Linear Hypothesis Testing for High Dimensional Tobit Models

Tate Jacobson[1] and Hui Zou[2]

[1]*Department of Statistics, Oregon State University*

[2]*School of Statistics, University of Minnesota*

Supplementary Material

S.1 Technical Proofs

In this section, we prove the main results of the paper. Supporting results used in these proofs can be found in Section S.2.

S.1.1 Proof of Theorem 1

Proof of Theorem 1. We will focus on proving the results for $\hat{\boldsymbol{\theta}}_0$ as the arguments for $\hat{\boldsymbol{\theta}}_a$ follow the same lines. This proof is divided into three parts.

(i) We will start by proving that there exists a local minimizer $\hat{\boldsymbol{\theta}}$ of $Q_n(\boldsymbol{\theta})$ with constraints $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$ such that $\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$. We define $\tilde{\boldsymbol{\theta}}^*$ as follows

$$\begin{cases} \tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} & = \boldsymbol{\delta}^*_{\mathcal{M}} - \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\gamma^*\boldsymbol{h}_n \\ \tilde{\boldsymbol{\delta}}^*_{\mathcal{M}^c} & = \boldsymbol{\delta}^*_{\mathcal{M}^c} \\ \tilde{\gamma}^* & = \gamma^*. \end{cases}$$

We see that $\mathbf{C}\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \tilde{\gamma}^*\boldsymbol{t} = \mathbf{C}\boldsymbol{\delta}^*_{\mathcal{M}} - \mathbf{C}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\gamma^*\boldsymbol{h}_n - \gamma^*\boldsymbol{t} = \mathbf{C}\boldsymbol{\delta}^*_{\mathcal{M}} - \gamma^*\boldsymbol{h}_n - \gamma^*\boldsymbol{t} = \mathbf{0}$, with the last equality following from (A1). We also find that

$$\begin{aligned} \left\|\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}\right\|_2^2 &= \left\|\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\gamma^*\boldsymbol{h}_n\right\|_2^2 \\ &= \boldsymbol{h}_n'\gamma^*(\mathbf{C}\mathbf{C}')^{-1}\mathbf{C}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\gamma^*\boldsymbol{h}_n \\ &= \gamma^{*2}\boldsymbol{h}_n'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ &\le \gamma^{*2}\boldsymbol{h}_n'\boldsymbol{h}_n\lambda_{\max}\{(\mathbf{C}\mathbf{C}')^{-1}\} \\ &= O(\|\boldsymbol{h}_n\|_2^2) \\ &= O\left(\frac{s+m-r+1}{n}\right) \end{aligned}$$

with the last two equalities following from (A1). As such, if we can show that there exists a local minimizer $\hat{\boldsymbol{\theta}}$ of $Q_n(\theta)$ with constraints $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$ such that $\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2 = O_p\left(\frac{s+m-r+1}{n}\right)$, then $\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$ and we've finished part (i) of the proof.

For any $\boldsymbol{\theta}$ such that $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = 0$, we see that $\mathbf{C}^*(\boldsymbol{\theta}_{\mathcal{M}'} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'}) = \mathbf{C}\boldsymbol{\delta}_{\mathcal{M}} - \gamma\boldsymbol{t} - (\mathbf{C}\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \tilde{\gamma}^*\boldsymbol{t}) = \mathbf{0}$. As such, $\boldsymbol{\theta}_{\mathcal{M}'} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'}$ belongs to the null space of $\mathbf{C}^*$. Let $\mathbf{Z} \in \mathbb{R}^{(m+1)\times(m-r+1)}$ be a basis matrix for the null space of $\mathbf{C}^*$ with orthogonal columns, meaning that $\mathbf{C}^*\mathbf{Z} = \mathbf{0}$ and $\mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m-r+1}$. Then for any $\boldsymbol{\theta}_{\mathcal{M}'}$ such that $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, there exists $\boldsymbol{v} \in \mathbb{R}^{m-r+1}$ such that $\boldsymbol{\theta}_{\mathcal{M}'} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'} = \mathbf{Z}\boldsymbol{v}$.

For any $\Delta \in \mathbb{R}^{s+m-r+1}$, we define $\boldsymbol{\theta}(\Delta)$ by

$$\begin{cases} \boldsymbol{\theta}(\Delta)_{\mathcal{M}'} & = \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'} + \mathbf{Z}\Delta_{1:m-r+1} \\ \boldsymbol{\theta}(\Delta)_{\mathcal{S}} & = \tilde{\boldsymbol{\theta}}^*_{\mathcal{S}} + \Delta_{m-r+2:m-r+s+1} \\ \boldsymbol{\theta}(\Delta)_{(\mathcal{M}'\cup\mathcal{S})^c} & = \tilde{\boldsymbol{\theta}}^*_{(\mathcal{M}'\cup\mathcal{S})^c} \end{cases}$$

and define $\overline{Q}_n(\Delta) := Q_n(\boldsymbol{\theta}(\Delta))$. Since $\|\mathbf{Z}\Delta_{1:m-r+1}\|_2^2 = \Delta'_{1:m-r+1}\mathbf{Z}'\mathbf{Z}\Delta_{1:m-r+1} = \Delta'_{1:m-r+1}\Delta_{1:m-r+1} = \|\Delta_{1:m-r+1}\|_2^2$, we see that $\left\|\boldsymbol{\theta}(\Delta)_{\mathcal{M}'\cup\mathcal{S}} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = \|\Delta\|_2$. As such, we need only show that there exists a local minimizer $\hat{\Delta}$ of $\overline{Q}_n(\Delta)$ such that $\left\|\hat{\Delta}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$.

For $\tau > 0$, we define $\mathcal{N}_\tau := \left\{\Delta \in \mathbb{R}^{s+m-r+1} : \|\Delta\|_2 \leq \tau\sqrt{\frac{s+m-r+1}{n}}\right\}$. Consider the event

$$\mathcal{E}_n := \left\{\overline{Q}_n(\mathbf{0}) < \min_{\Delta\in\partial\mathcal{N}_\tau} \overline{Q}_n(\Delta)\right\}$$

where $\partial\mathcal{N}_\tau$ denotes the boundary of $\mathcal{N}_\tau$. We can see that if $\mathcal{E}_n$ holds, then there exists a local minimizer of $\overline{Q}_n(\Delta)$ in $\mathcal{N}_\tau$. Therefore it suffices to show that $\mathcal{E}_n$ holds with probability close to 1 as $n \to \infty$ for large τ. For any $\Delta \in \mathbb{R}^{s+m-r+1}$, a second order Taylor expansion provides that

$$\overline{Q}_n(\Delta) = \overline{Q}_n(\mathbf{0}) + \Delta'\nabla\overline{Q}_n(\mathbf{0}) + \frac{1}{2}\Delta'\nabla^2\overline{Q}_n(\tilde{\Delta})\Delta \tag{S1.1}$$

where $\tilde{\Delta}$ lies on the line segment connecting $\mathbf{0}$ and Δ. We will use (S1.1) to show that $\overline{Q}_n(\Delta) > \overline{Q}_n(\mathbf{0})$ for all $\Delta \in \partial\mathcal{N}_\tau$ with high probability.

Let $\tau \leq \sqrt{\log n}/2$ and $\Delta \in \partial\mathcal{N}_\tau$. Clearly $\tilde{\Delta} \in \mathcal{N}_\tau$. By definition, we have $\boldsymbol{\theta}(\Delta)_{(\mathcal{M}'\cup\mathcal{S})^c} = \tilde{\boldsymbol{\theta}}^*_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$. Moreover, we see that for sufficiently large n

$$\left\|\boldsymbol{\theta}(\tilde{\Delta})_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \leq \left\|\boldsymbol{\theta}(\tilde{\Delta})_{\mathcal{M}'\cup\mathcal{S}} - \tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 + \left\|\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2$$

$$
\begin{aligned}
&\le \tau\sqrt{\frac{s+m-r+1}{n}} + O\left(\sqrt{\frac{s+m-r+1}{n}}\right) \\
&\le \tau\sqrt{\frac{s+m-r+1}{n}} + \frac{1}{2}\sqrt{\frac{(s+m+1)\log n}{n}} \\
&\le \sqrt{\frac{(s+m+1)\log n}{n}}
\end{aligned}
$$

and, consequently, $\boldsymbol{\theta}(\tilde{\Delta}) \in \mathcal{N}_0$.

We want to find a lower bound for $\Delta'\nabla^2\overline{Q}_n(\tilde{\Delta})\Delta$. One can show that for any $\Delta \in \mathbb{R}^{s+m-r+1}$

$$
\begin{aligned}
\nabla^2\overline{Q}_n(\Delta) =& \frac{1}{n}\begin{bmatrix} n_1(\tilde{\gamma}^* + \mathbf{Z}_{m+1}\Delta_{1:m-r+1})^{-2}\mathbf{Z}'_{m+1}\mathbf{Z}_{m+1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\
&+ \frac{1}{n}\begin{bmatrix} \mathbf{Z}'\begin{bmatrix} -\mathbf{X}'_{1,\mathcal{M}} \\ \mathbf{y}'_1 \end{bmatrix} \\ -\mathbf{X}'_{1,\mathcal{S}} \end{bmatrix} \begin{bmatrix} \begin{bmatrix} -\mathbf{X}_{1,\mathcal{M}} & \mathbf{y}_1 \end{bmatrix}\mathbf{Z} & -\mathbf{X}_{1,\mathcal{S}} \end{bmatrix} \\
&+ \frac{1}{n}\begin{bmatrix} \mathbf{Z}'\begin{bmatrix} -\mathbf{X}'_{0,\mathcal{M}} \\ \mathbf{y}'_0 \end{bmatrix} \\ -\mathbf{X}'_{0,\mathcal{S}} \end{bmatrix} \mathbf{D}\left(\boldsymbol{\delta}(\tilde{\Delta})\right) \begin{bmatrix} \begin{bmatrix} -\mathbf{X}_{0,\mathcal{M}} & \mathbf{y}_0 \end{bmatrix}\mathbf{Z} & -\mathbf{X}_{0,\mathcal{S}} \end{bmatrix} \\
&+ \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda(\boldsymbol{\theta}(\Delta)) \end{bmatrix}
\end{aligned}
$$

where $\Lambda(\boldsymbol{\theta}(\Delta))$ is a diagonal matrix with negative diagonal elements.

We will derive lower bounds for each of the terms in the above expression for $\nabla^2\overline{Q}_n(\Delta)$. Since $\boldsymbol{\theta}(\tilde{\Delta}) \in \mathcal{N}_0$, one can easily show using the definition of κ_0 that the smallest element of $\Lambda(\boldsymbol{\theta}(\tilde{\Delta}))$ is bounded below by $-\lambda_n\kappa_0$ and, therefore, $\lambda_{\min}\{\Lambda(\boldsymbol{\theta}(\tilde{\Delta}))\} > -\lambda_n\kappa_0$. Moreover, we see that $n_1(\tilde{\gamma}^* + \mathbf{Z}_{m+1}\Delta_{1:m_r+1})^{-2} > 0$.

Bounding the remaining terms will require a bit more work. Let $\mathbf{L} = \begin{bmatrix} \mathbf{Z}_{1:m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_s \\ \mathbf{Z}_{m+1} & \mathbf{0} \end{bmatrix}$. Since we assumed $\mathcal{M} = \{0, \ldots, m-1\}$, we see that $\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} = \begin{bmatrix}\mathbf{X}_{(\mathcal{M})} & \mathbf{X}_{(\mathcal{S})}\end{bmatrix}$. We find

$$\begin{aligned}
&\frac{1}{n}\begin{bmatrix}\mathbf{Z}'\begin{bmatrix}-\mathbf{X}'_{1,\mathcal{M}} \\ \mathbf{y}'_1\end{bmatrix} \\ -\mathbf{X}'_{1,\mathcal{S}}\end{bmatrix}\begin{bmatrix}\begin{bmatrix}-\mathbf{X}_{1,\mathcal{M}} & \mathbf{y}_1\end{bmatrix}\mathbf{Z} & -\mathbf{X}_{1,\mathcal{S}}\end{bmatrix} \\
&\qquad + \frac{1}{n}\begin{bmatrix}\mathbf{Z}'\begin{bmatrix}-\mathbf{X}'_{0,\mathcal{M}} \\ \mathbf{y}'_0\end{bmatrix} \\ -\mathbf{X}'_{0,\mathcal{S}}\end{bmatrix}\mathbf{D}\left(\boldsymbol{\delta}(\tilde{\Delta})\right)\begin{bmatrix}\begin{bmatrix}-\mathbf{X}_{0,\mathcal{M}} & \mathbf{y}_0\end{bmatrix}\mathbf{Z} & -\mathbf{X}_{0,\mathcal{S}}\end{bmatrix} \\
&\qquad = \frac{1}{n}\begin{bmatrix}\mathbf{Z}'\begin{bmatrix}-\mathbf{X}'_{(\mathcal{M})} \\ \mathbf{y}'\end{bmatrix} \\ -\mathbf{X}'_{(\mathcal{S})}\end{bmatrix}\begin{bmatrix}\mathbf{D}\left(\boldsymbol{\delta}(\tilde{\Delta})\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1}\end{bmatrix}\begin{bmatrix}\begin{bmatrix}-\mathbf{X}_{(\mathcal{M})} & \mathbf{y}\end{bmatrix}\mathbf{Z} & -\mathbf{X}_{(\mathcal{S})}\end{bmatrix} \\
&\qquad = \frac{1}{n}\mathbf{L}'\begin{bmatrix}-\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})} \\ \mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{D}\left(\boldsymbol{\delta}(\tilde{\Delta})\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1}\end{bmatrix}\begin{bmatrix}-\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & \mathbf{y}\end{bmatrix}\mathbf{L} \\
&\qquad = \frac{1}{n}\mathbf{L}'\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\mathbf{L}.
\end{aligned}$$

Since the columns of $\mathbf{Z}$ are orthogonal, we see that $\|\mathbf{Z}\boldsymbol{v}\|_2^2 = \boldsymbol{v}'\mathbf{Z}'\mathbf{Z}\boldsymbol{v} = \boldsymbol{v}'\boldsymbol{v} = \|\boldsymbol{v}\|_2^2$ for any $\boldsymbol{v} \in \mathbb{R}^{m-r+1}$ and, by extension, $\|\mathbf{L}\Delta\|_2 = \|\Delta\|_2$ for any $\Delta \in \mathbb{R}^{s+m-r+1}$. Therefore, for any $\Delta \in \mathbb{R}^{s+m-r+1}$ we have

$$\frac{1}{n}\Delta'\mathbf{L}'\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\mathbf{L}\Delta \geq \|\mathbf{L}\Delta\|_2^2\,\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right\} = \|\Delta\|_2^2\,\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right\}.$$

We know that $\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right\} = \min_{\|\boldsymbol{v}\|_2=1}\boldsymbol{v}'\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\boldsymbol{v}$. Let $\boldsymbol{v} \in \mathbb{R}^{s+m+1}$ with $\|\boldsymbol{v}\|_2 = 1$. We note that $\|\boldsymbol{v}\|_1 \leq \sqrt{s+m+1}\,\|\boldsymbol{v}\|_2 = \sqrt{s+m+1}$. Consequently, we have

$$\begin{aligned}
\left|\boldsymbol{v}'\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\boldsymbol{v} - \boldsymbol{v}'\,\mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\boldsymbol{v}\right| &= \left|\boldsymbol{v}'\left(\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta})) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\right)\boldsymbol{v}\right| \\
&\leq \left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta})) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\right\|_{\max}\|\boldsymbol{v}\|_1^2 \\
&\leq \left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta})) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\right\|_{\max}(s+m+1).
\end{aligned}$$

Since $\boldsymbol{\theta}(\tilde{\Delta}) \in \mathcal{N}_0$, (A3) implies that $\boldsymbol{v}' \operatorname{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\boldsymbol{v} \geq \inf_{\boldsymbol{\theta}\in\mathcal{N}_0} \lambda_{\min}\left\{\operatorname{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta})\right]\right\} \geq c_H$. Therefore if the event $\tilde{\mathcal{E}}_{n,1} := \left\{\left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta})) - \operatorname{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\right\|_{\max} \leq \frac{c_H}{2(s+m+1)}\right\}$ holds, then

$$\boldsymbol{v}'\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\boldsymbol{v} \geq \boldsymbol{v}' \operatorname{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right]\boldsymbol{v} - \frac{c_H}{2} \geq c_H - \frac{c_H}{2} = \frac{c_H}{2}$$

for all $\boldsymbol{v}$ such that $\|\boldsymbol{v}\|_2 = 1$ and, by extension, $\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right\} \geq \frac{c_H}{2}$.

Pulling this all together, we see that if $\tilde{\mathcal{E}}_{n,1}$ holds, then

$$\begin{aligned}
\Delta'\nabla^2\overline{Q}_n(\tilde{\Delta})\Delta \geq{}& \frac{n_1}{n}(\tilde{\gamma}^* + \mathbf{Z}_{m+1}\Delta_{1:m-r+1})^{-2}\left\|\mathbf{Z}_{m+1}\Delta_{1:m-r+1}\right\|_2 \\
&+ \|\Delta\|_2^2\,\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}(\tilde{\Delta}))\right\} + \|\Delta_{m-r+2:m-r+s+1}\|_2^2\,\lambda_{\min}\{\Lambda(\boldsymbol{\theta}(\tilde{\Delta}))\} \\
\geq{}& \|\Delta\|_2^2\frac{c_H}{2} - \|\Delta\|_2^2\lambda_n\kappa_0 \\
={}& \left(\frac{c_H}{2} - \lambda_n\kappa_0\right)\|\Delta\|_2^2 \\
\geq{}& \frac{c_H}{4}\|\Delta\|_2^2
\end{aligned}$$

for sufficiently large n due to (A2). Given $\Delta \in \partial\mathcal{N}_\tau$, if $\Delta'\nabla^2\overline{Q}_n(\tilde{\Delta})\Delta \geq \frac{c_H}{4}\|\Delta\|_2^2$ then by the Cauchy Schwarz inequality we have

$$\begin{aligned}
\overline{Q}_n(\Delta) ={}& \overline{Q}_n(\mathbf{0}) + \Delta'\nabla\overline{Q}_n(\mathbf{0}) + \frac{1}{2}\Delta'\nabla^2\overline{Q}_n(\tilde{\Delta})\Delta \\
\geq{}& \overline{Q}_n(\mathbf{0}) - \|\Delta\|_2\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2 + \frac{c_H}{8}\|\Delta\|_2^2 \\
={}& \overline{Q}_n(\mathbf{0}) - \tau\sqrt{\frac{s+m-r+1}{n}}\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2 + \frac{c_H}{8}\frac{\tau^2(s+m-r+1)}{n} \\
={}& \overline{Q}_n(\mathbf{0}) + \tau\sqrt{\frac{s+m-r+1}{n}}\left(-\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2 + \frac{c_H}{8}\tau\sqrt{\frac{s+m-r+1}{n}}\right).
\end{aligned}$$

This implies that $\overline{Q}_n(\Delta) > \overline{Q}_n(\mathbf{0})$ for any $\Delta \in \partial\mathcal{N}_\tau$ if $\tilde{\mathcal{E}}_{n,1}$ and $\tilde{\mathcal{E}}_{n,2} := \left\{\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2 < \frac{c_H}{8}\tau\sqrt{\frac{s+m-r+1}{n}}\right\}$ hold and n is sufficiently large.

Lemma S.1 provides that

$$\begin{aligned} P(\tilde{\mathcal{E}}_{n,1}^c) \leq\; & 2(s+m)^2 \exp\left(-\frac{nc_H^2}{2(s+m+1)^2 O(1)}\right) + 4(s+m)\exp\left(-\frac{nc_H^2 \gamma^{*2}}{2(s+m+1)^2 O(1)}\right) \\ & + 2\exp\left(-\frac{n}{2}\min\left\{\frac{c_H\gamma^{*2}}{16(s+m+1)}, \frac{c_H^2\gamma^{*4}}{4(s+m+1)^2 O(1)}\right\}\right) \\ \leq\; & 2\exp\left(-\frac{n^{1/3}c_H^2}{O(1)} + 2\log(s+m)\right) + 4\exp\left(-\frac{n^{1/3}c_H^2\gamma^{*2}}{O(1)} + \log(s+m)\right) \\ & + 2\exp\left(-\min\left\{\frac{n^{2/3}c_H\gamma^{*2}}{O(1)}, \frac{n^{1/3}c_H^2\gamma^{*4}}{O(1)}\right\}\right), \end{aligned}$$

if (A4) is satisfied and $(s+m)^3 = o(n)$. As such, $P(\tilde{\mathcal{E}}_{n,1}^c) \to 0$ under these assumptions. Applying Markov's inequality, we find

$$P(\tilde{\mathcal{E}}_{n,2}^c) = P\left(\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2 \geq \frac{c_H\tau}{8}\sqrt{\frac{s+m-r+1}{n}}\right) \leq \frac{64n\,\mathrm{E}\left[\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2^2\right]}{c_H^2\tau^2(s+m-r+1)}.$$

Suppose we can show that

$$\mathrm{E}\left[\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2^2\right] = O\left(\frac{s+m-r+1}{n}\right). \tag{S1.2}$$

Then the union bound yields

$$\begin{aligned} P(\mathcal{E}_n) &\geq P(\tilde{\mathcal{E}}_{n,1} \cap \tilde{\mathcal{E}}_{n,2}) \\ &\geq 1 - P(\tilde{\mathcal{E}}_{n,1}^c) - P(\tilde{\mathcal{E}}_{n,2}^c) \\ &\geq 1 - o(1) - \frac{64n\,\mathrm{E}\left[\left\|\nabla\overline{Q}_n(\mathbf{0})\right\|_2^2\right]}{c_H^2\tau^2(s+m-r+1)} \\ &= 1 - o(1) - O\left(\frac{1}{\tau^2}\right). \end{aligned}$$

This is sufficient to show that for any $\epsilon > 0$, we can find $N \in \mathbb{N}$ and $\tau \leq \sqrt{\log(N)}/2$ such that $P(\mathcal{E}_n) > 1 - \epsilon$ for all $n > N$. By our earlier argument, this implies that there exists a local minimizer $\hat{\boldsymbol{\theta}}$ of $Q_n(\boldsymbol{\theta})$ with constraints $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$ which satisfies

$\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$.

