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

Chengchun Shi, Rui Song and Wenbin Lu
Department of Statistics, North Carolina State University
and
Runze Li
Department of Statistics, Pennsylvania State University

In this supplementary article, we present proofs of Theorem 3.1, Theorem 3.2, Theorem 3.3, Lemma A.1, Lemma A.2, Lemma A.3 and Lemma B.1.

B Proofs

B.1 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 $t = 0, \dots, s_n - 1$. 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 $\mathcal{M} \subseteq \mathbb{I}_{j_0}$, define

$$
\begin{aligned}
&\boldsymbol{\omega}_{\mathcal{M},j_0} = \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1}\boldsymbol{\Sigma}_{\mathcal{M},j_0}, \qquad \sigma^2_{\mathcal{M},j_0} = \boldsymbol{\Sigma}_{j_0,j_0} - \boldsymbol{\omega}^T_{\mathcal{M},j_0}\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}^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},\\
&\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}^{*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},\\
&\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}}^{(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}),\\
&\widehat{Z}^*_{t+1,j_0} = X_{t+1,j_0} - \widehat{\boldsymbol{\omega}}^{*T}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}, \quad \widetilde{Z}_{t+1,j_0} = X_{t+1,j_0} - \widetilde{\boldsymbol{\omega}}^{T}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}},\\
&\hat{\xi}^{(j)}_{\mathcal{M},j_0} = \widehat{\boldsymbol{\Psi}}^{(j)}_{j_0,j_0} - \widehat{\boldsymbol{\omega}}^{*T}_{\mathcal{M},j_0}\left(2\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0} + \widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}\right),\\
&\tilde{\sigma}^2_{\mathcal{M},j_0} = \hat{\sigma}^{*2}_{\mathcal{M},j_0} + \sum_{j=1}^{p}\hat{\xi}^{(j)}_{\mathcal{M},j_0}(\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 B.1 is given in Section B.7 of the supplementary material.

Lemma B.1 *Under conditions in Theorem 3.1, we have*

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

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

$$
\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p} + \kappa_n}{\sqrt{n}} + \eta_n\right), \tag{B.2}
$$

$$
\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}|\le\kappa_n}} |\hat{\sigma}_{\mathcal{M},j_0} - \sigma_{\mathcal{M},j_0}| \le \bar{c}_0\left(\frac{\sqrt{\kappa_n\log p} + \kappa_n}{\sqrt{n}} + \eta_n\right), \tag{B.3}
$$

$$
\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 \le \bar{c}_0\eta_n^2, \qquad \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}|\le\kappa_n}} \left|\hat{\sigma}^2_{\mathcal{M},j_0} - \tilde{\sigma}^2_{\mathcal{M},j_0}\right| \le \bar{c}_0\eta_n^2, \tag{B.4}
$$

$$
\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\ |\mathcal{M}|\le\kappa_n}} \left|\hat{\sigma}^2_{\mathcal{M},j_0} - \hat{\sigma}^{*2}_{\mathcal{M},j_0}\right| \le \bar{c}_0\eta_n, \tag{B.5}
$$

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

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

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

In the following, we omit the superscript and write $\hat{\beta}_{j_0}^{(1)}$, $\Gamma_n^{*,(0)}$ 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

$$\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0} - \widetilde{\beta}_{j_0}) = \frac{1}{\sqrt{n}} \sum_{t=0}^{n-1} \frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)}, j_0}} \widehat{Z}_{t+1,j_0} \left\{ Y_{t+1} - \mu \left(X_{t+1} \tilde{\beta}_{0,j_0} + \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} \right) \right\}. \tag{B.6}$$

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

$$\mathcal{M}_{j_0} \subseteq \widehat{\mathcal{M}}_{j_0}^{(t)}, \quad |\widehat{\mathcal{M}}_{j_0}^{(t)}| \leq \kappa_n, \quad t = 0, \ldots, n-1. \tag{B.7}$$

Besides, similar to (28) and (31), we can show

$$\max_{\substack{\mathcal{M} \subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}| \leq \kappa_n}} |\hat{\sigma}_{\mathcal{M},j_0}^{*2} - \sigma_{\mathcal{M},j_0}^2| \leq \bar{c}_0 \left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} + \frac{\kappa_n}{\sqrt{n}} \right), \tag{B.8}$$

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

$$\min_{\substack{\mathcal{M} \subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}| \leq \kappa_n}} \hat{\sigma}_{\mathcal{M},j_0} \geq \sqrt{\bar{c}}/2 \quad \text{and} \quad \min_{\substack{\mathcal{M} \subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}| \leq \kappa_n}} \hat{\sigma}_{\mathcal{M},j_0}^* \geq \sqrt{\bar{c}}/2, \tag{B.9}$$

with probability tending to 1.

