Supplement to “Statistical Inference for High-Dimensional Models via Recursive Online-Score Estimation”

Chengchun Shi
Department of Statistics, London School of Economics and Political Science

Rui Song and Wenbin Lu
Department of Statistics, North Carolina State University

Runze Li
Department of Statistics, Pennsylvania State University

This supplementary article is organized as follows. In Section A, we present proofs of Theorem 2.1, Theorem 3.1, Theorem 3.2, Theorem 3.3, Lemma A.1, Lemma A.2, Lemma A.3 and Lemma A.4. In Section B, we provide detailed discussions on our technical conditions and compare them with those imposed in the existing literature. In Section C, we give more details on the extensions to the generic penalized M-estimators.

A Proofs

A.1 Proof of Theorem 2.1

Before proving Theorem 2.1, we present the following lemmas whose proofs are given in the supplementary material.

Lemma A.1. *Under conditions in Theorem 2.1, we have*

$$\min_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sigma_{\mathcal{M},j_0} \ge \sqrt{\bar{c}}, \qquad \max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \le (\bar{c})^{-1/2} c_0, \tag{A.1}$$

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} E\left\{\frac{1}{\sigma^4_{\mathcal{M},j_0}}\left(X_{0,j_0} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\boldsymbol{X}_{0,\mathcal{M}}\right)^4\right\} \le \frac{c_0^4}{\bar{c}^2}\left(1+\frac{c_0^2}{\bar{c}}\right)^2, \tag{A.2}$$

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ |\mathcal{M}|\le\kappa_n}} E\|X_{0,\mathcal{M}}\|_2^2 \le \kappa_n c_0^2, \tag{A.3}$$

where $\bar{c}$ *and* c_0 *are defined in Condition (A2) and (A3), respectively. Moreover, we have with probability tending to* 1 *that*

$$\max_{j\in[1,\dots,p]} |X_{0,j}| \le \sqrt{3c_0^2 \max(\log p, \log n)}. \tag{A.4}$$

Lemma A.2. *Under conditions in Theorem 2.1, the following events hold with probability tending to* 1*,*

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right), \tag{A.5}$$

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}_{\mathcal{M},j_0} - \sigma_{\mathcal{M},j_0}| \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right), \tag{A.6}$$

for some constant $\bar{c}_0 > 0$*, where* $\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}$ *and* $\hat{\sigma}^2_{\mathcal{M},j_0} = \widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \widehat{\boldsymbol{\Sigma}}^T_{\mathcal{M},j_0}\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}$*,*

$$\overline{\boldsymbol{\omega}}_{\mathcal{M},j_0} = \boldsymbol{\omega}_{\mathcal{M},j_0} + \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\{\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0} - (\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\}.$$

In addition, we have

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 = O_p\left(\frac{\kappa_n\log p}{n}\right). \tag{A.7}$$

Lemma A.3. *Under conditions in Theorem 2.1, the following events hold with probability* *tending to* 1*,*

$$\max_{j\in[1,\ldots,p]}\left|\sum_{t=s_n}^{n-1}\frac{X_{t+1,j}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\left(X_{t+1,j_0}-\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)I(j\in\widehat{\mathcal{M}}_{j_0}^{(t)})\right|\le\bar{c}_*\sqrt{\log p}, \quad (A.8)$$

$$\sum_{t=s_n}^{n-1}\left\|\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}-\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right\|_2^2\le\bar{c}_*n\eta_n^2, \quad (A.9)$$

for some constant $\bar{c}_*>0$*.*

Proof of Theorem 2.1: Under (A1), it follows from the Bonferroni's inequality that

$$\Pr\left(\mathcal{M}_{j_0}\subseteq\bigcap_{t=s_n}^{n-1}\widehat{\mathcal{M}}_{j_0}^{(t+1)}\right)\ge 1-O\left(\sum_{t=s_n}^{\infty}\frac{1}{t^{\alpha_0}}\right)\to 1. \quad (A.10)$$

Besides,

$$\Pr\left(\mathcal{M}_{j_0}\subseteq\widehat{\mathcal{M}}_{j_0}^{(-s_n)}\right)\to 1. \quad (A.11)$$

Under the events defined in the left-hand-side (LHS) of (A.10) and (A.11), we have

$$\begin{aligned}\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0}-\beta_{0,j_0}) &= \underbrace{\sum_{t=s_n}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}}_{I_1}-\underbrace{\sum_{t=s_n}^{n-1}\frac{\widehat{Z}_{t+1,j_0}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}-\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)}_{I_2}\\ &\quad \underbrace{-\sum_{t=0}^{s_n-1}\frac{\widehat{Z}_{t+1,j_0}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}\left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}-\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}\right)}_{I_3}+\underbrace{\sum_{t=0}^{s_n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}}_{I_4}.\end{aligned}$$

In the following, we break the proof into four steps. In the first three steps, we show $I_j=o_p(1)$, for $j=2,3,4$, respectively. In the last step, we prove

$$I_1\xrightarrow{d}N(0,\sigma_0^2).$$

By Assumption (A6) and Slutsky's theorem, we have

$$\frac{\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0}-\beta_{0,j_0})}{\hat{\sigma}}\xrightarrow{d}N(0,1).$$

The assertion therefore follows.

Step 1: Let

$$I_2^* = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right).$$

$|I_2 - I_2^*|$ is upper bounded by

$$\sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left| \left(\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right| \left| \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right|.$$

Under the given conditions, we have $\kappa_n \log p = o(n)$. Under the events defined in (A.1) and (A.6), we have for sufficiently large n,

$$\min_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \hat{\sigma}_{\mathcal{M},j_0} \ge \min_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \sigma_{\mathcal{M},j_0} - o(1) \ge \frac{\sqrt{\bar{c}}}{2},$$

and hence

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \frac{1}{\hat{\sigma}_{\mathcal{M},j_0}} \le \frac{2}{\sqrt{\bar{c}}}. \tag{A.12}$$

Under the events defined in Condition (A1) and (A.12), $|I_2 - I_2^*|$ can be upper bounded by

$$\begin{aligned} &\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \frac{1}{\hat{\sigma}_{\mathcal{M},j_0}} \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}} \left| \left(\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right| \left| \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right| \\ &\qquad \le \frac{2}{\sqrt{\bar{c}n}} \sum_{t=s_n}^{n-1} \left| \left(\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right| \left| \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right|. \end{aligned}$$

By Cauchy-Schwarz inequality, we have with probability tending to 1 that

$$|I_2 - I_2^*| \le 2(\bar{c})^{-1/2} \sqrt{n I_2^{(1)} I_2^{(2)}}, \tag{A.13}$$

where

$$\begin{aligned} I_2^{(1)} &= \frac{1}{n} \sum_{t=s_n}^{n-1} \left| \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right|^2, \\ I_2^{(2)} &= \frac{1}{n} \sum_{t=s_n}^{n-1} \left| \left(\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right|^2. \end{aligned}$$

It follows from (A.9) that

$$I_2^{(1)} = O(\eta_n^2), \tag{A.14}$$

with probability tending to 1.

For any $a, b \in \mathbb{R}$, we have by Cauchy-Schwarz inequality that $(a+b)^2 \le 2a^2 + 2b^2$. It follows that $I_2^{(2)} \le 2I_2^{(3)} + 2I_2^{(4)}$ where

$$\begin{aligned} I_2^{(3)} &= \frac{1}{n}\sum_{t=s_n}^{n-1}\left|\left(\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right|^2 \\ I_2^{(4)} &= \frac{1}{n}\sum_{t=s_n}^{n-1}\left|\left(\overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right|^2. \end{aligned}$$

By Condition (A1), (A.7) and Cauchy-Schwarz inequality, $I_2^{(3)}$ can be bounded by

$$\frac{1}{n}\sum_{t=s_n}^{n-1}\|\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 = O_p(n^{-2}\kappa_n^2\log^2 p)\frac{1}{n}\sum_{t=s_n}^{n-1}\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2.$$

In the following, we show

$$\frac{1}{n}\sum_{t=s_n}^{n-1}\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 = O_p(\kappa_n). \tag{A.15}$$

This further implies

$$I_2^{(3)} = O_p\left(\frac{\kappa_n^3\log^2 p}{n^2}\right) = O_p\left(\frac{\kappa_n\log p}{n}\right), \tag{A.16}$$

under the condition that $\kappa_n^2\log p = O(n/\log^2 n)$. To prove (A.15), it suffices to show

$$\frac{1}{n}\sum_{t=s_n}^{n-1}\mathrm{E}\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 = O(\kappa_n). \tag{A.17}$$

Since X_{t+1} and $\widehat{\mathcal{M}}_{j_0}^{(t)}$ is independent, we have by Condition (A1) that

$$\frac{1}{n}\sum_{t=s_n}^{n-1}\mathrm{E}\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 \le \frac{1}{n}\sum_{t=s_n}^{n-1}\sup_{\substack{\mathcal{M}\le[1,\ldots,p]\\|\mathcal{M}|\le\kappa_n}}\mathrm{E}\|\boldsymbol{X}_{t+1,\mathcal{M}}\|_2^2.$$

The RHS is $O(\kappa_n)$ by (A.3).

Consider $I_2^{(4)}$. For $i = 1, \ldots, n$ and any $\mathcal{M} \subseteq [1, \ldots, p]$, define

$$\overline{\boldsymbol{\omega}}_{\mathcal{M},j_0}^{(-i)} = \boldsymbol{\omega}_{\mathcal{M},j_0} + \frac{n-1}{n} \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \{ \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^{(-i)} - \boldsymbol{\Sigma}_{\mathcal{M},j_0} - (\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}^{(-i)} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}) \boldsymbol{\omega}_{\mathcal{M},j_0} \},$$

where

$$\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^{(-i)} = \frac{1}{n-1} \sum_{l \neq i} \boldsymbol{X}_{l,\mathcal{M}} X_{l,j_0} \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^{(-i)} = \frac{1}{n-1} \sum_{l \neq i} \boldsymbol{X}_{l,\mathcal{M}} \boldsymbol{X}_{l,\mathcal{M}}^T.$$

It follows that

$$\| \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0}^{(-i)} \|_2 \leq \frac{1}{n} \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \{ \boldsymbol{X}_{l,\mathcal{M}} X_{l,j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0} - (\boldsymbol{X}_{l,\mathcal{M}} \boldsymbol{X}_{l,\mathcal{M}}^T - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}) \boldsymbol{\omega}_{\mathcal{M},j_0} \}.$$

Under the event defined in (A.4), it follows from Condition (A2) and (A.1) that

$$\max_{\substack{\mathcal{M} \in [1,\ldots,p], i \in [1,\ldots,n] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \leq \kappa_n}} \| \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0}^{(-i)} \|_2 = O\left(\frac{\kappa_n \log p + \kappa_n \log n}{n} \right). \tag{A.18}$$

The condition $\kappa_n^2 \log p = O(n / \log^2 n)$ implies that $O_p(n^{-1} \kappa_n \log p) = o_p(n^{-1/2} \sqrt{\kappa_n \log p})$. By (A.7), we have with probability tending to 1 that

$$\max_{\substack{\mathcal{M} \subseteq [1,\ldots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \leq \kappa_n}} \| \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0} \|_2 \leq \left(\frac{\kappa_n \log p}{n} \right)^{1/2}.$$

Combining this together with (A.5), (A.18) and the condition $\kappa_n^2 \log p = O(n / \log^2 n)$ yields

$$\begin{aligned} & \max_{\substack{\mathcal{M} \in [1,\ldots,p], i \in [1,\ldots,n] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \leq \kappa_n}} \| \boldsymbol{\omega}_{\mathcal{M},j_0} - \overline{\boldsymbol{\omega}}_{\mathcal{M},j_0}^{(-i)} \|_2 \\ = & \; O\left(\frac{\kappa_n \log p + \kappa_n \log n}{n} + \frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} \right) = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} \right), \end{aligned} \tag{A.19}$$

with probability tending to 1. Define

$$\begin{aligned} I_2^{(5)} &= \frac{1}{n} \sum_{t=s_n}^{n-1} \left| \left(\overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^{(-t-1)} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right|^2, \\ I_2^{(6)} &= \frac{1}{n} \sum_{t=s_n}^{n-1} \left| \left(\overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \overline{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^{(-t-1)} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right|^2. \end{aligned}$$

By Cauchy-Schwarz inequality, we can similarly show $I_2^{(4)} \le 2I_2^{(5)} + 2I_2^{(6)}$. Using similar arguments in bounding $I_2^{(3)}$, we can similarly show that

$$I_2^{(6)} = O_p\left(\frac{\kappa_n^3(\log^2 p + \log^2 n)}{n^2}\right), \tag{A.20}$$

by (A.15) and (A.18). Under the events defined in Condition (A1) and (A.19), we have

$$I_2^{(5)} \le \frac{1}{n}\sum_{t=s_n}^{n-1}\left|\left(\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right|^2 \mathcal{I}\left\{\|\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n}\log p}{\sqrt{n}}\right)\right\},$$

where $\mathcal{I}\{\cdot\}$ denote the indicator function. Since $\boldsymbol{X}_{t+1}$ is independent of $\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ and $\widehat{\mathcal{M}}_{j_0}^{(t)}$, we have

$$\mathrm{E}I_2^{(5)} = \frac{1}{n}\sum_{t=s_n}^{n-1}\mathrm{E}\left\|\boldsymbol{\Sigma}^{1/2}_{\widehat{\mathcal{M}}_{j_0}^{(t)},\widehat{\mathcal{M}}_{j_0}^{(t)}}\left(\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\right)\right\|_2^2 \mathcal{I}\left\{\|\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n}\log p}{\sqrt{n}}\right)\right\}.$$

For any random variable Z, it follows from the definition of the Orlicz norm that

$$1 + \mathrm{E}\frac{Z^2}{\|Z\|_{\psi_2}^2} \le \mathrm{E}\exp\left(\frac{Z^2}{\|Z\|_{\psi_2}^2}\right) \le 2,$$

and hence

$$\mathrm{E}Z^2 \le \|Z\|_{\psi_2}^2. \tag{A.21}$$

Note that $\boldsymbol{\Sigma}$ is positive definite, we have by Condition (A3) that

$$\|\boldsymbol{\Sigma}\|_2 = \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^p\\ \|\boldsymbol{a}\|_2=1}} \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a} = \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^p\\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_0|^2 \le \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^p\\ \|\boldsymbol{a}\|_2=1}} \|\boldsymbol{a}^T\boldsymbol{X}_0\|_{\psi_2}^2 \le c_0^2, \tag{A.22}$$

It follows from (A.22) that

$$\begin{aligned}\mathrm{E}I_2^{(5)} \le \frac{1}{n}\sum_{t=s_n}^{n-1}\mathrm{E}\lambda_{\max}(\boldsymbol{\Sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},\widehat{\mathcal{M}}_{j_0}^{(t)}})\|\overline{\boldsymbol{\omega}}^{(-t-1)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2 \mathcal{I}\left\{\|\overline{\boldsymbol{\omega}}^{-t-1}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n}\log p}{\sqrt{n}}\right)\right\} \\ = O(n^{-1}\kappa_n\log p),\end{aligned}$$

with probability tending to 1. This further implies $I_2^{(5)} = O_p(n^{-1}\kappa_n\log p)$. Combining this together with (A.20) and the condition $\kappa_n^2 \log p = O(n/\log^2 n)$ yields that

$$I_2^{(4)} = O_p\left(\frac{\kappa_n \log p}{n} + \frac{\kappa_n^3(\log^2 p + \log^2 n)}{n^2}\right) = O_p\left(\frac{\kappa_n \log p}{n}\right).$$

This together with (A.13), (A.14) and (A.16) yields that

$$|I_2 - I_2^*| = O_p(\eta_n \sqrt{\kappa_n \log p}).$$

It follows from Condition (A4) that $|I_2 - I_2^*| = o_p(1)$.

Let

$$I_2^{**} = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right).$$

By definition, we have

$$|I_2^* - I_2^{**}| = \sum_{t=s_n}^{n-1} \frac{|\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}|}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left|\left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right|.$$

Using similar arguments in bounding $I_2 - I_2^*$, we can show

$$|I_2^* - I_2^{**}| \le \frac{2\bar{c}_0}{\bar{c}}\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) \sum_{t=s_n}^{n-1} \left|X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right| \left|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right|.$$

By Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\sum_{t=s_n}^{n-1} \left|X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right| \left|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right| \\
\le\ &\left(\sum_{t=s_n}^{n-1} \left|X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right|^2\right)^{1/2} \left(\sum_{t=s_n}^{n-1} \left|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right|^2\right)^{1/2}.
\end{aligned}$$

By (A.1), (A.21) and Condition (A3), we can show

$$\sum_{t=s_n}^{n-1} \mathrm{E}\left|X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right|^2 = O(n),$$

and hence

$$\sum_{t=s_n}^{n-1} \left| X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right|^2 = O_p(n). \tag{A.23}$$

This together with (A.9) yields

$$\sum_{t=s_n}^{n-1} \left| X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right| \left| \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}_{j_0}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right) \right| = O_p(\sqrt{n}\eta_n).$$

It follows that

$$|I_2^* - I_2^{**}| = O_p(\eta_n \sqrt{\kappa_n \log p}),$$

which is $o_p(1)$ under (A4).

Thus, to prove $I_2 = o_p(1)$, it suffices to show $I_2^{**} = o_p(1)$. Define

$$I_{2,j} = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right) X_{t+1,j} \mathcal{I}(j \in \widehat{\mathcal{M}}^{(t)}_{j_0}).$$

We have $I_2^{**} = \sum_{j=1}^p I_{2,j}(\widetilde{\beta}_j - \beta_{0,j})$. Therefore,

$$|I_2^{**}| \le \max_{j \in [1,\ldots,p]} |I_{2,j}| \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1. \tag{A.24}$$

Let $\sigma(\mathcal{F}_t)$ be the σ-algebra generated by $\{(\boldsymbol{X}_1, Y_1), \ldots, (\boldsymbol{X}_t, Y_t)\}$. Then, each $I_{2,j}$ forms a mean zero martingale with respect to $\sigma(\mathcal{F}_t)$. To see this, note that $\widehat{\mathcal{M}}^{(t)}_{j_0}$ is fixed given $\mathcal{F}_t$. If $j \notin \mathcal{M}^{(t)}_{j_0}$, we have

$$\begin{aligned}
& \mathrm{E}\left\{ \frac{1}{\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right) X_{t+1,j} \mathcal{I}(j \in \widehat{\mathcal{M}}^{(t)}_{j_0}) | \mathcal{F}_t \right\} \\
= \ & \mathrm{E}\left\{ \frac{1}{\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} \right) X_{t+1,j} | \mathcal{F}_t \right\} \mathcal{I}(j \in \widehat{\mathcal{M}}^{(t)}_{j_0}) = 0.
\end{aligned}$$

If $j \in \mathcal{M}_{j_0}^{(t)}$, then we have

$$\mathrm{E}\left\{\frac{X_{t+1,j}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\left(X_{t+1,j_0}-\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)|\mathcal{F}_t\right\}=\frac{\boldsymbol{\Sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}-\boldsymbol{\Sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}=0.$$

By some exponential inequalities for martingales, we show in (A.8) that $\Pr(\max_j |I_{2,j}| \geq \bar{c}_*\sqrt{\log p}) \to 0$. It follows from Condition (A4) that $\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1 \leq (k_0+1)\|\widetilde{\boldsymbol{\beta}}_{\mathcal{M}_0}-\boldsymbol{\beta}_{0,\mathcal{M}_0}\|_1 \leq \sqrt{|\mathcal{M}_0|}(k_0+1)\|\widetilde{\boldsymbol{\beta}}_{\mathcal{M}_0}-\boldsymbol{\beta}_{0,\mathcal{M}_0}\|_2 \leq \sqrt{|\mathcal{M}_0|}(k_0+1)\eta_n$, with probability tending to 1. Under (A1), we have $|\mathcal{M}_0| \leq \kappa_n - 1$. It follows that

$$\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1 = O(\sqrt{\kappa_n}\eta_n), \tag{A.25}$$

with probability tending to 1. Since $\eta_n\sqrt{\kappa_n \log p} = o(1)$, we have $\max_j |I_{2,j}|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1 = o_p(1)$. This together with (A.24) gives $I_2^{**} = o_p(1)$.