All that remains for part (i) is to prove (S1.2). Let $D(\boldsymbol{\theta}(\Delta))$ denote the Jacobian of $\boldsymbol{\theta}(\Delta)$ and $\nabla_\Delta \ell_n(\boldsymbol{\theta}(\Delta))$ and $\nabla_{\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}} \ell_n(\boldsymbol{\theta}(\Delta))$ denote the gradient of $\ell_n(\boldsymbol{\theta}(\Delta))$ with respect to Δ and $\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}$, respectively. Applying the chain rule, we have

$$\begin{aligned}
\nabla_\Delta \ell_n(\boldsymbol{\theta}(\Delta)) &= [D(\boldsymbol{\theta}(\Delta))]' \nabla_{\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}} \ell_n(\boldsymbol{\theta}(\Delta)) \\
&= \begin{bmatrix} \mathbf{Z}_{1:m} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_s \\ \mathbf{Z}_{m+1} & \mathbf{0} \end{bmatrix}' \nabla_{\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}} \ell_n(\boldsymbol{\theta}(\Delta)) \\
&= \mathbf{L}' \nabla_{\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}} \ell_n(\boldsymbol{\theta}(\Delta))
\end{aligned}$$

for any Δ. As such, $\nabla_\Delta \ell_n(\boldsymbol{\theta}(\mathbf{0})) = \mathbf{L}'\nabla_{\boldsymbol{\theta}_{\mathcal{M}'\cup\mathcal{S}}} \ell_n(\boldsymbol{\theta}(\mathbf{0})) = \mathbf{L}'\nabla_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\tilde{\boldsymbol{\theta}}^*)$. We find that

$$\begin{aligned}
\nabla_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\tilde{\boldsymbol{\theta}}^*) - \nabla_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*) &= \frac{1}{n}\begin{bmatrix} \mathbf{X}'_{1,\mathcal{M}\cup\mathcal{S}}\mathbf{X}_{1,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) + \mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}}(\boldsymbol{g}(\tilde{\boldsymbol{\delta}}^*) - \boldsymbol{g}(\boldsymbol{\delta}^*)) \\ -\mathbf{y}'_1\mathbf{X}_{1,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \end{bmatrix} \\
&= \frac{1}{n}\begin{bmatrix} \mathbf{X}'_{1,\mathcal{M}\cup\mathcal{S}} \\ -\mathbf{y}'_1 \end{bmatrix} \mathbf{X}_{1,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) + \frac{1}{n}\begin{bmatrix} \mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}} \\ -\mathbf{y}'_0 \end{bmatrix} (\boldsymbol{g}(\tilde{\boldsymbol{\delta}}^*) - \boldsymbol{g}(\boldsymbol{\delta}^*)) \\
&= \frac{1}{n}\begin{bmatrix} \mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \boldsymbol{g}(\tilde{\boldsymbol{\delta}}^*) - \boldsymbol{g}(\boldsymbol{\delta}^*) \\ \mathbf{X}_{1,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \end{bmatrix}.
\end{aligned}$$

Let $i \in \{1, \ldots, n_0\}$. By the Mean Value Theorem, there exists $\bar{\boldsymbol{\delta}}^{(i)}$ on the line between $\tilde{\boldsymbol{\delta}}^*$ and $\boldsymbol{\delta}^*$ such that $g(-\mathbf{x}'_i\tilde{\boldsymbol{\delta}}^*) - g(-\mathbf{x}'_i\boldsymbol{\delta}^*) = g'(-\mathbf{x}'_i\bar{\boldsymbol{\delta}}^{(i)})(-\mathbf{x}'_i(\tilde{\boldsymbol{\delta}}^* - \boldsymbol{\delta}^*)) = h(-\mathbf{x}'_i\bar{\boldsymbol{\delta}}^{(i)})\mathbf{x}'_{i,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}})$. Define $\mathbf{D}(\bar{\boldsymbol{\delta}}) := \mathrm{diag}(h(-\mathbf{x}'_1\bar{\boldsymbol{\delta}}^{(1)}), \ldots, h(-\mathbf{x}'_{n_0}\bar{\boldsymbol{\delta}}^{(n_0)}))$. Then we have

$$\begin{aligned}
\boldsymbol{g}(\tilde{\boldsymbol{\delta}}^*) - \boldsymbol{g}(\boldsymbol{\delta}^*) &= \begin{bmatrix} h(-\mathbf{x}'_1\bar{\boldsymbol{\delta}}^{(1)})\mathbf{x}'_{1,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \\ \vdots \\ h(-\mathbf{x}'_{n_0}\bar{\boldsymbol{\delta}}^{(n_0)})\mathbf{x}'_{n_0,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \end{bmatrix} \\
&= \mathbf{D}(\bar{\boldsymbol{\delta}})\mathbf{X}_{0,\mathcal{M}}(\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}).
\end{aligned}$$

We plug this into our previous expression for $\nabla \ell_n(\tilde{\boldsymbol{\theta}}^*) - \nabla \ell_n(\boldsymbol{\theta}^*)$ to find

$$\begin{aligned}
\nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\tilde{\boldsymbol{\theta}}^*) - \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) &= \frac{1}{n} \begin{bmatrix} \mathbf{X}'_{(\mathcal{M} \cup \mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \mathbf{D}(\bar{\boldsymbol{\delta}}) \mathbf{X}_{0,\mathcal{M}} (\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \\ \mathbf{X}_{1,\mathcal{M}} (\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \end{bmatrix} \\
&= \frac{1}{n} \begin{bmatrix} \mathbf{X}'_{(\mathcal{M} \cup \mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \mathbf{D}(\bar{\boldsymbol{\delta}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix} \mathbf{X}_{(\mathcal{M})} (\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}) \\
&= \frac{1}{n} \begin{bmatrix} \mathbf{X}'_{(\mathcal{M} \cup \mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \mathbf{D}(\bar{\boldsymbol{\delta}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(\mathcal{M} \cup \mathcal{S})} & -\mathbf{y} \end{bmatrix} (\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}' \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}' \cup \mathcal{S}}) \\
&=: \frac{1}{n} \mathbf{H}(\bar{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}' \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}' \cup \mathcal{S}}).
\end{aligned}$$

Based on the previous expression, we apply the Cauchy-Schwarz inequality to find

$$\begin{aligned}
\mathrm{E}\left[\left\| \nabla \bar{Q}_n(\mathbf{0}) \right\|_2^2 \right] &= \mathrm{E}\left[\left\| \mathbf{L}' \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\tilde{\boldsymbol{\theta}}^*) + \begin{bmatrix} \mathbf{0} \\ \lambda_n \bar{\boldsymbol{\rho}}(\boldsymbol{\delta}^*_{\mathcal{S}}; \lambda_n) \\ 0 \end{bmatrix} \right\|_2^2 \right] \\
&\leq 3\,\mathrm{E}\left[\| \mathbf{L}' \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \|_2^2 \right] + 3\,\mathrm{E}\left[\left\| \frac{1}{n} \mathbf{L}' \mathbf{H}(\bar{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}' \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}' \cup \mathcal{S}}) \right\|_2^2 \right] \\
&\quad + 3 \left\| \lambda_n \bar{\boldsymbol{\rho}}(\boldsymbol{\delta}^*_{\mathcal{S}}; \lambda_n) \right\|_2^2. \qquad \text{(S1.3)}
\end{aligned}$$

We will bound the three terms in (S1.3) in turn. Leveraging that the Tobit model is an exponential family, that $\|\mathbf{L}\boldsymbol{v}\|_2 = \|\boldsymbol{v}\|_2$, and Weyl's inequality, we find

$$\begin{aligned}
\mathrm{E}\left[\| \mathbf{L}' \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \|_2^2 \right] &= \mathrm{E}\left[\nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*)' \mathbf{L} \mathbf{L}' \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right] \\
&= \mathrm{E}\left[\mathrm{tr}\left\{ \mathbf{L}' \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*)' \mathbf{L} \right\} \right] \\
&= \mathrm{tr}\left\{ \mathbf{L}'\, \mathrm{E}\left[\nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \nabla_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*)' \right] \mathbf{L} \right\} \\
&= \frac{1}{n^2} \mathrm{tr}\left\{ \mathbf{L}'\, \mathrm{E}\left[-\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \log L_n(\boldsymbol{\theta}^*) \right] \mathbf{L} \right\} \\
&\leq \frac{s + m - r + 1}{n^2} \lambda_{\max} \left\{ \mathbf{L}'\, \mathrm{E}\left[-\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \log L_n(\boldsymbol{\theta}^*) \right] \mathbf{L} \right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{s+m-r+1}{n^2}\lambda_{\max}\left\{\mathrm{E}\left[-\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)\right]\right\} \\
&= \frac{s+m-r+1}{n^2}\lambda_{\max}\left\{\mathrm{E}\left[\mathbf{H}(\boldsymbol{\theta}^*)\right] + \mathrm{E}\left[\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & n_1\gamma^{*-2}\end{bmatrix}\right]\right\} \\
&\leq \frac{s+m-r+1}{n^2}\left(\lambda_{\max}\left\{\mathrm{E}\left[\mathbf{H}(\boldsymbol{\theta}^*)\right]\right\} + \lambda_{\max}\left\{E\left[\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & n_1\gamma^{*-2}\end{bmatrix}\right]\right\}\right) \\
&= \frac{s+m-r+1}{n^2}\left(\lambda_{\max}\left\{\mathrm{E}\left[\mathbf{H}(\boldsymbol{\theta}^*)\right]\right\} + \mathrm{E}[n_1\gamma^{*-2}]\right) \\
&= O\left(\frac{s+m-r+1}{n}\right)
\end{aligned}
$$

where the final equality follows from (A3) and the fact that $\mathrm{E}[n_1\gamma^{*-2}] = O(n)$.

Moving on to the second term in (S1.3), we see that

$$
\begin{aligned}
\mathrm{E}\left[\left\|\frac{1}{n}\mathbf{L}'\mathbf{H}(\bar{\boldsymbol{\theta}})(\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})\right\|_2^2\right] &\leq \frac{1}{n^2}\left\|\mathbf{L}'\right\|_2^2 \mathrm{E}\left[\left\|\mathbf{H}(\bar{\boldsymbol{\theta}})\right\|_2^2\right]\left\|\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2 \\
&\leq \frac{1}{n^2}\left\|\mathbf{L}'\right\|_2^2 \mathrm{E}\left[\lambda^2_{\max}\{\mathbf{H}(\bar{\boldsymbol{\theta}})\}\right]\left\|\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2.
\end{aligned}
$$

We find $\|\mathbf{L}'\|_2^2 = \lambda_{\max}\{\mathbf{L}'\mathbf{L}\} = \lambda_{\max}\{\mathbf{I}_{s+m-r+1}\} = 1$. We've already established that $\left\|\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2 = \left\|\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}\right\|_2^2 = O\left(\frac{s+m-r+1}{n}\right)$. Suppose that $\boldsymbol{v} \in \mathbb{R}^{p+2}$ with $\|\boldsymbol{v}\|_2 = 1$. We see that

$$
\begin{aligned}
&\boldsymbol{v}'\begin{bmatrix}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\\ -\mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y}\end{bmatrix}\boldsymbol{v} - \boldsymbol{v}'\mathbf{H}(\bar{\boldsymbol{\theta}})\boldsymbol{v} \\
&\qquad\qquad = \boldsymbol{v}'\begin{bmatrix}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\\ -\mathbf{y}'\end{bmatrix}\left(\mathbf{I}_n - \begin{bmatrix}\mathbf{D}(\bar{\boldsymbol{\delta}}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_1}\end{bmatrix}\right)\begin{bmatrix}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y}\end{bmatrix}\boldsymbol{v}.
\end{aligned}
$$

Sampford (1953) showed that $0 < h(s) < 1$ for all $s \in \mathbb{R}$. As such, $\mathbf{I}_n - \begin{bmatrix}\mathbf{D}(\bar{\boldsymbol{\delta}}) & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_{n_1}\end{bmatrix}$ is a diagonal matrix with non-negative diagonal entries and, by extension,

$$\boldsymbol{v}' \begin{bmatrix} \mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \left(\mathbf{I}_n - \begin{bmatrix} \mathbf{D}(\bar{\boldsymbol{\delta}}) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1} \end{bmatrix} \right) \begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix} \boldsymbol{v} \geq 0.$$

Since this holds for all $\boldsymbol{v}$ with $\|\boldsymbol{v}\|_2 = 1$, we see that

$$\lambda_{\max}\left\{\mathbf{H}(\bar{\boldsymbol{\theta}})\right\} \leq \lambda_{\max}\left\{ \begin{bmatrix} \mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix} \right\}.$$

Pulling these findings together, we see that if condition (A3) holds, then

$$\begin{aligned} \frac{1}{n^2}\|\mathbf{L}'\|_2^2 \,\mathrm{E}\left[\lambda^2_{\max}\{\mathbf{H}(\bar{\boldsymbol{\theta}})\}\right] \left\|\tilde{\boldsymbol{\theta}}^*_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2 & \\ \leq \frac{1}{n^2}\,\mathrm{E}\left[\lambda^2_{\max}\left\{ \begin{bmatrix} \mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})} \\ -\mathbf{y}' \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y} \end{bmatrix} \right\}\right] \left\|\tilde{\boldsymbol{\delta}}^*_{\mathcal{M}} - \boldsymbol{\delta}^*_{\mathcal{M}}\right\|_2^2 & \\ = O\left(\frac{s+m-r+1}{n}\right). & \end{aligned}$$

To bound the third term in (S1.3), we note that $\|\lambda_n \bar{\boldsymbol{\rho}}(\boldsymbol{\delta}^*_{\mathcal{S}}; \lambda_n)\|_2^2 \leq s(\lambda_n \rho'(d_n; \lambda_n))^2$ since $\rho'(t;\lambda_n)$ is non-increasing. Therefore if (A2) holds, then $\|\lambda_n \bar{\boldsymbol{\rho}}(\boldsymbol{\delta}^*_{\mathcal{S}}; \lambda_n)\|_2^2 = o\left(\frac{1}{n}\right)$. With that, we have bounded all three terms in (S1.3) and shown that (S1.2) holds, completing this portion of the proof.

(ii) We've shown that there exists a local minimizer $\hat{\boldsymbol{\theta}}$ of $Q_n(\boldsymbol{\theta})$ with constraints $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$ such that $\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$. We will now show that, with probability converging to 1, this minimizer $\hat{\boldsymbol{\theta}}$ is also a local minimizer of $Q_n(\boldsymbol{\theta})$ with only the constraint $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$.

Per Lemma S.2, it is sufficient to show that

$$\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\hat{\boldsymbol{\theta}})\right\|_{\max} < \lambda_n \rho'(0^+;\lambda_n)$$

with probability converging to 1, as (L2.1) and (L2.3) are clearly satisfied by $\hat{\boldsymbol{\theta}}$. In part (i), we showed that for some constant $\bar{c} > 0$ and any $\tau \leq \frac{\sqrt{\log n}}{2}$, $\hat{\boldsymbol{\theta}}$ satisfies

$$\mathcal{F}_n := \left\{ \left\| \hat{\boldsymbol{\theta}}_{\mathcal{M}' \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}' \cup \mathcal{S}} \right\|_2 \leq \bar{c}\tau \sqrt{\frac{s+m-r+1}{n}} \right\}$$

with probability at least $1 - o(1) - O\left(\frac{1}{\tau^2}\right)$. By the Mean Value Theorem, we know that for $j = 0, 1, \ldots, p+1$

$$\nabla_j \ell_n(\hat{\boldsymbol{\theta}}) = \nabla_j \ell_n(\boldsymbol{\theta}^*) + \left[\nabla^2 \ell_n(\boldsymbol{\theta}^*)\right]_j (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + R_j(\tilde{\boldsymbol{\theta}}^{(j)})$$

where $[\nabla^2 \ell_n(\boldsymbol{\theta}^*)]_j$ denotes the jth row of $\nabla^2 \ell_n(\boldsymbol{\theta}^*)$, $\tilde{\boldsymbol{\theta}}^{(j)}$ lies on the line segment joining $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}$, and $R_j(\tilde{\boldsymbol{\theta}}^{(j)}) = \left[\nabla^2 \ell_n(\tilde{\boldsymbol{\theta}}^{(j)}) - \nabla^2 \ell_n(\boldsymbol{\theta}^*)\right]_j (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$. We define $\mathbf{R}(\tilde{\boldsymbol{\theta}}) = (R_0(\tilde{\boldsymbol{\theta}}^{(0)}), \ldots, R_{p+1}(\tilde{\boldsymbol{\theta}}^{(p+1)}))'$ for ease of notation. This gives us

$$\nabla_{(\mathcal{M}' \cup \mathcal{S})^c} \ell_n(\hat{\boldsymbol{\theta}}) = \nabla_{(\mathcal{M}' \cup \mathcal{S})^c} \ell_n(\boldsymbol{\theta}^*) + \left[\nabla^2 \ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}' \cup \mathcal{S})^c} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) + \mathbf{R}_{(\mathcal{M}' \cup \mathcal{S})^c}(\tilde{\boldsymbol{\theta}}). \tag{S1.4}$$

We will bound (with high probability) the max-norm of each of the terms in (S1.4) in turn.

We start with $\left\| \mathbf{R}_{(\mathcal{M}' \cup \mathcal{S})^c}(\tilde{\boldsymbol{\theta}}) \right\|_{\max}$. For any $j \in \{0, 1, \ldots, p\}$, we see that

$$\begin{aligned}
R_j(\tilde{\boldsymbol{\theta}}^{(j)}) &= \frac{1}{n} \mathbf{X}'_{(j)} \begin{bmatrix} \mathbf{D}(\tilde{\boldsymbol{\delta}}^{(j)}) - \mathbf{D}(\boldsymbol{\delta}^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{X} & -\mathbf{y} \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\
&= \frac{1}{n} \mathbf{X}'_{(j)} \left[\mathbf{D}(\tilde{\boldsymbol{\delta}}^{(j)}) - \mathbf{D}(\boldsymbol{\delta}^*)\right] \begin{bmatrix} \mathbf{X}_0 & -\mathbf{y}_0 \end{bmatrix} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \\
&= \frac{1}{n} \mathbf{X}'_{(j)} \left[\mathbf{D}(\tilde{\boldsymbol{\delta}}^{(j)}) - \mathbf{D}(\boldsymbol{\delta}^*)\right] \mathbf{X}_{0, \mathcal{M} \cup \mathcal{S}} (\hat{\boldsymbol{\theta}}_{\mathcal{M} \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M} \cup \mathcal{S}})
\end{aligned}$$

since $\hat{\boldsymbol{\theta}}_{(\mathcal{M}' \cup \mathcal{S})^c} = \boldsymbol{\theta}^*_{(\mathcal{M}' \cup \mathcal{S})^c} = \mathbf{0}$ and $\mathbf{y}_0 = \mathbf{0}$. Applying the Mean Value Theorem again, we find

$$|R_j(\tilde{\boldsymbol{\theta}}^{(j)})| \leq \frac{1}{n} (\hat{\boldsymbol{\theta}}_{\mathcal{M} \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M} \cup \mathcal{S}})' \mathbf{X}'_{0, \mathcal{M} \cup \mathcal{S}} \operatorname{diag}\{|\mathbf{X}_{0,j}| \circ |\boldsymbol{g}''(\bar{\boldsymbol{\delta}}^{(j)})|\} \mathbf{X}_{0, \mathcal{M} \cup \mathcal{S}} (\hat{\boldsymbol{\theta}}_{\mathcal{M} \cup \mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M} \cup \mathcal{S}})$$

where $\bar{\boldsymbol{\delta}}^{(j)}$ is on the line segment joining $\boldsymbol{\delta}^*$ and $\tilde{\boldsymbol{\delta}}^{(j)}$ and $\boldsymbol{v} \circ \boldsymbol{w}$ denotes the element-wise vector product. Lemma S.6 of Jacobson and Zou (2023) establishes that $|g''(s)| < 4.3$ for all $s \in \mathbb{R}$. Combining this with our last expression, we have

$$|R_j(\tilde{\boldsymbol{\theta}}^{(j)})| \leq 4.3\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}}\operatorname{diag}\{|\mathbf{X}_{0,j}|\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\}\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}\cup\mathcal{S}}\right\|_2^2.$$

Jacobson and Zou (2023) show in the proof of their Theorem S.4 that

$$\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}}\operatorname{diag}\{|\mathbf{X}_{0,j}|\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\} \leq \lambda_{\max}\left\{\frac{1}{n}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}\right\}. \quad \text{(S1.5)}$$

As such, if condition (A3) is satisfied and $\mathcal{F}_n$ holds, then

$$\begin{aligned}|R_j(\tilde{\boldsymbol{\theta}}^{(j)})| &\leq 4.3\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}\right\}\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}\cup\mathcal{S}}\right\|_2^2 \\ &= O\left(\frac{\tau^2(s+m)}{n}\right) \quad \text{(S1.6)}\end{aligned}$$

for any $j \in (\mathcal{M}' \cup \mathcal{S})^c$ and, by extension, $\left\|\mathbf{R}_{(\mathcal{M}'\cup\mathcal{S})^c}(\tilde{\boldsymbol{\theta}})\right\|_{\max} = O\left(\frac{\tau^2(s+m)}{n}\right)$.