Under the events defined in (B.7), we have for $t = 0, 1, \ldots, n-1$,

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

Hence, using a second order Taylor expansion, we have

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

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 R_t^* be the second order Remainder term. Under the events defined in (B.7), we have

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

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

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

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

$$\max_{t=0,\dots,n-1} \left| b'''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_t^*\right)\right| \leq c_*. \tag{B.10}$$

By Condition (A4*) and (A5*), this further implies that we have with probability tending to 1,

$$\begin{aligned}
\max_{t\in\{0,\dots,n-1\}} |R_t^*| \leq \frac{c_*}{2} \max_{t\in\{0,\dots,n-1\}} \left|\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T (\boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \widetilde{\boldsymbol{\beta}}_t^*)\right|^2 \\
\leq \frac{c_* \omega_0^2}{2} \left\|\boldsymbol{\beta}_0 - \widetilde{\boldsymbol{\beta}}\right\|_1^2 \leq \frac{c_* \omega_0^2 \eta_n^2}{2}.
\end{aligned} \tag{B.11}$$

We now prove

$$\Pr\left\{\max_{t\in[0,\dots,n-1]} \left|\widehat{Z}_{t+1}\right| \leq \omega_0 \left(1 + 2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right)\right\} \to 1. \tag{B.12}$$

Note that

$$\max_{t\in[0,\dots,n-1]} \left|\widehat{Z}_{t+1,j_0}\right| \leq \max_{t\in[0,\dots,n-1]} |X_{t+1,j_0}| + \max_{t\in[0,\dots,n-1]} \left\|\widehat{\boldsymbol{\omega}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}\right\|_2 \|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\|_2.$$

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

$$\max_{t\in[0,\dots,n-1]} \left|\widehat{Z}_{t+1,j_0}\right| \leq \omega_0 + \sqrt{\kappa_n}\omega_0 \max_{\substack{\mathcal{M}\subseteq\mathbb{I} \\ |\mathcal{M}|\leq\kappa_n}} \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2. \tag{B.13}$$

By (B.1) and (B.2), we have with probability tending to 1,

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2\le 2\sqrt{c_0/\bar{c}}, \tag{B.14}$$

for sufficiently large n. Combining (B.14) together with (B.13) yields (B.12). Under the events defined in (B.7), (B.9), (B.11) and (B.12), we have

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\widehat{Z}_{t+1,j_0}\left\{Y_{t+1}-\mu\left(X_{t+1}\tilde{\beta}_{0,j_0}+\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right\}\right.\\
-&\ \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}-\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}b''\left(\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\\
\times&\ \left.\left(X_{t+1,j_0}(\beta_{0,j_0}-\tilde{\beta}_{j_0})+\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}(\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}-\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}})\right)\right|\\
\le&\ \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}|R_t^*||\widehat{Z}_{t+1,j_0}|\le\frac{\sqrt{n}c_*}{\sqrt{\bar{c}}}\omega_0^3\eta_n^2\left(1+2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right).
\end{aligned}$$

By Condition (A5*), we have $\sqrt{n\kappa_n}\eta_n^2=o(1)$. Hence, we've shown

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

In view of (B.6), we have

$$\begin{aligned}
&\sqrt{n}\Gamma_n^*(\hat{\beta}_{j_0}-\beta_{0,j_0})=o_p(1)+\underbrace{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}}_{I_1}\\
+&\ \underbrace{\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}b''\left(\boldsymbol{X}^T_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\boldsymbol{X}^T_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}}-\boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)}_{I_2}.
\end{aligned}$$

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 (B.7), using a first order Taylor expansion, we have

$$\begin{aligned} & b''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) = b''(\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0) \\ + & \underbrace{\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \left(\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}\right) b'''\left(\boldsymbol{X}_{t+1,\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}^T \widetilde{\boldsymbol{\beta}}_t^{**}\right)}_{R_t^{**}}, \end{aligned}$$

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 $\widetilde{\boldsymbol{\beta}}_{\{j_0\}\cup\widehat{\mathcal{M}}_{j_0}^{(t)}}$. Using similar arguments in (B.10) and (B.11), we can show the remainder term R_t^{**} satisfies

$$\max_{t\in[0,\dots,n-1]} |R_t^{**}| \le c_{**}\omega_0\eta_n, \tag{B.15}$$

for some constant $c_{**}>0$, with probability tending to 1. 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

$$|I_2 - I_2^*| \le \frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} \frac{|\widehat{Z}_{t+1,j_0}|}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} |R_t^{**}| \left\|\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right\|_\infty \left\|\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right\|_1.$$

By (B.9), (B.12), (B.15), Condition (A4*) and (A5*), we have with probability tending to 1,

$$|I_2 - I_2^*| \le \sqrt{n}\frac{2\omega_0}{\sqrt{\bar{c}}}\left(1+2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right) c_{**}\omega_0\eta_n^2 = o(1).$$

This implies $I_2 = I_2^* + o_p(1)$. It suffices to show $I_2^* = o_p(1)$.