Step 2: Using similar arguments in Step 1, we can show that I_3 is asymptotically equivalent to I_3^{**}, defined as

$$I_3^{**} = \sum_{t=0}^{s_n-1}\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} Z_{t+1,j_0}\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}\left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}-\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}\right),$$

where $Z_{t+1,j_0} = X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}$. Hence, it suffices to show $I_3^{**} = o_p(1)$. Note that $|I_3^{**}|$ is upper bounded by

$$|I_3^{**}| \leq \max_{j\in[1,\dots,p]} |I_{3,j}|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1, \tag{A.26}$$

where

$$I_{3,j} = \sum_{t=0}^{s_n-1}\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}\left(X_{t+1,j_0}-\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}}\right)X_{t+1,j}\mathcal{I}(j \in \widehat{\mathcal{M}}_{j_0}^{(-s_n)}).$$

Given $\{(\boldsymbol{X}_{s_n+1}, Y_{s_n+1}), \dots, (\boldsymbol{X}_n, Y_n)\}$, the set $\widehat{\mathcal{M}}_{j_0}^{(-s_n)}$ is fixed. For any $j \in [1,\dots,p]$, $I_{3,j}$ corresponds to a sum of mean zero i.i.d random variables. Similar to the proof of Lemma A.3, we can show

$$\Pr(\max_j |I_{3,j}| \leq c_*\sqrt{\log p}) \to 1,$$

for some constant $c_* > 0$ that is independent of $\widehat{\mathcal{M}}_{j_0}^{(-s_n)}$. By (A.26) and Condition (A4), we have $|I_3^{**}| \to 0$ with probability tending to 1. This proves $I_3 = o_p(1)$.

Step 3: Let

$$I_4^* = \sum_{t=0}^{s_n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \boldsymbol{X}_{t+1,\mathcal{M}_{j_0}^{(-s_n)}} \right) \varepsilon_{t+1},$$

we have

$$I_4 - I_4^* = \sum_{t=0}^{s_n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \left(\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}} - \widehat{\boldsymbol{\omega}}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}} \right) \varepsilon_{t+1}.$$

We first show $I_4 - I_4^* = o_p(1)$. Since $\varepsilon_1, \ldots, \varepsilon_{s_n}$ are independent of $\{\boldsymbol{X}_i\}_{i=1}^n$, it follows from the Chebyshev's inequality that

$$\begin{aligned}
& \Pr\left(|I_4 - I_4^*| > t^* | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \varepsilon_{s_n+1}, \ldots, \varepsilon_n\right) \\
\leq\ & \frac{1}{(t^*)^2} \mathrm{E}\{(I_4 - I_4^*)^2 | \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n, \varepsilon_{s_n+1}, \ldots, \varepsilon_n\} \\
\leq\ & \sum_{t=0}^{s_n-1} \frac{\sigma_0^2}{(t^*)^2 n \hat{\sigma}^2_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \left\{ \left(\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}} \right\}^2. \qquad \text{(A.27)}
\end{aligned}$$

By (A.5), (A.12) and (A.15), we can show that

$$\sum_{t=0}^{s_n-1} \frac{\sigma_0^2}{n \hat{\sigma}^2_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \left\{ \left(\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \right)^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}} \right\}^2 = O_p\left(\frac{\kappa_n^2 \log p}{n} \right).$$

In view of (A.27), this further implies that $|I_4 - I_4^*| = O_p(n^{-1/2}\kappa_n \sqrt{\log p})$. Under the given conditions, we obtain that $I_4 - I_4^* = o_p(1)$. Similarly, we can show $I_4^* - I_4^{**} = o_p(1)$, where

$$I_4^{**} = \sum_{t=0}^{s_n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(-s_n)}} \right) \varepsilon_{t+1}.$$

Thus, it suffices to show $I_4^{**} = o_p(1)$. Note that we have

$$\mathrm{E}[(I_4^{**})^2 | \{(\boldsymbol{X}_{s_n+1}, \varepsilon_{s_n+1}), \ldots, (\boldsymbol{X}_n, \varepsilon_n)\}] = \frac{s_n \sigma_0^2}{n},$$

and hence $\mathrm{E}(I_4^{**})^2 \le s_n\sigma_0^2/n$. Since $s_n = o(n)$, it follows from Chebyshev's inequality that $I_4^{**} = o_p(1)$.

Step 4: For $t = s_n, \ldots, n-1$, define

$$\mathcal{A}_t = \left\{ \hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \le \sqrt{\bar{c}}/2 \right\} \bigcap \left\{ \left\| \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \right\|_2 \le \bar{c}_0 \frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} \right\}.$$

By (A.5), (A.12) and Condition (A1), we have $\Pr(\cap_{t=s_n}^{n-1} \mathcal{A}_t) \to 1$. Hence, we have $\Pr(I_1 = I_1^*) \to 1$ and $\Pr(I_1^{**} = I_1^{***}) \to 1$ where

$$I_1^* = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \left(X_{t+1,j_0} - \widehat{\boldsymbol{\omega}}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \varepsilon_{t+1} \mathcal{I}(\mathcal{A}_t),$$
$$I_1^{**} = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \varepsilon_{t+1} \mathcal{I}(\mathcal{A}_t),$$
$$I_1^{***} = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \varepsilon_{t+1}.$$

In the following, we prove $I_1^* = I_1^{**} + o_p(1)$. This further implies $I_1 = I_1^{***} + o_p(1)$. For any $t_0 > 0$,

$$\begin{aligned}
\Pr\left(|I_1^* - I_1^{**}| > t_0\right) &\le \frac{1}{nt_0^2} \mathrm{E} \left\{ \sum_{t=s_n}^{n-1} \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \left(\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \right)^T \boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}} \varepsilon_{t+1} \mathcal{I}(\mathcal{A}_t) \right\}^2 \\
= \;& \frac{1}{nt_0^2} \mathrm{E} \sum_{t=s_n}^{n-1} \left\{ \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \left(\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} \right)^T \boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}} \varepsilon_{t+1} \mathcal{I}(\mathcal{A}_t) \right\}^2 \\
\le \;& \frac{\sigma_0^2}{nt_0^2} \mathrm{E} \sum_{t=s_n}^{n-1} \left(\frac{\|\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}\|_2^2}{\hat{\sigma}^2_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \mathcal{I}(\mathcal{A}_t) \|\boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 \right) \\
\le \;& \frac{4\bar{c}_0^2\sigma_0^2}{\bar{c}nt_0^2} \left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} \right)^2 \mathrm{E} \sum_{t=s_n}^{n-1} \|\boldsymbol{X}_{t+1, \widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2^2 = O(n^{-1}\kappa_n^2 \log p) = o(1),
\end{aligned}$$

where the second equality is due to (A.17) and the last equality is due to the condition that $\kappa_n^2 \log p = O(n/\log^2 n)$. This implies $I_1^* = I_1^{**} + o_p(1)$ and hence $I_1 = I_1^{***} + o_p(1)$.

Similarly, we can show I_1 is asymptotically equivalent to

$$I_1^{****} = \frac{1}{\sqrt{n}} \sum_{t=s_n}^{n-1} \frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \varepsilon_{t+1}.$$

Observe that I_1^{****} is a mean zero martingale with respect to the filtration $\{\sigma(\mathcal{F}_t)\}_t$. Since $s_n = o(n)$, we have

$$\sum_{t=s_n}^{n-1} \mathrm{E}\left[\left\{ \frac{\varepsilon_{t+1}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right\}^2 |\mathcal{F}_t \right] = \frac{n-s_n}{n}\sigma_0^2 \to \sigma_0^2.$$

Let $Z_{t+1,j_0} = X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}$ for $t \geq s_n$. It follows from Condition (A1) and (A.2) that

$$\mathrm{E}\left\{ \frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^4} \left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right)^4 |\mathcal{F}_t \right\} \leq \frac{c_0^4}{\bar{c}^2} \left(1 + \frac{c_0^2}{\bar{c}} \right)^2.$$

By Hölder's inequality, we have

$$\mathrm{E}\left(\frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^3} |Z_{t+1,j_0}|^3 |\mathcal{F}_t \right) \leq \left\{ \mathrm{E}\left(\frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^4} Z_{t+1,j_0}^4 |\mathcal{F}_t \right) \right\}^{3/4} \leq \frac{c_0^3}{\bar{c}^{3/2}} \left(1 + \frac{c_0^2}{\bar{c}} \right)^{3/2}.$$

By condition, $\mathrm{E}|\varepsilon_{t+1}|^3 = O(1)$. Since ε_0 and $\boldsymbol{X}_0$ are independent, we have

$$\mathrm{E}\left(\frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^3} |Z_{t+1}|^3 |\varepsilon_{t+1}|^3 |\mathcal{F}_t \right) \leq \mathrm{E}|\varepsilon_{t+1}|^3 \mathrm{E}\left(\frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^3} |Z_{t+1}|^3 |\mathcal{F}_t \right) \leq \bar{c}_{**},$$

for some constant $\bar{c}_{**} > 0$. Therefore, for any $\delta_0 > 0$, it follows from Markov's inequality that

$$\begin{aligned} &\sum_{t=s_n}^{n-1} \mathrm{E}\left\{ \left| \frac{\varepsilon_{t+1} Z_{t+1,j_0}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \right|^2 \mathcal{I}\left(\left| \frac{\varepsilon_{t+1} Z_{t+1,j_0}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \right| \geq \delta_0 \right) |\mathcal{F}_t \right\} \\ &\leq \sum_{t=s_n}^{n-1} \frac{1}{n^{3/2}\delta_0} \mathrm{E}\left\{ \left| \frac{Z_{t+1,j_0}\varepsilon_{t+1}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \right|^3 |\mathcal{F}_t \right\} \leq \frac{\bar{c}_{**}}{\sqrt{n}\delta_0} \to 0. \end{aligned}$$

This verifies the Lindeberg's condition for I_1^{****}. It follows from the martingale central limit theorem that

$$I_1^{****} \xrightarrow{d} N(0, \sigma_0^2).$$

As a result, we have $I_1 \xrightarrow{d} N(0, \sigma_0^2)$. This completes the proof.

A.2 Proof of Theorem 3.1

We use a shorthand and write $\widehat{\mathcal{M}}_{j_0}^{(t)} = \widehat{\mathcal{M}}_{j_0}^{(-s_n)}$ for $t = 0, \ldots, s_n - 1$. Let

$$\widehat{\boldsymbol{\Sigma}}^* = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i b''(\boldsymbol{X}_i^T \boldsymbol{\beta}_0) \boldsymbol{X}_i^T \quad \text{and} \quad \widehat{\boldsymbol{\Psi}}^{(j)} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{X}_i b'''(\boldsymbol{X}_i^T \boldsymbol{\beta}_0) \boldsymbol{X}_i^T X_{i,j},$$

for any $j \in \{1, 2, \ldots, p\}$. For any $\mathcal{M} \subseteq \mathbb{I}$, define

$$\begin{aligned}
&\boldsymbol{\omega}_{\mathcal{M},j_0} = \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \boldsymbol{\Sigma}_{\mathcal{M},j_0}, \qquad \sigma_{\mathcal{M},j_0}^2 = \boldsymbol{\Sigma}_{j_0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T \boldsymbol{\Sigma}_{\mathcal{M},j_0}, \\
&\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1} \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}, \qquad \hat{\sigma}_{\mathcal{M},j_0}^2 = \widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^T \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}, \\
&\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^* = \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}^{*-1} \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^*, \qquad \hat{\sigma}_{\mathcal{M},j_0}^{*2} = \widehat{\boldsymbol{\Sigma}}_{j_0,j_0}^* - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}^{*T} \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^*, \\
&\widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^* + \sum_{j=1}^{p} \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}^{*-1} \left(\widehat{\boldsymbol{\Psi}}_{\mathcal{M},j_0}^{(j)} + \widehat{\boldsymbol{\Psi}}_{\mathcal{M},\mathcal{M}}^{(j)} \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^* \right) (\widetilde{\beta}_j - \beta_{0,j}), \\
&\widehat{Z}_{t+1,j_0}^* = X_{t+1,j_0} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^{*T} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}, \quad \widetilde{Z}_{t+1,j_0} = X_{t+1,j_0} - \widetilde{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^{T} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}, \\
&\hat{\xi}_{\mathcal{M},j_0}^{(j)} = \widehat{\boldsymbol{\Psi}}_{j_0,j_0}^{(j)} - \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^{*T} \left(2\widehat{\boldsymbol{\Psi}}_{\mathcal{M},j_0}^{(j)} + \widehat{\boldsymbol{\Psi}}_{\mathcal{M},\mathcal{M}}^{(j)} \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}^* \right), \\
&\tilde{\sigma}_{\mathcal{M},j_0}^2 = \hat{\sigma}_{\mathcal{M},j_0}^{*2} + \sum_{j=1}^{p} \hat{\xi}_{\mathcal{M},j_0}^{(j)} (\widetilde{\beta}_j - \beta_{0,j}).
\end{aligned}$$

Here, $\widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0}$ and $\tilde{\sigma}_{\mathcal{M},j_0}$ correspond to first-order approximations of $\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}$ and $\hat{\sigma}_{\mathcal{M},j_0}$ around $\boldsymbol{\beta}_0$. We introduce the following lemmas before proving Theorem 3.1. The proof of Lemma A.4 is given in Section A.8.

Lemma A.4. *Under conditions in Theorem 3.1, we have*

$$\min_{\substack{\mathcal{M} \subseteq [1,\ldots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \le \kappa_n}} \sigma_{\mathcal{M},j_0} \ge \sqrt{\bar{c}}, \qquad \max_{\substack{\mathcal{M} \subseteq [1,\ldots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \le \kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \le (\bar{c})^{-1/2} c_0, \tag{A.28}$$

where $\bar{c}$ and c_0 are defined in Condition (A2) and (A3*). Besides, the following events* *hold with probability tending to* 1,

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}-\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}+\eta_n\right), \tag{A.29}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}_{\mathcal{M},j_0}-\sigma_{\mathcal{M},j_0}| \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}+\eta_n\right), \tag{A.30}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}-\widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\eta_n^2, \qquad \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left|\hat{\sigma}^2_{\mathcal{M},j_0}-\tilde{\sigma}^2_{\mathcal{M},j_0}\right| \le \bar{c}_0\eta_n^2, \tag{A.31}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left|\hat{\sigma}^2_{\mathcal{M},j_0}-\hat{\sigma}^{*2}_{\mathcal{M},j_0}\right| \le \bar{c}_0\eta_n, \tag{A.32}$$

for some constant $\bar{c}_0>0$*. Moreover, we have*

$$\sum_{t=0}^{n-1}\frac{\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(-sn)},j_0}}\right)=\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}+o_p(1).$$

Similar to (A.25), we have

$$\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1=O(\sqrt{\kappa_n}\eta_n), \tag{A.33}$$

with probability tending to 1, under Condition (A5*).

For simplicity, we only consider the case where $l=1$. When $l>1$, assume we've shown the asymptotic normality of $\hat{\beta}^{(l-1)}_{j_0}$. Under the given conditions, we can show $\Gamma_n^{*,(l-2)}$ is lower bounded by $\sqrt{\bar{c}}/2$, with probability tending to 1. This implies $\hat{\beta}^{(l-1)}_{j_0}$ converges to β_{0,j_0} at a rate of $O_p(n^{-1/2})$. As a result, the estimator $\widehat{\boldsymbol{\beta}}^{(l-1)}=\widetilde{\boldsymbol{\beta}}+\boldsymbol{e}_{j_0,p}(\hat{\beta}^{(l-1)}_{j_0}-\widetilde{\beta}_{j_0})$ also satisfies the conditions in (A5*). The asymptotic normality of $\hat{\beta}^{(l)}_{j_0}$ can be similarly derived.

In the following, we omit the superscript and write $\hat{\beta}^{(1)}_{j_0}$ and $\Gamma_n^{*,(0)}$ as $\hat{\beta}_{j_0}$ and Γ_n^*. Let $\varepsilon_i=Y_i-\mu(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)$ for $i=0,1,\ldots,n$. By definition, we have

$$\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0}-\widetilde{\beta}_{j_0})=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\widehat{Z}_{t+1,j_0}\left\{Y_{t+1}-\mu\left(X_{t+1}\tilde{\beta}_{0,j_0}+\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right\} \tag{A.34}$$

By Condition (A1), we can show the following events occur with probability tending to 1,

$$\mathcal{M}_{j_0}\subseteq\widehat{\mathcal{M}}_{j_0}^{(t)},\quad |\widehat{\mathcal{M}}_{j_0}^{(t)}|\le\kappa_n,\quad t=0,\ldots,n-1. \tag{A.35}$$

Besides, similar to (A.6) and (A.12), we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}^{*2}_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right), \tag{A.36}$$

for some constant $\bar{c}_0 > 0$, and

$$\min_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \hat{\sigma}_{\mathcal{M},j_0} \ge \sqrt{\bar{c}}/2 \quad\text{and}\quad \min_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \hat{\sigma}^*_{\mathcal{M},j_0} \ge \sqrt{\bar{c}}/2, \tag{A.37}$$

with probability tending to 1.