We will now bound $\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta}^*)\right\|_{\max}$ with high probability. Lemma S.1 in Jacobson and Zou (2023) provides that for $j = 0, \ldots, p$, $\nabla_j \log L_n(\boldsymbol{\theta}^*) \sim \operatorname{subG}\left(\left\|\mathbf{X}_{(j)}\right\|_2^2\right)$. Applying a Chernoff bound, we see that for any $c > 0$

$$P\left(|\nabla_j \ell_n(\boldsymbol{\theta}^*)| > c\right) = P\left(|\nabla_j \log L_n(\boldsymbol{\theta}^*)| > nc\right) \leq 2\exp\left(-\frac{n^2c^2}{2\left\|\mathbf{X}_{(j)}\right\|_2^2}\right).$$

The union bound then provides that for any constant $k > 0$

$$\begin{aligned}P\left(\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta}^*)\right\|_{\max} > \frac{k\sqrt{\log p}}{\sqrt{n}}\right) &\leq \sum_{j\in(\mathcal{M}'\cup\mathcal{S})^c} P\left(|\nabla_j \ell_n(\boldsymbol{\theta}^*)| > \frac{k\sqrt{\log p}}{\sqrt{n}}\right) \\ &\leq 2(p-s-m+1)\exp\left(-\frac{n^2\left(\frac{k\sqrt{\log p}}{\sqrt{n}}\right)^2}{2\max_j\left\|\mathbf{X}_{(j)}\right\|_2^2}\right)\end{aligned}$$

$$\leq 2\exp\left(-\frac{k^2\log p}{O(1)} + \log p\right)$$

if condition (A4) holds. As such, for sufficiently large k, $P\left(\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta}^*)\right\|_{\max} \leq \frac{k\sqrt{\log p}}{\sqrt{n}}\right)$ $\to 1$ as $n, p \to \infty$. We define this key event $\mathcal{G}_n := \left\{\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta}^*)\right\|_{\max} \leq \frac{k\sqrt{\log p}}{\sqrt{n}}\right\}$, where k is large enough that $P(\mathcal{G}_n) \to 1$, for later reference.

Lastly, we will bound $\left\|[\nabla^2\ell_n(\boldsymbol{\theta}^*)]_{(\mathcal{M}'\cup\mathcal{S})^c}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right\|_{\max}$ with high probability. We note that since $\hat{\boldsymbol{\theta}}_{(\mathcal{M}'\cup\mathcal{S})^c} = \boldsymbol{\theta}^*_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$, we have $[\nabla^2\ell_n(\boldsymbol{\theta}^*)]_{(\mathcal{M}'\cup\mathcal{S})^c}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) =$ $[\nabla^2\ell_n(\boldsymbol{\theta}^*)]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})$. We find

$$\begin{aligned}\left\|\left[\nabla^2\ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})\right\|_{\max} &\leq \left\|\left[\nabla^2\ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}}\right\|_{\infty}\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_{\max} \\ &\leq \left\|\left[\nabla^2\ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}}\right\|_{\infty}\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2.\end{aligned}$$

Define $\mathbf{W} := n\,[\nabla^2\ell_n(\boldsymbol{\theta}^*)]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}} = \mathbf{X}'_{((\mathcal{M}\cup\mathcal{S})^c)}\begin{bmatrix}\mathbf{D}(\boldsymbol{\delta}^*) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n_1}\end{bmatrix}\begin{bmatrix}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y}\end{bmatrix}$. Consider the event $\mathcal{H}_n := \left\{\left\|\frac{1}{n}\mathbf{W} - \mathrm{E}\left[\frac{1}{n}\mathbf{W}\right]\right\|_{\max} \leq (\log p)^{1/2}(s+m+1)^{-3/2}\right\}$. Under $\mathcal{H}_n$, we see that

$$\begin{aligned}\left\|\frac{1}{n}\mathbf{W}\right\|_{\infty} &\leq (s+m+1)\left\|\frac{1}{n}\mathbf{W} - \mathrm{E}\left[\frac{1}{n}\mathbf{W}\right]\right\|_{\max} + \left\|\mathrm{E}\left[\frac{1}{n}\mathbf{W}\right]\right\|_{\infty} \\ &\leq \frac{\sqrt{\log p}}{\sqrt{s+m+1}} + \left\|\mathrm{E}\left[\frac{1}{n}\mathbf{W}\right]\right\|_{\infty}.\end{aligned}$$

Therefore if (A3) and $\mathcal{F}_n$ hold, then

$$\begin{aligned}\left\|\left[\nabla^2\ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}'\cup\mathcal{S})^c,\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})\right\|_{\max} &\leq \frac{\sqrt{\log p}}{\sqrt{s+m+1}}\bar{c}\tau\sqrt{\frac{s+m}{n}} + O\left(\tau\sqrt{\frac{s+m}{n}}\right) \\ &= O\left(\frac{\tau\sqrt{\log p}}{\sqrt{n}}\right) + O\left(\frac{\tau\sqrt{s+m}}{\sqrt{n}}\right).\end{aligned}$$

Lemma S.4 of Jacobson and Zou (2023) can be adapted to show

$$P(\mathcal{H}_n^c) \leq 2(s+m)(p-m-s+1)\exp\left(-\frac{2n^2\left(\frac{(\log p)^{1/2}}{(s+m)^{3/2}}\right)^2}{\max_{j,k}\sum_{i=1}^n x_{ij}^2 x_{ik}^2}\right)$$

$$
\begin{aligned}
&+ 2(p-m-s+1)\exp\left(-\frac{2n^2\left(\frac{(\log p)^{1/2}}{(s+m)^{3/2}}\right)^2\gamma^{*2}}{\max_j\sum_{i=1}^n x_{ij}^2(2+\mathbf{x}_i'\boldsymbol{\delta}^*+g(-\mathbf{x}_i'\boldsymbol{\delta}^*))^2}\right)\\
&\leq 2(s+m)(p-m-s+1)\exp\left(-\frac{2\log p}{\frac{1}{n}\max_{j,k}\sum_{i=1}^n x_{ij}^2x_{ik}^2\frac{1}{n}(s+m)^3}\right)\\
&+ 2(p-m-s+1)\exp\left(-\frac{2\gamma^{*2}\log p}{\frac{1}{n}\max_j\sum_{i=1}^n x_{ij}^2(2+\mathbf{x}_i'\boldsymbol{\delta}^*+g(-\mathbf{x}_i'\boldsymbol{\delta}^*))^2\frac{1}{n}(s+m)^3}\right)\\
&\leq 2\exp\left(-\frac{2\log p}{o(1)}+2\log p\right)+2\exp\left(-\frac{2\gamma^{*2}\log p}{o(1)}+\log p\right)
\end{aligned}
$$

if condition (A4) holds and $(s+m)^3 = o(n)$. As such, $P(\mathcal{H}_n) \to 1$ as $n, p \to \infty$.

Let τ_n be a diverging sequence satisfying $\tau_n \leq \min\{\sqrt{\log n}/2, \sqrt{\log p}\}$ and $\tau_n \max\{\sqrt{s+m}, \sqrt{\log p}\}/\sqrt{n} = o(\lambda_n)$. Returning to (S1.4), we've shown that together $\mathcal{F}_n$, $\mathcal{G}_n$, and $\mathcal{H}_n$ imply

$$
\begin{aligned}
\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\hat{\boldsymbol{\theta}})\right\|_{\max} &\leq \left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta}^*)\right\|_{\max} + \left\|\left[\nabla^2\ell_n(\boldsymbol{\theta}^*)\right]_{(\mathcal{M}'\cup\mathcal{S})^c}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)\right\|_{\max} + \left\|\mathbf{R}_{(\mathcal{M}'\cup\mathcal{S})^c}(\tilde{\boldsymbol{\theta}})\right\|_{\max}\\
&= O\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right) + O\left(\frac{\tau_n\sqrt{\log p}}{\sqrt{n}}\right) + O\left(\frac{\tau_n\sqrt{s+m}}{\sqrt{n}}\right) + O\left(\frac{\tau_n^2(s+m)}{n}\right)\\
&= o(\lambda_n) + o(\lambda_n) + o(\lambda_n) + O\left(\frac{\tau_n\sqrt{\log p}}{n^{2/3}}\right)\\
&= o(\lambda_n)
\end{aligned}
$$

provided that (A2) is satisfied and $(s+m)^3 = o(n)$. By extension, $\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\hat{\boldsymbol{\theta}})\right\|_{\max} < \lambda_n\rho'(0^+;\lambda_n)$ for sufficiently large n. Therefore

$$
P\left(\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\hat{\boldsymbol{\theta}})\right\|_{\max} < \lambda_n\rho'(0^+;\lambda_n)\right) \geq P(\mathcal{F}_n\cap\mathcal{G}_n\cap\mathcal{H}_n) \geq 1 - o(1) - O\left(\frac{1}{\tau_n^2}\right) \to 1,
$$

as $n, p \to \infty$ if our assumptions hold. As such, we can conclude that $\hat{\boldsymbol{\theta}}$ is a local minimizer of $Q_n(\boldsymbol{\theta})$ subject to $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$ with probability converging to 1 as $n \to \infty$.

Note: In part (ii) we've assumed that p diverges with n. The proof can easily be adapted to handle the case where p does not diverge: simply replace $\log p$ with $\log n$ in $\mathcal{G}_n$, $\mathcal{H}_n$, (A2) and the requirements for τ_n.

(iii) Together, parts (i) and (ii) imply that there exists a local minimizer $\hat{\boldsymbol{\theta}}_0$ of (3.4) satisfying $\left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m-r+1}{n}}\right)$. In particular, $\hat{\boldsymbol{\theta}}_0 = \hat{\boldsymbol{\theta}}$ under $\mathcal{F}_n \cap \mathcal{G}_n \cap \mathcal{H}_n$.

Define $\mathcal{I}_n := \left\{\left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)\right]\right\|_{\max} \le \frac{c_H}{2(s+m+1)}\right\}$. We see that $P(\mathcal{I}_n) \to 1$ by Lemma S.1. Likewise, define $\mathcal{J}_n := \left\{\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_{\max} \le k\sqrt{\frac{\log(s+m)}{n}}\right\}$ where k is large enough that $P(\mathcal{J}_n) \to 1$ per (S2.54) in the proof of Lemma S.3. We will now show that if $\mathcal{F}_n$, $\mathcal{G}_n$, $\mathcal{H}_n$, $\mathcal{I}_n$, and $\mathcal{J}_n$ hold, with τ in $\mathcal{F}_n$ satisfying $0 < \tau \le \min\{\sqrt{\log n}/2, \sqrt{\log p}\}$, then

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{0,\mathcal{M}'\cup\mathcal{S}}) =& \frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}(\mathbf{I}_{s+m+1} - \mathbf{P}_n)\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) \\ &- \sqrt{n}\gamma^*\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\begin{bmatrix}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ \mathbf{0}\end{bmatrix} + o(1)\end{aligned} \tag{S1.7}$$

under the ℓ_2 norm. This is sufficient to prove (3.7) since $P(\mathcal{F}_n \cap \mathcal{G}_n \cap \mathcal{H}_n) \ge 1 - o(1) - O\left(\frac{1}{\tau^2}\right)$ as we showed in part (ii), implying that for any $\epsilon > 0$ we can find $N \in \mathbb{N}$ and a corresponding τ such that $P(\mathcal{F}_n \cap \mathcal{G}_n \cap \mathcal{H}_n \cap \mathcal{I}_n \cap \mathcal{J}_n) > 1 - \epsilon$ for all $n > N$.

For the remainder of this proof, we suppose that $\mathcal{F}_n$, $\mathcal{G}_n$, $\mathcal{H}_n$, $\mathcal{I}_n$, and $\mathcal{J}_n$ hold and that $0 < \tau \le \min\{\sqrt{\log n}/2, \sqrt{\log p}\}$. In parts (i) and (ii) we showed that, under $\mathcal{F}_n \cap \mathcal{G}_n \cap \mathcal{H}_n$, $\hat{\boldsymbol{\theta}}_0$ is the local minimizer of $Q_n(\boldsymbol{\theta})$ with constraints $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$, $\boldsymbol{\theta}_{(\mathcal{M}'\cap\mathcal{S})^c} = \mathbf{0}$. As a consequence of this, there exists $\boldsymbol{\nu} \in \mathbb{R}^r$ such that

$$\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \begin{bmatrix}\frac{1}{\sqrt{n}}\mathbf{C}'\boldsymbol{\nu} \\ -\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n) \\ -\frac{1}{\sqrt{n}}\boldsymbol{t}'\boldsymbol{\nu}\end{bmatrix}. \tag{S1.8}$$

Applying the Mean Value Theorem componentwise as in (S1.4), we find that

$$\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) + \mathbf{R} \tag{S1.9}$$

where $R_j = [\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\tilde{\boldsymbol{\theta}}^{(j)}) - \nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)](\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}})$ and $\tilde{\boldsymbol{\theta}}^{(j)}$ lies on the line segment joining $\boldsymbol{\theta}^*$ and $\hat{\boldsymbol{\theta}}_0$.

We see that $|R_j| = O\left(\frac{\tau^2(s+m)}{n}\right) = O\left(\frac{s+m}{n}\right)$ for $j = 0, 1, \ldots, p$ by the same argument we used to prove (S1.6). Note that $R_{p+1} = \frac{n_1}{n}((\tilde{\gamma}^{(p+1)})^{-2} - \gamma^{*-2})(\hat{\gamma}_0 - \gamma^*)$ for some $\tilde{\gamma}^{(p+1)}$ between $\hat{\gamma}_0$ and γ^*. Applying the Mean Value Theorem to this expression, we know that there exists $\bar{\gamma}^{(p+1)}$ between $\tilde{\gamma}^{(p+1)}$ and γ^* such that

$$\begin{aligned}
|R_{p+1}| &\leq \left| \frac{2n_1}{n} (\bar{\gamma}^{(p+1)})^{-3} (\tilde{\gamma}^{(p+1)} - \gamma^*)(\hat{\gamma}_0 - \gamma^*) \right| \\
&\leq 2 \left| (\bar{\gamma}^{(p+1)})^{-3} \right| (\hat{\gamma}_0 - \gamma^*)^2 \\
&= \frac{O\left(\frac{s+m}{n}\right)}{\left| \left(\gamma^* + O\left(\sqrt{\frac{s+m}{n}}\right)\right)^3 \right|}. \qquad \text{(S1.10)}
\end{aligned}$$

Since $s + m = o(n^{1/3})$, we have that $\gamma^* + O\left(\sqrt{\frac{s+m}{n}}\right) = \gamma^* + o(1) \geq \frac{\gamma^*}{2}$ for large n. As such, we see that $|R_{p+1}| = O\left(\frac{s+m}{n}\right)$ and, consequently, $\|\mathbf{R}\|_2 \leq (s + m + 1)^{1/2} \|\mathbf{R}\|_{\max} = O\left(\frac{(s+m)^{3/2}}{n}\right) = o\left(\frac{1}{\sqrt{n}}\right)$ since $(s + m)^3 = o(n)$.

Turning to the second term in (S1.9), we find

$$\begin{aligned}
\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) &= \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) \\
&\quad + \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right] (\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}).
\end{aligned}$$

$\mathcal{J}_n$ implies $\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_2 \leq (s + m + 1) \|\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_{\max} = O\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$. Since $\left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O\left(\sqrt{\frac{s+m}{n}}\right)$ and $(s + m)^3 \log(s + m) = o(n)$, this implies

$$\begin{aligned}
\left\| \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right] (\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) \right\|_2 &= O\left(\frac{(s+m)^{3/2}(\log(s+m))^{1/2}}{n}\right) \\
&= o\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}$$

Applying this to (S1.9), we find

$$\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) + \mathbf{R}_0, \tag{S1.11}$$

where $\|\mathbf{R}_0\|_2 = o\left(\frac{1}{\sqrt{n}}\right)$.

Combining (S1.8) and (S1.11), we see that

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\mathbf{C}'\boldsymbol{\nu} \\ -\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n) \\ -\frac{1}{\sqrt{n}}\boldsymbol{t}'\boldsymbol{\nu} \end{bmatrix} = \nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) + \mathbf{R}_0.$$

Multiplying both sides of this expression by $\sqrt{n}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}$ and rearranging terms, we find

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) &= -\sqrt{n}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) \\ &\quad + \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\begin{bmatrix} \mathbf{C}'\boldsymbol{\nu} \\ -\sqrt{n}\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n) \\ -\boldsymbol{t}'\boldsymbol{\nu} \end{bmatrix} - \sqrt{n}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{R}_0. \end{aligned} \tag{S1.12}$$

Under $\mathcal{F}_n$ and assumption (A2), we have $\left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{S}}\right\|_{\max} \leq \bar{c}\tau\sqrt{\frac{s+m}{n}} = o(\lambda_n) = o(d_n)$. As such, for sufficiently large n we have $\min_{j\in\mathcal{S}}|\hat{\boldsymbol{\theta}}_{0,j}| > \min_{j\in\mathcal{S}}|\boldsymbol{\theta}^*_{0,j}| - d_n = d_n$. Because $\rho'(t;\lambda_n)$ is non-increasing, (A2) provides that $\left\|\sqrt{n}\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n)\right\|_2 \leq \sqrt{sn}|\lambda_n\rho'(d_n;\lambda_n)| = o(1)$. We see that (L3.3) from Lemma S.3 implies $\left\|\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\begin{bmatrix}\mathbf{0}' & \sqrt{n}\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n)' & \mathbf{0}'\end{bmatrix}'\right\|_2 \leq \lambda_{\max}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\}\left\|\sqrt{n}\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{S}};\lambda_n)\right\|_2 = o(1)$ and $\left\|\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{R}_0\right\|_2 \leq \lambda_{\max}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\}\|\mathbf{R}_0\|_2 = o\left(\frac{1}{\sqrt{n}}\right)$. As such, (S1.12) simplifies to

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) = -\sqrt{n}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\nu} + o(1). \tag{S1.13}$$

Since $\widetilde{\mathbf{C}}\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} = \mathbf{C}\hat{\boldsymbol{\delta}}_{0,\mathcal{M}} - \hat{\gamma}_{0\mathcal{M}}\boldsymbol{t} = \mathbf{0}$ and $\widetilde{\mathbf{C}}\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}} = \mathbf{C}\boldsymbol{\delta}^*_{\mathcal{M}} - \gamma^*\boldsymbol{t} = \gamma^*\boldsymbol{h}_n$, we have $\widetilde{\mathbf{C}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) = -\gamma^*\boldsymbol{h}_n$. As such, by multiplying both sides of (S1.13) by $\widetilde{\mathbf{C}}$ and rearranging terms, we find that

$$\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\nu} = -\sqrt{n}\gamma^*\boldsymbol{h}_n + \sqrt{n}\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \widetilde{\mathbf{C}}\mathbf{R}^*$$

where $\mathbf{R}^* \in \mathbb{R}^{s+m+1}$ satisfies $\|\mathbf{R}^*\|_2 = o(1)$. Recall that $\mathbf{\Psi} = \widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'$. Multiplying both sides of the previous expression by $\mathbf{\Psi}^{-1}$, we find

$$\boldsymbol{\nu} = -\sqrt{n}\gamma^*\mathbf{\Psi}^{-1}\boldsymbol{h}_n + \sqrt{n}\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{R}^*.$$

Plugging this expression for $\boldsymbol{\nu}$ back into (S1.13), we have

$$\begin{aligned}
&\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) \\
&\quad = -\sqrt{n}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) \\
&\qquad + \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\left(-\sqrt{n}\gamma^*\mathbf{\Psi}^{-1}\boldsymbol{h}_n + \sqrt{n}\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{R}^*\right) \\
&\qquad + o(1) \\
&\quad = \frac{1}{\sqrt{n}}\left(\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} - \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\right)\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) \\
&\qquad - \sqrt{n}\gamma^*\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n + \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{R}^* + o(1) \\
&\quad = \frac{1}{\sqrt{n}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\left(\mathbf{I} - \mathbf{P}_n\right)\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) \\
&\qquad - \sqrt{n}\gamma^*\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n + \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{R}^* + o(1). \qquad \text{(S1.14)}
\end{aligned}$$

From here, we aim to simplify the terms involving $\boldsymbol{h}_n$ and $\mathbf{R}^*$. Starting with the $\boldsymbol{h}_n$ term, we note that $\boldsymbol{h}_n = \mathbf{C}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n = \widetilde{\mathbf{C}}\begin{bmatrix}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ \mathbf{0}\end{bmatrix}$. Substituting this into our $\boldsymbol{h}_n$ term, we find

$$\sqrt{n}\gamma^*\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n = \sqrt{n}\gamma^*\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\begin{bmatrix}\mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ \mathbf{0}\end{bmatrix}$$

$$= \sqrt{n}\gamma^* \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2} \begin{bmatrix} \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ \mathbf{0} \end{bmatrix}$$

$$= \sqrt{n}\gamma^* \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{P}_n \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2} \begin{bmatrix} \mathbf{C}'(\mathbf{C}\mathbf{C}')^{-1}\boldsymbol{h}_n \\ \mathbf{0} \end{bmatrix}. \tag{S1.15}$$

Moving on to the $\mathbf{R}^*$ term, we apply (L3.1) and (L3.3) from Lemma S.3 to find

$$\begin{aligned}
&\left\| \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \widetilde{\mathbf{C}} \mathbf{R}^* \right\|_2 \\
&\quad = \left\| \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \widetilde{\mathbf{C}}' \left(\widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \right)^{-1} \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2} \mathbf{R}^* \right\|_2 \\
&\quad \le \left\| \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \right\|_2 \left\| \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \widetilde{\mathbf{C}}' \left(\widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \right)^{-1} \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \right\|_2 \left\| \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2} \right\|_2 \|\mathbf{R}^*\|_2 \\
&\quad = \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\} \lambda_{\max}^{1/2} \left\{ \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \widetilde{\mathbf{C}}' \left(\widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \right)^{-1} \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \right\} \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\} \|\mathbf{R}^*\|_2 \\
&\quad = \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\} \lambda_{\max}^{1/2} \left\{ \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \left(\widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \right)^{-1} \right\} \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\} \|\mathbf{R}^*\|_2 \\
&\quad = \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\} \lambda_{\max}^{1/2} \{\mathbf{I}_r\} \lambda_{\max}^{1/2}\{\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\} \|\mathbf{R}^*\|_2 \\
&\quad = o(1).
\end{aligned}$$

That is, $\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \widetilde{\mathbf{C}} \mathbf{R}^* = o(1)$ under the ℓ_2 norm. Substituting these simplified expressions into (S1.14), we arrive at (S1.7), completing the proof. □

S.1.2 Proof of Theorem 2

Proof of Theorem 2. We begin by defining a few key terms. Let $\boldsymbol{\omega}_n := \frac{1}{\sqrt{n}} \widetilde{\mathbf{C}} \mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*)$ and $T_0 := (\boldsymbol{\omega}_n + \sqrt{n}\gamma^* \boldsymbol{h}_n)' \mathbf{\Psi}^{-1} (\boldsymbol{\omega}_n + \sqrt{n}\gamma^* \boldsymbol{h}_n) = \left\| \mathbf{\Psi}^{-1/2} (\boldsymbol{\omega}_n + \sqrt{n}\gamma^* \boldsymbol{h}_n) \right\|_2^2$. Note that $\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}$ is positive definite by (A3). This proof is divided into five parts. We will prove that $T_W = T_0 + o_p(r)$, $T_S = T_0 + o_p(r)$, and $T_L = T_0 + o_p(r)$ in parts (i), (ii), and (iii), respectively. In part (iv) we will show that

$\sup_x |P(T_0 \le x) - P(\chi^2(r,\nu) \le x)| \to 0$ as $n, p \to \infty$. We will then combine our findings in parts (i) - (iv) to finish the proof.