Similar to the proof of Theorem 2.1, by (B.2), (B.7) and (B.9), we can show the following event occurs with probability tending to 1,

$$\begin{aligned} & \left|\frac{1}{\sqrt{n}}\sum_{t=0}^{n-1} \frac{\widehat{Z}_{t+1,j_0} - Z_{t+1,j_0}}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} b''\left(\boldsymbol{X}_{t+1}^T\boldsymbol{\beta}_0\right) \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \left(\widetilde{\boldsymbol{\beta}}_{\widehat{\mathcal{M}}_{j_0}^{(t)}} - \boldsymbol{\beta}_{0,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)\right| \qquad \text{(B.16)} \\ \le\ & c_{***}\frac{2\sqrt{n}}{\sqrt{\bar{c}}}\left(\frac{\sqrt{\kappa_n\log p}+\kappa_n}{\sqrt{n}} + \eta_n\right)\sqrt{\kappa_n}\omega_0^2\eta_n, \end{aligned}$$

for some constant $c_{***} > 0$, where $Z_{t+1,j_0} = X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}$.

Under the conditions $(\kappa_n\sqrt{\log p} + \kappa_n^{3/2})\eta_n \to 0$ and $\sqrt{n\kappa_n}\eta_n^2 \to 0$ in (A5*), (B.16) is $o_p(1)$. Similarly, we can show

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

This together with (B.16) implies $I_2^* = I_2^{**} + o_p(1)$, where

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

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

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

Similar to Lemma A.3, we can show $\max_j |I_{2,j}| = O_p(\log p)$. This together with 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 $\mathcal{M} \subseteq \mathbb{I}_{j_0}$,

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

for some $0 < \rho_{\mathcal{M}} < 1$. By (B.5) and (B.9), the second-order remainder term satisfies

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left|\frac{(\hat{\sigma}^2_{\mathcal{M},j_0} - \hat{\sigma}^{*2}_{\mathcal{M},j_0})^2}{2\{\rho_{\mathcal{M}}\hat{\sigma}_{\mathcal{M},j_0} + (1-\rho_{\mathcal{M}})\hat{\sigma}^*_{\mathcal{M},j_0}\}^5}\right| \le \frac{16\bar{c}_0^2\eta_n^2}{\bar{c}^{5/2}}, \tag{B.17}$$

with probability tending to 1.

Besides, it follows from (B.4) and (B.9) 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 (B.17) 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*) and (B.12), we have

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

$$\leq \sqrt{n}\bar{c}_1\eta_n^2 \max_t |\widehat{Z}_{t+1,j_0}|\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}| \leq \sqrt{n}\bar{c}_1\eta_n^2\omega_0\left(1 + 2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right)\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}|,$$

with probability tending to 1. By Condition (A6*) and Hölder's inequality, we have

$$\mathrm{E}\left(\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}|\right) = \mathrm{E}|\varepsilon_0| \leq (\mathrm{E}|\varepsilon_0|^3)^{1/3} = O(1).$$

Hence, it follows from Markov's inequality that $\sum_{t=0}^{n-1}|\varepsilon_{t+1}|/n = O_p(1)$. As a result, we have

$$\left|\sum_{t=0}^{n-1}\frac{\widehat{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}}\left(\frac{1}{\hat{\sigma}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} - \frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} + \frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j - \beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\right)\right| = O_p(\sqrt{n\kappa_n}\eta_n^2) = o_p(1), \tag{B.18}$$

where the last equality is due to Condition (A5*).

By (B.4), (B.8), (B.9) and Condition (A1*), (A4*), (A5*), (A6*), we can similarly show

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

This together with (B.18) yields

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

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 B.1, 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.

B.2 Proof of Theorem 3.2

Recall that $\mathbb{I} = [1, \dots, p]$ and $\mathbb{I}_{j_0} = \mathbb{I} - \{j_0\}$. By (17) and Lemma B.2, we have

$$\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{DL}, \alpha) = 2z_{\frac{\alpha}{2}}\sqrt{\phi_0 \boldsymbol{e}_{j_0,p}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_{j_0,p}} + o_p(1) = \frac{2z_{\frac{\alpha}{2}}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} + o_p(1). \tag{B.19}$$

It follows from (14) that

$$\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}, \alpha) = \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} + o_p(1). \tag{B.20}$$

With some calculations, we have

$$\begin{aligned} & \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \\ = \quad & 2z_{\alpha/2}\sqrt{\phi_0} \frac{s_n\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n}{\sigma_{\mathbb{I}_{j_0},j_0}\{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n\}}. \end{aligned} \tag{B.21}$$

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

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

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 (B.21) 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,

$$\sigma^2_{\mathcal{M},j_0} = \arg\min_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \mathrm{E}|X_{0,j_0} - \boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2 b''(\boldsymbol{X}_0^T\boldsymbol{\beta}_0) \leq \bar{k}\mathrm{E}|X_{0,j_0}^2| \leq \bar{k}\sqrt{\mathrm{E}|X_{0,j_0}^4} = \bar{k}\sqrt{c_0}. \tag{B.22}$$

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

$$\begin{aligned}
&\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \\
\geq\ & \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\bar{k}\sqrt{c_0}}\left(s_n\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma_{\mathbb{I}_{j_0},j_0}\}/n\right).
\end{aligned} \tag{B.23}$$

