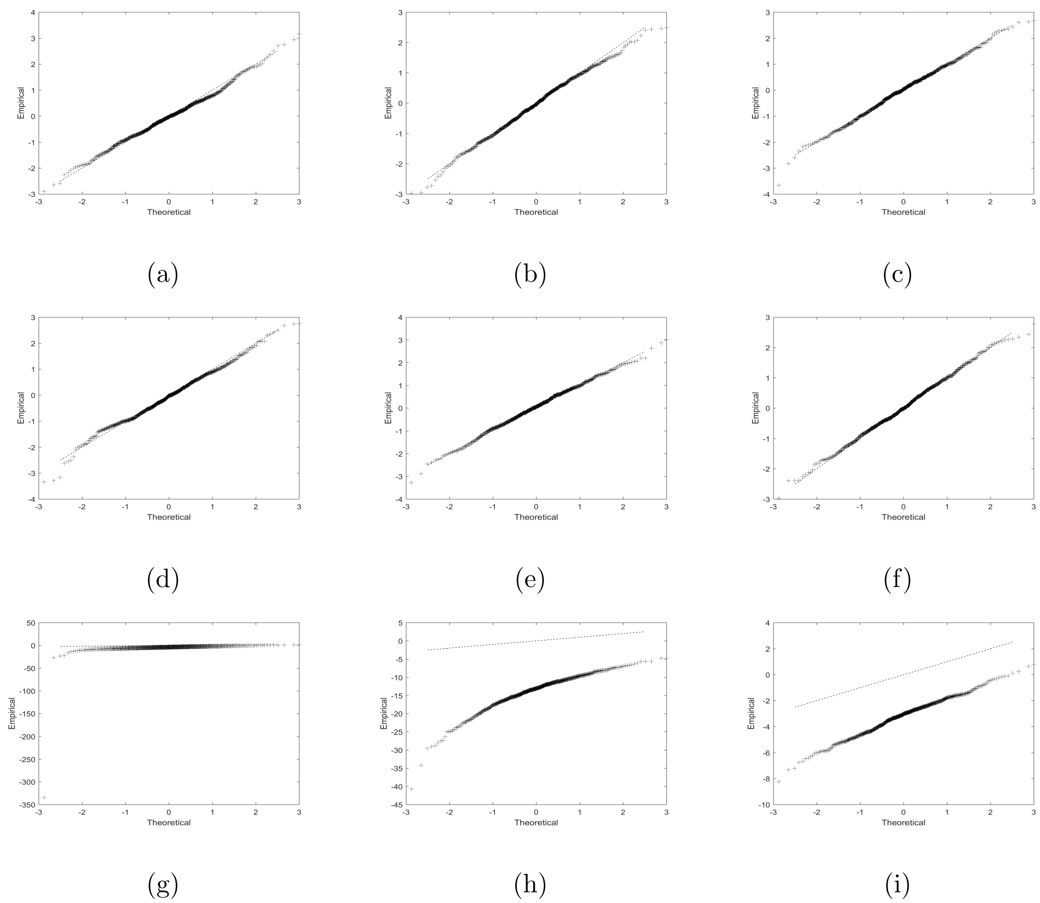
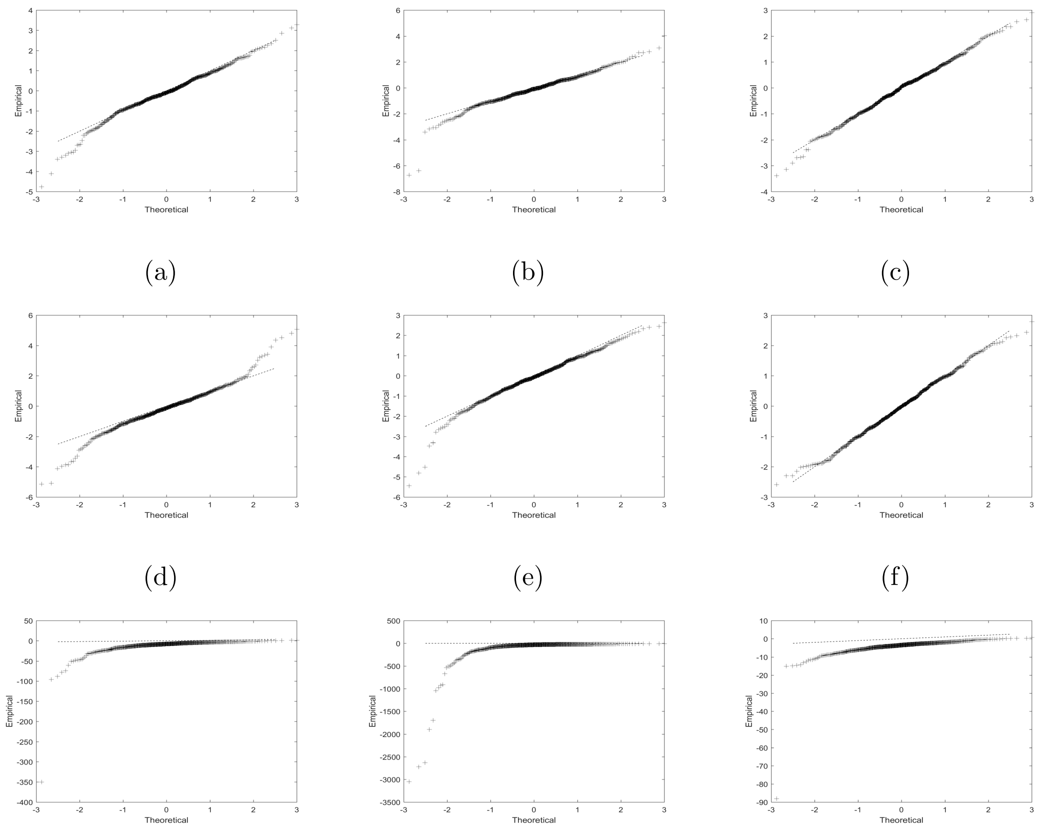