(i) We know from (3.6) in Theorem 1 that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right) = \frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) + \boldsymbol{R}_a$$

where $\|\boldsymbol{R}_a\|_2 = o_p(1)$. Left-multiplying both sides of this expression by $\widetilde{\mathbf{C}}$, we have

$$\sqrt{n}\widetilde{\mathbf{C}}\left(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}\right) = \frac{1}{\sqrt{n}}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) + \widetilde{\mathbf{C}}\boldsymbol{R}_a = \boldsymbol{\omega}_n + \widetilde{\mathbf{C}}\boldsymbol{R}_a.$$

We know that $\widetilde{\mathbf{C}}\boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}} = \mathbf{C}\boldsymbol{\delta}^*_{\mathcal{M}} - \gamma^*\boldsymbol{t} = \gamma^*\boldsymbol{h}_n$ by (A1). Moreover, we note that $\widetilde{\mathbf{C}}\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} = \mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}$. Therefore we have

$$\sqrt{n}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'} = \boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n + \widetilde{\mathbf{C}}\boldsymbol{R}_a.$$

Left multiplying both sides of this expression by $\boldsymbol{\Psi}^{-1/2}$, we find

$$\sqrt{n}\boldsymbol{\Psi}^{-1/2}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'} = \boldsymbol{\Psi}^{-1/2}\left(\boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n\right) + \boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{R}_a.$$

We find $\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{R}_a\right\|_2 \le \left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\right\|_2 \|\boldsymbol{R}_a\|_2 = o_p(1)$ by (L3.5) in Lemma S.3. Combining this with the previous expression, we have

$$\sqrt{n}\boldsymbol{\Psi}^{-1/2}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'} = \boldsymbol{\Psi}^{-1/2}\left(\boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n\right) + o_p(1). \tag{S1.16}$$

Because the Tobit model is an exponential family, we find

$$\begin{aligned}\mathrm{E}\left[\left\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n\right\|_2^2\right] &= \mathrm{E}\left[\mathrm{tr}\left\{\boldsymbol{\omega}_n'\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n\right\}\right] \\ &= \mathrm{E}\left[\mathrm{tr}\left\{\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n\boldsymbol{\omega}_n'\boldsymbol{\Psi}^{-1/2}\right\}\right]\end{aligned}$$

$$
\begin{aligned}
&= \operatorname{tr}\left\{\boldsymbol{\Psi}^{-1/2}\,\mathrm{E}\left[\boldsymbol{\omega}_n\boldsymbol{\omega}_n'\right]\boldsymbol{\Psi}^{-1/2}\right\} \\
&= \operatorname{tr}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\,\mathrm{E}\left[-\frac{1}{n}\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)\right]\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} \\
&= \operatorname{tr}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} \\
&= \operatorname{tr}\left\{\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1/2}\right\} \\
&= \operatorname{tr}\{\mathbf{I}_r\} = r.
\end{aligned}
$$

Therefore by Markov's inequality

$$
\left\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n\right\|_2 = O_p(\sqrt{r}). \tag{S1.17}
$$

Moreover, by (A1) and (L3.4) we see that

$$
\left\|\sqrt{n}\gamma^*\boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\right\|_2 \le \left\|\boldsymbol{\Psi}^{-1/2}\right\|_2\left\|\sqrt{n}\gamma^*\boldsymbol{h}_n\right\|_2 = O(\sqrt{r}). \tag{S1.18}
$$

Applying these bounds to (S1.16) , we find

$$
\left\|\sqrt{n}\boldsymbol{\Psi}^{-1/2}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}\right\|_2 = O_p(\sqrt{r}). \tag{S1.19}
$$

Let $T_{W,0} := n\mathbf{C}^{*\prime}\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}'\boldsymbol{\Psi}^{-1}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}$. When $\hat{\mathcal{S}}_a = \mathcal{S}$, we have $\nabla^2_{\mathcal{M}'\cup\hat{\mathcal{S}}_a}\ell_n(\hat{\boldsymbol{\theta}}_a) = \nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)$ and $T_W = n(\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'})'\left(\widetilde{\mathbf{C}}\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1}\widetilde{\mathbf{C}}'\right)^{-1}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}$. By extension,

$$
\begin{aligned}
|T_W - T_{W,0}| &\le \left\|\sqrt{n}\boldsymbol{\Psi}^{-1/2}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'}\right\|_2^2\left\|\boldsymbol{\Psi}^{1/2}\left(\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\right)^{-1}\boldsymbol{\Psi}^{1/2} - \mathbf{I}_r\right\|_2 \\
&= O_p\left(\frac{r(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right) \\
&= o_p(r)
\end{aligned}
$$

where the last two equalities follow from (S1.19), (L3.6), and the fact that $(s+m)^3\log(s+$

$m) = o(n)$. Since $\hat{\mathcal{S}}_a = \mathcal{S}$ with probability converging to 1 by Theorem 1, the previous expression implies that $T_W = T_{W,0} + o_p(r)$. In addition, we apply (S1.16) and the Cauchy-Schwarz inequality to find

$$\begin{aligned}
T_{W,0} &= \left\| \sqrt{n}\boldsymbol{\Psi}^{-1/2}\mathbf{C}^*\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'} \right\|_2^2 \\
&= \left\| \boldsymbol{\Psi}^{-1/2}\left(\boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n\right) + o_p(1) \right\|_2^2 \\
&= \left\| \boldsymbol{\Psi}^{-1/2}\left(\boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n\right) \right\|_2^2 + o_p(1) + o_p\left(\boldsymbol{\Psi}^{-1/2}\left(\boldsymbol{\omega}_n + \sqrt{n}\gamma^*\boldsymbol{h}_n\right)\right) \\
&= T_0 + o_p(1) + O_p(\sqrt{r}) \\
&= T_0 + o_p(r),
\end{aligned}$$

where the penultimate equality follows from (S1.17) and (S1.18). Thus $T_W = T_{W,0} + o_p(r) = T_0 + o_p(r)$, completing this portion of the proof.

(ii) We established in (S1.11) in the proof of Theorem 1 that

$$\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) + \mathbf{R}_1$$

where $\|\mathbf{R}_1\|_2 = o_p\left(\frac{1}{\sqrt{n}}\right)$. Multiplying both sides of the previous expression by $\sqrt{n}$, we find

$$\sqrt{n}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \sqrt{n}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) + \sqrt{n}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) + o_p(1). \tag{S1.20}$$

By combining (3.7) from Theorem 1 and (S1.15), we find that

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) = \sqrt{n}\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}(\mathbf{P}_n - \mathbf{I}_{s+m+1})\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) \\
- \sqrt{n}\gamma^*\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n + \mathbf{R}_2
\end{aligned}$$

where $\|\mathbf{R}_2\|_2 = o_p(1)$. By (L3.1), we see that $\|\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{R}_2\|_2 \leq \|\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_2 \|\mathbf{R}_2\|_2 = o_p(1)$. As such, left-multiplying both sides of the previous expression by $\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}$ yields

$$\begin{aligned}\sqrt{n}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}'\cup\mathcal{S}}) &= \sqrt{n}\mathbf{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}(\mathbf{P}_n - \mathbf{I}_{s+m+1})\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) \\ &\quad - \sqrt{n}\gamma^*\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n + o_p(1). \end{aligned} \tag{S1.21}$$

Plugging (S1.21) into (S1.20), we see that

$$\sqrt{n}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \sqrt{n}\mathbf{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \sqrt{n}\gamma^*\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n + o_p(1).$$

Left-multiplying both sides of the previous expression by $\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}$, we find

$$\sqrt{n}\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0) = \sqrt{n}\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \sqrt{n}\gamma^*\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\boldsymbol{h}_n + o_p(1). \tag{S1.22}$$

since $\left\|\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O(1)$ by (L3.3).

We see that

$$\begin{aligned}&\mathrm{E}\left[\left\|\sqrt{n}\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right\|_2^2\right] \\ &\qquad = \mathrm{E}\left[\mathrm{tr}\left\{n\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\left(\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right)'\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\right\}\right] \\ &\qquad = \mathrm{tr}\left\{\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\,\mathrm{E}\left[n\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\left(\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right)'\right]\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\right\} \\ &\qquad = \mathrm{tr}\left\{\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\,\mathrm{E}\left[-\frac{1}{n}\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)\right]\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\right\} \\ &\qquad = \mathrm{tr}\left\{\mathbf{P}_n\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\mathbf{P}_n\right\} \\ &\qquad = \mathrm{tr}\left\{\mathbf{P}_n\mathbf{P}_n\right\} \\ &\qquad = \mathrm{tr}\left\{\mathbf{P}_n\right\} \\ &\qquad = \mathrm{tr}\left\{\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\widetilde{\mathbf{C}}\mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\right\} \\ &\qquad = \mathrm{tr}\left\{\widetilde{\mathbf{C}}\mathbf{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\mathbf{\Psi}^{-1}\right\}\end{aligned}$$

$$= \operatorname{tr}\{\mathbf{I}_r\} = r.$$

As a result,

$$\left\|\sqrt{n}\mathbf{P}_n\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right\|_2 = O_p(\sqrt{r}) \tag{S1.23}$$

by Markov's inequality. In addition, we find

$$\left\|\sqrt{n}\gamma^*\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n\right\|_2 \le \sqrt{n}\gamma^*\left\|\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\right\|_2\left\|\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\|_2\left\|\boldsymbol{\Psi}^{-1/2}\right\|_2\|\boldsymbol{h}_n\|_2 = O_p(\sqrt{r}) \tag{S1.24}$$

by (L3.3), (L3.4), (L3.5), and (A1). Applying these bounds to (S1.22), we now have

$$\left\|\sqrt{n}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right\|_2 = O_p(\sqrt{r}). \tag{S1.25}$$

We know from Theorem 1 that $\hat{\mathcal{S}}_0 = \mathcal{S}$ with probability converging to 1 and, by extension, $T_S = \frac{1}{n}\left(\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)\right)'\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)$ with probability converging to 1. We define $T_{S,0} := \frac{1}{n}\left(\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)\right)'\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)$. When $\hat{\mathcal{S}}_0 = \mathcal{S}$, we see that (S1.25) and (L3.7) imply

$$\begin{aligned}
|T_S - T_{S,0}| &= \left|\frac{1}{n}\left(\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)\right)'\left(\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right]^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right)\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)\right| \\
&= \left\|\frac{1}{\sqrt{n}}\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\hat{\boldsymbol{\theta}}_0)\right\|_2^2\left\|\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right]^{-1}\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}} - \mathbf{I}_{s+m+1}\right\|_2 \\
&= O_p\left(\frac{r(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right) \\
&= o_p(r) \qquad \text{(S1.26)}
\end{aligned}$$

since $(s+m)^3\log(s+m) = o(n)$. As such, we have that $T_S = T_{S,0} + o_p(r)$.

Define $T_{S,1} := \left\|\frac{1}{\sqrt{n}}\mathbf{P}_n\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^*\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n\right\|_2^2$. From (S1.22), we derive

$$\begin{aligned}
T_{S,0} &= \left\| \frac{1}{\sqrt{n}} \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\hat{\boldsymbol{\theta}}_0) \right\|_2^2 \\
&= \left\| \frac{1}{\sqrt{n}} \mathbf{P}_n \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^* \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \boldsymbol{h}_n + o_p(1) \right\|_2^2 \\
&= \left\| \frac{1}{\sqrt{n}} \mathbf{P}_n \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^* \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \boldsymbol{h}_n \right\|_2^2 + o_p(1) + O_p(\sqrt{r}) \\
&= T_{S,1} + o_p(r), \qquad \text{(S1.27)}
\end{aligned}$$

where the penultimate equality follows from (S1.23) and (S1.24). Next, we find

$$\begin{aligned}
T_{S,1} &= \left\| \frac{1}{\sqrt{n}} \mathbf{P}_n \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^* \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1} \boldsymbol{h}_n \right\|_2^2 \\
&= \left\| \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1/2} \left(\frac{1}{\sqrt{n}} \mathbf{\Psi}^{-1/2} \widetilde{\mathbf{C}} \mathbf{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right) \right\|_2^2 \\
&= \left\| \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1/2} \left(\mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right) \right\|_2^2 \\
&= \left(\mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right)' \mathbf{\Psi}^{-1/2} \widetilde{\mathbf{C}} \mathbf{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}} \widetilde{\mathbf{C}}' \mathbf{\Psi}^{-1/2} \left(\mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right) \\
&= \left(\mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right)' \mathbf{\Psi}^{-1/2} \mathbf{\Psi} \mathbf{\Psi}^{-1/2} \left(\mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right) \\
&= \left\| \mathbf{\Psi}^{-1/2} \boldsymbol{\omega}_n + \sqrt{n}\gamma^* \mathbf{\Psi}^{-1/2} \boldsymbol{h}_n \right\|_2^2 = T_0. \qquad \text{(S1.28)}
\end{aligned}$$

Thus we have $T_S = T_{S,0} + o_p(r) = T_{S,1} + o_p(r) = T_0 + o_p(r)$, completing this portion of the proof.

(iii) By combining (3.6) and (3.7) from Theorem 1, we find

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}) =& \frac{1}{\sqrt{n}} \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \mathbf{P}_n \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) \\
&+ \sqrt{n}\gamma^* \mathbf{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}} \mathbf{P}_n \mathbf{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}} \begin{bmatrix} \mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-1} \boldsymbol{h}_n \\ \mathbf{0} \end{bmatrix} + o_p(1).
\end{aligned}$$

We showed in (S1.15) in the proof of Theorem 1 that

$$\sqrt{n}\gamma^* \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{P}_n \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2} \begin{bmatrix} \mathbf{C}' \left(\mathbf{C}\mathbf{C}'\right)^{-1} \boldsymbol{h}_n \\ \mathbf{0} \end{bmatrix} = \sqrt{n}\gamma^* \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1} \boldsymbol{h}_n.$$

As such, we have

$$\begin{aligned} \sqrt{n}(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}) &= \frac{1}{\sqrt{n}} \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \mathbf{P}_n \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2} \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_n(\boldsymbol{\theta}^*) \\ &\quad + \sqrt{n}\gamma^* \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1} \boldsymbol{h}_n + o_p(1). \end{aligned} \tag{S1.29}$$

Note the similarity to (S1.22). Following the argument we used to prove (S1.25) and leveraging (L3.3), one can show

$$\left\| \sqrt{n}(\hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}}) \right\|_2 = O_p(\sqrt{r}). \tag{S1.30}$$

Taking a Taylor series expansion, we find that

$$\begin{aligned} \ell_n(\hat{\boldsymbol{\theta}}_0) - \ell_n(\hat{\boldsymbol{\theta}}_a) &= (\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_a)' \nabla \ell_n(\hat{\boldsymbol{\theta}}_a) + \frac{1}{2}(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_a)' \nabla^2 \ell_n(\hat{\boldsymbol{\theta}}_a)(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_a) \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_a)' \left(\nabla^2 \ell_n(\tilde{\boldsymbol{\theta}}) - \nabla^2 \ell_n(\hat{\boldsymbol{\theta}}_a) \right) (\hat{\boldsymbol{\theta}}_0 - \hat{\boldsymbol{\theta}}_a) \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ lies on the line segment between $\hat{\boldsymbol{\theta}}_0$ and $\hat{\boldsymbol{\theta}}_a$. Theorem 1 establishes that $\hat{\boldsymbol{\theta}}_{0,(\mathcal{M}'\cup\mathcal{S})^c} = \hat{\boldsymbol{\theta}}_{0,(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$ with probability converging to 1. When $\hat{\boldsymbol{\theta}}_{0,(\mathcal{M}'\cup\mathcal{S})^c} = \hat{\boldsymbol{\theta}}_{0,(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}$, the previous expression simplifies to

$$\begin{aligned} \ell_n(\hat{\boldsymbol{\theta}}_0) - \ell_n(\hat{\boldsymbol{\theta}}_a) &= (\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})' \nabla_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})' \nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a)(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}) \\ &\quad + (\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})' \mathbf{R} \end{aligned} \tag{S1.31}$$

where $\mathbf{R} = \frac{1}{2}\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\tilde{\boldsymbol{\theta}}) - \nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right) (\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})$.

Note that $\mathbf{R}$ is similar to the remainder terms in (S1.4) and (S1.9). Taking the same approach as in (S1.6), we can show that for $j \in \mathcal{M} \cup \mathcal{S}$,

$$|R_j| \leq 4.3\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})}\right\}\left\|\hat{\boldsymbol{\theta}}_{\mathcal{M}\cup\mathcal{S}} - \boldsymbol{\theta}^*_{\mathcal{M}\cup\mathcal{S}}\right\|_2^2 = O_p\left(\frac{r}{n}\right)$$

by (A3) and (S1.30). Following the same lines as in (S1.10), we can show that (S1.30) implies

$$|R_{p+1}| \leq 2\left|\tilde{\gamma}^{-3}\right|(\hat{\gamma}_0 - \hat{\gamma}_a)^2 = \frac{O_p\left(\frac{r}{n}\right)}{\left|\left(\gamma^* + O_p\left(\sqrt{\frac{\max\{s+m,r\}}{n}}\right)\right)^3\right|} = O_p\left(\frac{r}{n}\right)$$

since $r \leq m$ and $(s+m)^3 = o(n)$, so that $\gamma^* + O_p\left(\sqrt{\frac{\max\{s+m,r\}}{n}}\right) \geq \frac{\gamma^*}{2}$ for large n. Therefore $\|\mathbf{R}\|_{\max} = O_p\left(\frac{r}{n}\right)$ and, by extension,

$$\begin{aligned}\left\|(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\mathbf{R}\right\|_2 &\leq \left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}\right\|_2\|\mathbf{R}\|_2 \\ &= O_p\left(\sqrt{\frac{r}{n}}\frac{r}{n}\sqrt{s+m+1}\right) = o_p\left(\frac{\sqrt{r}}{n}\right) \qquad \text{(S1.32)}\end{aligned}$$

since $r \leq m$ and $(s+m)^3 = o(n)$. By a similar argument, we can show

$$\begin{aligned}&\left\|(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a) - \nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right)(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})\right\|_2 \\ &\qquad = o_p\left(\frac{\sqrt{r}}{n}\right). \qquad \text{(S1.33)}\end{aligned}$$

Together (S1.30) and (S2.55) from the proof of Lemma S.3 imply

$$\begin{aligned}&\left\|(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right)(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})\right\|_2 \\ &\qquad \leq \left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}\right\|_2^2\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \\ &\qquad = O_p\left(\frac{r}{n}\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right) \\ &\qquad = o_p\left(\frac{\sqrt{r}}{n}\right) \qquad \text{(S1.34)}\end{aligned}$$

since $r \leq m$ and $(s+m)^3 \log(s+m) = o(n)$.