Besides, it follows from (B.22) that

$$\sigma_{\mathcal{M},j_0} - \sigma_{\mathbb{I}_{j_0},j_0} = \frac{\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}}{\sigma_{\mathcal{M},j_0} + \sigma_{\mathbb{I}_{j_0},j_0}} \geq \frac{\sigma^2_{\mathcal{M},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}}{2\sqrt{\bar{k}}c_0^{1/4}},$$

for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$. This together with (B.23) gives

$$\begin{aligned}
&\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \\
\geq\ & \frac{z_{\alpha/2}\sqrt{\phi_0}}{\bar{k}^{3/2}c_0^{3/4}}\left(s_n\{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}\}/n + \sum_{t=s_n}^{n-1}\{\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0} - \sigma^2_{\mathbb{I}_{j_0},j_0}\}/n\right).
\end{aligned} \tag{B.24}$$

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 B.2 that

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

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

$$\lambda_{\min}(\boldsymbol{\Omega}_{\mathcal{M},j_0}) \geq \lambda_{\min}\left\{\left(\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c}\right)^{-1}\right\} = \left\{\lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathbb{I}_{j_0}\cap\mathcal{M}^c,\mathbb{I}_{j_0}\cap\mathcal{M}^c}\right)\right\}^{-1} \geq \{\lambda_{\max}(\boldsymbol{\Sigma})\}^{-1} = \frac{1}{k_0},$$

and hence

$$\boldsymbol{\Sigma}_{\mathbb{I}_{j_0},j_0}^T \boldsymbol{\Sigma}_{\mathbb{I}_{j_0},\mathbb{I}_{j_0}}^{-1} \boldsymbol{\Sigma}_{\mathbb{I}_{j_0},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0}^T \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1} \boldsymbol{\Sigma}_{\mathcal{M},j_0} \ge \frac{1}{k_0} \|\boldsymbol{\xi}_{\mathcal{M},j_0}\|_2^2.$$

Note that we have

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

This further implies

$$\sigma_{\mathcal{M},j_0}^2 - \sigma_{\mathbb{I}_{j_0},j_0}^2 \ge \frac{1}{k_0} \|\boldsymbol{\xi}_{\mathcal{M},j_0}\|_2^2,$$

for any $\mathcal{M} \subseteq \mathbb{I}_{j_0}$. By (B.24), we have

$$\frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathbb{I}_{j_0},j_0}} - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \ge \frac{\sqrt{\phi_0} z_{\alpha/2}}{\bar{k}^{3/2} c_0^{3/4} k_0} \left(\frac{s_n}{n} \|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\|_2^2 + \frac{1}{n} \sum_{t=s_n}^{n-1} \|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2 \right).$$

In view of (B.19) and (B.20), we've shown

$$\sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{DL}, \alpha) \ge \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}, \alpha) + \frac{\sqrt{\phi_0} z_{\alpha/2}}{\bar{k}^{3/2} c_0^{3/4} k_0} \left(\frac{s_n}{n} \|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}\|_2^2 + \frac{1}{n} \sum_{t=s_n}^{n-1} \|\boldsymbol{\xi}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2^2 \right) + o_p(1).$$

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

B.3 Proof of Theorem 3.3

Under the given conditions, using similar arguments in (29), 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

$$\frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} = \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathcal{M}_{j_0},j_0}}. \tag{B.25}$$

By (14) and (21), for any sufficiently small $\varepsilon_0 > 0$, the following events occur with probability tending to 1,

$$\limsup_n \left| \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{(l)}, \alpha) - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{s_n \sigma_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}/n + \sum_{t=s_n}^{n-1} \sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}/n} \right| \le \frac{\varepsilon_0}{2}, \tag{B.26}$$

$$\limsup_n \left| \sqrt{n}\mathrm{L}(\hat{\beta}_{j_0}^{oracle}, \alpha) - \frac{2z_{\alpha/2}\sqrt{\phi_0}}{\sigma_{\mathcal{M}_{j_0},j_0}} \right| \le \frac{\varepsilon_0}{2}. \tag{B.27}$$

Conditional on the events defined in (B.25)-(B.27), 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.

B.4 Proof of Lemma A.1

We first prove (24). Condition (A2) states that

$$\min_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \inf_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|+1}\\ \|\boldsymbol{a}\|_2\ge 1}} \boldsymbol{a}^T \boldsymbol{\Sigma}_{j_0\cup\mathcal{M}, j_0\cup\mathcal{M}} \boldsymbol{a} \ge \bar{c}. \tag{B.28}$$

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 (B.28), this implies

$$\min_{\substack{\mathcal{M}\subseteq[1,\dots,p]\\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} \sigma_{\mathcal{M},j_0}^2 \ge \bar{c}, \tag{B.29}$$

and hence

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

By Cauchy-Schwarz inequality, Assumption (A3) implies that

$$\boldsymbol{\Sigma}_{j_0,j_0} = \mathrm{E}X_{0,j_0}^2 \le \sqrt{\mathrm{E}X_{0,j_0}^4} \le c_0.$$

In view of (B.29), this further implies that

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

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

$$\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2^2 \le \frac{c_0}{\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})} \le \frac{c_0}{\bar{c}},$$

where the last inequality is due to Condition (A2). Therefore, (24) is proven.