Under the events defined in (A.35), we have for $t = 0, 1, \dots, n-1$,

$$\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0 = X_{t+1,j_0}\beta_{0,j_0} + \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}_{j_0}}.$$

Hence, using a second order Taylor expansion, we have

$$\begin{aligned}
&\mu(\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0) = \mu\left(X_{t+1,j_0}\beta_{0,j_0} + \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)\\
= \;& \mu\left(X_{t+1,j_0}\widetilde{\beta}_{j_0} + \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}_{j_0}}\right) + b''\left(X_{t+1,j_0}\widetilde{\beta}_{j_0} + \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)\\
\times \;& \left(X_{t+1,j_0}(\beta_{0,j_0} - \widetilde{\beta}_{j_0}) + \boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}(\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}_{j_0}} - \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}_{j_0}})\right)\\
+ \;& \frac{1}{2}b'''\left(\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}\widetilde{\boldsymbol{\beta}}^*_t\right)\left(\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}(\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}})\right)^2,
\end{aligned}$$

for some $\widetilde{\boldsymbol{\beta}}^*_t \in \mathbb{R}^{1+|\widehat{\mathcal{M}}^{(t)}_{j_0}|}$ lying on the line segment joining $\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}$ and $\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}$. Let R^*_t be the second order Remainder term. Under the events defined in (A.35), we have

$$\begin{aligned}
&\left|\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}\widetilde{\boldsymbol{\beta}}^*_t\right| \le \left|\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}\right| + \left|\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}(\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}} - \widetilde{\boldsymbol{\beta}}^*_t)\right|\\
= \;& |\boldsymbol{X}^T_{t+1}\boldsymbol{\beta}_0| + \left|\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}}(\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}} - \widetilde{\boldsymbol{\beta}}^*_t)\right| \le \bar{\omega} + \omega_0\left\|\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}^{(t)}_{j_0}} - \widetilde{\boldsymbol{\beta}}^*_t\right\|_1\\
\le \;& \bar{\omega} + \omega_0\left\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\right\|_1,
\end{aligned}$$

where the second inequality is due to Condition (A4*). By Condition (A5*) and (A.33), we have with probability tending to 1,

$$\omega_0 \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 \leq \omega_0 \eta_n^{(1)} \leq \bar{\omega}.$$

Since $b'''(\cdot)$ is continuous, $\sup_{|z|\leq 2\bar{\omega}} |b'''(z)|$ is upper bounded by some constant $c_* > 0$. Therefore, we have with probability tending to 1 that

$$\max_{t=0,\dots,n-1} \left| b''' \left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_t^* \right) \right| \leq c_*. \tag{A.38}$$

Under the event defined in (A.38), we have

$$|R_t^*| \leq \frac{c_*}{2} \left| \boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right|^2, \tag{A.39}$$

for any t. Note that

$$\left| \widehat{Z}_{t+1,j_0} \right| \leq |X_{t+1,j_0}| + \left\| \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \right\|_2 \|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2 + |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}|.$$

By Condition (A1*) and (A4*), we have almost surely,

$$|\widehat{Z}_{t+1,j_0}| \leq \omega_0 + \sqrt{\kappa_n}\omega_0 \max_{\substack{\mathcal{M}\subseteq\mathbb{I} \\ |\mathcal{M}|\leq\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 + |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}|. \tag{A.40}$$

The second term on the RHS of (A.40) is $o(1)$ with probability tending to 1, by (A.29) and Condition (A5*). Thus, we have with probability tending to 1 that

$$|\widehat{Z}_{t+1,j_0}| \leq 2\omega_0 + |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}|, \qquad \forall t. \tag{A.41}$$

Under the events defined in (A.35), (A.37), (A.39) and (A.41), we have

$$\begin{aligned}
&\left| \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \widehat{Z}_{t+1,j_0} \left\{ Y_{t+1} - \mu\left(X_{t+1}\tilde{\beta}_{0,j_0} + \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right\} \right. \\
- &\quad \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} - \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \\
\times &\quad \left. \left(X_{t+1,j_0}(\beta_{0,j_0} - \tilde{\beta}_{j_0}) + \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right) \right| \le \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} |R_t^*||\widehat{Z}_{t+1,j_0}| \\
\le &\quad \frac{c_*}{\sqrt{n\bar{c}}} \sum_{t=0}^{n-1} (2\omega_0 + |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}|) \left| \boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right|^2. \quad \text{(A.42)}
\end{aligned}$$

Similar to (A.9), we can show

$$\sum_{t=0}^{n-1} \left| \boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right|^2 = O(n\eta_n^2),$$

$$\sum_{t=0}^{n-1} |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}| \left| \boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right|^2 = O(n\eta_n^2),$$

with probability tending to 1. It follows that

$$\frac{c_*}{\sqrt{n\bar{c}}} \sum_{t=0}^{n-1} (2\omega_0 + |\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}|) \left| \boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}) \right|^2 = o_p(1), \quad \text{(A.43)}$$

under the condition $\sqrt{n}\eta_n^2 = o(1)$ in (A5*). Hence, we've shown

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \widehat{Z}_{t+1,j_0} \left\{ Y_{t+1} - \mu\left(X_{t+1}\tilde{\beta}_{0,j_0} + \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right\} \\
= &\quad \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \\
+ &\quad \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} + \sqrt{n}\Gamma_n^*(\beta_{0,j_0} - \tilde{\beta}_{j_0}) + o_p(1).
\end{aligned}$$

In view of (A.34), we have

$$\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0} - \beta_{0,j_0}) = o_p(1) + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}}_{I_1}$$

$$+ \quad \underbrace{\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)}_{I_2}.$$

In the following, we break the proof into two steps. In the first step, we prove $I_2 = o_p(1)$. In the second step, we show $I_1 \xrightarrow{d} N(0, \phi_0)$. This implies $\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0} - \beta_{0,j_0}) \xrightarrow{d} N(0, \phi_0)$. By Condition (A7*), $\hat{\phi}$ is consistent to ϕ_0. It follows from Slutsky's theorem that

$$\frac{\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0} - \beta_{0,j_0})}{\hat{\phi}^{1/2}} \xrightarrow{d} N(0, 1).$$

The proof is hence completed.

Step 1: Under the events defined in (A.35), using a first order Taylor expansion, we have

$$b''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) = b''(\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0)$$

$$+ \quad \underbrace{\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) b'''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_t^{**}\right)}_{R_t^{**}},$$

for some $\widetilde{\boldsymbol{\beta}}_t^{**} \in \mathbb{R}^{1+|\widehat{\mathcal{M}}_{j_0}^{(t)}|}$ lying on the line segment joining $\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}$ and $\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}$. Let

$$I_2^* = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b''\left(\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0\right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right),$$

we have

$$|I_2 - I_2^*| \le \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{|\widehat{Z}_{t+1,j_0}|}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} |R_t^{**}| \left|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right|_2.$$

Similar to (A.39), (A.42) and (A.43), we can show

$$|I_2 - I_2^*| \leq \frac{c_*}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{|\widehat{Z}_{t+1,j_0}|}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left| \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right|_2^2 = o(1),$$

with probability tending to 1. Thus, to prove $I_2 = o_p(1)$, it suffices to show $I_2^* = o_p(1)$.

Similar to the proof of Theorem 2.1, we can show under the given conditions that

$$\left| \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0} - Z_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b'' \left(\boldsymbol{X}_{t+1}^T \boldsymbol{\beta}_0 \right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right| = o_p(1),$$

where $Z_{t+1,j_0} = X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}$, and

$$\frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \left(\frac{Z_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} - \frac{Z_{t+1,j_0}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \right) b'' \left(\boldsymbol{X}_{t+1}^T \boldsymbol{\beta}_0 \right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) = o_p(1).$$

This implies $I_2^* = I_2^{**} + o_p(1)$, where

$$I_2^{**} = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{Z_{t+1,j_0}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b'' \left(\boldsymbol{X}_{t+1}^T \boldsymbol{\beta}_0 \right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right).$$

Note that I_2^{**} can be further bounded from above by $\max_j |I_{2,j}| \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1$ where

$$I_{2,j} = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{Z_{t+1,j_0}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b'' \left(\boldsymbol{X}_{t+1}^T \boldsymbol{\beta}_0 \right) \boldsymbol{X}_{t+1,j} \mathcal{I} \left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)} \right).$$

Similar to Lemma A.3, we can show $\max_j |I_{2,j}| = O_p(\sqrt{\log p})$. This together with (A.33) and Condition (A5*) implies $\max_j |I_{2,j}| \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1 = o_p(1)$ and hence $I_2^{**} = o_p(1)$. This proves $I_2 = o_p(1)$.

Step 2: By Taylor's theorem, we have for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$,

$$\frac{1}{\hat{\sigma}_{\mathcal{M},j_0}} - \frac{1}{\hat{\sigma}_{\mathcal{M},j_0}^*} + \frac{\hat{\sigma}_{\mathcal{M},j_0}^2 - \hat{\sigma}_{\mathcal{M},j_0}^{*2}}{\hat{\sigma}_{\mathcal{M},j_0}^{*3}} = \frac{(\hat{\sigma}_{\mathcal{M},j_0}^2 - \hat{\sigma}_{\mathcal{M},j_0}^{*2})^2}{2\{\rho_{\mathcal{M}} \hat{\sigma}_{\mathcal{M},j_0} + (1-\rho_{\mathcal{M}}) \hat{\sigma}_{\mathcal{M},j_0}^*\}^5},$$

for some $0 < \rho_{\mathcal{M}} < 1$. By (A.32) and (A.37), the second-order remainder term satisfies

$$\max_{\mathcal{M} \subseteq \mathbb{I}_{j_0}, |\mathcal{M}| \leq \kappa_n} \left| \frac{(\hat{\sigma}_{\mathcal{M},j_0}^2 - \hat{\sigma}_{\mathcal{M},j_0}^{*2})^2}{2\{\rho_{\mathcal{M}} \hat{\sigma}_{\mathcal{M},j_0} + (1-\rho_{\mathcal{M}}) \hat{\sigma}_{\mathcal{M},j_0}^*\}^5} \right| \leq \frac{16 \bar{c}_0^2 \eta_n^2}{\bar{c}^{5/2}}, \tag{A.44}$$

with probability tending to 1.

Besides, it follows from (A.31) and (A.37) that

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left|\frac{\hat{\sigma}^2_{\mathcal{M},j_0}-\hat{\sigma}^{*2}_{\mathcal{M},j_0}-\sum_j\hat{\xi}^{(j)}_{\mathcal{M},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\mathcal{M},j_0}}\right|=\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left|\frac{\hat{\sigma}^2_{\mathcal{M},j_0}-\tilde{\sigma}^2_{\mathcal{M},j_0}}{\hat{\sigma}^{*3}_{\mathcal{M},j_0}}\right|\le\frac{8\bar{c}_0\eta_n^2}{\bar{c}^{3/2}},$$

with probability tending to 1. Combining this together with (A.44) yields

$$\Pr\left(\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left|\frac{1}{\hat{\sigma}_{\mathcal{M},j_0}}-\frac{1}{\hat{\sigma}^*_{\mathcal{M},j_0}}+\frac{\sum_j\hat{\xi}^{(j)}_{\mathcal{M},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\mathcal{M},j_0}}\right|\le\bar{c}_1\eta_n^2\right)\to 1,$$

for some constant $\bar{c}_1>0$. By Condition (A1*), we have

$$\frac{1}{\sqrt{n}}\left|\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}-\sum_{t=0}^{n-1}\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}-\frac{\sum_j\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}\right)\right|$$
$$\le\sqrt{n}\bar{c}_1\eta_n^2\max_t\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}||\widehat{Z}_{t+1,j_0}|,$$

with probability tending to 1. Similar to (A.23), we can show $\sum_{t=0}^{n-1}|\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0}}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}|^2=O_p(n)$. This together with (A.41) and Cauchy-Schwarz inequality yields

$$\sum_{t=0}^{n-1}|\widehat{Z}_{t+1,j_0}|^2\le 8n\omega_0^2+2\sum_{t=0}^{n-1}|\boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0}}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}|^2=O_p(n).\tag{A.45}$$

In addition, we have $\sum_{t=0}^{n-1}\varepsilon^2_{t+1}=O_p(n)$, under (A6*). It follows from Cauchy-Schwarz inequality that

$$\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}||\widehat{Z}_{t+1,j_0}|\le\left(\frac{1}{n}\sum_{t=0}^{n-1}|\widehat{Z}_{t+1,j_0}|^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=0}^{n-1}\varepsilon^2_{t+1}\right)^{1/2}=O_p(1),$$

and hence

$$\frac{1}{\sqrt{n}}\left|\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}-\sum_{t=0}^{n-1}\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}-\frac{\sum_j\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}\right)\right|=O_p(\sqrt{n}\eta_n^2)$$
$$=o_p(1),$$

under Condition (A5*).

Using similar arguments in bounding $I_2^{(2)}$ in the proof of Theorem 2.1, we can show

$$\frac{1}{\sqrt{n}}\left|\sum_{t=0}^{n-1}(\widehat{Z}_{t+1,j_0}-\widetilde{Z}_{t+1,j_0})\varepsilon_{t+1}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\right)\right|=o_p(1).$$

It follows that

$$\left|\sum_{t=0}^{n-1}\left\{\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\right)\right\}\right|=o_p(\sqrt{n}).$$

Therefore, we've shown $I_1=I_1^*+o_p(1)$ where

$$I_1^*=\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}\right).$$

In Lemma A.4, we further show I_1^* is equivalent to

$$I_1^{**}\equiv\sqrt{n}\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}.$$

Hence, we have $I_1=I_1^{**}+o_p(1)$. Unlike $\widetilde{Z}_{t+1,j_0}$ and $\tilde{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$, $\widehat{Z}^*_{t+1,j_0}$ and $\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ didn't depend on the initial estimator $\widetilde{\boldsymbol{\beta}}$. As a result, $\widehat{Z}^*_{t+1,j_0}$ and $\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ are fixed given $\{\boldsymbol{X}_1,\dots,\boldsymbol{X}_n\}$ and $\widehat{\mathcal{M}}_{j_0}^{(t)}$. Following the arguments in the proof of Theorem 2.1, we can show

$$I_1^{**}=\frac{1}{\sqrt{n}}\sum_{t=s_n}^{n-1}\frac{Z_{t+1,j_0}\varepsilon_{t+1}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}+o_p(1)\ \text{ and }\ \frac{1}{\sqrt{n}}\sum_{t=s_n}^{n-1}\frac{Z_{t+1,j_0}\varepsilon_{t+1}}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\xrightarrow{d}N(0,\phi_0).$$

By Slutsky's theorem, we have $I_1\xrightarrow{d}N(0,\phi_0)$. The proof is hence completed.

A.3 Proof of Theorem 3.2

Recall that $\mathbb{I}=[1,\dots,p]$ and $\mathbb{I}_{j_0}=\mathbb{I}-\{j_0\}$. By (19) and Lemma A.5, we have

$$\sqrt{n}\mathrm{L}(\hat{\beta}^{DL}_{j_0},\alpha)=2z_{\frac{\alpha}{2}}\sqrt{\phi_0\boldsymbol{e}^T_{j_0,p}\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_{j_0,p}}+o_p(1)=\frac{2z_{\frac{\alpha}{2}}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}}+o_p(1). \tag{A.46}$$

It follows from (16) that

$$\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0},\alpha) = \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} + o_p(1). \tag{A.47}$$

With some calculations, we have

$$\begin{aligned} &\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \\ = \;& 2z_{\alpha/2}\sqrt{\phi_0}\frac{s_n\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n}{\sigma_{\mathbb{I}_{j_0},j_0}\{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n\}}. \end{aligned} \tag{A.48}$$

For any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$, we have

$$\begin{aligned} \sigma^2_{\mathcal{M},j_0} &= \mathrm{E}|X_{0,j_0} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\boldsymbol{X}_{0,\mathcal{M}}|^2 b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) = \arg\min_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}}\mathrm{E}|X_{0,j_0} - \boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2 b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) \\ &\geq \arg\min_{\boldsymbol{a}\in\mathbb{R}^{p-1}}\mathrm{E}|X_{0,j_0} - \boldsymbol{a}^T\boldsymbol{X}_{0,\mathbb{I}_{j_0}}|^2 b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) = \sigma^2_{\mathbb{I}_{j_0},j_0}. \end{aligned}$$

This shows $\sigma_{\mathcal{M},j_0} \geq \sigma_{\mathbb{I}_{j_0},j_0}$ for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$. Hence, the numerator of the RHS of (A.48) is nonnegative.

On the other hand, by Condition (A4*), we have $|\boldsymbol{X}_0^T\boldsymbol{\beta}_0| \leq \bar{\omega}$ and hence $b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) \leq \bar{k}$. Therefore,

$$\sigma^2_{\mathcal{M},j_0} = \arg\min_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}}\mathrm{E}|X_{0,j_0} - \boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2 b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) \leq \bar{k}\mathrm{E}|X^2_{0,j_0}| \leq \bar{k}\|X_{0,j_0}\|^2_{\psi_2} = \bar{k}c_0^2, \tag{A.49}$$

where the last inequality is due to Condition (A3*). This implies

$$\begin{aligned} &\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \qquad\qquad (\text{A.50}) \\ \geq\;& \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\bar{k}c_0^2}\left(s_n\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n\right). \end{aligned}$$

Besides, it follows from (A.49) that

$$\sigma_{\mathcal{M},j_0} - \sigma_{\mathbb{I}_{j_0},j_0} = \frac{\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}}{\sigma_{\mathcal{M},j_0} + \sigma_{\mathbb{I}_{j_0},j_0}} \geq \frac{\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}}{2\sqrt{\bar{k}}c_0},$$

for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$. This together with (A.50) gives

$$\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \tag{A.51}$$
$$\geq \frac{z_{\alpha/2}\sqrt{\phi_0}}{\bar{k}^{3/2}c_0^3}\left(s_n\{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}\}/n\right).$$

For any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$, define

$$\boldsymbol{\Omega}_{\mathcal{M},j_0} = \left(\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c} - \boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\right)^{-1}.$$

It follows from Lemma A.5 that

$$\begin{pmatrix} \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}} & \boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c} \\ \boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}} & \boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c}\end{pmatrix}^{-1} - \begin{pmatrix}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O}\end{pmatrix}$$
$$= \begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\boldsymbol{\Omega}_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} & -\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\boldsymbol{\Omega}_{\mathcal{M},j_0} \\ -\boldsymbol{\Omega}_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} & \boldsymbol{\Omega}_{\mathcal{M},j_0}\end{pmatrix}.$$

Therefore,

$$\begin{aligned}
&\boldsymbol{\Sigma}^T_{\mathbb{I}_{j_0},j_0}\boldsymbol{\Sigma}^{-1}_{\mathbb{I}_{j_0},\mathbb{I}_{j_0}}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0},j_0} - \boldsymbol{\Sigma}^T_{\mathcal{M},j_0}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},j_0} = (\boldsymbol{\Sigma}_{j_0,\mathcal{M}}, \boldsymbol{\Sigma}_{j_0,\mathbb{I}_{j_0}\cap\mathcal{M}^c}) \\
\times & \begin{pmatrix} \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\boldsymbol{\Omega}_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} & -\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\boldsymbol{\Omega}_{\mathcal{M},j_0} \\ -\boldsymbol{\Omega}_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathcal{M}}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} & \boldsymbol{\Omega}_{\mathcal{M},j_0}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Sigma}_{\mathcal{M},j_0} \\ \boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,j_0}\end{pmatrix} \\
= & (\boldsymbol{\Sigma}_{j_0,\mathbb{I}_{j_0}\cap\mathcal{M}^c} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c})\boldsymbol{\Omega}_{\mathcal{M},j_0}(\boldsymbol{\Sigma}_{j_0,\mathbb{I}_{j_0}\cap\mathcal{M}^c} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c})^T \\
\geq & \lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M},j_0})\|\boldsymbol{\Sigma}_{j_0,\mathbb{I}_{j_0}\cap\mathcal{M}^c} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\boldsymbol{\Sigma}_{\mathcal{M},\mathbb{I}_{j_0}\cap\mathcal{M}^c}\|_2^2 = \lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M},j_0})\|\boldsymbol{\xi}_{\mathcal{M},j_0}\|_2^2.
\end{aligned}$$

By definition, we have

$$\lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M},j_0}) \geq \lambda_{\min}\left\{\left(\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c}\right)^{-1}\right\} = \left\{\lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c}\right)\right\}^{-1} \geq \{\lambda_{\max}(\boldsymbol{\Sigma})\}^{-1} \geq \frac{1}{c_0^2},$$

where the last inequality follows from (A.22). It follows that

$$\boldsymbol{\Sigma}^T_{\mathbb{I}_{j_0},j_0}\boldsymbol{\Sigma}^{-1}_{\mathbb{I}_{j_0},\mathbb{I}_{j_0}}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0},j_0} - \boldsymbol{\Sigma}^T_{\mathcal{M},j_0}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},j_0} \geq \frac{1}{c_0^2}\|\boldsymbol{\xi}_{\mathcal{M},j_0}\|_2^2.$$

Note that we have

$$\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0} = \boldsymbol{\Sigma}_{j_0,j_0} - \boldsymbol{\Sigma}^c_{\mathcal{M},j_0}\boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\boldsymbol{\Sigma}_{\mathcal{M},j_0} - (\boldsymbol{\Sigma}_{j_0,j_0} - \boldsymbol{\Sigma}^c_{\mathbb{I}_{j_0},j_0}\boldsymbol{\Sigma}^{-1}_{\mathbb{I}_{j_0},\mathbb{I}_{j_0}}\boldsymbol{\Sigma}_{\mathbb{I}_{j_0},j_0}).$$

This further implies

$$\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0} \geq \frac{1}{c_0^2}\|\boldsymbol{\xi}_{\mathcal{M},j_0}\|_2^2,$$

for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$. By (A.51), we have

$$\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \geq \frac{\sqrt{\phi_0}z_{\alpha/2}}{\bar{k}^{3/2}c_0^5}\left(\frac{s_n}{n}\|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\|_2^2 + \frac{1}{n}\sum_{t=s_n}^{n-1}\|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2\right).$$

In view of (A.46) and (A.47), we've shown

$$\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{DL},\alpha) \geq \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0},\alpha) + \frac{\sqrt{\phi_0}z_{\alpha/2}}{\bar{k}^{3/2}c_0^5}\left(\frac{s_n}{n}\|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\|_2^2 + \frac{1}{n}\sum_{t=s_n}^{n-1}\|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2\right) + o_p(1).$$

The proof is completed by noting that $\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{DL},\alpha) = \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{DS},\alpha) + o_p(1)$.