Because $\hat{\boldsymbol{\theta}}_a$ is the minimizer of $\ell_n(\boldsymbol{\theta}) + \sum_{j\in\mathcal{M}^c} p_{\lambda_n}(|\delta_j|)$, we know that

$$\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a) = \begin{bmatrix} \mathbf{0} \\ -\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\delta}}_{a,\mathcal{S}};\lambda_n) \\ 0 \end{bmatrix}$$

Theorem 1 implies that $\min_{j\in\mathcal{S}} |\hat{\delta}_{a,j}| \geq \min_{j\in\mathcal{S}} |\delta_j^*| - d_n \geq d_n$ with probability converging to 1. Therefore $\left\|\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2 \leq \left\|\lambda_n\bar{\boldsymbol{\rho}}(\hat{\boldsymbol{\delta}}_{a,\mathcal{S}};\lambda_n)\right\|_2 \leq \sqrt{s}(\lambda_n\rho'(d_n;\lambda_n)) = o_p\left(\frac{1}{\sqrt{n}}\right)$ by (A2). Combining this with (S1.30), we find

$$\begin{aligned}\left\|(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2 &\leq \left\|\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}\right\|_2 \left\|\nabla_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2 \\ &= o_p\left(\frac{\sqrt{r}}{n}\right). \end{aligned} \tag{S1.35}$$

Applying our findings in (S1.32) - (S1.35) to (S1.31), we have

$$\ell_n(\hat{\boldsymbol{\theta}}_0) - \ell_n(\hat{\boldsymbol{\theta}}_a) = \frac{1}{2}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}) + o_p\left(\frac{\sqrt{r}}{n}\right). \tag{S1.36}$$

From here, we leverage (S1.29) and (L3.1) to show

$$\begin{aligned} n(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} &- \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})'\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}}) \\ &= \left\|\sqrt{n}\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}(\hat{\boldsymbol{\theta}}_{0,\mathcal{M}'\cup\mathcal{S}} - \hat{\boldsymbol{\theta}}_{a,\mathcal{M}'\cup\mathcal{S}})\right\|_2 \\ &= \left\|\frac{1}{\sqrt{n}}\mathbf{P}_n\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\nabla_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) + \sqrt{n}\gamma^*\boldsymbol{\Sigma}^{-1/2}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1}\boldsymbol{h}_n + o_p(1)\right\|_2 \\ &= T_0 + o_p(r), \end{aligned}$$

as we showed in (S1.27) and (S1.28) in part (ii). Together, this and (S1.36) imply that $T_L = 2n(\ell_n(\hat{\boldsymbol{\theta}}_0) - \ell_n(\hat{\boldsymbol{\theta}}_a)) = T_0 + o_p(r)$, completing this portion of the proof.

(iv) Let $\log L_1(\boldsymbol{\theta}; y_i, \mathbf{x}_i)$ denote the log-likelihood for a single observation $(\mathbf{x}_i, y_i)$. For $i = 1, \ldots, n$, we define $\boldsymbol{\xi}_i := \frac{1}{\sqrt{n}} \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \nabla_{\mathcal{M}' \cup \mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)$. We see that the $\boldsymbol{\xi}_i$ are independent and that $\mathrm{E}[\boldsymbol{\xi}_i] = \frac{1}{\sqrt{n}} \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \mathrm{E}[\nabla_{\mathcal{M}' \cup \mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)] = \mathbf{0}$ since the Tobit model is an exponential family. In addition, we find

$$\begin{aligned}
\sum_{i=1}^n \mathrm{Var}(\boldsymbol{\xi}_i) &= \mathrm{Var}\left(\sum_{i=1}^n \boldsymbol{\xi}_i\right) \\
&= \frac{1}{n} \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \mathrm{Var}(\nabla_{\mathcal{M}' \cup \mathcal{S}} \log L_n(\boldsymbol{\theta}^*)) \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \\
&= \frac{1}{n} \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \mathrm{E}[-\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \log L_n(\boldsymbol{\theta}^*)] \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \\
&= \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \\
&= \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \\
&= \boldsymbol{\Psi}^{-1/2} \boldsymbol{\Psi} \boldsymbol{\Psi}^{-1/2} = \mathbf{I}_r.
\end{aligned}$$

Since the $\boldsymbol{\xi}_i$ are independent, $\mathrm{E}[\boldsymbol{\xi}_i] = \mathbf{0}$, and $\sum_{i=1}^n \mathrm{Var}(\boldsymbol{\xi}_i) = \mathbf{I}_r$, we can apply Lemma S.6 of Shi et al. (2019) to conclude

$$\sup_{\mathcal{C}} \left| P\left(\sum_{i=1}^n \boldsymbol{\xi}_i \in \mathcal{C}\right) - P(\boldsymbol{Z} \in \mathcal{C}) \right| \le c_0\, r^{1/4} \sum_{i=1}^n \mathrm{E}\left[\|\boldsymbol{\xi}_i\|_2^3\right] \tag{S1.37}$$

where c_0 is a constant, $\boldsymbol{Z} \sim N(\mathbf{0}, \mathbf{I}_r)$, and the supremum is taken over all convex sets $\mathcal{C}$ in $\mathbb{R}^r$.

Our next goal is to show $\lim_{n\to\infty} r^{1/4} \sum_{i=1}^n \mathrm{E}\left[\|\boldsymbol{\xi}_i\|_2^3\right] = 0$. Let $i \in \{1, \ldots, n\}$. We see that

$$\mathrm{E}\left[\|\boldsymbol{\xi}_i\|_2^3\right] \le \frac{1}{n^{3/2}} \left\|\boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}}\right\|_2^3 \left\|\boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1}\right\|_2^3 \mathrm{E}\left[\|\nabla_{\mathcal{M}' \cup \mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)\|_2^3\right]. \tag{S1.38}$$

We know from (L3.3) and (L3.5) in Lemma S.3 that there exists $b_1 > 0$ such that $\left\|\boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}}\right\|_2^3 \left\|\boldsymbol{\Sigma}_{\mathcal{M}' \cup \mathcal{S}}^{-1}\right\|_2^3 < b_1$ for all n.

Define $\boldsymbol{v}_i = \nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)$ for $i = 1, \ldots, n$. Lemmas S.1 and S.2 of Jacobson and Zou (2023) establish that for all i, j, $v_{ij} \sim \text{subExp}(\sigma_{ij}^2, 4)$ where $\sigma_{ij}^2 < b_2$ for some constant $b_2 > 0$ by (A4). As an immediate corollary, Proposition 2.7.1 of Vershynin (2018) implies that there exists some $K > 0$ such that for all i, j, $\mathrm{E}[e^{t|v_{ij}|}] \leq e^{tK}$ for $t \in [0, \frac{1}{K}]$. Applying the generalized Hölder Inequality, we find that for $i = 1, \ldots, n$ and $t \in \left[0, \frac{1}{K(m+s+1)}\right]$,

$$\begin{aligned}
\mathrm{E}\left[e^{t\|\boldsymbol{v}_i\|_1}\right] &= \mathrm{E}\left[e^{t\sum_j |v_{ij}|}\right] \\
&= \mathrm{E}\left[\prod_j e^{t|v_{ij}|}\right] \\
&\leq \prod_j \left(\mathrm{E}\left[e^{t|v_{ij}|(m+s+1)}\right]\right)^{\frac{1}{m+s+1}} \\
&\leq \prod_j \left(e^{tK(m+s+1)}\right)^{\frac{1}{m+s+1}} \\
&= \prod_j e^{tK} \\
&= e^{tK(m+s+1)}.
\end{aligned}$$

By Proposition 2.7.1 of Vershynin (2018), this implies that there exists some $\widetilde{K} > 0$ such that $\mathrm{E}\left[\|\boldsymbol{v}_i\|_1^3\right] \leq (3\widetilde{K})^3$. Thus we find

$$\mathrm{E}\left[\|\nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)\|_2^3\right] \leq \mathrm{E}\left[\|\nabla_{\mathcal{M}'\cup\mathcal{S}} \log L_1(\boldsymbol{\theta}^*; y_i, \mathbf{x}_i)\|_1^3\right] \leq (3\widetilde{K})^3$$

for $i = 1, \ldots, n$. Returning to (S1.38), we have shown that $\mathrm{E}\left[\|\boldsymbol{\xi}_i\|_2^3\right] \leq \frac{1}{n^{3/2}} b_1 (3\widetilde{K})^3$ for $i = 1, \ldots, n$. Since $r \leq m = o(n^2)$, the previous expression and (S1.37) imply that as

$n \to \infty$,

$$\sup_{\mathcal{C}} \left| P\left(\sum_{i=1}^{n} \boldsymbol{\xi}_i \in \mathcal{C}\right) - P(\boldsymbol{Z} \in \mathcal{C}) \right| \leq c_0\, r^{1/4} \sum_{i=1}^{n} \frac{1}{n^{3/2}} b_1 (3\widetilde{K})^3 = c_0 b_1 (3\widetilde{K})^3 \cdot \frac{r^{1/4}}{n^{1/2}} \to 0. \tag{S1.39}$$

For $x \in \mathbb{R}$, we define the set $\mathcal{C}_x := \left\{\boldsymbol{z} \in \mathbb{R}^r : \left\|z + \sqrt{n}\gamma^* \boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\right\|_2^2 \leq x\right\}$. Since each set $\mathcal{C}_x$ is convex, (S1.39) implies

$$\sup_{x} \left| P\left(\sum_{i=1}^{n} \boldsymbol{\xi}_i \in \mathcal{C}_x\right) - P(\boldsymbol{Z} \in \mathcal{C}_x) \right| \to 0.$$

Note that $\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n = \sum_{i=1}^{n} \boldsymbol{\xi}_i$. We see that $\sum_{i=1}^{n} \boldsymbol{\xi}_i \in \mathcal{C}_x$ if and only if $T_0 = \left\|\boldsymbol{\Psi}^{-1/2}\boldsymbol{\omega}_n + \sqrt{n}\gamma^* \boldsymbol{\Psi}^{-1/2}\boldsymbol{h}_n\right\|_2^2 \leq x$. By the definition of the non-central χ^2 distribution, we also know that $\boldsymbol{Z} \in \mathcal{C}_x$ if and only if $\chi^2(r, \nu_n) \leq x$. As such, we have

$$\sup_{x} \left| P\left(T_0 \leq x\right) - P(\chi^2(r, \nu_n) \leq x) \right| \to 0 \tag{S1.40}$$

as $n, p \to \infty$, completing this portion of the proof.

(v) Let T be a test statistic satisfying $T = T_0 + o_p(r)$. This implies that for any $x \in \mathbb{R}$ and $\epsilon > 0$

$$P(T_0 \leq x - \epsilon r) + o(1) \leq P(T \leq x) \leq P(T_0 \leq x + \epsilon r) + o(1).$$

At the same time, (S1.40) in part (iv) implies that $P(\chi^2(r, \nu_n) \leq x - \epsilon r) = P(T_0 \leq x - \epsilon r) + o(1)$ and $P(\chi^2(r, \nu_n) \leq x + \epsilon r) = P(T_0 \leq x + \epsilon r) + o(1)$. As such, we see that

$$P(\chi^2(r, \nu_n) \leq x - \epsilon r) + o(1) \leq P(T \leq x) \leq P(\chi^2(r, \nu_n) \leq x + \epsilon r) + o(1). \tag{S1.41}$$

Lemma S.7 of Shi et al. (2019) provides that

$$\lim_{\epsilon \to 0^+} \limsup_n \left| P(\chi^2(r, \nu_n) \le x + \epsilon r) - P(\chi^2(r, \nu_n) \le x - \epsilon r) \right| = 0.$$

Together this and (S1.41) imply

$$\sup_x \left| P(T \le x) - P(\chi^2(r, \nu_n) \le x) \right| \to 0$$

as $n, p \to \infty$. Since $T_W = T_0 + o_p(r)$, $T_S = T_0 + o_p(r)$, and $T_L = T_0 + o_p(r)$, as we showed in parts (i) - (iii), this completes the proof. □

S.2 Supporting Results

In this section, we state and prove supporting results used in our theoretical study. Our technical results rely on the properties of sub-Gaussian and sub-exponential random variables, which we define as follows:

Definition S.1 (Sub-Gaussian). *We say that a random variable X with $\mathrm{E}\, X = \mu$ is* sub-Gaussian *with variance proxy $\sigma^2 \ge 0$ if its moment generating function satisfies*

$$\mathrm{E}\left[e^{t(X-\mu)}\right] \le e^{\frac{\sigma^2 t^2}{2}} \qquad \forall t \in \mathbb{R}.$$

We denote this by $X \sim \mathrm{subG}(\sigma^2)$.

Definition S.2 (Sub-Exponential). *We say that a random variable X with $\mathrm{E}\, X = \mu$ is* sub-exponential *with parameters $\sigma^2 \ge 0$, $\alpha \ge 0$ if its moment generating function satisfies*

$$\mathrm{E}\left[e^{t(X-\mu)}\right] \le e^{\frac{\sigma^2 t^2}{2}} \qquad \forall |t| < \frac{1}{\alpha}.$$

We denote this by $X \sim \mathrm{subExp}(\sigma^2, \alpha)$.

Lemma S.1. *Suppose* $Y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i$ *where* $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ *and define* $Y_i = Y_i^* \mathbb{1}_{Y_i^*>0}$ *for* $i = 1, \dots, n$. *Then for* $c > 0$ *and any* $\boldsymbol{\theta} \in \mathbb{R}^{p+2}$

$$P\left(\left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta})\right]\right\|_{\max} > c\right)$$
$$\leq 2(s+m)^2 \exp\left(-n \cdot \frac{2c^2}{\frac{1}{n}\sum_{i=1}^n x_{ij}^2 x_{ik}^2}\right)$$
$$+ 4(s+m) \exp\left(-n \cdot \frac{2c^2\gamma^2}{\frac{1}{n}\sum_{i=1}^n x_{ij}^2 (2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2}\right)$$
$$+ 2\exp\left(-\frac{n}{2}\min\left\{\frac{c\gamma^2}{8}, \frac{c^2\gamma^4}{\frac{65}{2} + \frac{1}{n}\sum_{i=1}^n \frac{1}{2}(\mathbf{x}_i'\boldsymbol{\delta})^2(2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2 + 8(\mathbf{x}_i'\boldsymbol{\delta})^2}\right\}\right).$$

Proof of Lemma S.1. We see that $\mathbf{H}(\boldsymbol{\theta})$ has three types of entries.

(i) $\sum_{i=1}^n x_{ij}x_{ik}[d_i + (1-d_i)h(-\mathbf{x}_i'\boldsymbol{\delta})]$ for $j, k \in \mathcal{M} \cup \mathcal{S}$.

Following the proof of Lemma S.3 in Jacobson and Zou (2023), one can show $\sum_{i=1}^n x_{ij}x_{ik}[d_i + (1-d_i)h(-\mathbf{x}_i'\boldsymbol{\delta})] \sim \text{subG}\left(\frac{1}{4}\sum_{i=1}^n x_{ij}^2 x_{ik}^2\right)$. By applying a Chernoff bound, we find

$$P\left(\left|\sum_{i=1}^n x_{ij}x_{ik}[d_i + (1-d_i)h(-\mathbf{x}_i'\boldsymbol{\delta})] - \mathrm{E}\left[\sum_{i=1}^n x_{ij}x_{ik}[d_i + (1-d_i)h(-\mathbf{x}_i'\boldsymbol{\delta})]\right]\right| > c\right)$$
$$\leq 2\exp\left(-\frac{2c^2}{\sum_{i=1}^n x_{ij}^2 x_{ik}^2}\right)$$

for any $j, k \in \mathcal{M} \cup \mathcal{S}$.

(ii) $-\sum_{i=1}^n d_i Y_i^* x_{ij}$ for $j \in \mathcal{M} \cup \mathcal{S}$

Again following the proof of Lemma S.3 in Jacobson and Zou (2023), one can show $-\sum_{i=1}^n d_i Y_i^* x_{ij} \sim \text{subG}\left(\frac{1}{4\gamma^2}\sum_{i=1}^n x_{ij}^2(2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2\right)$ and, consequently,

$$P\left(\left|-\sum_{i=1}^n d_i Y_i^* x_{ij} - \mathrm{E}\left[-\sum_{i=1}^n d_i Y_i^* x_{ij}\right]\right| > c\right) \leq 2\exp\left(-\frac{2c^2\gamma^2}{\sum_{i=1}^n x_{ij}^2(2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2}\right)$$

for any $j \in \mathcal{M} \cup \mathcal{S}$.

(iii) $\sum_{i=1}^n d_i Y_i^{*2}$

Following the proof of Lemma S.5 in Jacobson and Zou (2023), one can show that $\sum_{i=1}^n d_i \gamma^2 Y_i^{*2} \sim \text{subExp}(\frac{65}{2}n + \sum_{i=1}^n \frac{1}{2}(\mathbf{x}_i'\boldsymbol{\delta})^2(2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2 + 8(\mathbf{x}_i'\boldsymbol{\delta})^2, 8)$. As such, we find

$$P\left(\left|\sum_{i=1}^n d_i Y_i^{*2} - \mathrm{E}\left[\sum_{i=1}^n d_i Y_i^{*2}\right]\right| > c\right) \leq 2\exp\left(-\frac{1}{2}\min\left\{\frac{c\gamma^2}{8}, \frac{c^2\gamma^4}{\frac{65}{2}n + \sum_{i=1}^n \frac{1}{2}(\mathbf{x}_i'\boldsymbol{\delta})^2(2 + \mathbf{x}_i'\boldsymbol{\delta} + g(-\mathbf{x}_i'\boldsymbol{\delta}))^2 + 8(\mathbf{x}_i'\boldsymbol{\delta})^2}\right\}\right).$$

We apply the union bound to arrive at our final result. □

Lemma S.2. *Let $p_\lambda(t)$ be a folded-concave penalty function. Then $\hat{\boldsymbol{\theta}}$ is a strict local minimizer of $Q_n(\boldsymbol{\theta})$ subject to the constraint $\mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}$ if*

(L2.1) $\nabla_{\mathcal{M}'\cup\mathcal{S}} Q_n(\hat{\boldsymbol{\theta}}) = \begin{bmatrix} \mathbf{C}'\boldsymbol{\nu} \\ \mathbf{0} \\ \boldsymbol{t}'\boldsymbol{\nu} \end{bmatrix}$ *for some* $\boldsymbol{\nu} \in \mathbb{R}^r$, $\mathbf{C}^*\hat{\boldsymbol{\theta}}_{\mathcal{M}'} = \mathbf{0}$

(L2.2) $\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c} \ell_n(\hat{\boldsymbol{\theta}})\right\|_{\max} < \lambda_n \rho'(0^+; \lambda_n)$, *and*

(L2.3) $\lambda_{\min}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}})\right\} > \lambda_n \kappa(\rho, \hat{\boldsymbol{\theta}}, \lambda_n)$.

Proof of Lemma S.2. Define $\mathcal{A} := \{\boldsymbol{\theta} : \mathbf{C}^*\boldsymbol{\theta}_{\mathcal{M}'} = \mathbf{0}\}$ and $\mathcal{B} := \left\{\boldsymbol{\theta} : \boldsymbol{\theta}_{(\mathcal{M}'\cup\mathcal{S})^c} = \mathbf{0}\right\}$. Under condition (L2.3), we have $\lambda_{\min}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}} Q_n(\hat{\boldsymbol{\theta}})\right\} > 0$, meaning that $Q_n(\boldsymbol{\theta})$ is strictly concave in a ball $\mathcal{N}_0 \subseteq \mathcal{B}$ centered at $\hat{\boldsymbol{\theta}}$. Additionally, condition (L2.1) provides that $\hat{\boldsymbol{\theta}}$ is a stationary point of the Lagrangian. As a result, we have that $\hat{\boldsymbol{\theta}}$ is the unique minimizer of $Q_n(\boldsymbol{\theta})$ in $\mathcal{N}_0 \cap \mathcal{A}$.

We will now show that $\hat{\boldsymbol{\theta}}$ is a strict local minimizer of $Q_n(\boldsymbol{\theta})$ on $\mathcal{A}$. We know that $\rho'(t; \lambda_n)$

is decreasing in t because $\rho(t;\lambda_n)$ is concave. Since $\nabla_j \ell_n(\boldsymbol{\theta})$ is continuous for $j = 0, 1, \ldots, p$ and $\rho'(t;\lambda_n)$ is continuous, we know from (L2.2) that there exists $\epsilon > 0$ such that for all $\boldsymbol{\theta}$ in a ball with radius ϵ centered at $\hat{\boldsymbol{\theta}}$, $\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\boldsymbol{\theta})\right\|_{\max} < \lambda_n\rho'(\epsilon;\lambda_n)$.