Consider (25). For any $a, b \in \mathbb{R}$, we have $(a+b)^4 \le 8a^4 + 8b^4$. Therefore,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \mathrm{E}|X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T \boldsymbol{X}_{0,\mathcal{M}}|^4 \le 8\mathrm{E}X_{0,j_0}^4 + \min_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} 8\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T \boldsymbol{X}_{0,\mathcal{M}}|^4.$$

Under (A3), we have $\max_j \mathrm{E}X_{0,j}^4 \le c_0$, $\mathrm{E}X_{0,j_0}^4 \le c_0$. By (B.46), we have $\mathrm{E}|\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|^4 \le c_0^3/\bar{c}^2$ for all $\mathcal{M}$ such that $|\mathcal{M}| \le \kappa_n$. Moreover, by (24),

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

Thus, we have

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{1}{\sigma_{\mathcal{M},j_0}^4} \mathrm{E}|X_{0,j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}^T \boldsymbol{X}_{0,\mathcal{M}}|^4 \le 8\left(c_0 + \frac{c_0^3}{\bar{c}^2}\right) \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{1}{\sigma_{\mathcal{M},j_0}^4} \le \frac{8}{\bar{c}^2}\left(c_0 + \frac{c_0^3}{\bar{c}^2}\right).$$

For any random variable Z with $\|Z\|_{\psi_2} \le \omega$, it follows from the definition of the Orlicz norm that $\|Z\|_{\psi_1} \le \omega^2$. Under Condition (A5), this implies

$$\max_j \|X_{0,j}^2\|_{\psi_1} \le \max_j (\|X_{0,j}\|_{\psi_2})^2 \le \omega_0^2. \tag{B.30}$$

For any random variable Z, we have $\mathrm{E}|Z| \le \|Z\|_{\psi_1}$. This implies

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}\\ |\mathcal{M}|\le\kappa_n}} \mathrm{E} \max_{j\in\mathcal{M}} X_{0,j}^2 \le \max_{\substack{\mathcal{M}\subseteq\mathbb{I}\\ |\mathcal{M}|\le\kappa_n}} \|\max_{j\in\mathcal{M}} X_{0,j}^2\|_{\psi_1}.$$

By (B.30) and Lemma 2.2.2 in van der Vaart and Wellner (1996), we have

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}\\ |\mathcal{M}|\le\kappa_n}} \|\max_{j\in\mathcal{M}} X_{0,j}^2\|_{\psi_1} \le \max_{\substack{\mathcal{M}\subseteq\mathbb{I}\\ |\mathcal{M}|\le\kappa_n}} K_0 \log(1+\kappa_n) \max_{j\in\mathcal{M}} \|X_{0,j}^2\|_{\psi_1} \le K_0\omega_0^2 \log(1+\kappa_n).$$

for some constant $K_0 > 0$. Note that $\kappa_n = o(n)$, we have $\log(1+\kappa_n) \le \log(n)$ for sufficiently large n. This completes the proof.

B.5 Proof of Lemma A.2

We first prove (27). Note that

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

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

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

Upper bound for η_1: Since

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

it suffices to show with probability tending to 1 that,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\|_2 = O(1), \tag{B.31}$$

and

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0})\|_2 = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right). \tag{B.32}$$

Note that $\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}$ is symmetric. To prove (B.31), it is equivalent to show that the eigenvalues of $\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}$ are uniformly bounded with probability tending to 1. Hence, it suffices to prove

$$\min_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right) > \frac{\bar{c}}{2}, \tag{B.33}$$

with probability tending to 1.

Observe that

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

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

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \left\|\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right\|_2. \tag{B.34}$$

For any symmetric matrix $\boldsymbol{S}$, we have $|\boldsymbol{S}|_2 \leq \sqrt{\|\boldsymbol{S}\|_1 \|\boldsymbol{S}\|_\infty} = \|\boldsymbol{S}\|_\infty$, where $\|\|_1$ and $\|\|_\infty$ stand for the ℓ_1 and ℓ_∞ induced matrix norms. Hence, (B.34) is upper bounded by

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \left\| \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}} \right\|_\infty \leq \kappa_n \max_{j_1,j_2\in[1,\ldots,p]} |\widehat{\boldsymbol{\Sigma}}_{j_1,j_2} - \boldsymbol{\Sigma}_{j_1,j_2}|.$$

To summarize, we've shown

$$\min_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \lambda_{\min}\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right) \geq \bar{c} - \kappa_n \max_{j_1,j_2\in[1,\ldots,p]} |\widehat{\boldsymbol{\Sigma}}_{j_1,j_2} - \boldsymbol{\Sigma}_{j_1,j_2}|. \tag{B.35}$$

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 (B.30) with Cauchy-Schwarz inequality, we have

$$\|X_{0,j_1}X_{0,j_2}\|_{\psi_1} \leq \frac{\|X_{0,j_1}^2 + X_{0,j_2}^2\|_{\psi_1}}{2} \leq \frac{\|X_{0,j_1}^2\|_{\psi_1}}{2} + \frac{\|X_{0,j_2}^2\|_{\psi_1}}{2} \leq \omega_0^2, \tag{B.36}$$

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

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

This implies $\|\mathrm{E}X_{0,j_0}X_{0,j_2}\|_{\psi_1} \leq \omega_0^2$, $\forall j_1, j_2$. Combining this together with (B.36) gives

$$\|X_{0,j_1}X_{0,j_2} - \mathrm{E}X_{0,j_0}X_{0,j_2}\|_{\psi_1} \leq \|X_{0,j_1}X_{0,j_2}\|_{\psi_1} + \|\mathrm{E}X_{0,j_1}X_{0,j_2}\|_{\psi_1} \leq 2\omega_0^2.$$

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

$$\max_{1\leq j_1,j_2\leq p} \Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2} - \boldsymbol{\Sigma}_{j_1,j_2})\right| \geq t\right) \leq 2\exp\left(-c_1 \min\left(\frac{t^2}{4n\omega_0^2}, \frac{t}{2\omega_0}\right)\right), \tag{B.37}$$

for any $t > 0$ and some constant $c_1 > 0$.