A.4 Proof of Theorem 3.3

Under the given conditions, using similar arguments in (A.10), we can show the following event occurs with probability tending to 1,

$$\widehat{\mathcal{M}}_{j_0}^{(-s_n)} = \widehat{\mathcal{M}}_{j_0}^{(s_n)} = \cdots = \widehat{\mathcal{M}}_{j_0}^{(n)} = \mathcal{M}_{j_0}.$$

Under these events, we have

$$\frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} = \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathcal{M}_{j_0},j_0}}. \tag{A.52}$$

By (16) and (23), for any sufficiently small $\varepsilon_0 > 0$, the following events occur with probability tending to 1,

$$\limsup_n \left|\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{(l)},\alpha) - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n}\right| \leq \frac{\varepsilon_0}{2}, \tag{A.53}$$

$$\limsup_n \left|\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{oracle},\alpha) - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathcal{M}_{j_0},j_0}}\right| \leq \frac{\varepsilon_0}{2}. \tag{A.54}$$

Conditional on the events defined in (A.52)-(A.54), we have

$$\limsup_n \left| \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{(l)}, \alpha) - \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{oracle}, \alpha) \right| \le \varepsilon_0.$$

The proof is hence completed.

A.5 Proof of Lemma A.1

We first prove (A.1). Condition (A2) states that

$$\min_{\substack{\mathcal{M} \subseteq [1,\dots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \le \kappa_n}} \inf_{\substack{\boldsymbol{a} \in \mathbb{R}^{|\mathcal{M}|+1} \\ \|\boldsymbol{a}\|_2 \ge 1}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{j_0 \cup \mathcal{M}, j_0 \cup \mathcal{M}} \boldsymbol{a} \ge \bar{c}. \tag{A.55}$$

Note that

$$\begin{aligned} \sigma_{\mathcal{M},j_0}^2 = \boldsymbol{\Sigma}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,\mathcal{M}} \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \boldsymbol{\Sigma}_{\mathcal{M},j_0} = (1, -\boldsymbol{\omega}_{\mathcal{M},j_0}) \begin{pmatrix} \boldsymbol{\Sigma}_{j_0,j_0} & \boldsymbol{\Sigma}_{j_0,\mathcal{M}} \\ \boldsymbol{\Sigma}_{\mathcal{M},j_0} & \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}} \end{pmatrix} \begin{pmatrix} 1 \\ -\boldsymbol{\omega}_{\mathcal{M},j_0}^T \end{pmatrix} \\ \ge \inf_{\substack{\boldsymbol{a} \in \mathbb{R}^{|\mathcal{M}|+1} \\ \|\boldsymbol{a}\|_2 \ge 1}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{j_0 \cup \mathcal{M}, j_0 \cup \mathcal{M}} \boldsymbol{a}. \end{aligned}$$

By (A.55), this implies

$$\min_{\substack{\mathcal{M} \subseteq [1,\dots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \le \kappa_n}} \sigma_{\mathcal{M},j_0}^2 \ge \bar{c}, \tag{A.56}$$

and hence

$$\min_{\substack{\mathcal{M} \subseteq [1,\dots,p] \\ j_0 \notin \mathcal{M}, |\mathcal{M}| \le \kappa_n}} \sigma_{\mathcal{M},j_0} \ge \sqrt{\bar{c}}.$$

It follows from (A.21) and Assumption (A3) implies that

$$\boldsymbol{\Sigma}_{j_0,j_0} = \mathrm{E} X_{0,j_0}^2 \le \|X_{0,j_0}\|_{\psi_2}^2 \le c_0^2.$$

In view of (A.56), this further implies that

$$\boldsymbol{\Sigma}_{j_0,\mathcal{M}} \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \boldsymbol{\Sigma}_{\mathcal{M},j_0} = \boldsymbol{\Sigma}_{j_0,j_0} - \sigma_{\mathcal{M},j_0}^2 \le c_0^2.$$

Note that $\mathbf{\Sigma}_{j_0,\mathcal{M}}\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}\mathbf{\Sigma}_{\mathcal{M},j_0} = \boldsymbol{\omega}_{\mathcal{M},j_0}^T\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}\boldsymbol{\omega}_{\mathcal{M},j_0}$. Hence, we have

$$\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2^2 \leq \frac{c_0^2}{\lambda_{\min}(\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})} \leq \frac{c_0^2}{\bar{c}}, \tag{A.57}$$

where the last inequality is due to Condition (A2). Therefore, (A.1) is proven.

Similar to (A.21), we can show for any random variable Z,

$$\mathrm{E}Z^4 \leq 2\|Z\|_{\psi_2}^4. \tag{A.58}$$

It follows from Condition (A3) that

$$\begin{aligned}\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \mathrm{E}|X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4 &\leq 2 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}\|_{\psi_2}^4 \\ &\leq 2c_0^4 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|(1,\boldsymbol{\omega}_{\mathcal{M},j_0}^T)^T\|_2^4 \leq 2c_0^4(1+\bar{c}^{-1}c_0^2)^2.\end{aligned} \tag{A.59}$$

Moreover, by (A.1),

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \frac{1}{\sigma_{\mathcal{M},j_0}} \leq \frac{1}{\sqrt{\bar{c}}}.$$

Thus, we have

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \frac{1}{\sigma_{\mathcal{M},j_0}^4}\mathrm{E}|X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4 \leq \frac{2c_0^4}{\bar{c}^2}\left(1+\frac{c_0^2}{\bar{c}}\right)^2.$$

For any random variable Z with $\|Z\|_{\psi_2} \leq \omega$, it follows from the definition of the Orlicz norm that $\|Z\|_{\psi_1} \leq \omega^2$. Under Condition (A3), this implies

$$\max_j \|X_{0,j}^2\|_{\psi_1} \leq \max_j (\|X_{0,j}\|_{\psi_2})^2 \leq c_0^2. \tag{A.60}$$

For any random variable Z, we have $\mathrm{E}|Z| \leq \|Z\|_{\psi_1}$. This together with (A.60) yields that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ |\mathcal{M}|\leq\kappa_n}} \mathrm{E}\|X_{0,\mathcal{M}}\|_2^2 = \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ |\mathcal{M}|\leq\kappa_n}} \sum_{j\in\mathcal{M}} \mathrm{E}X_{0,j}^2 \leq \kappa_n \max_j \mathrm{E}X_{0,j}^2 \leq \kappa_n \max_j \|X_{0,j}^2\|_{\psi_1} \leq \kappa_n c_0^2. \tag{A.61}$$

Finally, notice that

$$\Pr\left(|X_{i,j}| > \sqrt{3c_0^2 \max(\log p, \log n)}\right) \le \exp\{-3\max(\log p, \log n)\} \mathrm{E}\exp(|X_{i,j}|^2/c_0^2)$$
$$\le 2\exp\{-3\max(\log p, \log n)\} \le 2\min(p^{-3}, n^{-3}),$$

where the first inequality follows from Markov's inequality and the second inequality follows from the definition of the Orlicz norm. Now it follows from Bonferroni's inequality that

$$\Pr\left(\max_{i,j} |X_{i,j}| > \sqrt{3c_0^2 \max(\log p, \log n)}\right) \le 2pn\min(p^{-3}, n^{-3}) = 2\min(np^{-2}, pn^{-2}) \to 0.$$

The proof is hence completed.

A.6 Proof of Lemma A.2

We first prove (A.5). Note that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 = \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}} \boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2$$
$$\le \underbrace{\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} (\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0})\|_2}_{\eta_1} + \underbrace{\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|(\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}) \boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2}_{\eta_2}.$$

Hence, it suffices to show that with probability tending to 1,

$$\eta_1 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) \quad \text{and} \quad \eta_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right).$$

Upper bound for η_1: Since

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} (\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0})\|_2$$
$$\le \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\|_2 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0})\|_2,$$

it suffices to show with probability tending to 1 that,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\|_2 = O(1), \tag{A.62}$$

and

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0} - \mathbf{\Sigma}_{\mathcal{M},j_0})\|_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right). \tag{A.63}$$

Note that $\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}$ is symmetric. To prove (A.62), it is equivalent to show that the eigenvalues of $\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}$ are uniformly bounded with probability tending to 1. Hence, it suffices to prove

$$\min_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \lambda_{\min}\left(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}\right) > \frac{\bar{c}}{2}, \tag{A.64}$$

with probability tending to 1.

Observe that

$$\begin{aligned} \min_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \lambda_{\min}\left(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}\right) &= \inf_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \inf_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \boldsymbol{a}^T\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}\boldsymbol{a} \\ \ge \min_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \min_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \boldsymbol{a}^T\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}\boldsymbol{a} &- \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \max_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \left|\boldsymbol{a}^T\left(\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} - \widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{a}\right|. \end{aligned}$$

By Condition (A2), the first term on the second line is greater than or equal to $\bar{c}$. Since $\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} - \widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}$ is symmetric, the second term can be bounded by

$$\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^p\\ \|\boldsymbol{a}\|_2=1,\|\boldsymbol{a}\|_0\le\kappa_n}} |\boldsymbol{a}^T(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})\boldsymbol{a}|. \tag{A.65}$$

Define the stochastic process

$$\mathbb{X}(\boldsymbol{a}) = \boldsymbol{a}^T(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\boldsymbol{a}.$$

For any $\boldsymbol{a}_1, \boldsymbol{a}_2 \in \mathbb{R}^p$ with $\|\boldsymbol{a}_1\|_2, \|\boldsymbol{a}_2\|_2 \le 1$, we have

$$|\mathbb{X}(\boldsymbol{a}_1) - \mathbb{X}(\boldsymbol{a}_2)| \le |(\boldsymbol{a}_1 - \boldsymbol{a}_2)^T(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})(\boldsymbol{a}_1 + \boldsymbol{a}_2)|,$$

since $\boldsymbol{a}_2^T(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\boldsymbol{a}_1 = \boldsymbol{a}_1^T(\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma})\boldsymbol{a}_2$, by the symmetricity of the matrix $\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}$. Recall that $\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma} = \sum_{i=1}^n (\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathrm{E}\boldsymbol{X}_0\boldsymbol{X}_0^T)/n$. It follows from Condition (A3) and Cauchy-Schwarz inequality that

$$
\begin{aligned}
\|(\boldsymbol{a}_1-\boldsymbol{a}_2)^T\boldsymbol{X}_0\boldsymbol{X}_0^T(\boldsymbol{a}_1+\boldsymbol{a}_2)\|_{\psi_1} &\le \sqrt{2}\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2 \sup_{\substack{\boldsymbol{a}_3,\boldsymbol{a}_4\in\mathbb{R}^p\\ \|\boldsymbol{a}_3\|_2,\|\boldsymbol{a}_4\|_2\le 1}} \|\boldsymbol{a}_3^T\boldsymbol{X}_0\boldsymbol{X}_0^T\boldsymbol{a}_4\|_{\psi_1}\\
&\le \sqrt{2}\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2 \sup_{\substack{\boldsymbol{a}_3,\boldsymbol{a}_4\in\mathbb{R}^p\\ \|\boldsymbol{a}_3\|_2,\|\boldsymbol{a}_4\|_2\le 1}} \frac{\|(\boldsymbol{a}_3^T\boldsymbol{X}_0)^2+(\boldsymbol{a}_4^T\boldsymbol{X}_0)^2\|_{\psi_1}}{2}\\
&\le \sqrt{2}\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2 \sup_{\substack{\boldsymbol{a}_3,\boldsymbol{a}_4\in\mathbb{R}^p\\ \|\boldsymbol{a}_3\|_2,\|\boldsymbol{a}_4\|_2\le 1}} \frac{\|(\boldsymbol{a}_3^T\boldsymbol{X}_0)^2\|_{\psi_1}+\|(\boldsymbol{a}_4^T\boldsymbol{X}_0)^2\|_{\psi_1}}{2}\\
&\le \sqrt{2}\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2 \sup_{\substack{\boldsymbol{a}_3,\boldsymbol{a}_4\in\mathbb{R}^p\\ \|\boldsymbol{a}_3\|_2,\|\boldsymbol{a}_4\|_2\le 1}} \frac{\|\boldsymbol{a}_3^T\boldsymbol{X}_0\|_{\psi_2}^2+\|\boldsymbol{a}_4^T\boldsymbol{X}_0\|_{\psi_2}^2}{2} \le \sqrt{2}c_0^2\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2. \qquad (\text{A.66})
\end{aligned}
$$

By Jensen's inequality, we have

$$
\begin{aligned}
\|(\boldsymbol{a}_1-\boldsymbol{a}_2)^T(\boldsymbol{X}_0\boldsymbol{X}_0^T-\mathrm{E}\boldsymbol{X}_0\boldsymbol{X}_0^T)(\boldsymbol{a}_1+\boldsymbol{a}_2)\|_{\psi_1} &\le 2\|(\boldsymbol{a}_1-\boldsymbol{a}_2)^T\boldsymbol{X}_0\boldsymbol{X}_0^T(\boldsymbol{a}_1+\boldsymbol{a}_2)\|_{\psi_1}\\
&\le 2\sqrt{2}c_0^2\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2.
\end{aligned}
$$

It follows from Bernstein's inequality (Theorem 3.1, Klartag and Mendelson, 2005) that

$$
\Pr(|\mathbb{X}(\boldsymbol{a}_1)-\mathbb{X}(\boldsymbol{a}_2)|>t)\le 2\exp\left\{-O(1)\min\left(\frac{nt^2}{\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2^2},\frac{nt}{\|\boldsymbol{a}_1-\boldsymbol{a}_2\|_2}\right)\right\},
$$

for some positive constant $O(1)$ that is independent of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$. Let $\mathbb{S}=\{\boldsymbol{a}\in\mathbb{R}^p : \|\boldsymbol{a}\|_2\le 1, \|\boldsymbol{a}\|_0\le\kappa_n\}$. It follows from Theorem 1.2.7 of Talagrand (2005) that

$$
\mathrm{E}\sup_{\boldsymbol{a}_1,\boldsymbol{a}_2\in\mathbb{S}}|\mathbb{X}(\boldsymbol{a}_1)-\mathbb{X}(\boldsymbol{a}_2)| = O\{\gamma_2(\mathbb{S},n^{-1/2}\|\cdot\|_2)+\gamma_1(\mathbb{S},n^{-1}\|\cdot\|_2)\},
$$

where the definitions of the γ_p-functionals are given in Definition 1.2.5 of Talagrand (2005). Since $\mathbb{X}(\mathbf{0}_p)=0$, we have

$$
\mathrm{E}\sup_{\boldsymbol{a}\in\mathbb{S}}|\mathbb{X}(\boldsymbol{a})| = O\{\gamma_2(\mathbb{S},n^{-1/2}\|\cdot\|_2)+\gamma_1(\mathbb{S},n^{-1}\|\cdot\|_2)\}. \qquad (\text{A.67})
$$

By Lemma 2.3 of Mendelson et al. (2008), for any $0\le\varepsilon\le 1/2$, there exists an ε-cover of $\mathbb{S}$ with cardinality at most $(5/2\varepsilon)^{\kappa_n}\binom{p}{\kappa_n}$. Using similar arguments in proving Lemma G.8

of Shi et al. (2018), we can show that

$$\gamma_2(\mathbb{S}, n^{-1/2}\|\cdot\|_2) \le n^{-1/2}\gamma_2(\mathbb{S}, \|\cdot\|_2) = O(n^{-1/2}\sqrt{\kappa_n \log p}),$$
$$\gamma_1(\mathbb{S}, n^{-1}\|\cdot\|_2) \le n^{-1}\gamma_1(\mathbb{S}, \|\cdot\|_2) = O(n^{-1}\kappa_n \log p).$$

Under the given conditions, it follows from (A.67) that $\mathrm{E}\sup_{\boldsymbol{a}\in\mathbb{S}} |\mathbb{X}(\boldsymbol{a})| = O(n^{-1/2}\sqrt{\kappa_n \log p})$. By Markov's inequality, we obtain $\sup_{\boldsymbol{a}\in\mathbb{S}} |\mathbb{X}(\boldsymbol{a})| = O_p(n^{-1/2}\sqrt{\kappa_n \log p}) = o_p(1)$ and hence (A.65) is $o_p(1)$. Under Condition (A2), we have

$$\min_{\substack{\mathcal{M}\subseteq[1,\dots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \min_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}} \boldsymbol{a} \ge \bar{c}.$$

Assertion (A.62) thus follows.

Recall that $\widehat{\boldsymbol{\Sigma}}_{j_1,j_2} - \boldsymbol{\Sigma}_{j_1,j_2} = \sum_i (X_{i,j_1}X_{i,j_2} - \mathrm{E}X_{0,j_1}X_{0,j_2})/n$. Combining (A.60) with Cauchy-Schwarz inequality, we have

$$\|X_{0,j_1}X_{0,j_2}\|_{\psi_1} \le \frac{\|X_{0,j_1}^2 + X_{0,j_2}^2\|_{\psi_1}}{2} \le \frac{\|X_{0,j_1}^2\|_{\psi_1}}{2} + \frac{\|X_{0,j_2}^2\|_{\psi_1}}{2} \le c_0^2, \tag{A.68}$$

for all $j_1, j_2 \in [1,\dots,p]$. By Jensen's inequality, we have

$$\mathrm{E}\exp\left(\frac{\mathrm{E}|X_{0,j_1}X_{0,j_2|}}{\omega_0^2}\right) \le \mathrm{E}\exp\left(\frac{|X_{0,j_1}X_{0,j_2|}}{\omega_0^2}\right) \le 2.$$

This implies $\|\mathrm{E}X_{0,j_0}X_{0,j_2}\|_{\psi_1} \le c_0^2$, $\forall j_1, j_2$. Combining this together with (A.68) gives

$$\|X_{0,j_1}X_{0,j_2} - \mathrm{E}X_{0,j_0}X_{0,j_2}\|_{\psi_1} \le \|X_{0,j_1}X_{0,j_2}\|_{\psi_1} + \|\mathrm{E}X_{0,j_1}X_{0,j_2}\|_{\psi_1} \le 2c_0^2.$$

Therefore, it follows from Bernstein's inequality that

$$\max_{1\le j_1,j_2\le p} \Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2} - \boldsymbol{\Sigma}_{j_1,j_2})\right| \ge t\right) \le 2\exp\left(-O(1)\min\left(\frac{t^2}{4nc_0^2}, \frac{t}{2c_0}\right)\right), \tag{A.69}$$

for any $t > 0$, where $O(1)$ denotes some positive constant.

Take $t_0 = 3\sqrt{n\log p}c_0/\sqrt{c_1}$. Since $\log p = o(n)$, we have for sufficiently large n,

$$\frac{t_0^2}{4nc_0^2} = \frac{9\log p}{4c_1} \ll \frac{3\sqrt{n\log p}}{2\sqrt{c_1}} = \frac{t_0}{2c_0}.$$

It follows from (A.69) that

$$\max_{j_1,j_2} \Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2} - \mathbf{\Sigma}_{j_1,j_2})\right| \geq t_0\right) \leq 2\exp\left(-\frac{c_1 t_0^2}{4nc_0^2}\right) \leq 2\exp\left(-\frac{9\log p}{4}\right).$$

By Bonferroni's inequality, we have

$$\begin{aligned} & \Pr\left(\max_{j_1,j_2\in[1,\ldots,p]} \left|\sum_i (X_{i,j_1}X_{i,j_2} - \mathbf{\Sigma}_{j_1,j_2})\right| \geq t_0\right) \qquad \text{(A.70)} \\ \leq & \sum_{j_1,j_2\in[1,\ldots,p]} \Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2} - \mathbf{\Sigma}_{j_1,j_2})\right| \geq t_0\right) \\ \leq & \; p^2 2\exp\left(-\frac{9\log p}{4}\right) = 2\exp\left(-\frac{9\log p}{4} + 2\log p\right) = 2\exp\left(-\frac{\log p}{4}\right) \to 0. \end{aligned}$$

Under the event defined in (A.70), we have

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0} - \mathbf{\Sigma}_{\mathcal{M},j_0})\|_2 \leq \sqrt{\kappa_n} \max_{j_1,j_2\in[1,\ldots,p]} \left|\widehat{\mathbf{\Sigma}}_{j_1,j_2} - \mathbf{\Sigma}_{j_1,j_2}\right| \leq \frac{\sqrt{\kappa_n}t_0}{n}.$$

This proves (A.63).