Let $\mathcal{N}_1 \subseteq \mathcal{A}$ be a ball centered at $\hat{\boldsymbol{\theta}}$ with a radius which is less than ϵ and is also small enough that $\mathcal{N}_1 \cap \mathcal{B} \subseteq \mathcal{N}_0 \cap \mathcal{A}$. Our goal is to show that $Q_n(\hat{\boldsymbol{\theta}}) < Q_n(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathcal{N}_1$. It suffices to prove that $Q_n(\hat{\boldsymbol{\theta}}) < Q_n(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathcal{N}_1 \cap \mathcal{N}_0^c$, as we already know that $Q_n(\hat{\boldsymbol{\theta}}) < Q_n(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \mathcal{N}_1 \cap \mathcal{N}_0 \subseteq \mathcal{N}_0 \cap \mathcal{A}$.

Let $\boldsymbol{\theta}_1 \in \mathcal{N}_1 \cap \mathcal{N}_0^c$ and let $\boldsymbol{\theta}_2$ be the projection of $\boldsymbol{\theta}_1$ onto $\mathcal{B}$. Then $\boldsymbol{\theta}_2 \in \mathcal{N}_0 \cap \mathcal{A}$, so $Q_n(\hat{\boldsymbol{\theta}}) < Q_n(\boldsymbol{\theta}_2)$. By the Mean Value Theorem, there exists $\tilde{\boldsymbol{\theta}}$ on the line segment between $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ such that

$$Q_n(\boldsymbol{\theta}_2) - Q_n(\boldsymbol{\theta}_1) = (\nabla Q_n(\tilde{\boldsymbol{\theta}}))'(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1).$$

Since $\boldsymbol{\theta}_2$ is the projection of $\boldsymbol{\theta}_1$ onto $\mathcal{B}$, we have that $\theta_{1,j} - \theta_{2,j} = 0$ for $j \in \mathcal{M}' \cup \mathcal{S}$, $\operatorname{sgn}(\tilde{\theta}_j) = \operatorname{sgn}(\theta_{1,j})$ for $j \in (\mathcal{M}' \cup \mathcal{S})^c$, and $\tilde{\boldsymbol{\theta}} \in \mathcal{N}_1$. Therefore

$$\nabla Q_n(\tilde{\boldsymbol{\theta}})'(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) = -(\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\tilde{\boldsymbol{\theta}}))'\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c} - \sum_{j\in(\mathcal{M}'\cup\mathcal{S})^c} \lambda_n\rho'(|\tilde{\theta}_j|;\lambda_n)|\theta_{1,j}|$$

Because $\tilde{\boldsymbol{\theta}} \in \mathcal{N}_1$, we see that $|\tilde{\theta}_j| \leq |\theta_{1,j}| < \epsilon$ for all $j \in (\mathcal{M}' \cup \mathcal{S})^c$ and $\left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\tilde{\boldsymbol{\theta}})\right\|_{\max} < \lambda_n\rho'(\epsilon;\lambda_n)$. Leveraging these properties and the fact that $\rho'(t;\lambda_n)$ is decreasing, we have

$$\begin{aligned}
-(\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}&\ell_n(\tilde{\boldsymbol{\theta}}))'\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c} - \sum_{j\in(\mathcal{M}'\cup\mathcal{S})^c} \lambda_n\rho'(|\tilde{\theta}_j|;\lambda_n)|\theta_{1,j}| \\
&< \left\|\nabla_{(\mathcal{M}'\cup\mathcal{S})^c}\ell_n(\tilde{\boldsymbol{\theta}})\right\|_{\max} \left\|\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c}\right\|_1 - \lambda_n\rho'(\epsilon;\lambda_n)\left\|\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c}\right\|_1 \\
&< \lambda_n\rho'(\epsilon;\lambda_n)\left\|\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c}\right\|_1 - \lambda_n\rho'(\epsilon;\lambda_n)\left\|\boldsymbol{\theta}_{1,(\mathcal{M}'\cup\mathcal{S})^c}\right\|_1 = 0,
\end{aligned}$$

giving us $Q_n(\boldsymbol{\theta}_2) < Q_n(\boldsymbol{\theta}_1)$ and, by extension, $Q_n(\hat{\boldsymbol{\theta}}) < Q_n(\boldsymbol{\theta}_1)$. □

Lemma S.3. *If* (A1) - (A5) *are satisfied and* $(s+m)^3\log(s+m) = o(n)$, *then the following hold*

(L3.1) $\lambda_{\max}\left\{\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\} = O(1)$

(L3.2) $\liminf_n \lambda_{\min}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\} > 0$

(L3.3) $\lambda_{\max}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\} = O(1)$

(L3.4) $\lambda_{\max}\{\boldsymbol{\Psi}^{-1}\} = O(1)$

(L3.5) $\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\right\|_2 = O(1)$

(L3.6) $\left\|\boldsymbol{\Psi}^{1/2}\left(\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\right)^{-1}\boldsymbol{\Psi}^{1/2} - \mathbf{I}_r\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$

(L3.7) $\left\|\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1}\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}} - \mathbf{I}_{s+m+1}\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$.

Proof of Lemma S.3. We see that (L3.1) follows immediately from (A3), as Weyl's inequality provides that $\lambda_{\max}\left\{\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\} \le \lambda_{\max}\left\{\mathrm{E}[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)]\right\} + \mathrm{E}\left[\frac{n_1}{n}\gamma^{*-2}\right] \le \lambda_{\max}\left\{\mathrm{E}[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)]\right\} + \gamma^{*-2} = O(1)$. Because $\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}$ is positive definite, we know that $\liminf_n \lambda_{\min}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\} = \liminf_n \left(\lambda_{\max}\left\{\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\}\right)^{-1}$. As such, (L3.2) follows immediately from (L3.1). Under (A3), we apply Weyl's inequality to show that $\lambda_{\min}\left\{\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\} \ge \lambda_{\min}\left\{\mathrm{E}[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)]\right\} \ge c_H$ for all n. Therefore we have that $\lambda_{\max}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\} = \left(\lambda_{\min}\left\{\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\}\right)^{-1} \le c_H^{-1}$ for all n, proving (L3.3).

We will show that

$$\lambda_{\max}\left\{\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right)^{-1}\right\} = O_p(1) \tag{S2.42}$$

as it is a helpful intermediate result. Let $n \in \mathbb{N}$ and define

$$\mathcal{E}_n := \left\{\left\|\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*) - \mathrm{E}\left[\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)\right]\right\|_{\max} \le \frac{c_H}{2(s+m+1)}\right\}.$$

As in the proof of Theorem 1, we can show that if (A3), (A4), and $\mathcal{E}_n$ hold, then $\lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)\right\} \geq \frac{c_H}{2}$. This implies that $\lambda_{\min}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right\} \geq \lambda_{\min}\left\{\frac{1}{n}\mathbf{H}(\boldsymbol{\theta}^*)\right\} \geq \frac{c_H}{2}$ and, by extension, $\lambda_{\max}\left\{(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1}\right\} = \left(\lambda_{\min}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\right\}\right)^{-1} \leq \frac{2}{c_H}$. As in the proof of Theorem 1, one can show that if (A4) holds and $s+m = o(n^{1/3})$, then $P(\mathcal{E}_n) \to 1$ as $n \to \infty$. As such, if (A3) and (A4) are satisfied and $s+m = o(n^{1/3})$, then (S2.42) holds.

We will prove (L3.4) next. Recall that $\boldsymbol{\Psi} = \widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'$. Because $\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}$ is positive definite and $\widetilde{\mathbf{C}}$ has full row rank, we see that $\boldsymbol{\Psi}$ is positive definite as well. As such, we have

$$\lambda_{\max}\left\{\boldsymbol{\Psi}^{-1}\right\} = \lambda_{\max}\left\{\left(\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\right)^{-1}\right\} = \left(\lambda_{\min}\left\{\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\right\}\right)^{-1}. \tag{S2.43}$$

Define $a := \liminf_n \left(\lambda_{\max}\left\{(\mathbf{C}\mathbf{C}')^{-1}\right\}\right)^{-1}$. Under (A1), we have that $a > 0$. Since $\mathbf{C}$ has full rank, $\mathbf{C}\mathbf{C}'$ is positive definite and, by extension, $\liminf_n \lambda_{\min}\left\{\mathbf{C}\mathbf{C}'\right\} = \liminf_n \left(\lambda_{\max}\left\{(\mathbf{C}\mathbf{C}')^{-1}\right\}\right)^{-1} = a$. We see that $\lambda_{\min}\left\{\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\right\} \geq \lambda_{\min}\left\{\mathbf{C}\mathbf{C}'\right\} + \lambda_{\min}\left\{\boldsymbol{t}\boldsymbol{t}'\right\} \geq \lambda_{\min}\left\{\mathbf{C}\mathbf{C}'\right\}$ for all n, so $\liminf_n \lambda_{\min}\left\{\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\right\} \geq a$. By the min-max theorem, we have that $\lambda_{\min}\left\{\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\right\} = \min_{\boldsymbol{v}\neq\mathbf{0}} \frac{\boldsymbol{v}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}}$. Therefore for sufficiently large n, $\min_{\boldsymbol{v}\neq\mathbf{0}} \frac{\boldsymbol{v}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} > \frac{a}{2}$ and, by extension, $\boldsymbol{v}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\boldsymbol{v} > \frac{a}{2}\boldsymbol{v}'\boldsymbol{v}$ for all $\boldsymbol{v} \neq 0$. Thus we find that

$$\begin{aligned}
\liminf_n \lambda_{\min}\left\{\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\right\} &= \liminf_n \min_{\boldsymbol{v}\neq\mathbf{0}} \frac{\boldsymbol{v}'\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{v}}{\boldsymbol{v}'\boldsymbol{v}} \\
&\geq \liminf_n \min_{\boldsymbol{v}\neq\mathbf{0}} \frac{\boldsymbol{v}'\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{v}}{\frac{2}{a}\boldsymbol{v}'\widetilde{\mathbf{C}}\widetilde{\mathbf{C}}'\boldsymbol{v}} \\
&\geq \frac{a}{2} \liminf_n \min_{\boldsymbol{w}\neq\mathbf{0}} \frac{\boldsymbol{w}'\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\boldsymbol{w}}{\boldsymbol{w}'\boldsymbol{w}} \\
&= \frac{a}{2} \liminf_n \lambda_{\min}\left\{\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\} \\
&> 0
\end{aligned}$$

by (L3.2). Combining this with (S2.43), we see that (L3.4) holds.

Moving on to (L3.5), we see that $\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\right\|_2 \leq \left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\right\|_2 \left\|\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{1/2}\right\|_2$. We find

$$\begin{aligned}\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1/2}\right\|_2 &= \lambda_{\max}^{1/2}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\}\\ &= \lambda_{\max}^{1/2}\left\{\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1/2}\right\}\\ &= \lambda_{\max}^{1/2}\left\{\mathbf{I}_r\right\} = 1.\end{aligned}$$

Combining this with (L3.1) yields (L3.5).

It will take us several steps to prove (L3.6). We will start by showing that

$$\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2 = O_p\left(\frac{s+m}{\sqrt{n}}\right), \tag{S2.44}$$

a helpful intermediate result for the rest of our proof. Since $\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)$ and $\nabla^2_{\mathcal{M}\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)$ are symmetric, Lemma S.8 of Shi et al. (2019) provides that

$$\begin{aligned}&\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2\\ &\quad\leq \left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_\infty\\ &\quad= \left\|\frac{1}{n}\begin{bmatrix}\mathbf{X}'_{(\mathcal{M}\cup\mathcal{S})}\\ -\mathbf{y}'\end{bmatrix}\begin{bmatrix}\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a) & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\mathbf{X}_{(\mathcal{M}\cup\mathcal{S})} & -\mathbf{y}\end{bmatrix} + \frac{1}{n}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & n_1(\gamma^{*-2} - \hat{\gamma}_a^{-2})\end{bmatrix}\right\|_\infty\\ &\quad= \max\left\{\left\|\frac{1}{n}\mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}}\left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right]\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_\infty, \left|\frac{n_1}{n}(\gamma^{*-2} - \hat{\gamma}_a^{-2})\right|\right\}\end{aligned} \tag{S2.45}$$

We will start by bounding the first term in (S2.45). Applying the Cauchy Schwarz inequality, we find

$$\begin{aligned}&\left\|\frac{1}{n}\mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}}\left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right]\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_\infty\\ &\qquad= \max_{j\in\mathcal{M}\cup\mathcal{S}}\left\|\frac{1}{n}\mathbf{X}'_{0,j}\left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right]\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_1\\ &\qquad\leq \max_{j\in\mathcal{M}\cup\mathcal{S}}\sqrt{s+m+1}\left\|\frac{1}{n}\mathbf{X}'_{0,j}\left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right]\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_2.\end{aligned} \tag{S2.46}$$

Let $j \in \mathcal{M} \cup \mathcal{S}$. By the Fundamental Theorem of Calculus, we have

$$\frac{1}{n}\mathbf{X}_{0,j}' \left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right] \mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}} = (\boldsymbol{\delta}^* - \hat{\boldsymbol{\delta}}_a)' \int_0^1 \frac{1}{n}\mathbf{X}_0' \operatorname{diag}\{\mathbf{X}_{0,j} \circ \boldsymbol{g}''(t\boldsymbol{\delta}^* + (1-t)\hat{\boldsymbol{\delta}}_a)\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\, dt$$

where the integral is applied componentwise. Theorem 1 establishes that $\hat{\boldsymbol{\delta}}_{a,(\mathcal{M}\cup\mathcal{S})^c} = \boldsymbol{\delta}^*_{(\mathcal{M}\cup\mathcal{S})^c} = \mathbf{0}$ with probability converging to 1, which implies $(\boldsymbol{\delta}^* - \hat{\boldsymbol{\delta}}_a)'\mathbf{X}_0' = (\boldsymbol{\delta}^*_{\mathcal{M}\cup\mathcal{S}} - \hat{\boldsymbol{\delta}}_{a,\mathcal{M}\cup\mathcal{S}})'\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}'$ with probability converging to 1. As such, the previous expression implies that for all $j \in \mathcal{M} \cup \mathcal{S}$

$$\begin{aligned}
&\left\|\frac{1}{n}\mathbf{X}_{0,j}' \left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right] \mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_2 \\
&\leq \left\|\boldsymbol{\delta}^*_{\mathcal{M}\cup\mathcal{S}} - \hat{\boldsymbol{\delta}}_{a,\mathcal{M}\cup\mathcal{S}}\right\|_2 \sup_{t\in[0,1]} \left\|\frac{1}{n}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}' \operatorname{diag}\{\mathbf{X}_{0,j} \circ \boldsymbol{g}''(t\boldsymbol{\delta}^* + (1-t)\hat{\boldsymbol{\delta}}_a)\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_2 \quad \text{(S2.47)}
\end{aligned}$$

with probability converging to 1. Lemma S.6 of Jacobson and Zou (2023) provides that $|g''(s)| < 4.3$ for all $s \in \mathbb{R}$. Therefore, we see that

$$\begin{aligned}
&\sup_{t\in[0,1]} \left\|\frac{1}{n}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}' \operatorname{diag}\{\mathbf{X}_{0,j} \circ \boldsymbol{g}''(t\boldsymbol{\delta}^* + (1-t)\hat{\boldsymbol{\delta}}_a)\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_2 \\
&\qquad = \sup_{t\in[0,1]} \lambda_{\max}\left\{\frac{1}{n}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}' \operatorname{diag}\{\mathbf{X}_{0,j} \circ \boldsymbol{g}''(t\boldsymbol{\delta}^* + (1-t)\hat{\boldsymbol{\delta}}_a)\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\} \\
&\qquad \leq 4.3\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}' \operatorname{diag}\{|\mathbf{X}_{0,j}|\}\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\} \\
&\qquad \leq 4.3\lambda_{\max}\left\{\frac{1}{n}\mathbf{X}_{\mathcal{M}\cup\mathcal{S}}' \operatorname{diag}\{|\mathbf{X}_{(j)}|\}\mathbf{X}_{\mathcal{M}\cup\mathcal{S}}\right\} \\
&\qquad = O(1)
\end{aligned}$$

where the final two lines follow from (S1.5) and (A3). Moreover, Theorem 1 provides that $\left\|\boldsymbol{\delta}^*_{\mathcal{M}\cup\mathcal{S}} - \hat{\boldsymbol{\delta}}_{a,\mathcal{M}\cup\mathcal{S}}\right\|_2 = O_p\left(\sqrt{\frac{s+m}{n}}\right)$. Applying these findings to (S2.47), we have

$$\left\|\frac{1}{n}\mathbf{X}_{0,j}' \left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a)\right]\mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}}\right\|_2$$

$$\leq 4.3 \left\| \boldsymbol{\delta}^*_{\mathcal{M}\cup\mathcal{S}} - \hat{\boldsymbol{\delta}}_{a,\mathcal{M}\cup\mathcal{S}} \right\|_2 \lambda_{\max} \left\{ \frac{1}{n} \mathbf{X}'_{\mathcal{M}\cup\mathcal{S}} \operatorname{diag}\{|\mathbf{X}_{(j)}|\} \mathbf{X}_{\mathcal{M}\cup\mathcal{S}} \right\}$$
$$= O_p\left(\sqrt{\frac{s+m}{n}}\right)$$

for all $j \in \mathcal{M}\cup\mathcal{S}$. Applying this to (S2.46), we find $\left\| \frac{1}{n} \mathbf{X}'_{0,\mathcal{M}\cup\mathcal{S}} \left[\mathbf{D}(\boldsymbol{\delta}^*) - \mathbf{D}(\hat{\boldsymbol{\delta}}_a) \right] \mathbf{X}_{0,\mathcal{M}\cup\mathcal{S}} \right\|_\infty = O_p\left(\frac{s+m}{\sqrt{n}}\right)$.

Turning to the second term in (S2.45), the Mean Value Theorem guarantees that there exists $\tilde{\gamma}$ between γ^* and $\hat{\gamma}_a$ such that $\left|\frac{n_1}{n}(\gamma^{*-2} - \hat{\gamma}_a^{-2})\right| = \left|\frac{2n_1}{n}\tilde{\gamma}^{-3}(\gamma^* - \hat{\gamma}_a)\right|$. By Theorem 1, we have that $\gamma^* - \hat{\gamma}_a = O_p\left(\sqrt{\frac{s+m}{n}}\right)$ and, by extension, $\left|\frac{2n_1}{n}\tilde{\gamma}^{-3}(\gamma^* - \hat{\gamma}_a)\right| \leq O_p\left(\sqrt{\frac{s+m}{n}}\right) \left| \left(\gamma^* + O_p\left(\sqrt{\frac{s+m}{n}}\right)\right)^{-3} \right|$. Since $(s+m)^3 = o(n)$, $\gamma^* + O_p\left(\sqrt{\frac{s+m}{n}}\right) = \gamma^* + o_p(1) \geq \frac{\gamma^*}{2}$ for sufficiently large n. Thus we have $\left|\frac{n_1}{n}(\gamma^{*-2} - \hat{\gamma}_a^{-2})\right| = O_p\left(\sqrt{\frac{s+m}{n}}\right)$. Our bounds for the two terms in (S2.45) are sufficient to establish (S2.44).

We now focus on (L3.6) directly. We see that

$$\begin{aligned}
&\left\| \boldsymbol{\Psi}^{1/2} \left(\widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' \right)^{-1} \boldsymbol{\Psi}^{1/2} - \mathbf{I}_r \right\|_2 \\
&\quad = \left\| \left\{ \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \right\}^{-1} - \mathbf{I}_r \right\|_2 \\
&\quad \leq \left\| \left\{ \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \right\}^{-1} \right\|_2 \\
&\qquad \times \left\| \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} - \mathbf{I}_r \right\|_2 . \qquad \text{(S2.48)}
\end{aligned}$$

We will bound the second term on the right hand side of (S2.48) first. We find that

$$\begin{aligned}
&\left\| \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} - \mathbf{I}_r \right\|_2 \\
&\quad = \left\| \boldsymbol{\Psi}^{-1/2} \left(\widetilde{\mathbf{C}} \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} \widetilde{\mathbf{C}}' - \boldsymbol{\Psi} \right) \boldsymbol{\Psi}^{-1/2} \right\|_2
\end{aligned}$$

$$
\begin{aligned}
&= \left\| \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \left(\left[\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}' \cup \mathcal{S}} \right) \widetilde{\mathbf{C}}' \boldsymbol{\Psi}^{-1/2} \right\|_2 \\
&\leq \left\| \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \right\|_2^2 \left\| \left[\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right]^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}' \cup \mathcal{S}} \right\|_2 . \qquad \text{(S2.49)}
\end{aligned}
$$

We have that $\left\| \boldsymbol{\Psi}^{-1/2} \widetilde{\mathbf{C}} \right\|_2 = O(1)$ from (L3.5). As for the second term in (S2.49), we see that

$$
\begin{aligned}
\left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}' \cup \mathcal{S}} \right\|_2 &\leq \left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} - \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\|_2 \\
&\quad + \left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}' \cup \mathcal{S}} \right\|_2 . \qquad \text{(S2.50)}
\end{aligned}
$$

Focusing on the first term in (S2.50), we find

$$
\begin{aligned}
&\left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} - \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\|_2 \\
&= \left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right) \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\|_2 \\
&\leq \left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} \right\|_2 \left\| \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right\|_2 \left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\|_2 . \qquad \text{(S2.51)}
\end{aligned}
$$

Since $\left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1}$ is symmetric, $\left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\|_2 = \lambda_{\max} \left\{ \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right)^{-1} \right\} = O_p(1)$ by (S2.42). Turning our attention to $\left\| \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right)^{-1} \right\|_2$, we see that

$$
\begin{aligned}
\lambda_{\min} \left\{ \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right\} &= \min_{\|\boldsymbol{v}\|_2 = 1} \boldsymbol{v}' \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \boldsymbol{v} \\
&= \min_{\|\boldsymbol{v}\|_2 = 1} \boldsymbol{v}' \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) + \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right) \boldsymbol{v} \\
&\geq \min_{\|\boldsymbol{v}\|_2 = 1} \boldsymbol{v}' \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \boldsymbol{v} - \sup_{\|\boldsymbol{v}\|_2 = 1} \left| \boldsymbol{v}' \left(\nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right) \boldsymbol{v} \right| \\
&= \lambda_{\min} \left\{ \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right\} - \lambda_{\max} \left\{ \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right\} \\
&= \lambda_{\min} \left\{ \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) \right\} - \left\| \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}' \cup \mathcal{S}} \ell_n(\hat{\boldsymbol{\theta}}_a) \right\|_2 \qquad \text{(S2.52)}
\end{aligned}
$$

where the second to last equality follows from Lemma S.5 of Shi et al. (2019). We established in our proof of (S2.42) that if (A3) and (A4) are satisfied, then $\lambda_{\min}\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*)\} \geq \frac{c_H}{2}$ with probability converging to 1. Under the assumption that $(s+m)^3 = o(n)$, we see that (S2.44) implies $\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\|_2 \leq \frac{c_H}{4}$ for sufficiently large n with probability converging to 1. Combining these findings with (S2.52), we see that $\lambda_{\min}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\} \geq \frac{c_H}{4}$ for sufficiently large n and, by extension,

$$\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1}\right\|_2 = \lambda_{\max}\left\{\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1}\right\} = \lambda_{\min}^{-1}\left\{\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right\} \leq \frac{4}{c_H}$$

with probability converging to 1. Thus, $\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1}\right\|_2 = O_p(1)$. Combining this with (S2.44), we return to (S2.51) and find $\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1} - (\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1}\right\|_2 = O_p\left(\frac{s+m}{\sqrt{n}}\right)$.