Figure S1: QQ-plots of the test statistic $\widehat{T}_i$ based on the two-stage IV model for several randomly selected variables to demonstrate the validity of its asymptotic distribution. The panels in the first and second row correspond to selected variables whose true value are zero and the third row are variables that are not zero. For different columns, (a)(d)(g), (b)(e)(h) and (c)(f)(i) correspond to different (n, p, q) values as $(200, 100, 100)$, $(400, 200, 200)$ and $(200, 500, 500)$.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure S2: Selected QQ-plots of the test statistics $\widehat{T}_i$ developed for fitting naive high dimensional regression models. The panels in the first and second row corresponds to selected variables whose true value are zero and the third row are variables that are not zero. For different columns, (a)(d)(g), (b)(e)(h) and (c)(f)(i) correspond to different (n, p, q) values as $(200, 100, 100)$, $(400, 200, 200)$ and $200, 500, 500)$.

using Ridge regression in the first stage. The result indicates that our method could still successfully control the FDR and FDV in this scenario. When p is relatively large, the simulation demonstrate that the procedure

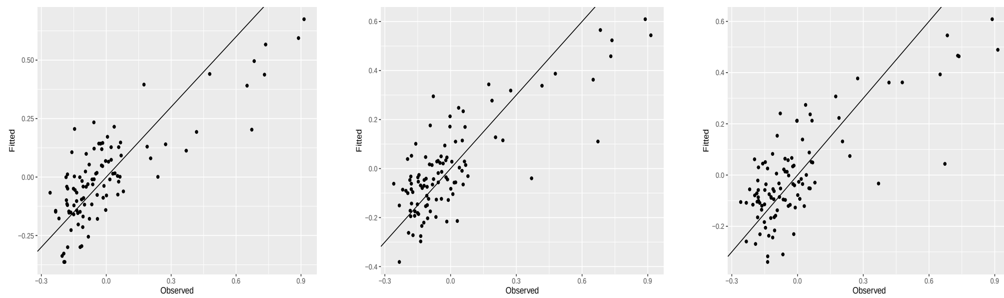
Table S1: Sensitivity analysis results based on 500 replications. Ridge regression is applied in the first stage regression. The eFDR and eFDV for multiple testing procedures based on IV regression for different combinations of (n, p, q) and different α, k levels.

(n, p, q)	α-level	eFDR	k-level	eFDV
	0.05	0.055	2	1.95
$(n, p, q) = (200, 100, 100)$	0.1	0.10	3	2.96
	0.2	0.20	4	3.75
	0.05	0.020	2	1.02
$(n, p, q) = (400, 200, 200)$	0.1	0.045	3	1.61
	0.2	0.10	4	2.29

is slightly conservative as the eFDR and eFDV is much smaller than the desired α and k levels. This indicates that Ridge regression could be used in practice but may lead to hypothesis testing procedure that is conservative and hence lack of power.

S4 Additional results of yeast data set

We further compared the the fitted versus the observed yeast growth yields using three different scenarios in Figure S3. The first model was to use the 15 genes selected using our proposed multiple testing method and to refit a linear model with the estimated $\widehat{\mathbf{X}}$. The second model used the 34 genes selected based on nominal p-value< 0.05 and refitted a linear model using



(a) (b) (c)

Figure S3: Scatter-plots of the fitted versus the observed yeast growth yields. (a): refitted model using the estimated expression levels of the 15 genes selected by our proposed method; (b): refitted model using expression levels of 34 genes selected with nominal p-value< 0.05. (c): the refitted model using expression levels of the genes selected by Lasso.

the original $\mathbf{X}$. The last model used the genes selected by Lasso using $\mathbf{X}$ with refitted coefficients using the original $\mathbf{X}$. Overall, we observe that the proposed IV regression gave better fit than those based on linear regression with gene expressions as the covariates.

S5 Software and data sets used in the paper

The codes and data sets are available upon request.