Upper bound for η_2: Observe that

$$\begin{aligned} & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|(\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}})\mathbf{\Sigma}_{\mathcal{M},j_0}\|_2 \\ = & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\mathbf{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}}\mathbf{\Sigma}_{\mathcal{M},j_0}\|_2 \\ \leq & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\|_2 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2. \end{aligned}$$

By (A.62), it suffices to show

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}} \|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right), \qquad \text{(A.71)}$$

with probability tending to 1.

LHS of (A.71) can be upper bounded by

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |\boldsymbol{a}^T(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}|.$$

For any subset $\mathcal{M}$ such that $j_0\notin\mathcal{M}$, $|\mathcal{M}|\le\kappa_n$, define the stochastic process

$$T_{\mathcal{M}}(\boldsymbol{a})=\frac{1}{n}\sum_{i=1}^{n} g_{\mathcal{M}}(\boldsymbol{X}_i,\boldsymbol{a})=\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{a}^T\left(\boldsymbol{X}_{i,\mathcal{M}}\boldsymbol{X}_{i,\mathcal{M}}^T-\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}.$$

Using similar arguments in bounding (A.65), we can show

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})|=O(n^{-1/2}\sqrt{\kappa_n}). \tag{A.72}$$

The envelope function of |g| is bounded by

$$G_{\mathcal{M}}(\boldsymbol{X}_i)\triangleq\|\boldsymbol{X}_{i,\mathcal{M}}\|_2|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|_2+\|\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\|_2\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|.$$

Combing (A.22) together with (A.1), we have

$$G_{\mathcal{M}}(\boldsymbol{X}_i)\le\|\boldsymbol{X}_{i,\mathcal{M}}\|_2|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|_2+c_0^3/\sqrt{\bar{c}}.$$

The $\|\cdot\|_{\psi_1}$ Orlicz norm of G can be upper bounded by

$$\begin{aligned}\|G_{\mathcal{M}}(\boldsymbol{X}_i)\|_{\psi_1}&\le\|c_0^3/\sqrt{\bar{c}}\|_{\psi_1}+\|\|\boldsymbol{X}_{i,\mathcal{M}}\|_2|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|\|_{\psi_1}\\&\le c_0^3/\sqrt{\bar{c}}+\|\|\boldsymbol{X}_{i,\mathcal{M}}\|_2|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|\|_{\psi_1}.\end{aligned} \tag{A.73}$$

Notice that

$$\begin{aligned}\|\|\boldsymbol{X}_{i,\mathcal{M}}\|_2|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|_2\|_{\psi_1}&\le\left\|\frac{\|\boldsymbol{X}_{i,\mathcal{M}}\|_2^2}{2\sqrt{\kappa_n}}+\frac{\sqrt{\kappa_n}|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|^2}{2}\right\|_{\psi_1}\\&\le\frac{\|\|\boldsymbol{X}_{i,\mathcal{M}}\|_2^2\|_{\psi_1}}{2\sqrt{\kappa_n}}+\frac{\sqrt{\kappa_n}\|(\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0})^2\|_{\psi_1}}{2}\\&\le\frac{\sum_{j\in\mathcal{M}}\|X_{i,j}^2\|_{\psi_1}}{2\sqrt{\kappa_n}}+\frac{\sqrt{\kappa_n}\|\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}\|_{\psi_2}^2}{2}=O(\sqrt{\kappa_n}),\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality, the equality follows from

(A.60), (A.1) and Condition (A3). This together with (A.73) yields that

$$\max_{i\in[1,\ldots,n]} \|G_{\mathcal{M}}(\boldsymbol{X}_i)\|_{\psi_1} = O(\sqrt{\kappa_n}).$$

Hence, it follows from Lemma 2.2.2 in van der Vaart and Wellner (1996) that

$$\left\| \max_{i\in[1,\ldots,n]} |G_{\mathcal{M}}(\boldsymbol{X}_i)| \right\|_{\psi_1} \leq K_1 \log(1+n) \max_{i\in[1,\ldots,n]} \|G_{\mathcal{M}}(\boldsymbol{X}_i)\|_{\psi_1} = O(\sqrt{\kappa_n}\log n), \quad \text{(A.74)}$$

for some constant K_1 that is independent of $\mathcal{M}$.

Moreover, it follows from Cauchy-Schwarz inequality that

$$\begin{aligned}
\sigma_*^2 \equiv & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E} g_{\mathcal{M}}(\boldsymbol{X}_0, \boldsymbol{a})^2 \\
\leq & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T(\boldsymbol{X}_{0,\mathcal{M}}\boldsymbol{X}_{0,\mathcal{M}}^T - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}|^2 \\
\leq & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}\boldsymbol{X}_{0,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|^2 \\
\leq & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4} \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sqrt{\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4}.
\end{aligned}$$

Using similar arguments in proving (A.59), we can show

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4} \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sqrt{\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4} = O(1),$$

and hence $\sigma_*^2 = O(1)$.

Therefore, it follows from Theorem 4 in Adamczak (2008) that there exists some constant $K_2, K_3 > 0$ such that

$$\begin{aligned}
& \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \Pr\left(\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \geq \frac{t}{n} \right) \\
\leq & \exp\left(-\frac{t^2}{3n\sigma_*^2}\right) + 3\exp\left(-\frac{t}{K_2\sqrt{\kappa_n}\log n}\right) \\
\leq & \exp\left(-\frac{t^2}{3K_3 n}\right) + 3\exp\left(-\frac{t}{K_2\sqrt{\kappa_n}\log n}\right), \qquad \forall t > 0.
\end{aligned}$$

Define

$$t_0 = \max\left(2\sqrt{K_3 n\kappa_n \log p}, \frac{4}{3}K_2\kappa_n^{3/2}\log p \log n\right),$$

we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \Pr\left(\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \ge \frac{t_0}{n}\right)\\
\le\ & \exp\left(-\frac{4K_3 n\kappa_n \log p}{3nK_3}\right) + 3\exp\left(-\frac{4K_2\kappa_n^2\log p\log(n+1)}{3K_2\kappa_n\log(1+n)}\right) \le 4\exp\left(-\frac{4}{3}\kappa_n\log p\right).
\end{aligned}$$

The number of subset $\mathcal{M}$ with less than or equal to κ_n elements is upper bounded by $C_p^{\kappa_n} \le p^{\kappa_n}$. Hence, it follows from Bonferroni's inequality that

$$\begin{aligned}
&\Pr\left(\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \ge \frac{t_0}{n}\right)\\
\le\ & p^{\kappa_n} \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \Pr\left(\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \ge \frac{t_0}{n}\right)\\
\le\ & 4p^{\kappa_n}\exp\left(-\frac{4}{3}\kappa_n\log p\right) = 4\exp\left(-\frac{4}{3}\kappa_n\log p + \kappa_n\log p\right) = 4\exp\left(-\frac{1}{3}\kappa_n\log p\right) \to 0.
\end{aligned}$$

This together with (A.72) implies that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \le O(1)n^{-1/2}\sqrt{\kappa_n} + \frac{t_0}{n}, \tag{A.75}$$

with probability tending to 1, where $O(1)$ denotes some positive constant.

Under the given conditions, we have

$$\frac{4}{3}K_2\kappa_n^{3/2}\log p\log n = O(\sqrt{n\kappa_n\log p}),$$

and hence $t_0 = O(\sqrt{n\kappa_n\log p})$. Under the event defined in (A.75), we have for sufficiently large n,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right).$$

This proves (A.71). The upper bound for η_2 is thus given.

Consider (A.6). Assume for now, we've shown

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right), \tag{A.76}$$

with probability tending to 1. Then, under the event defined in (A.76), we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}_{\mathcal{M},j_0} - \sigma_{\mathcal{M},j_0}| = \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{|\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}|}{|\hat{\sigma}_{\mathcal{M},j_0} + \sigma_{\mathcal{M},j_0}|}\\
\le\ & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{|\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}|}{|\sigma_{\mathcal{M},j_0}|} \le \frac{1}{\sqrt{\bar{c}}} \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right),
\end{aligned}$$

where the last inequality follows from (A.1) and the last equality is due to (A.76). Hence, it suffices to show (A.76).

By definition, we have

$$\begin{aligned}
&|\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| \le |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| + |\widehat{\boldsymbol{\Sigma}}^T_{\mathcal{M},j_0}\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}^T_{\mathcal{M},j_0}\boldsymbol{\omega}_{\mathcal{M},j_0}| \qquad (A.77)\\
\le\ & |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| + |\boldsymbol{\Sigma}^T_{\mathcal{M},j_0}(\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0})| + |(\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0})^T\boldsymbol{\omega}_{\mathcal{M},j_0}|\\
+\ & |(\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0})^T(\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0})| \le |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}|\\
+\ & \|\boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2\|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 + \|\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2\\
+\ & \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2\|\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2.
\end{aligned}$$

It follows from (A.1), (A.5), (A.63) and (A.70) that with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right), \tag{A.78}$$

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right), \tag{A.79}$$

$$|\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| = O\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right), \quad \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 = O(1). \tag{A.80}$$

By Condition (A2), $\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}$ is invertible for any subset $\mathcal{M}$ such that $|\mathcal{M}| \le \kappa_n$. Hence, it follows from (A.22) that

$$\min_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}) \ge c_0^{-1/2}.$$

Using similar arguments in proving (A.1), we can show that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2^2 \le \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{c_0^2}{\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1})} \le c_0^{5/2}.$$

Under the given conditions, we have $\kappa_n \log p = o(n)$. Under the events defined in (A.77)-(A.79), we obtain that

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} |\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| \le O\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right) + c_0^{3/2} O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) \\
+ \quad & O(1) O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) + O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right).
\end{aligned}$$

This proves (A.76).

We now focus on proving (A.7). By definition, we have

$$\begin{aligned}
&\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0}) + (\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}^{-1}_{\mathcal{M},\mathcal{M}})\boldsymbol{\Sigma}_{\mathcal{M},j_0} \\
= \quad & \widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\{\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0} - (\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\},
\end{aligned}$$

for any $\mathcal{M}$ and hence the LHS of (A.7) can be upper bounded by

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1})\{\widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0}-\mathbf{\Sigma}_{\mathcal{M},j_0}-(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\}\right\|_2. \tag{A.81}$$

It suffices to provide an upper bound for (A.81). Using similar arguments in bounding η_1 and η_2, we can show the following event occurs with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0}-\mathbf{\Sigma}_{\mathcal{M},j_0}-(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\right\|_2 = O\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right). \tag{A.82}$$

Notice that

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}\right\|_2\\
\le\ &\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1}\right\|_2 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}\right\|_2 \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}\right\|_2.
\end{aligned}$$

To bound η_2, we have shown that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}\right\|_2 = O_p\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right).$$

By (A.64) and Condition (A2), we obtain

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}\right\|_2 = O_p\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right). \tag{A.83}$$

Combining this together with (A.82) and Cauchy-Schwarz inequality yields that

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}^{-1}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1})\{\widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0}-\mathbf{\Sigma}_{\mathcal{M},j_0}-(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\}\right\|_2\\
&\qquad = O_p\left(\frac{\kappa_n\log p}{n}\right).
\end{aligned}$$

This completes the proof.

A.7 Proof of Lemma A.3

Let

$$I_{2,j} = \frac{1}{\sqrt{n}} \sum_{t=s_n}^{n-1} \frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) X_{t+1,j} \mathcal{I}(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}).$$

Similar to (A.66), we can show

$$\left\| \frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) X_{t+1,j} \mathcal{I}(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}) \right\|_{\psi_1|\mathcal{F}_t} \le \frac{c_0^2}{\sqrt{\bar{c}}} \left(1 + \frac{c_0^2}{\bar{c}} \right),$$

almost surely, under Condition (A1). For any random variable Z, it follows from the definition of the Orlicz norm that $1 + \mathrm{E}|Z|^k / \|Z\|_{\psi_1}^k \le \mathrm{E}\exp(|Z|/\|Z\|_{\psi_1}) = 2$ for any integer $k > 0$ and hence $\mathrm{E}|Z|^k \le k! \|Z\|_{\psi_1}^k$. As a result, we have

$$\mathrm{E}\left\{ \left| \frac{1}{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) X_{t+1,j} \mathcal{I}(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}) \right|^k \middle| \mathcal{F}_t \right\} \le k! \frac{c_0^{2k}}{\bar{c}^{k/2}} \left(1 + \frac{c_0^2}{\bar{c}} \right)^k, \tag{A.84}$$

almost surely, for any $k \ge 1$.

Let $c_0^* = \bar{c}^{-1/2} c_0^2 (1 + \bar{c} c_0^2)$. It follows from Theorem 9.12 in de la Peña et al. (2009) that

$$\Pr\left(|I_{2,j}^{**}| > z, \sum_{t=s_n}^{n-1} \mathrm{E} \frac{1}{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left\{ \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right)^2 X_{t+1,j}^2 | \mathcal{F}_t \right\} \le 2n(c_0^*)^2 \right) \le 2\exp\left(-\frac{z^2}{2(2(c_0^*)^2 + c_0^* z/\sqrt{n})} \right), \quad \forall z > 0.$$

In view of (A.84), we have

$$\Pr\left(|I_{2,j}^{**}| > z \right) \le 2\exp\left(-\frac{z^2}{2(2(c_0^*)^2 + c_0^* z/\sqrt{n})} \right).$$

Let $z_0 = 3c_0^* \sqrt{\log p}$, we have by the condition $\log p = o(n)$ that

$$\Pr\left(|I_{2,j}^{**}| > z_0 \right) \le 2\exp\left(-\frac{9\log p}{4 + 6n^{-1/2}\sqrt{\log p}} \right) \le 2\exp\left(-\frac{3}{2}\log p \right),$$

for sufficiently large n. It follows from Bonferroni's inequality that

$$\Pr\left(\max_j |I_{2,j}^{**}| > z_0\right) \le \sum_j \Pr\left(|I_{2,j}^{**}| > z_0\right) \le 2p\exp\left(-\frac{3}{2}\log p\right) = 2\exp\left(-\frac{1}{2}\log p\right) \to 0.$$

This completes the first part of the proof.

For $t \in [s_n, \dots, n-1]$ and $j \in [1, \dots, p]$, let $W_{t+1,j} = X_{t+1,j}\mathcal{I}(j \in \widehat{\mathcal{M}}^{(t)})$ and $\boldsymbol{W}_{t+1} = (W_{t+1,1}, \cdots, W_{t+1,p})^T$. It follows that

$$\begin{aligned}
\frac{1}{n}\sum_{t=s_n}^{n-1} \|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}}(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}})\|_2^2 &= \frac{1}{n}\sum_{t=s_n}^{n-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \boldsymbol{W}_{t+1}\boldsymbol{W}_{t+1}^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\le \underbrace{\frac{1}{n}\sum_{t=s_n}^{n-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathrm{E}(\boldsymbol{W}_{t+1}\boldsymbol{W}_{t+1}^T|\mathcal{F}_t)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)}_{\eta_3} \\
&+ \underbrace{\left|\frac{1}{n}\sum_{t=s_n}^{n-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \{\boldsymbol{W}_{t+1}\boldsymbol{W}_{t+1}^T - \mathrm{E}(\boldsymbol{W}_{t+1}\boldsymbol{W}_{t+1}^T|\mathcal{F}_t)\}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right|}_{\eta_4}.
\end{aligned}$$

Notice that

$$\begin{aligned}
\eta_3 &= \frac{1}{n}\sum_{t=s_n}^{n-1}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathrm{E}(\boldsymbol{W}_{t+1}\boldsymbol{W}_{t+1}^T|\mathcal{F}_t)(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \frac{1}{n}\sum_{t=s_n}^{n-1}(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}})^T \boldsymbol{\Sigma}_{\widehat{\mathcal{M}}^{(t)},\widehat{\mathcal{M}}^{(t)}}(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}}) \le \lambda_{\max}(\boldsymbol{\Sigma}_{\widehat{\mathcal{M}}^{(t)},\widehat{\mathcal{M}}^{(t)}})\|\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}^{(t)}}\|_2^2.
\end{aligned}$$

It follows from Condition (A1), (A4) and (A.22) that $\eta_3 = O(\eta_n^2)$, with probability tending to 1. As for η_4, we have

$$\eta_4 \le \|\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|_1^2 \max_{j_1,j_2}\left|\frac{1}{n}\sum_{t=s_n}^{n-1}\{W_{t+1,j_1}W_{t+1,j_2} - \mathrm{E}(W_{t+1,j_1}W_{t+1,j_2}|\mathcal{F}_t)\}\right|. \tag{A.85}$$

For any j_1, j_2, $\sum_{t=s_n}^{n-1}\{W_{t+1,j_1}W_{t+1,j_2} - \mathrm{E}(W_{t+1,j_1}W_{t+1,j_2}|\mathcal{F}_t)$ forms a mean zero martingale with respect to the filtration $\{\sigma(\mathcal{F}_t)\}$. Using similar arguments in bounding $\max_j |\mathcal{I}_{2,j}^{**}|$, we can show the following holds with probability tending to 1,

$$\max_{j_1,j_2}\left|\frac{1}{n}\sum_{t=s_n}^{n-1}\{W_{t+1,j_1}W_{t+1,j_2} - \mathrm{E}(W_{t+1,j_1}W_{t+1,j_2}|\mathcal{F}_t)\}\right| = O(n^{-1/2}\sqrt{\log p}). \tag{A.86}$$

Combining (A.25) with (A.85), (A.86) and the condition $\kappa_n^2 \log p = O(n/\log^2 n)$ yields that

$$\eta_4 = O\left(\eta_n^2 n^{-1/2} \kappa_n \sqrt{\log p}\right) = O(\eta_n^2),$$

with probability tending to 1. (A.9) is hence proven.