Moving on to the second term in (S2.50), we see that

$$\begin{aligned}\left\|(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 &= \left\|(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1}\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right)\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \\ &\leq \left\|(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1}\right\|_2 \left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \left\|\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \\ &= O_p\left(\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2\right), \end{aligned} \tag{S2.53}$$

by (L3.3) and (S2.42). We know

$$\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \leq (s+m+1)\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_{\max}.$$

Lemma S.3 of Jacobson and Zou (2023) establishes that for $k > 0$

$$\begin{aligned}&P\left(\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_{\max} > k\sqrt{\frac{\log(s+m)}{n}}\right) \\ &\quad = P\left(\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*) - \mathrm{E}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\log L_n(\boldsymbol{\theta}^*)\right]\right\|_{\max} > k\sqrt{\log(s+m)n}\right) \\ &\quad \leq 2(s+m+1)^2\exp\left(-\frac{k^2\log(s+m)}{O(1)}\right) + 4(s+m+1)\exp\left(-\frac{k^2\log(s+m)}{O(1)}\right)\end{aligned}$$

$$+ 2\exp\left(-\frac{1}{2}\min\left\{\frac{k\sqrt{\log(s+m)n}}{O(1)}, \frac{k\log(s+m)}{O(1)}\right\}\right) \tag{S2.54}$$

if (A4) is satisfied. As such, we see that if k is sufficiently large, then $P\Bigg(\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\|_{\max} \leq k\sqrt{\frac{\log(s+m)}{n}}\Bigg) \to 1$ as $n \to \infty$. Therefore

$$\begin{aligned}\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 &\leq (s+m+1)\left\|\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*) - \boldsymbol{\Sigma}_{\mathcal{M}'\cup\mathcal{S}}\right\|_{\max} \\ &= O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right).\end{aligned} \tag{S2.55}$$

and, by (S2.53), $\left\|(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\boldsymbol{\theta}^*))^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$.

Having bounded both terms in (S2.50), we can conclude

$$\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right). \tag{S2.56}$$

As an immediate consequence, (S2.49) simplifies to

$$\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2} - \mathbf{I}_r\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right),$$

giving us a bound for the second term in (S2.48).

Shifting our focus to the first term of (S2.48), we take a similar approach to (S2.52) to show that for any n

$$\begin{aligned}\lambda_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} &\geq \lambda_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} \\ &\quad - \left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\right\|_2^2\left\|\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}} - \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\right\|_2.\end{aligned}$$

We see that $\lambda_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} = \lambda_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Psi}\boldsymbol{\Psi}^{-1/2}\right\} = \lambda_{\min}\{\mathbf{I}_r\} = 1$. We have already established that $\left\|\boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}} - \left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right) = o_p(1)$

and $\left\|\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\right\|_2 = O(1)$ if $(s+m)^3\log(s+m) = o(n)$. Therefore with probability converging to 1 as $n \to \infty$, we have

$$\liminf_n \lambda_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} > 0.$$

By extension, we see that

$$\begin{aligned}\left\|\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\}^{-1}\right\|_2 &= \lambda^{-1}_{\min}\left\{\boldsymbol{\Psi}^{-1/2}\widetilde{\mathbf{C}}\left[\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_a)\right]^{-1}\widetilde{\mathbf{C}}'\boldsymbol{\Psi}^{-1/2}\right\} \\ &= O_p(1).\end{aligned}$$

Having shown that the two terms on the right hand side of (S2.48) are $O_p(1)$ and $O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$, respectively, our proof of (L3.6) is complete.

Moving on to (L3.7), we see that

$$\begin{aligned}&\left\|\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1}\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}} - \mathbf{I}_{s+m+1}\right\|_2 \\ &\qquad = \left\|\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\left(\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right)\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \\ &\qquad = \left\|\boldsymbol{\Sigma}^{1/2}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2^2\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 \\ &\qquad = O\left(\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2\right)\end{aligned}$$

by (L3.1). By the same argument we used to prove (S2.56), we can show that $\left\|\left(\nabla^2_{\mathcal{M}'\cup\mathcal{S}}\ell_n(\hat{\boldsymbol{\theta}}_0)\right)^{-1} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M}'\cup\mathcal{S}}\right\|_2 = O_p\left(\frac{(s+m)\sqrt{\log(s+m)}}{\sqrt{n}}\right)$, establishing (L3.7). □

S.3 Null Distributions of p-values

In this section, we examine the empirical distributions of the p-values for the partial penalized Wald, score, and likelihood ratio tests in simulations where the null hypothesis is true. Based on Corollary 1, we would expect these p-values to approximately follow a standard uniform distribution. As such, we plot the quantiles of the observed p-values against the theoretical quantiles of a Uniform$(0, 1)$ distribution.

We examine the same simulation settings as in Section 6 of the main paper: we consider every combination of $\Sigma = \mathbf{I}_p$ and $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j and $p \in \{50, 250, 400\}$. As in those simulations, we test the following four hypotheses:

- $\mathrm{H}_0^{(1)} : \beta_1 + \beta_2 = 0$
- $\mathrm{H}_0^{(2)} : \beta_2 = -2$
- $\mathrm{H}_0^{(3)} : \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$
- $\mathrm{H}_0^{(4)} : \beta_1 + \beta_2 = 0$, $\beta_2 = -2$, and $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$.

In these simulations, we set $\beta_0 = 1$ and $\boldsymbol{\beta} = (2, -2, \mathbf{0}_{p-2})$. As such, for $i = 1, 2, 3, 4$, $\mathrm{H}_0^{(i)}$ is true in all of these results.

Figures S.1 - S.6 present QQ-plots of the observed p-values from 600 simulation replications against the Uniform$(0, 1)$ distribution. Each figure contains results for simulations with one combination of Σ and p (for example, Figure S.1 contains results for simulations with $\Sigma = \mathbf{I}_p$, $p = 50$). We see that across all of the simulation settings examined and for all four null hypotheses, the quantiles of the observed p-values for the partial penalized tests align well with the theoretical quantiles of the Uniform$(0, 1)$ distribution. Thus we see that our finite sample simulation results agree with the theoretical results in Corollary 1.

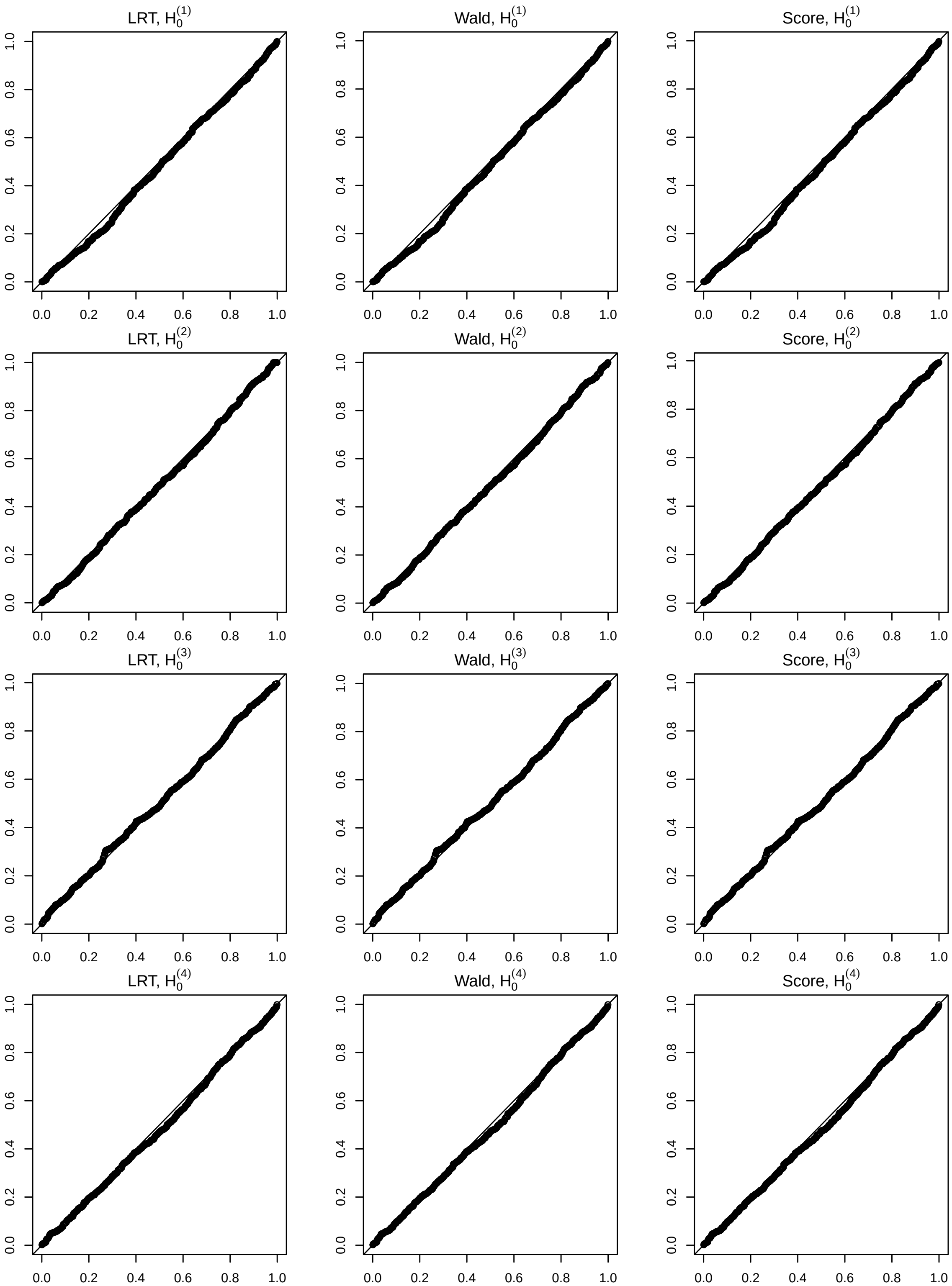

Figure S.1: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform$(0, 1)$ distribution when $\Sigma = \mathbf{I}_p$, $p = 50$.

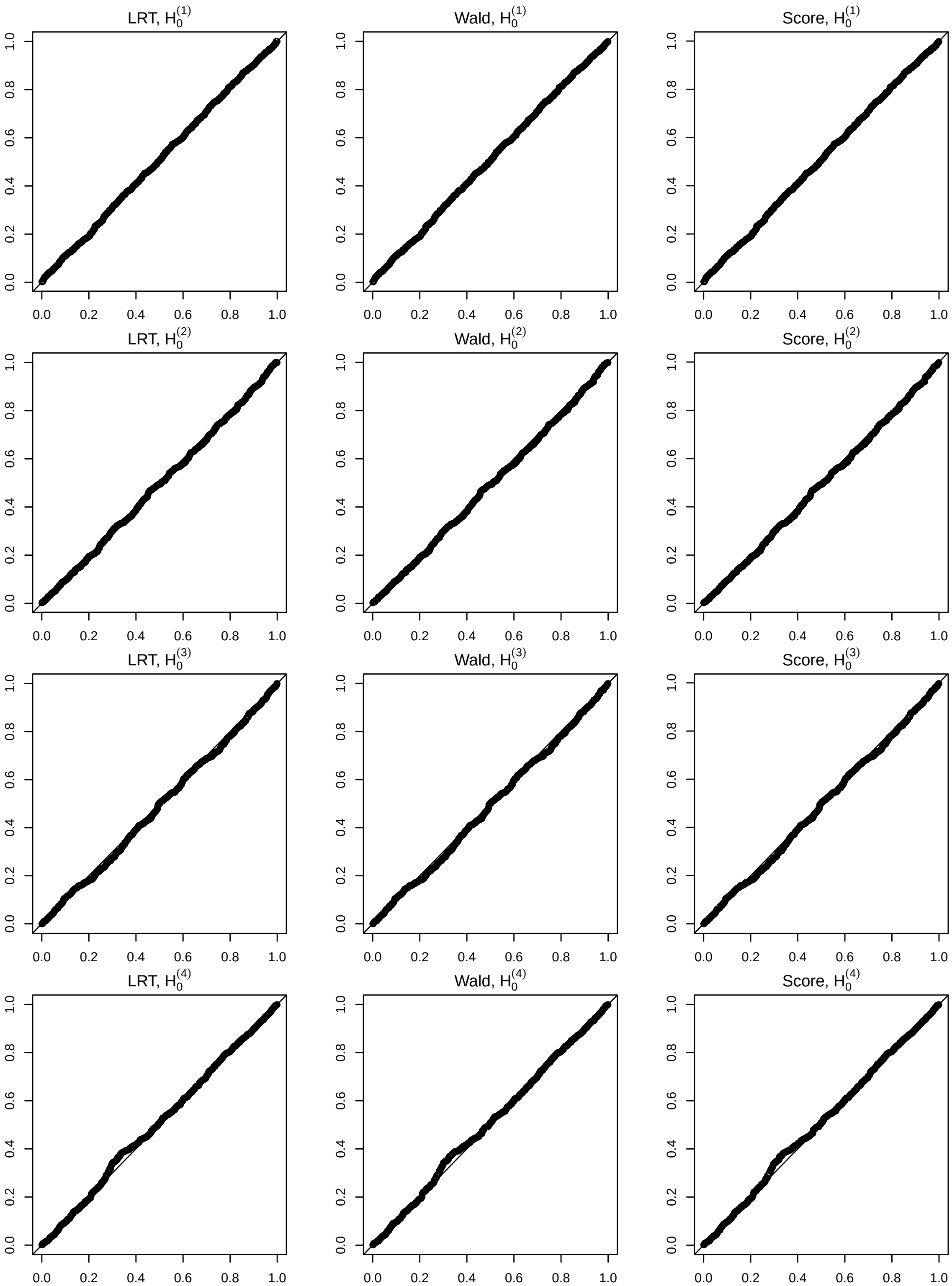

Figure S.2: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform$(0, 1)$ distribution when $\Sigma = \mathbf{I}_p$, $p = 250$.

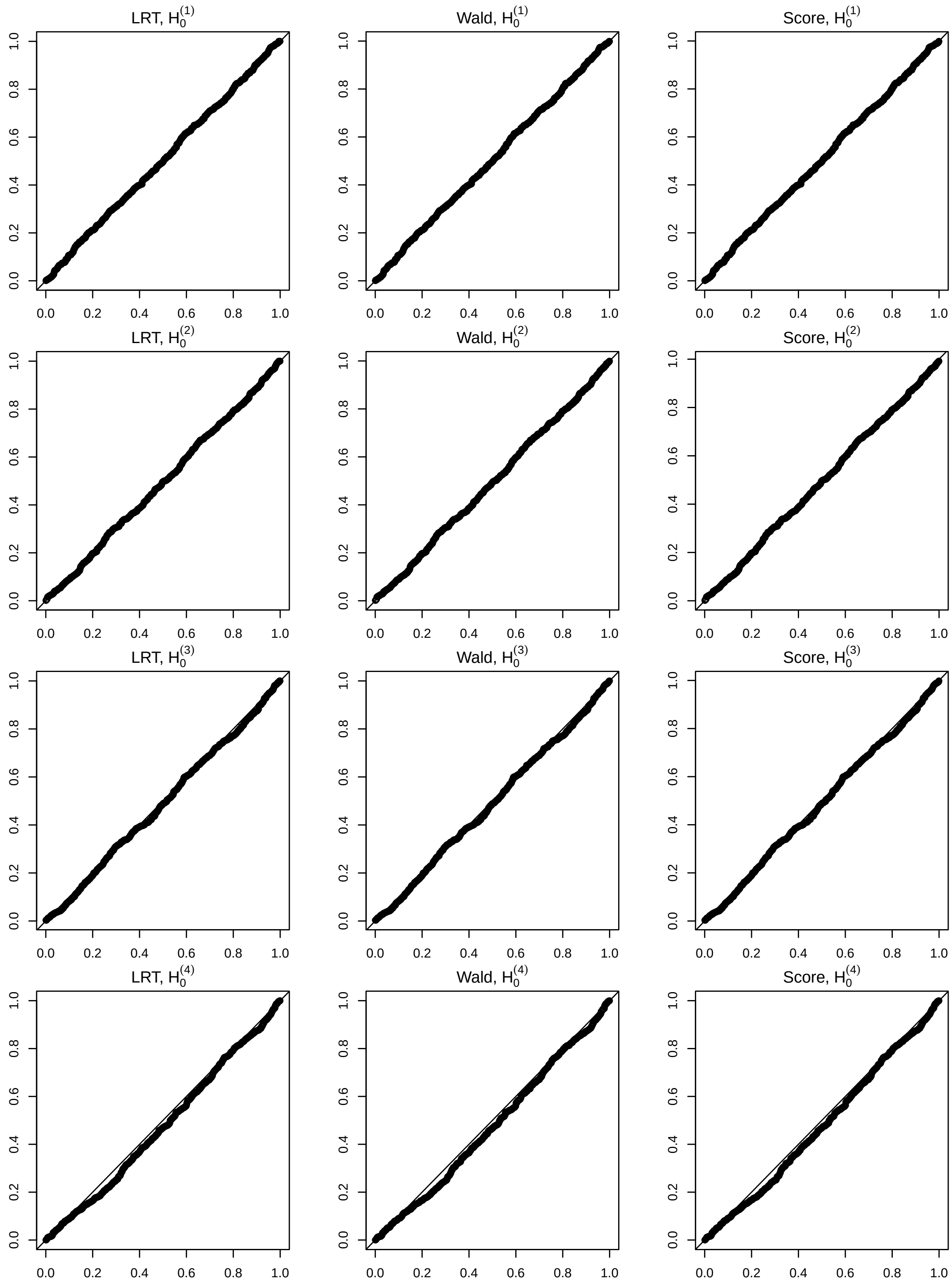

Figure S.3: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform$(0, 1)$ distribution when $\Sigma = \mathbf{I}_p$, $p = 400$.