Take $t_0 = 3\sqrt{n\log p}\,\omega_0/\sqrt{c_1}$. Since $\log p = o(n)$, we have for sufficiently large n,

$$\frac{t_0^2}{4n\omega_0^2} = \frac{9\log p}{4c_1} \ll \frac{3\sqrt{n\log p}}{2\sqrt{c_1}} = \frac{t_0}{2\omega_0}.$$

It follows from (B.37) that

$$\max_{j_1,j_2} \Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2} - \boldsymbol{\Sigma}_{j_1,j_2})\right| \geq t_0\right) \leq 2\exp\left(-\frac{c_1 t_0^2}{4n\omega_0^2}\right) \leq 2\exp\left(-\frac{9\log p}{4}\right).$$

By Bonferroni's inequality, we have

$$
\begin{aligned}
& \Pr\left(\max_{j_1,j_2\in[1,\ldots,p]}\left|\sum_i (X_{i,j_1}X_{i,j_2}-\mathbf{\Sigma}_{j_1,j_2})\right|\geq t_0\right) \qquad \text{(B.38)}\\
\leq\ & \sum_{j_1,j_2\in[1,\ldots,p]}\Pr\left(\left|\sum_i (X_{i,j_1}X_{i,j_2}-\mathbf{\Sigma}_{j_1,j_2})\right|\geq t_0\right)\\
\leq\ & p^2 2\exp\left(-\frac{9\log p}{4}\right)=2\exp\left(-\frac{9\log p}{4}+2\log p\right)=2\exp\left(-\frac{\log p}{4}\right)\to 0.
\end{aligned}
$$

Under the given conditions, we have that $\kappa_n t_0/n = O(\kappa_n\sqrt{\log p/n}) = o(1)$. In view of (B.35), we've shown (B.33) holds for sufficiently large n.

Moreover, under the event defined in (B.38), 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 (B.32).

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 (B.31), it suffices to show

$$
\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}}\|(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2=O\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right)+O\left(\frac{\kappa_n}{\sqrt{n}}\right), \qquad \text{(B.39)}
$$

with probability tending to 1.

LHS of (B.39) can be upper bounded by

$$
\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\leq\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}}|\boldsymbol{a}^T(\widehat{\mathbf{\Sigma}}_{\mathcal{M},\mathcal{M}}-\mathbf{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}|. \qquad \text{(B.40)}
$$

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

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

The envelope function of |g| is bounded by

$$G_{\mathcal{M}}(\boldsymbol{X}_i) \triangleq \|\boldsymbol{X}_{i,\mathcal{M}}\|_2 |\boldsymbol{X}_{i,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2 + \|\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\|_2 \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|. \tag{B.41}$$

Note that $\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}$ is positive definite, we have

$$\|\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\|_2 = \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \boldsymbol{a}^T \boldsymbol{\Sigma} \boldsymbol{a} = \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}}|^2 \le \sqrt{\mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}}|^4} \le \sqrt{c_0}, \tag{B.42}$$

where the first inequality follows from Cauchy-Schwarz inequality and the last inequality is due to Condition (A3). Combing this together with (B.41) and (24), we have

$$G_{\mathcal{M}}(\boldsymbol{X}_i) \le \|\boldsymbol{X}_{i,\mathcal{M}}\|_2 |\boldsymbol{X}_{i,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2 + c_0/\sqrt{\bar{c}}. \tag{B.43}$$

Given $\mathcal{M}$, the class of functions $\{\boldsymbol{a}^T(x_{\mathcal{M}} x_{\mathcal{M}}^T - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}\}$ has finite VC index $|\mathcal{M}|+2 \le \kappa_n + 2 \le 3\kappa_n$, since $\kappa_n \ge 1$. Therefore, it follows from Lemma 2.14.1 in van der Vaart and Wellner (1996) that

$$\mathrm{E} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \le K_0 \sqrt{\frac{3\kappa_n}{n}} \sqrt{\mathrm{E} G_{\mathcal{M}}^2(\boldsymbol{X}_0)}. \tag{B.44}$$

By (B.43), we have

$$\begin{aligned} & \mathrm{E} G_{\mathcal{M}}^2(\boldsymbol{X}_0) \le \mathrm{E}(\|\boldsymbol{X}_{0,\mathcal{M}}\|_2 |\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2 + c_0/\sqrt{\bar{c}})^2 \qquad \text{(B.45)} \\ \le\ & 2c_0^2/\bar{c} + 2\mathrm{E}\|\boldsymbol{X}_{0,\mathcal{M}}\|_2^2 |\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2^2 \le 2c_0^2/\bar{c} + 2\sqrt{\mathrm{E}\|\boldsymbol{X}_{0,\mathcal{M}}\|_2^4} \sqrt{\mathrm{E}|\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2^4}, \end{aligned}$$

where the last two inequalities are due to Cauchy-Schwarz inequality.