A.8 Proof of Lemma A.4

Assertion (A.28) can be proven in a similar manner as (A.1). We omit its proof for brevity. To prove (A.29) and (A.30), we first show the following events occur with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\le \kappa_n}} \left\| \widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} \right\|_2 \le \frac{\bar{c}_* \sqrt{\kappa_n \log p}}{\sqrt{n}} + \bar{c}_* \eta_n, \tag{A.87}$$

$$\max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\le \kappa_n}} \left\| \widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0} - \mathbf{\Sigma}_{\mathcal{M},j_0} \right\|_2 \le \frac{\bar{c}_* \sqrt{\kappa_n \log p}}{\sqrt{n}} + \bar{c}_* \eta_n, \tag{A.88}$$

$$\max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\le \kappa_n}} \left\| \left(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} \right) \boldsymbol{\omega}_{\mathcal{M},j_0} \right\|_2 \le \frac{\bar{c}_* \sqrt{\kappa_n \log p}}{\sqrt{n}} + \bar{c}_* \eta_n. \tag{A.89}$$

Using similar arguments in the proof of Lemma A.2, we can show that there exists some constant $\bar{c}_{**} > 0$ such that the following events occur with probability tending to 1,

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \widehat{\mathbf{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} \right\|_2 \le \frac{\bar{c}_{**} \sqrt{\kappa_n \log p}}{\sqrt{n}},$$

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \widehat{\mathbf{\Sigma}}^*_{\mathcal{M},j_0} - \mathbf{\Sigma}_{\mathcal{M},j_0} \right\|_2 \le \frac{\bar{c}_{**} \sqrt{\kappa_n \log p}}{\sqrt{n}},$$

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \left(\widehat{\mathbf{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \mathbf{\Sigma}_{\mathcal{M},\mathcal{M}} \right) \boldsymbol{\omega}_{\mathcal{M},j_0} \right\|_2 \le \frac{\bar{c}_{**} \sqrt{\kappa_n \log p}}{\sqrt{n}}.$$

Therefore, it suffices to show the following events occur with probability tending to 1,

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \left(\widehat{\mathbf{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} \right) \boldsymbol{\omega}_{\mathcal{M},j_0} \right\|_2 \le \bar{c}_{***} \eta_n, \tag{A.90}$$

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \widehat{\mathbf{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}} \right\|_2 \le \bar{c}_{***} \eta_n, \tag{A.91}$$

$$\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\le \kappa_n} \left\| \widehat{\mathbf{\Sigma}}^*_{\mathcal{M},j_0} - \widehat{\mathbf{\Sigma}}_{\mathcal{M},j_0} \right\|_2 \le \bar{c}_{***} \eta_n, \tag{A.92}$$

for some constant $\bar{c}_{***} > 0$.

Similar to (A.39), we can show that with probability tending to 1 that

$$|b''(\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}) - b''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)| \leq c_*|\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)|,$$

where the constant c_* is defined in (A.38). With some calculations, we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \left\|\left(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right\|_2 \qquad\qquad (A.93)\\
\leq\ & \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\leq 1}} \left|\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right|\\
\leq\ & \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\leq 1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}||b''(\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}) - b''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)|\\
\leq\ & c_*\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\leq 1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)| \leq c_*\sqrt{\eta_5\eta_6},
\end{aligned}$$

where the last inequality follows from Cauchy-Schwarz inequality and

$$\eta_5 = \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\leq 1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}|,$$

$$\eta_6 = \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)|^2.$$

Consider η_5. By (A.28), (A3*) and (A4*), we have for any $\mathcal{M}$ and $\boldsymbol{a} \in \mathbb{R}^{|\mathcal{M}|}$ that

$$\||\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2|\boldsymbol{X}_{0,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|\|_{\psi_1} \leq \bar{c}^{-1/2}c_0\sqrt{\kappa_n}\omega_0\||\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2\|_{\psi_1} \leq \bar{c}^{-1/2}c_0^3\sqrt{\kappa_n}\omega_0.$$

Using similar arguments in bounding (A.65), we can show for any $\mathcal{M}$ with $|\mathcal{M}| \leq \kappa_n$, we have that

$$\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\leq 1}}\left(\frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}| - \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|\right) = o(1),$$

under the given conditions on κ_n. Hence, using similar arguments in proving (A.75), we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\le 1}}\left(\frac{1}{n}\sum_{i=1}^n|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}| - \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|\right) \tag{A.94}$$
$$= O_p\left(\sqrt{\frac{\kappa_n\log p}{n}}+\frac{\kappa_n^2\log p\log n}{n}\right),$$

which is $o_p(1)$ under the condition that $\kappa_n^{5/2}\log p = O(n/\log^2 n)$. In addition, similar to (A.59), we can show

$$\sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4}(\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4)^{1/4} = O(1), \quad \forall \boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \text{ with } \|\boldsymbol{a}\|_2 = 1,$$

by (A.4) and Condition (A3*). It follows from Cauchy-Schwarz inequality that

$$\begin{aligned} &\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}| \le \sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4}\sqrt{\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^2}\\ \le\ & \sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4}(\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4)^{1/4} = O(1), \quad \forall \boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \text{ with } \|\boldsymbol{a}\|_2 = 1.\end{aligned}$$

This together with (A.94) yields that

$$\eta_5 = O(1), \tag{A.95}$$

with probability tending to 1.

Recall that s^* is the number of nonzero elements in $\boldsymbol{\beta}_0$. Under Condition (A5*), it follows from Lemma G.9 of Shi et al. (2018) that

$$\eta_6 \le (k_0+2)^2\max_{|\mathcal{M}|\le s^*}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{s^*}\\\|\boldsymbol{a}\|_2\le 1}}\frac{1}{n}\sum_{i=1}^n|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2^2,$$

with probability tending to 1. Condition (A1*) implies that $s^*\le\kappa_n$. It follows that

$$\eta_6 \le (k_0+2)^2\max_{|\mathcal{M}|\le \kappa_n}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{\kappa_n}\\\|\boldsymbol{a}\|_2\le 1}}\frac{1}{n}\sum_{i=1}^n|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2^2,$$

with probability tending to 1. Similar to (A.95), we can show

$$\eta_6 = O(\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2^2),$$

with probability tending to 1. Under (A5*), we obtain

$$\eta_6 = O(\eta_n^2), \tag{A.96}$$

with probability tending to 1. This together with (A.93) and (A.95) proves (A.90). Similarly, we can show (A.91) and (A.92) hold. We omit the technical details to save space.

This further implies (A.87)-(A.89) hold. Based on these results, following the arguments in the proof of Lemma A.2, we can show (A.29) and (A.30) hold. Besides, based on (A.90)-(A.92), we can similarly show (A.32) holds.

Now, we focus on proving (A.31). Similar to (A.83), we can show

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\right\|_2 = O\left(\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\boldsymbol{\Sigma}}^{*}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right\|_2\right),$$

with probability tending to 1. In view of (A.91), we obtain

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\right\|_2 \le \bar{c}_0^*\eta_n, \tag{A.97}$$

for some constant $\bar{c}_0^* > 0$, with probability tending to 1.

For any $\mathcal{M}\subseteq\mathbb{I}_{j_0}$, we have

$$\begin{aligned}
&\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0} = \underbrace{\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0})}_{I_1^*}+(\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}})\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0}\\
+\ &\underbrace{(\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}})(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0})}_{I_2^*} = I_1^*+\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}})\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}\\
+\ &I_2^* = I_1^*+I_2^*+\underbrace{(\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}})(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}})\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}}_{I_3^*}+\underbrace{\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}})\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}}_{I_4^*}
\end{aligned}$$

By (A.92) and (A.97), it is immediate to see that $|I_2^*|$ is upper bounded by $\bar{c}_0^*\bar{c}_{***}\eta_n^2$, with probability tending to 1.

Similar to (A.5), we can show

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\|\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}-\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 = O_p\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right) = o_p(1). \tag{A.98}$$

By (A.28), this further implies that

$$\Pr\left(\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\|\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}\|_2 \le 2(\bar{c})^{-1/2}c_0\right) \to 1. \tag{A.99}$$

This together with (A.91) and (A.97) yields that

$$\Pr\left(|I_3^*| \le \frac{4\bar{c}_{***}^2\eta_n^2}{\bar{c}^2}2(\bar{c})^{-1/2}c_0\right) \to 1.$$

Recall that

$$\widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0} + \sum_{j=1}^{p}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\left(\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0} + \widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}\right)(\widetilde{\beta}_j - \beta_{0,j}).$$

Hence, in order to prove

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\eta_n^2, \tag{A.100}$$

it suffices to show the following events occur with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|I_1^* - \sum_{j=1}^{p}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}(\tilde{\beta}_j - \beta_{0,j})\right\|_2 = O(\eta_n^2), \tag{A.101}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|I_4^* - \sum_{j=1}^{p}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}(\tilde{\beta}_j - \beta_{0,j})\right\|_2 = O(\eta_n^2). \tag{A.102}$$

We first prove (A.101). Similar to (A.62), we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\|_2 = O(1),$$

with probability tending to 1. By the definition of $\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}$, it suffices to show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0} - \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{i,\mathcal{M}}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}\right\|_2 \le c_1^*\eta_n^2, \tag{A.103}$$

for some constant $c_1^* > 0$, with probability tending to 1. This is equivalent to show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0}-\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}_{i,\mathcal{M}}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}\right)\right|\le c_1^*\eta_n^2,$$

with probability tending to 1. For any $\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}$, it follows from Taylor's theorem that

$$\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0}\right)=\frac{1}{n}\sum_{i=1}^n\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}^*_{\boldsymbol{a}})X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\},$$

for some $\boldsymbol{\beta}^*_a$ lying on the line segment joining $\boldsymbol{\beta}_0$ and $\widetilde{\boldsymbol{\beta}}$. The function b''' is Lipschitz continuous. Similar to (A.39), we can show

$$\sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}}|b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}^*_a)-b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)|\le L_0|\boldsymbol{X}_i^T(\boldsymbol{\beta}_0-\boldsymbol{\beta}^*_a)|\le L_0|\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}_0-\boldsymbol{\beta}_0)|,$$

for some constant $L_0 > 0$, with probability tending to 1. This together with Condition (A4*) yields that

$$\begin{aligned}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}&\left\|\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0}-\frac{1}{n}\sum_{i=1}^n\boldsymbol{X}_{i,\mathcal{M}}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}\right\|_2\\
&\le L_0\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^n|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||X_{i,j_0}|\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}^2\right|\\
&\le L_0\omega_0\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^n|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}^2\right|,\quad\text{(A.104)}\end{aligned}$$

with probability tending to 1. Now (A.103) can be proven in a similar manner as (A.96). This further implies (A.101) holds.

The proof of (A.102) is more involved. Define

$$I_4^{**}=\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}$$

Using similar arguments in proving (A.101), we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|I_4^{**}-\sum_{j=1}^{p}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\boldsymbol{\omega}_{\mathcal{M},j_0}(\tilde{\beta}_j-\beta_{0,j})\right\| \tag{A.105}$$
$$\le O(1)\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}\boldsymbol{X}_{i,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}^2\right|,$$

with probability tending to 1, where $O(1)$ denotes some positive constant. Using similar arguments in proving (A.75) and (A.96), we can show the last term is upper bounded by $O(\eta_n^2)$ with probability tending to 1, under the condition that $\kappa_n^3=O(n),\kappa_n^{5/2}\log p=O(n/\log^2 n)$. Hence, to prove (A.102), it suffice to show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|I_4^{*}-I_4^{**}-\sum_{j=1}^{p}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}(\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}-\boldsymbol{\omega}_{\mathcal{M},j_0})(\tilde{\beta}_j-\beta_{0,j})\right\|=o_p(\eta_n^2). \tag{A.106}$$

Using similar arguments in proving (A.105), we have by (A.98) that the LHS of (A.106) can be upper bounded by

$$|R_0|\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}^2\right|.$$

for some random variable $R_0=O_p(n^{-1/2}\kappa_n\sqrt{\log p})$. Under the condition that $\kappa_n^{5/2}\log p=O(n/\log^2 n)$, we can show similarly that the above expression is $o_p(\eta_n^2)$. This proves (A.106). As a result, (A.102) and (A.100) are proven. Similarly, we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\left\|\hat{\sigma}^2_{\mathcal{M},j_0}-\tilde{\sigma}^2_{\mathcal{M},j_0}\right\|_2\le\bar{c}_0\eta_n^2.$$

This together with (A.100) proves (A.31). We omit the details to save space.

Finally, we show

$$\sum_{t=0}^{n-1}\frac{\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(-sn)},j_0}}\right)=\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}+o_p(1). \tag{A.107}$$

With some calculations, we have

$$
\begin{aligned}
&\sum_{t=0}^{n-1}\frac{\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}}\left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j-\beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\right)-\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\\
= &\underbrace{\sum_{j=1}^{p}\left(\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}(\tilde{\beta}_j-\beta_{0,j})\right)}_{\eta_1^*}+\underbrace{\sum_{t=0}^{n-1}\frac{(\widetilde{Z}_{t+1,j_0}-\widehat{Z}^*_{t+1,j_0})\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}}_{\eta_2^*}\\
+ &\underbrace{\sum_{j=1}^{p}\left(\sum_{t=0}^{n-1}\frac{(\widetilde{Z}_{t+1,j_0}-\widehat{Z}^*_{t+1,j_0})\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}(\tilde{\beta}_j-\beta_{0,j})\right)}_{\eta_3^*}.
\end{aligned}
$$

In the following, we first prove $\eta_1^*=o_p(1)$. Note that $|\eta_1^*|\le \max_j|\eta_{1,j}^*|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1$ where

$$
\eta_{1,j}^*=\sum_{t=0}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}.
$$

By Condition (A5*), it suffices to show $\max_j|\eta_{1,j}^*|=O_p(\sqrt{\log p})$.

It follows from the Lipschitz continuity of $b'''(\cdot)$ that

$$
|b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)-b'''(0)|\le L_0|\boldsymbol{X}_i^T\boldsymbol{\beta}_0|.
$$

Hence, under Condition (A4*), $\max_{1\le i\le n}|b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)|$ is bounded by some universal constant. Since $X_{0,j}$'s are uniformly bounded, we obtain

$$
\max_{1\le j\le p}|\widehat{\boldsymbol{\Psi}}^{(j)}_{j_0,j_0}|=\max_{1\le j\le p}\left|\frac{1}{n}\sum_i X_{i,j}^3 b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)\right|=O(1). \tag{A.108}
$$

Similarly, we can show

$$
\begin{aligned}
\max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\|_2 &= \max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}|\boldsymbol{a}^T\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\boldsymbol{a}|\\
&\le O(1)\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^n(\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}})^2\right|,
\end{aligned}
$$

where $O(1)$ denotes some positive constant. Using similar arguments in bounding η_5, we can show

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}})^2\right| = O(1),$$

with probability tending to 1 and hence,

$$\max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\|_2 = O(1), \tag{A.109}$$

with probability tending to 1. Similarly, we can show

$$\max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}\|_2 = O(1), \tag{A.110}$$

with probability tending to 1. This together with (A.99), (A.108) and (A.109) yields

$$\max_{1\le j\le p}\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}|\hat{\xi}^{(j)}_{\mathcal{M},j_0}| = O(1), \tag{A.111}$$

with probability tending to 1.

Note that

$$\eta^*_{1,j} = \underbrace{\sum_{t=0}^{s_n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}}_{\eta^{**}_{1,j}} + \underbrace{\sum_{t=s_n}^{n-1}\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}}_{\eta^{***}_{1,j}}.$$

We first prove $\max_j|\eta^{***}_{1,j}| = O_p(\sqrt{\kappa_n\log p})$.

Define $\sigma(\mathcal{F}^*_t) = \sigma(\boldsymbol{X}_1,\boldsymbol{X}_2,\dots,\boldsymbol{X}_n,Y_1,Y_2,\dots,Y_t)$, $\eta^{***}_{1,j}$ corresponds to a mean-zero martingale with respect to the filtration $\{\sigma(\mathcal{F}^*_t): t\ge s_n\}$. By Condition (A1*) and (A4*), we have for any $t=0,\dots,n-1$,

$$|\widehat{Z}^*_{t+1,j_0}| \le \omega_0\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}(1+\sqrt{\kappa_n}\|\widehat{\omega}^*_{\mathcal{M},j_0}\|_2). \tag{A.112}$$

Let

$$\bar{c}_n^{(j)} \equiv \omega_0 n^{-1/2} \max_{1\le j\le p} \max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\le \kappa_n}} \frac{|(1+\sqrt{\kappa}_n \|\widehat{\omega}^*_{\mathcal{M},j_0}\|_2)|}{|\hat{\sigma}^*_{\mathcal{M},j_0}|} |\hat{\xi}^{(j)}_{\mathcal{M},j_0}|.$$

Under Condition (A6*), $\|\varepsilon_{t+1}\|_{\psi_1|\mathcal{F}_t^*} \le L^*$ for some constant $L^*>0$. Similar to (A.84), we can show

$$\mathrm{E}\{|\varepsilon_{t+1}|^k | \mathcal{F}_t^*\} \le k!(L^*)^k,$$

for any k and t. By Condition (A1*) and (A.112), we have

$$\mathrm{E}\left\{ \left| \frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right|^k \middle| \mathcal{F}_t^* \right\} \le \left(\frac{\widehat{Z}^*_{t+1,j_0}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right)^2 k!(L^*)^k (\bar{c}_n^{(j)})^{k-2}, \tag{A.113}$$

for any j,t and $k\ge 2$.

Let

$$V_n^{(j)} = 2\sum_{t=s_n}^{n-1} \left(\frac{\widehat{Z}^*_{t+1,j_0}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right)^2.$$

Similar to (A.41) and (A.12), we can show $\sum_{t=s_n}^{n-1}(\widehat{Z}^*_{t+1,j_0})^2 = O_p(n)$ and $\min_t \hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \le 2/\sqrt{\bar{c}}$, respectively. It follows from (A.111) that

$$\max_j V_n^{(j)} = O_p(1). \tag{A.114}$$

It follows from Theorem 9.12 in de la Peña et al. (2009) that for any $1\le j\le p$,

$$\Pr(|\eta^{***}_{1,j}| > z, V_n^{(j)} \le \bar{z}) \le 2\exp\left(-\frac{z^2}{2(L^*)^2\bar{z} + 2L^*\bar{c}_n^{(j)} z} \right) \le 2\exp\left(-\frac{z^2}{\max\{4(L^*)^2\bar{z}, 4L^*\bar{c}_n^{(j)} z\}} \right).$$

Take $z_0^{(j)} = \max(3L^*\sqrt{\bar{z}\log p}, 8L^*\bar{c}_n^{(j)}\log p)$, we have

$$\Pr(|\eta^{***}_{1,j}| > z_0^{(j)}, V_n^{(j)} \le n\bar{z}) \le 2\exp(-2\log p) = \frac{2}{p^2}.$$

It follows from Bonferroni's inequality that

$$\Pr\left(\bigcap_{j=1}^{p}\left\{|\eta_{1,j}^{***}|>z_0^{(j)}\right\},\max_j V_n^{(j)}\le n\bar{z}\right)\le\sum_{j=1}^{p}\Pr(|\eta_{1,j}^{***}|>z_0^{(j)})=\frac{2}{p}\to 0.$$

By (A.114), for any $\epsilon>0$, there exists some $\bar{z}>0$ such that $\Pr(\max_j V_n^{(j)}\le n\bar{z})\ge 1-\epsilon$.

This implies that

$$\max_{j=1}^{p}|\eta_{1,j}^{***}|\le\max_{j=1}^{p} z_0^{(j)}, \tag{A.115}$$

with probability tending to $1-\epsilon$. By (A.37) and (A.99), we have $\max_j \bar{c}_n^{(j)}=O_p(\sqrt{\kappa_n})$ and hence

$$\max_j|\eta_{1,j}^{***}|=O_p(\sqrt{\log p}), \tag{A.116}$$

by (A.115) and the condition that $\kappa_n^{5/2}\log p=O(n/\log^2 n)$.

Recall that

$$\eta_{1,j}^{**}=\sum_{t=0}^{s_n-1}\frac{\widehat{Z}_{t+1,j_0}^{*}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}.$$

Given $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$ and $Y_{s_n+1},\dots,Y_n$, each term in $\eta_{1,j}^{**}$ is independent of others. Using similar arguments, we can show $\max_j|\eta_{1,j}^{**}|=O_p(\sqrt{\log p})$. This together with (A.116) gives $\max_j|\eta_{1,j}^{*}|=O_p(\sqrt{\log p})$. By Condition (A5*), we obtain $|\eta_1^*|\le\max_j\max_j|\eta_{1,j}^*|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1=o_p(1)$. Similarly, we can show $\eta_2^*=o_p(1)$. It remains to show $\eta_3^*=o_p(1)$.