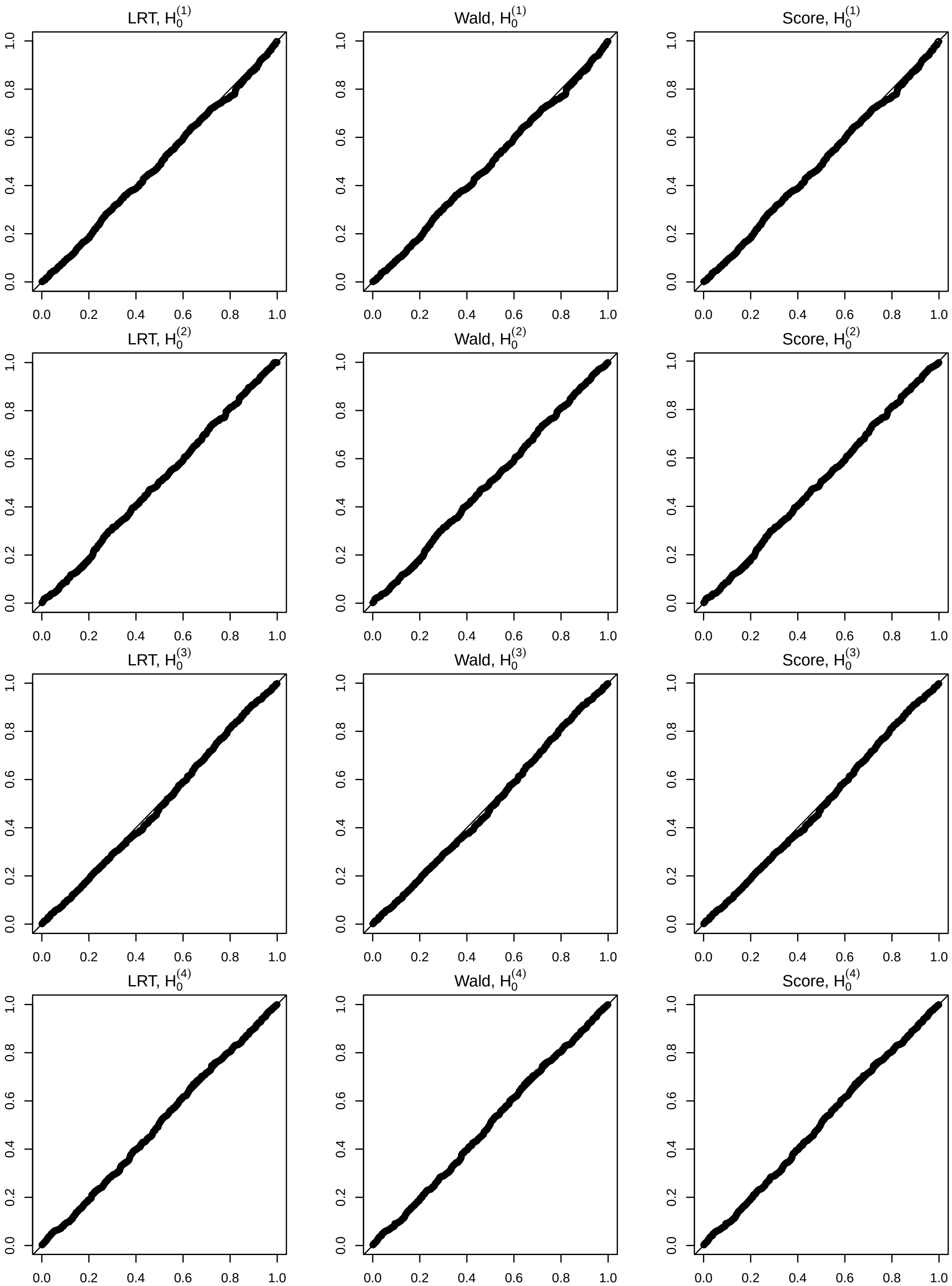

Figure S.4: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform(0, 1) distribution when $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j, $p = 50$.

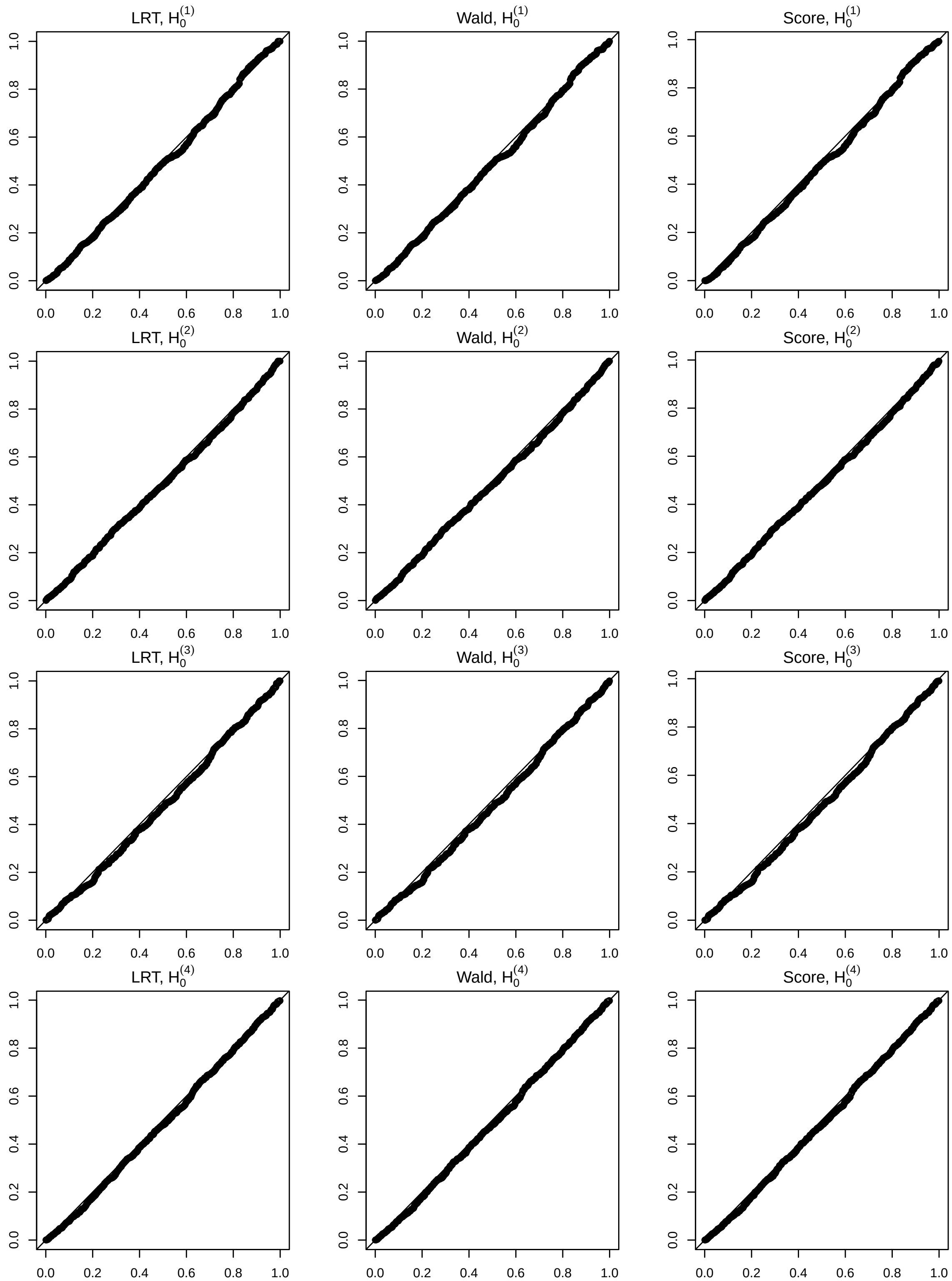

Figure S.5: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform(0, 1) distribution when $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j, $p = 250$.

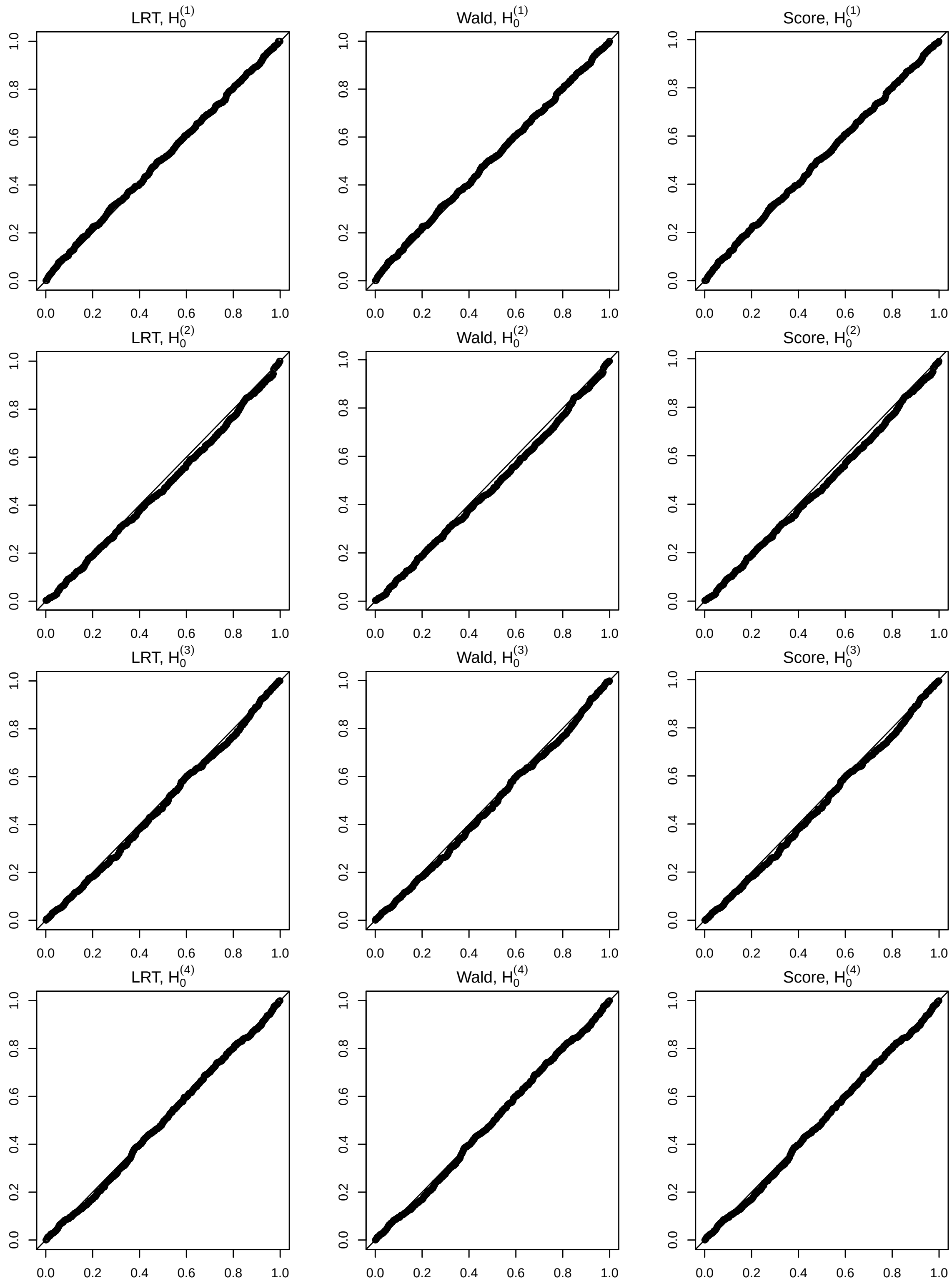

Figure S.6: QQ-plots of observed partial penalized test p-values from 600 simulation replications against the Uniform(0, 1) distribution when $\Sigma_{ij} = 0.5^{|i-j|}$ for all i, j, $p = 400$.

S.4 The Effects of n and ρ

In this section, we conduct additional simulation studies to examine how the sample size n and the correlation coefficient ρ in the AR1 model structure impact the power of the partial penalized Tobit tests. As in Section 6 of the main paper, we test the following four hypotheses:

- $\mathrm{H}_0^{(1)} : \beta_1 + \beta_2 = 0$
- $\mathrm{H}_0^{(2)} : \beta_2 = -2$
- $\mathrm{H}_0^{(3)} : \beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$
- $\mathrm{H}_0^{(4)} : \beta_1 + \beta_2 = 0$, $\beta_2 = -2$, and $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$

and generate the data with $\sigma = 1$, $\beta_0 = 1$, and $\boldsymbol{\beta} = (2, -2 - h_1, \mathbf{0}_{p-2})$, varying h_1 to create different test cases. We run 600 replications in each simulation setting to estimate the power of the tests.

We first examine the effect of n on the power of the partial penalized Tobit tests. We fix $p = 50$ and $\Sigma = \mathbf{I}_p$ and vary $n \in \{100, 200, 400\}$ across simulation settings. Table S.1 reports the estimated rejection probabilities by n. As in our simulation study in the main paper, we see that the partial penalized tests all achieve rejection probabilities close to their nominal size of $\alpha = 0.05$ when the null hypothesis is true, that the tests have similar rejection probabilities to each other within each simulation setting, and that the power of the tests steadily increases with h_1. In addition, we see that the power of the tests increases with n when the null is false, as one would expect.

Table S.1: Estimated rejection probabilities by n

	n = 100			n = 200			n = 400		
	LRT	Wald	Score	LRT	Wald	Score	LRT	Wald	Score
h_1					$H_0^{(1)}$				
0.0	4.33 (0.83)	4.17 (0.82)	4.67 (0.86)	6 (0.97)	6 (0.97)	6.17 (0.98)	4.67 (0.86)	4.67 (0.86)	4.67 (0.86)
0.1	11.5 (1.3)	11.5 (1.3)	11.5 (1.3)	13.67 (1.4)	13.5 (1.4)	13.5 (1.4)	22 (1.69)	21.83 (1.69)	21.83 (1.69)
0.2	23.33 (1.73)	23.17 (1.72)	23 (1.72)	36.5 (1.97)	36.5 (1.97)	36.5 (1.97)	70 (1.87)	70 (1.87)	70 (1.87)
0.4	67 (1.92)	66.5 (1.93)	66.5 (1.93)	92 (1.11)	92.17 (1.1)	92.17 (1.1)	99.67 (0.24)	99.67 (0.24)	99.67 (0.24)
h_1					$H_0^{(2)}$				
0.0	5.67 (0.94)	5.33 (0.92)	5.5 (0.93)	3.67 (0.77)	4 (0.8)	3.83 (0.78)	4.83 (0.88)	4.83 (0.88)	4.83 (0.88)
0.1	11 (1.28)	10.83 (1.27)	10.83 (1.27)	18.5 (1.59)	18.17 (1.57)	18.17 (1.57)	30 (1.87)	29.33 (1.86)	29.33 (1.86)
0.2	32.5 (1.91)	31.67 (1.9)	31.5 (1.9)	54.67 (2.03)	54.17 (2.03)	54.17 (2.03)	81.67 (1.58)	81.17 (1.6)	81.33 (1.59)
0.4	77.33 (1.71)	77.17 (1.71)	77.17 (1.71)	98 (0.57)	98 (0.57)	98 (0.57)	100 (0)	100 (0)	100 (0)
h_1					$H_0^{(3)}$				
0.0	6.83 (1.03)	6.83 (1.03)	6.83 (1.03)	5.83 (0.96)	5.83 (0.96)	6 (0.97)	5.83 (0.96)	5.83 (0.96)	5.83 (0.96)
0.1	7 (1.04)	7 (1.04)	7 (1.04)	9.83 (1.22)	9.67 (1.21)	9.67 (1.21)	11.17 (1.29)	11.17 (1.29)	11.17 (1.29)
0.2	13 (1.37)	12.83 (1.37)	13.17 (1.38)	23.83 (1.74)	23.83 (1.74)	23.83 (1.74)	43 (2.02)	42.83 (2.02)	42.83 (2.02)
0.4	40.83 (2.01)	40.83 (2.01)	40.67 (2.01)	62.33 (1.98)	62.33 (1.98)	62.33 (1.98)	93.17 (1.03)	93.17 (1.03)	93.17 (1.03)
h_1					$H_0^{(4)}$				
0.0	7.33 (1.06)	6.83 (1.03)	6.83 (1.03)	4.5 (0.85)	4 (0.8)	4 (0.8)	6 (0.97)	6 (0.97)	6 (0.97)
0.1	9.5 (1.2)	8.5 (1.14)	8.5 (1.14)	13.83 (1.41)	13.33 (1.39)	13.17 (1.38)	23.83 (1.74)	23.83 (1.74)	23.83 (1.74)
0.2	25.33 (1.78)	23.83 (1.74)	23.83 (1.74)	36.17 (1.96)	35.5 (1.95)	35.5 (1.95)	68.5 (1.9)	68.33 (1.9)	68.33 (1.9)
0.4	67.83 (1.91)	66.67 (1.92)	66.67 (1.92)	96.17 (0.78)	96 (0.8)	95.83 (0.82)	99.83 (0.17)	99.83 (0.17)	99.83 (0.17)

Next we examine the effect of ρ on the power of the partial penalized Tobit tests. We fix $p = 50$ and $n = 200$ and vary $\rho \in \{0.7, 0.8, 0.9\}$ in the predictor covariance matrix $\Sigma_{ij} = \rho^{|i-j|}$ across simulation settings. Table S.2 reports the estimated rejection probabilities by ρ. As in our previous simulation studies, we see that the rejection probabilities of the tests are close to their nominal size of $\alpha = 0.05$ when the null hypothesis is true and that the power of the tests increases with h_1. While the rejection probabilities of the three tests are similar to each other in most cases, we note that the powers of the LRT and (to a lesser extent) the score test are lower than that of the Wald test when $\rho = 0.9$ and $h_1 = 0.4$ or 0.2, though the differences are smaller in the latter case. We also see that the powers of the tests of $\mathrm{H}_0^{(2)}$ decrease as ρ increases when the null is false. This is likely because $\mathrm{H}_0^{(2)}$ tests a single coefficient and the high correlation between predictors is masking the predictor's signal. Conversely, the powers of the tests of $\mathrm{H}_0^{(3)}$ appear to slightly increase as ρ increases when $h_1 = 0.1$ or 0.2, perhaps because $\mathrm{H}_0^{(3)}$ tests four highly correlated predictors.

Table S.2: Estimated rejection probabilities by ρ

	$\rho = 0.7$			$\rho = 0.8$			$\rho = 0.9$		
	LRT	Wald	Score	LRT	Wald	Score	LRT	Wald	Score
h_1					$H_0^{(1)}$				
0.0	5.5 (0.93)	5.67 (0.94)	5.67 (0.94)	5.83 (0.96)	6 (0.97)	5.83 (0.96)	4.83 (0.88)	4.5 (0.85)	5 (0.89)
0.1	22.33 (1.7)	22.5 (1.7)	22.33 (1.7)	23.5 (1.73)	23.67 (1.74)	23.5 (1.73)	24.83 (1.76)	25.67 (1.78)	25 (1.77)
0.2	60.67 (1.99)	61.5 (1.99)	61 (1.99)	66 (1.93)	67 (1.92)	66.83 (1.92)	67.33 (1.91)	72.67 (1.82)	68.17 (1.9)
0.4	97.33 (0.66)	99.5 (0.29)	99.33 (0.33)	94.5 (0.93)	99.83 (0.17)	98.83 (0.44)	78.83 (1.67)	99.67 (0.24)	85.5 (1.44)
h_1					$H_0^{(2)}$				
0.0	5.33 (0.92)	5.33 (0.92)	5.33 (0.92)	4.83 (0.88)	4.83 (0.88)	4.83 (0.88)	4.5 (0.85)	4.5 (0.85)	4.83 (0.88)
0.1	12.17 (1.33)	11.33 (1.29)	11.17 (1.29)	12.17 (1.33)	11.83 (1.32)	11.5 (1.3)	8.67 (1.15)	8.67 (1.15)	8.67 (1.15)
0.2	32.5 (1.91)	32.17 (1.91)	32.17 (1.91)	30.17 (1.87)	29.67 (1.86)	29.67 (1.86)	17.67 (1.56)	17.83 (1.56)	17.33 (1.55)
0.4	88.33 (1.31)	88.67 (1.29)	88.67 (1.29)	79.67 (1.64)	79.83 (1.64)	79.5 (1.65)	58 (2.01)	59.17 (2.01)	58.67 (2.01)
h_1					$H_0^{(3)}$				
0.0	5.67 (0.94)	5.67 (0.94)	5.83 (0.96)	5.83 (0.96)	5.33 (0.92)	5.67 (0.94)	6.33 (0.99)	5.83 (0.96)	6.17 (0.98)
0.1	19.67 (1.62)	19.67 (1.62)	19.83 (1.63)	19.67 (1.62)	19.67 (1.62)	19.83 (1.63)	24.33 (1.75)	25.17 (1.77)	25 (1.77)
0.2	54.17 (2.03)	54.17 (2.03)	54 (2.03)	58.17 (2.01)	58.67 (2.01)	58.33 (2.01)	68.83 (1.89)	72.33 (1.83)	69.83 (1.87)
0.4	97.67 (0.62)	98.33 (0.52)	98.33 (0.52)	95.5 (0.85)	99 (0.41)	98.17 (0.55)	81.5 (1.59)	99.83 (0.17)	88.17 (1.32)
h_1					$H_0^{(4)}$				
0.0	4.5 (0.85)	4.5 (0.85)	4.33 (0.83)	6 (0.97)	5.83 (0.96)	5.83 (0.96)	4.5 (0.85)	4.5 (0.85)	4.67 (0.86)
0.1	16.17 (1.5)	15.67 (1.48)	15.5 (1.48)	16.83 (1.53)	16.5 (1.52)	16.5 (1.52)	16.33 (1.51)	16.33 (1.51)	16.33 (1.51)
0.2	49.67 (2.04)	49.17 (2.04)	49.17 (2.04)	53 (2.04)	52.5 (2.04)	52.5 (2.04)	52.5 (2.04)	54 (2.03)	52.67 (2.04)
0.4	98.83 (0.44)	99 (0.41)	99 (0.41)	97 (0.7)	99.17 (0.37)	98.17 (0.55)	89 (1.28)	99 (0.41)	92.5 (1.08)

References

Jacobson, T. and Zou, H. (2023). High-dimensional Censored Regression via the Penalized Tobit Likelihood. *Journal of Business & Economic Statistics* .

Sampford, M. R. (1953). Some Inequalities on Mill's Ratio and Related Functions. *The Annals of Mathematical Statistics* **24**, 130–132.

Shi, C., Song, R., Chen, Z. and Li, R. (2019). Linear Hypothesis Testing for High Dimensional Generalized Linear Models. *Annals of Statistics* **47**, 2671–2703.

Vershynin, R. (2018). *High Dimensional Probability. An Introduction with Applications in Data Science*. Cambridge University Press.