Note that

$$\mathrm{E}\|\boldsymbol{X}_{0,\mathcal{M}}\|_2^4 \le \mathrm{E}|\mathcal{M}| \sum_{j\in\mathcal{M}} \boldsymbol{X}_{0,j}^4 \le \kappa_n^2 \max_{j\in[1,\dots,p]} \mathrm{E} X_{0,j}^4 \le \kappa_n^2 \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{\kappa_n} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}}|^4,$$

$$\mathrm{E}|\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2^4 \le \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2^4 \mathrm{E}\left|\boldsymbol{X}_{0,\mathcal{M}}^T \frac{\boldsymbol{\omega}_{\mathcal{M},j_0}}{\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2}\right|_2^4 \le \frac{c_0^2}{\bar{c}^2} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{\kappa_n} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T \boldsymbol{X}_{0,\mathcal{M}}|^4,$$

where the last inequality is due to (24). Under Condition (A3), we have that

$$\mathrm{E}\|\boldsymbol{X}_{0,\mathcal{M}}\|_2^4 \le \kappa_n^2 c_0 \quad \text{and} \quad \mathrm{E}|\boldsymbol{X}_{0,\mathcal{M}}^T \boldsymbol{\omega}_{\mathcal{M},j_0}|_2^4 \le \frac{c_0^3}{\bar{c}^2}. \tag{B.46}$$

This together with (B.45) gives that

$$\mathrm{E}G_{\mathcal{M}}^2(\boldsymbol{X}_0) \leq \frac{2c_0^2}{\bar{c}} + 2\sqrt{\kappa_n^2 c_0}\sqrt{\frac{c_0^3}{\bar{c}^2}} = \frac{2c_0^2}{\bar{c}} + \frac{2\kappa_n c_0^2}{\bar{c}} \leq \frac{4\kappa_n c_0^2}{\bar{c}},$$

where the last inequality is due to that $\kappa_n \geq 1$. Hence, by (B.44),

$$\mathrm{E} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \leq K_0\sqrt{\frac{3\kappa_n}{n}}\sqrt{\frac{4\kappa_n c_0^2}{\bar{c}}} \leq K_0 c_0 \kappa_n \sqrt{\frac{12}{n\bar{c}}}. \tag{B.47}$$

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

$$\begin{aligned}
&\|G_{\mathcal{M}}(\boldsymbol{X}_0)\|_{\psi_1} \leq \|c_0/\sqrt{\bar{c}}\|_{\psi_1} + \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \left\|\|\boldsymbol{X}_{0,\mathcal{M}}\|_2^2\right\|_{\psi_1} \\
\leq\ & c_0/\sqrt{\bar{c}} + \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \left\|\sum_{j\in\mathcal{M}} X_{0,j}^2\right\|_{\psi_1} \leq c_0/\sqrt{\bar{c}} + \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \kappa_n \max_j \|X_{0,j}^2\|_{\psi_1} \\
\leq\ & c_0/\sqrt{\bar{c}} + \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \kappa_n \omega_0^2 \leq c_0/\sqrt{\bar{c}}\kappa_n + \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \kappa_n \omega_0^2 \leq \sqrt{\frac{c_0}{\bar{c}}}\kappa_n(\sqrt{c_0} + \omega_0^2),
\end{aligned}$$

where the fourth inequality follows from (B.30), and the last inequality is due to (24).

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

$$\left\|\max_{i\in[1,\ldots,n]} |G_{\mathcal{M}}(\boldsymbol{X}_i)|\right\|_{\psi_1} \leq K_1 \log(1+n) \max_{i\in[1,\ldots,n]} \|G_{\mathcal{M}}(\boldsymbol{X}_i)\|_{\psi_1} = K_1\sqrt{\frac{c_0}{\bar{c}}}(\sqrt{c_0} + \omega_0^2)\kappa_n \log(1+n).$$

Let $c_2 = K_1\sqrt{c_0/\bar{c}}(\sqrt{c_0} + \omega_0^2)$, we've shown $\|\max_{i\in[1,\ldots,n]} |G_{\mathcal{M}}(\boldsymbol{X}_i)|\|_{\psi_1} \leq c_2\kappa_n \log(1+n)$. Here, the constant c_2 is independent of $\mathcal{M}$.