Note that $|\eta_3^*|$ can be upper bounded by $\max_j|\eta_{3,j}^*|\|\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_1$ where

$$\eta_{3,j}^{*}=\sum_{t=0}^{n-1}\frac{(\widetilde{Z}_{t+1,j_0}-\widehat{Z}_{t+1,j_0}^{*})\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}.$$

Since

$$\widehat{Z}_{t+1,j_0}^{*}-\widetilde{Z}_{t+1,j_0}=\sum_{j=1}^{p}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^{T}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\widehat{\mathcal{M}}_{j_0}^{(t)},\widehat{\mathcal{M}}_{j_0}^{(t)}}\left(\widehat{\boldsymbol{\Psi}}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}+\widehat{\boldsymbol{\Psi}}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},\widehat{\mathcal{M}}_{j_0}^{(t)}}\widehat{\boldsymbol{\omega}}^{*}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\right)(\widetilde{\beta}_j-\beta_{0,j}).$$

Using similar arguments in proving $\max_j \eta^{**}_{1,j} = O_p(\sqrt{\log p})$, we can show

$$\max_{j_1,j_2}\left|\sum_{t=0}^{n-1}\frac{\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\widehat{\boldsymbol{\Sigma}}^{*-1}_{\widehat{\mathcal{M}}^{(t)}_{j_0},\widehat{\mathcal{M}}^{(t)}_{j_0}}\left(\widehat{\boldsymbol{\Psi}}^{(j_1)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}+\widehat{\boldsymbol{\Psi}}^{(j_1)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},\widehat{\mathcal{M}}^{(t)}_{j_0}}\widehat{\boldsymbol{\omega}}^*_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}\right)\varepsilon_{t+1}\hat{\xi}^{(j_2)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}\right| = O_p(\sqrt{\log p}).$$

Hence, we have $|\eta^*_3| = O_p(\sqrt{\log p}(\sqrt{\kappa_n}\eta_n)^2) = o_p(1)$, by (A5*). The proof is hence completed.

A.9 Technical lemmas

Lemma A.5. *For any positive definite matrix*

$$\boldsymbol{\Psi} = \begin{pmatrix} \boldsymbol{\Psi}_{11} & \boldsymbol{\Psi}_{12} \\ \boldsymbol{\Psi}_{21} & \boldsymbol{\Psi}_{22} \end{pmatrix},$$

denote its inverse matrix as $\boldsymbol{\Omega}$ *and partition it into* $\boldsymbol{\Omega}_{11},\dots,\boldsymbol{\Omega}_{22}$ *accordingly. Then,*

$$\boldsymbol{\Omega}_{11} = (\boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21})^{-1}.$$

Besides, let $\boldsymbol{\Psi}_* = \boldsymbol{\Psi}_{22} - \boldsymbol{\Psi}_{21}\boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}$, *we have*

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Psi}_{11}^{-1} + \boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_*^{-1}\boldsymbol{\Psi}_{21}\boldsymbol{\Psi}_{11}^{-1} & -\boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_*^{-1} \\ -\boldsymbol{\Psi}_*^{-1}\boldsymbol{\Psi}_{21}\boldsymbol{\Psi}_{11}^{-1} & \boldsymbol{\Psi}_* \end{pmatrix}.$$

B More on the technical conditions

B.1 More on (A1) and (A1*)

The validity of the sure screening property assumed in (A1) or (A1*) relies typically on the following minimum-signal-strength condition:

$$\min_{j\in\mathcal{M}_{j_0}} |\beta_{0,j}| \ge \sigma^*_n, \tag{B.1}$$

for some monotonically nonincreasing sequence $\{\sigma^*_n\}_n$ that satisfies $\sigma^*_n \gg n^{-1/2}$ and $\sigma^*_n \to 0$ as $n\to\infty$. Although such conditions are not assumed in van de Geer et al. (2014) or Ning and Liu (2017), these authors imposed some additional assumptions on the design matrix. For instance, consider the decorrelated score statistic proposed by Ning and Liu (2017). For linear regression models, its validity depends on the sparsity of a high-dimensional vector $\boldsymbol{w}^*$. When the covariates follow a Gaussian graphical model, the sparsity assumption on $\boldsymbol{w}^*$ requires the degree of a particular node in the graph to be relatively small. See Remark 6 of Ning and Liu (2017) for details.

B.1.1 A counterexample

Assume $\boldsymbol{X}_0 \sim N(0, \{\rho^{|i-j|}\}_{i,j=1,\dots,p})$ for some $0 < \rho < 1$, $Y_0 = \boldsymbol{X}_0^T \boldsymbol{\beta}_0 + \varepsilon_0$ where $\varepsilon_0 \sim N(0,1)$ that is independent of $\boldsymbol{X}_0$ and $\beta_{0,1} = 0$, $\beta_{0,2} = n^{-1/2}$, $\beta_{0,j} = 0$ for all $j > 2$. The minimum-signal-strength condition (B.1) is thus violated. Our goal is to construct a CI for $\beta_{0,1}$.

Suppose we use SIS to determine the set of important variables based on their marginal correlations with the response. Specifically, set

$$\begin{aligned} \widehat{\mathcal{M}}_1^{(t)} &= \left\{ j \geq 2 : t^{-1} \sum_{i=1}^{t} |Y_i X_{i,j}| \geq \sigma_t \right\}, \qquad \forall s_n \leq t < n, \\ \widehat{\mathcal{M}}_1^{(-s_n)} &= \left\{ j \geq 2 : (n - s_n)^{-1} \sum_{i=s_n+1}^{n} |Y_i X_{i,j}| \geq \sigma_{n-s_n} \right\}, \end{aligned}$$

for some sequence $\{\sigma\}_n$ that satisfies $\sigma_n \gg n^{-1/2} \log^{1/2} n$.

Notice that for any $j \geq 2$, we have $\mathrm{E} Y_0 X_{0,j} = n^{-1/2} \mathrm{E} X_{0,2} X_{0,j} = n^{-1/2} \rho^{j-2}$. Using Bernstein's inequality, we can show that the following events occur with probability tending to 1 that

$$\begin{aligned} t^{-1} \sum_{i=1}^{t} |Y_i X_{i,j}| \leq O(1) t^{-1/2} \sqrt{\log t + \log p}, \qquad \forall s_n \leq t < n, 2 \leq j \leq p, \\ (n - s_n)^{-1} \sum_{i=s_n+1}^{n} |Y_i X_{i,j}| \leq O(1) (n - s_n)^{-1/2} \sqrt{\log(n - s_n) + \log p}, \quad \forall 2 \leq j \leq p, \end{aligned}$$

where $O(1)$ denotes some positive constant. Suppose $p = O(n)$ and the sequence s_n is set to be $\lfloor \epsilon n \rfloor$ for some $0 < \epsilon < 1$. It follows that

$$\begin{aligned} \max_{s_n \leq t \leq n, 2 \leq j \leq p} t^{-1/2} \sqrt{\log t + \log p} = O(n^{-1/2} \log^{1/2} n), \\ (n - s_n)^{-1/2} \sqrt{\log(n - s_n) + \log p} = O(n^{-1/2} \log^{1/2} n). \end{aligned}$$

Hence, for sufficiently large n, we have $\widehat{\mathcal{M}}_1^{(-s_n)} = \widehat{\mathcal{M}}_1^{(s_n)} = \widehat{\mathcal{M}}_1^{(s_n+1)} = \widehat{\mathcal{M}}_1^{(n-1)} = \emptyset$, with probability tending to 1.

As a result, our score equation for $\beta_{0,1}$ is given by

$$\sum_{i=1}^n X_{i,1}(Y_i - X_{i,1}^T \beta_{0,1}) = 0,$$

with probability tending to 1. Therefore, the proposed CI for $\beta_{0,1}$ equals

$$\left[\hat{\beta}_1 - z_{\alpha/2} n^{-1/2} \left(\frac{1}{n} \sum_{i=1}^n X_{i,1}^2 \right)^{-1/2}, \hat{\beta}_1 + z_{\alpha/2} n^{-1/2} \left(\frac{1}{n} \sum_{i=1}^n X_{i,1}^2 \right)^{-1/2} \right],$$

where

$$\hat{\beta}_1 = \left(\sum_{i=1}^n X_{i,1}^2 \right)^{-1} \left(\sum_{i=1}^n X_{i,1} Y_i \right).$$

It follows that

$$\begin{aligned} \left(\sum_{i=1}^n X_{i,1}^2 \right)^{1/2} (\hat{\beta}_1 - \beta_{0,1}) &= \left(\sum_{i=1}^n X_{i,1}^2 \right)^{-1/2} \left(\sum_{i=1}^n X_{i,1}(Y_i - X_{i,1}\beta_{0,1}) \right) \\ = \left(\sum_{i=1}^n X_{i,1}^2 \right)^{-1/2} \left(\sum_{i=1}^n X_{i,1}\varepsilon_i \right) &+ n^{-1/2} \left(\sum_{i=1}^n X_{i,1}^2 \right)^{-1/2} \left(\sum_{i=1}^n X_{i,1} X_{i,2} \right). \end{aligned} \tag{B.2}$$

By the central limit theorem, the first term on the RHS of (B.2) converges to $N(0, 1)$ in distribution. The second term converges to ρ, according to the law of large numbers. Hence, our CI is not valid as long as $\rho > 0$. This implies that the minimal-signal-strength condition is necessary to guarantee the validity of our procedure.

B.1.2 Extension to many small but weak signals

Moreover, one could relax the minimum-signal-strength condition in (B.1) by assuming there are many small but weak signals in $\boldsymbol{\beta}_0$. Specifically, assume $\mathcal{M}_{j_0}$ is a union of two disjoint subsets $\mathcal{M}_{j_0}^*$ and $\mathcal{M}_{j_0}^{**}$ such that

$$\mathcal{M}_{j_0}^* = \{j \in \mathbb{I}_{j_0} : |\beta_{0,j}| \geq \sigma_n^*\}, \tag{B.3}$$

and

$$\mathcal{M}_{j_0}^{**} = \mathcal{M}_{j_0} \cap (\mathcal{M}_{j_0}^{*})^c \quad \text{with} \quad \|\boldsymbol{\beta}_{0,\mathcal{M}_{j_0}^{**}}\|_2 = O(n^{-\kappa^*}), \tag{B.4}$$

for some sequence $n^{-1/2} \ll \sigma_n^* \ll 1$ and some constant $\kappa^* > 1/2$. We require $|\mathcal{M}_{j_0}^*|$ is much smaller than n while $|\mathcal{M}_{j_0}^{**}|$ can be much larger than the sample size. Such conditions are very similar to the zonal assumption imposed by Bühlmann and Mandozzi (2014). When (B.3) and (B.4) hold, Condition (A1) or (A1*) can then be replaced by the following:

(A1**) Assume $\widehat{\mathcal{M}}_{j_0}^{(n)}$ satisfies $\Pr(|\widehat{\mathcal{M}}_{j_0}^{(n)}| \le \kappa_n) = 1$ for some $1 \le \kappa_n = o(n)$. Besides,

$$\Pr\left(\mathcal{M}_{j_0}^* \subseteq \widehat{\mathcal{M}}_{j_0}^{(n)}\right) \ge 1 - O\left(\frac{1}{n^{\alpha_0}}\right),$$

for some constant $\alpha_0 > 1$.

That is, we require the selected model will contain all those strong signals with probability tending to 1. This assumption can be satisfied under the condition in (B.3). In the following, we sketch a few lines to show the proposed method works. For simplicity, we focus on linear regression models.

By (A1**) and Bonferroni's inequality, the following event occurs with probability tending to 1,

$$\mathcal{M}_{j_0}^* \subseteq \bigcap_{t=s_n}^{n-1} \widehat{\mathcal{M}}_{j_0}^{(t+1)}. \tag{B.5}$$

Under the event defined in (B.5), we have

$$\begin{aligned} &\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0} - \beta_{0,j_0}) = I_1 + I_2 + I_3 + I_4 \\ &+ \sum_{t=0}^{s_n-1} \frac{\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} + \sum_{t=s_n}^{n-1} \frac{\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}, \end{aligned}$$

and

$$\left|\sum_{t=0}^{s_n-1}\frac{\widehat{Z}_{t+1,j_0}\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}\right| \le \sum_{t=0}^{s_n-1}\frac{|\widehat{Z}_{t+1,j_0}||\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}|}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}},$$

$$\left|\sum_{t=s_n}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}}\right| \le \sum_{t=s_n}^{n-1}\frac{|\widehat{Z}_{t+1,j_0}||\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(-s_n)})^c}|}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}},$$

where I_1, I_2, I_3 and I_4 are defined in Section A.1, $\widehat{Z}_{t+1,j_0}$, $\widehat{\mathcal{M}}_{j_0}^{(-s_n)}$, $\widehat{\mathcal{M}}_{j_0}^{(t)}$, $\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}$ and $\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ are defined in Section 2.

Note that we have shown in Section A.1 that $I_1+I_2+I_3+I_4$ is asymptotically normal. It suffices to show

$$\sum_{t=0}^{n}\frac{|\widehat{Z}_{t+1,j_0}||\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}|}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} = o_p(1), \tag{B.6}$$

where $\mathcal{M}_{j_0}^{(t)} = \mathcal{M}_{j_0}^{(-s_n)}$, for $t=0,\dots,s_n-1$. Under the event defined in (A.12) and (A1**), the LHS of (B.6) is upper bounded by

$$I_5 \equiv \sum_{t=0}^{n-1}\frac{2|\widehat{Z}_{t+1,j_0}||\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}|}{\sqrt{\bar{c}n}}.$$

By Cauchy-Schwarz inequality, we have

$$I_5 \le \frac{2}{\sqrt{\bar{c}n}}\left(\sum_{t=0}^{n-1}|\widehat{Z}_{t+1,j_0}|^2\right)^{1/2}\left(\sum_{t=0}^{n-1}|\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}|^2\right)^{1/2}. \tag{B.7}$$

Similar to (A.45), we can show $\sum_{t=0}^{n-1}|\widehat{Z}_{t+1,j_0}|^2 = O_p(n)$ under the given conditions in Theorem 2.1. Under (A1**), we have $\|\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\|_2 \le \|\boldsymbol{\beta}_{0,\mathcal{M}_{j_0}^{**}}\|_2$, almost surely for any $t=0,1,\dots,n-1$. This together with (A.22) and (B.4) yields that

$$\mathrm{E}\sum_{t=0}^{n-1}|\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}|^2 \le n\lambda_{\max}(\boldsymbol{\Sigma})\mathrm{E}\|\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\|_2^2 = O(n^{1-2\kappa^*}) = o(1). \tag{B.8}$$

By Markov's inequality, we obtain $\sum_{t=0}^{n-1}|\boldsymbol{X}^T_{t+1,\mathcal{M}_{j_0}^{**}}\boldsymbol{\beta}_{0,\mathcal{M}_{j_0}^{**}}|^2 = o_p(1)$. In view of (B.7), we have shown $I_5 = o_p(1)$. The proof is hence completed.

B.1.3 Additional details regarding the doubly-robust procedure

To better understand the proposed algorithm in Section 5.4, we decompose $\mathcal{M}_{j_0}$ into $\mathcal{M}_{j_0}^*$ and $\mathcal{M}_{j_0}^{**}$ as in Section B.1.2, where $\mathcal{M}_{j_0}^*$ denotes the set of strong signals that satisfies (B.3) and $\mathcal{M}_{j_0}^{**} = \mathcal{M}_{j_0} \cap (\mathcal{M}_{j_0}^*)^c$ is the set of weak signals.

In case the set $\mathcal{M}_{j_0}^{**}$ is nonempty, we can apply another model selection procedure to estimate the support of $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0}$ (denoted by $\mathcal{M}_{\boldsymbol{\omega}}$), in order to gain some robustness. For linear regression models, $\mathcal{M}_{\boldsymbol{\omega}}$ can be estimated by (I)SIS or regularized regression, with X_{i,j_0}'s being the responses and $\boldsymbol{X}_{i,\mathbb{I}_{j_0}}$'s being the covariates. Similarly, we decompose $\mathcal{M}_{\boldsymbol{\omega}}$ into $\mathcal{M}_{\boldsymbol{\omega}}^*$ and $\mathcal{M}_{\boldsymbol{\omega}}^{**}$, corresponding to the set of strong and weak signals in $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0}$, respectively.

Let $\widehat{\mathcal{M}}_{j_0}^{1,(t)}, s_n \le t < n$ and $\widehat{\mathcal{M}}_{j_0}^{1,(-s_n)}$ denote the estimated supports of $\boldsymbol{\beta}_{0,\mathbb{I}_{j_0}}$, and $\widehat{\mathcal{M}}_{j_0}^{2,(t)}, s_n \le t < n$ and $\widehat{\mathcal{M}}_{j_0}^{2,(-s_n)}$ the estimated supports of $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0}$. We will assume the following occurs with probability tending to 1,

$$\mathcal{M}_{j_0}^* \subseteq \left\{ \bigcap_{t=s_n}^{n} \widehat{\mathcal{M}}_{j_0}^{1,(t)} \right\} \bigcap \widehat{\mathcal{M}}_{j_0}^{1,(-s_n)} \quad \text{and} \quad \mathcal{M}_{\boldsymbol{\omega}}^* \subseteq \left\{ \bigcap_{t=s_n}^{n} \widehat{\mathcal{M}}_{j_0}^{2,(t)} \right\} \bigcap \widehat{\mathcal{M}}_{j_0}^{2,(-s_n)}.$$

Set $\widehat{\mathcal{M}}_{j_0}^{(t)} = \widehat{\mathcal{M}}_{j_0}^{1,(t)} \cup \widehat{\mathcal{M}}_{j_0}^{2,(t)}$, for $s_n \le t < n$ and $\widehat{\mathcal{M}}_{j_0}^{(-s_n)} = \widehat{\mathcal{M}}_{j_0}^{1,(-s_n)} \cup \widehat{\mathcal{M}}_{j_0}^{2,(-s_n)}$. We propose to use the union of these two sets in our algorithm to construct the CI for β_{0,j_0}. The number of elements in $\widehat{\mathcal{M}}_{j_0}^{(-s_n)}, \widehat{\mathcal{M}}_{j_0}^{(s_n)}, \cdots, \widehat{\mathcal{M}}_{j_0}^{(n-1)}$ shall be bounded by κ_n, almost surely. We require $\eta_n \sqrt{\kappa_n \log p} = o(1)$, $\kappa_n^2 \log p = O(n/\log^2 n)$ and $\kappa_n^2 \log^2 p = O(n)$. In the following, we focus on linear regression models and show the resulting CI for β_{0,j_0} is valid as long as either one of the following two conditions holds:

(i) $\mathcal{M}_{j_0}^{**} = \emptyset$.

(ii) $\|\boldsymbol{\beta}_{0,\mathcal{M}_{j_0}^{**}}\|_2 = o(n^{-1/4})$ and $\|\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}_{j_0}^{**}}\|_2 = o(n^{-1/4})$, where $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}_{j_0}^{**}}$ is the sub-vector of $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0}$ formed by elements in $\mathcal{M}_{j_0}^{**}$.

When (i) holds, the assertion can be proven in similar manner as Theorem 2.1. Consider the case where (ii) holds. Using similar arguments in Section B.1.2, it suffices to show

$$\sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0} \boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o_p(1). \tag{B.9}$$

We decompose the LHS of (B.9) into $I_6 + I_7 + I_8$ where

$$
\begin{aligned}
I_6 &= \sum_{t=0}^{n-1} \left(\frac{\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} - \frac{\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right), \\
I_7 &= \sum_{t=0}^{n-1} \frac{\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} (\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}) \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}, \\
I_8 &= \sum_{t=0}^{n-1} \frac{Z_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}.
\end{aligned}
$$

Under the event defined in (A.6) and (A.12), we have almost surely that

$$
|I_6| \leq \frac{2\bar{c}_0}{\bar{c}\sqrt{n}} \left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} \right) \sum_{t=0}^{n-1} |\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}|.
$$

Using similar arguments in bounding I_5 in Section B.1.2, we can show

$$
\sum_{t=0}^{n-1} |\widehat{Z}_{t+1,j_0} \boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}| = o_p(n^{3/4}).
$$

Under the condition $\kappa_n^2 \log^2 p = O(n)$, it follows that $I_6 = o_p(1)$.

Using similar arguments in bounding $I_2^{(2)}$ in the proof of Theorem 2.1, we have

$$
\sum_{t=0}^{n-1} |\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}} (\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0})|^2 = O_p(\kappa_n \log p).
$$

In addition, similar to (B.8), we can show

$$
\sum_{t=0}^{n-1} |\boldsymbol{X}^T_{t+1,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c} \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}^{(t)}_{j_0})^c}|^2 = o_p(n^{-1/2}). \tag{B.10}
$$

By (A.1) and Cauchy-Schwarz inequality, we obtain $I_7 = o_p(n^{-1/4}\sqrt{\kappa_n \log p}) = o_p(1)$, under the condition that $\kappa_n^2 \log^2 p = O(n)$.

It remains to show $I_8 = o_p(1)$. Since $\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}^T_{\mathbb{I}_{j_0},j_0} \boldsymbol{X}_{0,\mathbb{I}_{j_0}}) \boldsymbol{X}_{0,\mathbb{I}_{j_0}} = 0$, we have for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$ that $\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}^T_{\mathbb{I}_{j_0},j_0} \boldsymbol{X}_{0,\mathbb{I}_{j_0}}) \boldsymbol{X}_{0,\mathcal{M}} = 0$. Thus, for any $\mathcal{M}$ that contains $\mathcal{M}_{\boldsymbol{\omega}}$, we have

$$\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}}^T \boldsymbol{X}_{0,\mathcal{M}})\boldsymbol{X}_{0,\mathcal{M}} = 0,$$

where $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}}$ is the sub-vector of $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0}$ formed by elements in $\mathcal{M}$. This further implies $\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}} = \boldsymbol{\omega}_{\mathcal{M},j_0}$, for any $\mathcal{M}$ that contains $\mathcal{M}_{\boldsymbol{\omega}}$, and hence

$$\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T \boldsymbol{X}_{0,\mathcal{M}})\boldsymbol{X}_{0,\mathbb{I}_{j_0}} = 0.$$

For an arbitrary set $\mathcal{M}^*$ that contains $\mathcal{M}_{\boldsymbol{\omega}}^*$, define $\mathcal{M}^{**} = \mathcal{M}^* \cup \mathcal{M}_{\boldsymbol{\omega}}^{**}$. It follows that

$$\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}}^T \boldsymbol{X}_{0,\mathcal{M}^{**}})\boldsymbol{X}_{0,\mathbb{I}_{j_0}} = 0, \tag{B.11}$$

and hence $\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^*} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}-\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^{**}-\mathcal{M}^*})\boldsymbol{X}_{0,\mathcal{M}^*} = 0$. By (A.22), (ii) and Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}^*|},\|\boldsymbol{a}\|_2=1} \mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}^*}||\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}-\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^{**}-\mathcal{M}^*}| \\
\le &\left(\sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}^*|},\|\boldsymbol{a}\|_2=1} \mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}^*}|^2\right)^{1/2} \left(\mathrm{E}|\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}-\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^{**}-\mathcal{M}^*}|^2\right)^{1/2} \\
&\le \lambda_{\max}(\boldsymbol{\Sigma})\|\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}-\mathcal{M}^*}\|_2 \le \lambda_{\max}(\boldsymbol{\Sigma})\|\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}_{\boldsymbol{\omega}}^{**}}\|_2 = o(n^{-1/4}).
\end{aligned}$$

This yields $\|\mathrm{E}\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^{**}-\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^{**}-\mathcal{M}^*}\boldsymbol{X}_{0,\mathcal{M}^*}\|_2 = o(n^{-1/4})$ and hence

$$\|\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^*})\boldsymbol{X}_{0,\mathcal{M}^*}\|_2 = o(n^{-1/4}).$$

Notice that $\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M}^*,j_0}^T \boldsymbol{X}_{0,\mathcal{M}^*})\boldsymbol{X}_{0,\mathcal{M}^*} = 0$. For any $\mathcal{M}^*$ that satisfies $|\mathcal{M}^*| \le \kappa_n$, it follows from Condition (A2) that

$$\begin{aligned}
\|\boldsymbol{\omega}_{\mathcal{M}^*,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*}\|_2 &\le \frac{\|\mathrm{E}(\boldsymbol{\omega}_{\mathcal{M}^*,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*})^T \boldsymbol{X}_{0,\mathcal{M}^*(\boldsymbol{X}_{0,\mathcal{M}^*})^T}\|_2}{\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M}^*,\mathcal{M}^*})} \\
&= \frac{\|\mathrm{E}(X_{0,j_0} - \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*}^T \boldsymbol{X}_{0,\mathcal{M}^*})\boldsymbol{X}_{0,\mathcal{M}^*}\|_2}{\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M}^*,\mathcal{M}^*})} = o(n^{-1/4}).
\end{aligned}$$

To summarize, we have shown that

$$\max_{\substack{\mathcal{M}^*\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}^*|\le\kappa_n,\mathcal{M}^*_{\boldsymbol{\omega}}\subseteq\mathcal{M}^*}} \|\boldsymbol{\omega}_{\mathcal{M}^*,j_0}-\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\mathcal{M}^*}\|_2 = o(n^{-1/4}).$$

Under the given conditions, we obtain

$$\max_{t\in\{0,1,\dots,n-1\}} \|\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{(t)},j_0}-\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}}\|_2 = o(n^{-1/4}),$$

almost surely. By (A.22), this yields

$$\sum_{t=0}^{n-1}\mathrm{E}|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}}^T(\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{(t)},j_0}-\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}})|^2 = o(\sqrt{n}).$$

Similarly, we can show

$$\sum_{t=0}^{n-1}\mathrm{E}|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}}-\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}|^2 = o(\sqrt{n}).$$

This together with (A.1), (B.10) and Cauchy-Schwarz inequality yields that

$$\sum_{t=0}^{n-1}\frac{(\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}}-\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}})\boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o_p(1).$$

Thus, to show $I_8 = o_p(1)$, it suffices to show

$$\sum_{t=0}^{n-1}\frac{(X_{t+1,j_0}-\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}})\boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o_p(1).$$

We first show

$$\sum_{t=s_n}^{n-1}\frac{(X_{t+1,j_0}-\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}})\boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o_p(1). \tag{B.12}$$

By (B.11), the LHS of (B.12) forms a mean zero martingale with respect to the filtration $\{\sigma(\mathcal{F}_t) : t \ge s_n\}$. Moreover, it follows from (ii) that $\|\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}\|_2 = o(1)$ and hence

$$\frac{1}{n}\sum_{t=s_n}^{n-1}\frac{\mathrm{E}\{(X_{t+1,j_0}-\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}}^T\boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}^{**}_{\boldsymbol{\omega}}})^2(\boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T\boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c})^2|\mathcal{F}_t\}}{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o(1).$$

This proves (B.12). Similarly, we can show

$$\sum_{t=0}^{s_n-1} \frac{(X_{t+1,j_0} - \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}_{\boldsymbol{\omega}}^{**}}^T \boldsymbol{\omega}_{\mathbb{I}_{j_0},j_0,\widehat{\mathcal{M}}_{(t)}\cup\mathcal{M}_{\boldsymbol{\omega}}^{**}}) \boldsymbol{X}_{t+1,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}^T \boldsymbol{\beta}_{0,(\widehat{\mathcal{M}}_{j_0}^{(t)})^c}}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} = o_p(1).$$

The proof is hence completed.

B.2 More on (A2) and (A2*)

Condition (A2) requires $\lambda_{\min}(\boldsymbol{\Sigma}_{j_0\cup\mathcal{M},j_0\cup\mathcal{M}}) \geq \bar{c}$ for some constant $\bar{c} > 0$ and any $\mathcal{M} \subseteq \mathbb{I}$ and $|\mathcal{M}| \leq \kappa_n$, where $\boldsymbol{\Sigma} = \mathrm{E}\boldsymbol{X}_0\boldsymbol{X}_0^T$. This condition is similar to the restricted eigenvalue condition (Bickel et al., 2009) used to derive the oracle inequalities of the Lasso estimator and the Dantzig selector. Notice that this condition is weaker compared to the one used in van de Geer et al. (2014) or Ning and Liu (2017), which requires the minimum eigenvalue of $\boldsymbol{\Sigma}$ to be strictly positive. See Section 4.1 of Ning and Liu (2017), Condition (A2) and (B3) in van de Geer et al. (2014) for details.

B.3 More on (A5)

In this section, we provide a consistent estimator for σ_0^2. Specifically, define

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}})^2.$$

In the following, we show $|\hat{\sigma}^2 - \sigma_0^2| = o_p(1)$. Notice that

$$\begin{aligned}
|\hat{\sigma}^2 - \sigma_0^2| &= \left|\frac{1}{n}\sum_{i=1}^{n}(Y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}_0 + \boldsymbol{X}_i^T\boldsymbol{\beta}_0 - \boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}})^2 - \sigma_0^2\right| \\
= &\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 + \frac{2}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{X}_i^T(\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}) + \frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}^2 - \sigma_0^2\right| \\
\leq &\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma_0^2\right| + \left|\frac{2}{n}\sum_{i=1}^{n}\varepsilon_i\boldsymbol{X}_i^T(\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}})\right| + \frac{1}{n}\sum_{i=1}^{n}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\}^2. \qquad \text{(B.13)}
\end{aligned}$$

Under the condition $\mathrm{E}|\varepsilon_0|^3 = O(1)$, the first term on the RHS of (B.13) is $o_p(1)$ by the law of large numbers.

Suppose we can show

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{X}_i \right\|_\infty = O_p(\sqrt{\log p}). \tag{B.14}$$

It follows from Condition (A4) and (A.25) that the second term on the RHS of (B.13) is $o_p(1)$, since

$$\left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{X}_i^T (\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}) \right| \le \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{X}_i \right\|_\infty \|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\|_1 = O_p(\eta_n \sqrt{\kappa_n \log p}) = o_p(1).$$

The third term is $O_p(\eta_n^2)$ by (A.9). Under the given conditions, it is $o_p(1)$.

Therefore, to complete the proof, it suffices to show (B.14), or equivalently,

$$\mathrm{E} \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{X}_i \right\|_\infty = O(\sqrt{\log p}), \tag{B.15}$$

by Markov's inequality. It follows from Lemma A.3 in Chernozhukov et al. (2013) that

$$\mathrm{E} \left\| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i \boldsymbol{X}_i \right\|_\infty \le O(1)(\sigma \sqrt{\log p} + \mathbb{M} \log p),$$

where $O(1)$ denotes some positive constant, $\sigma^2 = \max_{j \in \{1,\dots,p\}} \sum_{i=1}^n \mathrm{E}\varepsilon_i^2 X_{i,j}^2 / n^2$ and $\mathbb{M}^2 = \mathrm{E} \max_{1 \le j \le p} \max_{1 \le i \le n} X_{i,j}^2 \varepsilon_i^2 / n^2$. Notice that

$$\sigma^2 = \max_{1 \le j \le p} n^{-1} \mathrm{E}\varepsilon_0^2 X_{0,j}^2 = n^{-1} \sigma_0^2 \max_{1 \le j \le p} \mathrm{E} X_{0,j}^2 \le n^{-1} \sigma_0^2 \max_{1 \le j \le p} \|X_{0,j}\|_{\psi_2}^2 \le n^{-1} \sigma_0^2 c_0^2,$$

where the last equality is due to the independence between ε_0 and $\boldsymbol{X}_0$, the first inequality is due to the fact that $\mathrm{E}|Z|^2 \le \|Z^2\|_{\psi_1} = \|Z\|_{\psi_2}^2$ for any random variable Z and the last inequality is due to (A3).

Similarly, we can show

$$\begin{aligned} n^2 \mathbb{M}^2 = \mathrm{E} \max_{1 \le i \le n} \varepsilon_i^2 \mathrm{E} \max_{\substack{1 \le i \le n \\ 1 \le j \le p}} X_{i,j}^2 &\le \left(\mathrm{E} \max_{1 \le i \le n} |\varepsilon_i|^3 \right)^{2/3} \mathrm{E} \max_{\substack{1 \le i \le n \\ 1 \le j \le p}} X_{i,j}^2 \\ &\le \left(\mathrm{E} \sum_{1 \le i \le n} |\varepsilon_i|^3 \right)^{2/3} \mathrm{E} \max_{\substack{1 \le i \le n \\ 1 \le j \le p}} X_{i,j}^2 \le (\mathrm{E}|\varepsilon_0|^3)^{2/3} n^{2/3} \mathrm{E} \max_{\substack{1 \le i \le n \\ 1 \le j \le p}} X_{i,j}^2, \end{aligned}$$

where the first equality is due to the independence between ε_0 and $\boldsymbol{X}_0$ and the first inequality follows from Hölder's inequality. Using similar arguments in (A.60), (A.61) and (A.74), we can show

$$\begin{aligned}\mathrm{E}\max_{\substack{1\le i\le n\\1\le j\le p}} X_{i,j}^2 \le \|\max_{\substack{1\le i\le n\\1\le j\le p}} X_{i,j}^2\|_{\psi_1} = K_1\{\log(1+pn)\}\max_{\substack{1\le i\le n\\1\le j\le p}}\|X_{i,j}^2\|_{\psi_1}\\ = K_1\{\log(1+pn)\}\max_{\substack{1\le i\le n\\1\le j\le p}}\|X_{i,j}\|_{\psi_2}^2 = O(\log p+\log n),\end{aligned}$$

by (A3). Under the condition $\mathrm{E}|\varepsilon_0|^3 = O(1)$, it follows that $\mathbb{M}^2 = O(n^{-4/3}\log p + n^{-4/3}\log n)$. Therefore, we obtain

$$\mathrm{E}\left\|\frac{1}{n}\sum_{i=1}^n \varepsilon_i \boldsymbol{X}_i\right\|_\infty = O(n^{-1/2}\log^{1/2}p) + O(n^{-2/3}\log^{3/2}p) + O(n^{-2/3}\log p\sqrt{\log n}).$$

Under the condition that $\log p = O(n^{2/3})$, we have $\mathrm{E}\|\sum_{i=1}^n \varepsilon_i\boldsymbol{X}_i/n\|_\infty = O(\sqrt{\log p})$. This proves (B.15). The proof is hence completed.

C Additional details regarding extensions to generic M-estimators

It this section, we sketch a few lines to show that the CI proposed in Section 5.3 is valid. It suffices to show that

$$\sqrt{n}\Gamma_n^{*,(l-1)}(\hat\beta_{j_0}^{(l)} - \beta_{0,j_0}) \xrightarrow{d} N(0,1).$$

It follows from Taylor's theorem that

$$\begin{aligned}&\sum_{t=0}^{n-1}\frac{1}{n\hat\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\left(\frac{\partial\ell(\boldsymbol{U}_{t+1},\mathrm{h}(\beta_{0,j_0},\widehat{\mathcal{M}}_{j_0}^{(t)},\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial\beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T\frac{\partial\ell(\boldsymbol{U}_{t+1},\mathrm{h}(\beta_{0,j_0},\widehat{\mathcal{M}}_{j_0}^{(t)},\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial\boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}}\right)\\ =\;&\sum_{t=0}^{n-1}\frac{1}{n\hat\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\left(\frac{\partial\ell(\boldsymbol{U}_{t+1},\mathrm{h}(\hat\beta_{j_0}^{(l-1)},\widehat{\mathcal{M}}_{j_0}^{(t)},\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial\beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T\frac{\partial\ell(\boldsymbol{U}_{t+1},\mathrm{h}(\hat\beta_{j_0}^{(l-1)},\widehat{\mathcal{M}}_{j_0}^{(t)},\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial\boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}}\right)\\ +\;&\Gamma_n^{*,(l-1)}(\beta_{0,j_0} - \hat\beta_{j_0}^{l-1}) + \mathrm{Rem},\end{aligned}$$

where the second-order remainder term satisfies $\text{Rem} = o_p(n^{-1/2})$ under certain local smoothness assumption on the loss function ℓ.

By the definition of $\hat{\beta}_{j_0}^{(l)}$, we obtain that

$$\begin{aligned}
&\sqrt{n}\Gamma_n^{*,(l-1)}(\hat{\beta}_{j_0}^{(l)} - \beta_{0,j_0}) = o_p(1) \\
+ & \sum_{t=0}^{n-1} \frac{1}{n\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right).
\end{aligned}$$

It suffices to show

$$\sum_{t=0}^{n-1} \frac{1}{n\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) \xrightarrow{d} N(0,1).$$

Under certain local smoothness assumptions on ℓ, it follows from Taylor's theorem that

$$\begin{aligned}
& \sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) \\
= & \sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) \\
+ & \sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial^2 \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0} \partial \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial^2 \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \partial \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T} \right) \\
\times & (\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}) + o_p(n^{-1/2}).
\end{aligned}$$

Suppose the model selection procedure satisfies the sure screening property. Then we have

$\mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}) = \boldsymbol{\beta}_0$, $\forall t$ and hence

$$
\begin{aligned}
&\sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) \\
+ & \underbrace{\sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right)}_{\zeta_1} \\
+ & \underbrace{\sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial^2 \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \beta_{j_0} \partial \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial^2 \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \partial \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T} \right) (\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}})}_{\zeta_2} + o_p(n^{-1/2}).
\end{aligned}
$$

Using similar arguments in the proof of Theorem 3.1, we can show that $\zeta_2 = o_p(1)$ when $\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ and $\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}$ satisfy certain uniform convergence rates, and

$$
\zeta_1 = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \beta_{j_0}} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \boldsymbol{\beta}_0)}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) + o_p(1).
$$

The first term on the RHS of the above expression is asymptotically normal under certain regularity conditions, according to the martingale central limit theorem. Thus, we obtain

$$
\sum_{t=0}^{n-1} \frac{1}{\sqrt{n}\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \left(\frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \beta_{j_0}} - \widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t_0)},j_0}^T \frac{\partial \ell(\boldsymbol{U}_{t+1}, \mathrm{h}(\beta_{0,j_0}, \widehat{\mathcal{M}}_{j_0}^{(t)}, \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}))}{\partial \boldsymbol{\beta}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}} \right) \xrightarrow{d} N(0,1),
$$

by Slutsky's theorem.

References

Adamczak, R. a. (2008). A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab. 13*, no. 34, 1000–1034.

Bühlmann, P. and J. Mandozzi (2014). High-dimensional variable screening and bias in subsequent inference, with an empirical comparison. *Comput. Statist. 29*(3-4), 407–430.

Chernozhukov, V., D. Chetverikov, and K. Kato (2013). Testing many moment inequalities. *arXiv preprint arXiv:1312.7614*.

de la Peña, V. H., T. L. Lai, and Q.-M. Shao (2009). *Self-normalized processes*. Probability and its Applications (New York). Springer-Verlag, Berlin. Limit theory and statistical applications.

Klartag, B. and S. Mendelson (2005). Empirical processes and random projections. *J. Funct. Anal. 225*(1), 229–245.

Mendelson, S., A. Pajor, and N. Tomczak-Jaegermann (2008). Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx. 28*(3), 277–289.

Ning, Y. and H. Liu (2017). A general theory of hypothesis tests and confidence regions for sparse high dimensional models. *Ann. Statist. 45*(1), 158–195.

Shi, C., A. Fan, R. Song, and W. Lu (2018). High-dimensional A-learning for optimal dynamic treatment regimes. *Ann. Statist. 46*(3), 925–957.

Talagrand, M. (2005). *The generic chaining*. Springer Monographs in Mathematics. Springer-Verlag, Berlin. Upper and lower bounds of stochastic processes.

van de Geer, S., P. Bühlmann, Y. Ritov, and R. Dezeure (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *Ann. Statist. 42*(3), 1166–1202.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York: Springer.