Moreover, it follows from Cauchy-Schwarz inequality that

$$\begin{aligned}
&\sigma_*^2 \equiv \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}g_{\mathcal{M}}(\boldsymbol{X}_0, \boldsymbol{a})^2 \\
\leq\ & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T(\boldsymbol{X}_{0,\mathcal{M}}\boldsymbol{X}_{0,\mathcal{M}}^T - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}})\boldsymbol{\omega}_{\mathcal{M},j_0}|^2 \\
\leq\ & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}\boldsymbol{X}_{0,\mathcal{M}}^T\boldsymbol{\omega}_{\mathcal{M},j_0}|^2 \\
\leq\ & \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2=1}} \sqrt{\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4} \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\leq\kappa_n}} \sqrt{\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4} \leq \frac{c_0^2}{\bar{c}},
\end{aligned}$$

where the last inequality follows by (B.46) and Condition (A3).

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

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \Pr\left(\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \ge \frac{t}{n}\right)\\
\le\ & \exp\left(-\frac{t^2}{3n\sigma_*^2}\right) + 3\exp\left(-\frac{t}{K_2 c_2\kappa_n\log(1+n)}\right)\\
\le\ & \exp\left(-\frac{t^2\bar{c}}{3nc_0^2}\right) + 3\exp\left(-\frac{t}{K_2 c_2\kappa_n\log(1+n)}\right), \qquad \forall t>0.
\end{aligned}$$

Define

$$t_0 = \max\left(\frac{2c_0\sqrt{n\kappa_n\log p}}{\sqrt{\bar{c}}}, \frac{4}{3}K_2c_2\kappa_n^2\log p\log(n+1)\right),$$

we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \Pr\left(\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| - \frac{3}{2}\mathrm{E}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \ge \frac{t_0}{n}\right)\\
\le\ & \exp\left(-\frac{4\bar{c}c_0^2n\kappa_n\log p}{3n\bar{c}c_0^2}\right) + 3\exp\left(-\frac{4K_2c_2\kappa_n^2\log p\log(n+1)}{3K_2c_2\kappa_n\log(1+n)}\right) \le 4\exp\left(-\frac{4}{3}\kappa_n\log p\right).
\end{aligned}$$

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 (B.44) implies that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\ \|\boldsymbol{a}\|_2=1}} |T_{\mathcal{M}}(\boldsymbol{a})| \le K_0c_0\kappa_n\sqrt{\frac{27}{n\bar{c}}} + \frac{t_0}{n}, \tag{B.48}$$

with probability tending to 1.

Under the given conditions, we have that for sufficiently large n,

$$\frac{2c_0\sqrt{n\kappa_n \log p}}{\sqrt{\bar{c}}} \gg \frac{4}{3}K_2 c_2 \kappa_n^2 \log p \log(n+1),$$

and hence $t_0 = 2c_0\sqrt{n\kappa_n \log p}/\sqrt{\bar{c}}$. Under the event defined in (B.48), we have for sufficiently large n,

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

This proves (B.39). The upper bound for η_2 is thus given.

Consider (28). Assume for now, we've shown

$$\max_{\substack{\mathcal{M}\subseteq[1,\dots,p] \\ j_0\notin\mathcal{M}, |\mathcal{M}|\le\kappa_n}} |\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}}\right) + O\left(\frac{\kappa_n}{\sqrt{n}}\right), \tag{B.49}$$

with probability tending to 1. Then, under the event defined in (B.49), we have

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

where the last inequality follows from (24) and the last equality is due to (B.49). Hence, it suffices to show (B.49).

By definition, we have

$$\begin{aligned}
& |\hat{\sigma}^2_{\mathcal{M},j_0} - \sigma^2_{\mathcal{M},j_0}| \le |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| + |\widehat{\boldsymbol{\Sigma}}^T_{\mathcal{M},j_0}\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}^T_{\mathcal{M},j_0}\boldsymbol{\omega}_{\mathcal{M},j_0}| \qquad \text{(B.50)} \\
\le\ & |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| + |\boldsymbol{\Sigma}^T_{\mathcal{M},j_0}(\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0})| + |(\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0})^T\boldsymbol{\omega}_{\mathcal{M},j_0}| \\
+\ & |(\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0})^T(\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0})| \le |\widehat{\boldsymbol{\Sigma}}_{j_0,j_0} - \boldsymbol{\Sigma}_{j_0,j_0}| \\
+\ & \|\boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2\|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 + \|\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2\|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 \\
+\ & \|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0} - \boldsymbol{\omega}_{\mathcal{M},j_0}\|_2\|\boldsymbol{\Sigma}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2.
\end{aligned}$$

It follows from (24), (27), (B.32) and (B.38) that with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\Sigma}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right), \tag{B.51}$$

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0}-\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2 = O\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}\right)+O\left(\frac{\kappa_n}{\sqrt{n}}\right), \tag{B.52}$$

$$|\widehat{\boldsymbol{\Sigma}}_{j_0,j_0}-\boldsymbol{\Sigma}_{j_0,j_0}| = O\left(\frac{\sqrt{\log p}}{\sqrt{n}}\right), \quad \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\omega}_{\mathcal{M},j_0}\|_2 = O(1). \tag{B.53}$$

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 (B.42) 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 (24), we can show that

$$\max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \|\boldsymbol{\Sigma}_{\mathcal{M},j_0}\|_2^2 \le \max_{\substack{\mathcal{M}\subseteq[1,\ldots,p]\\ j_0\notin\mathcal{M},|\mathcal{M}|\le\kappa_n}} \frac{c_0}{\lambda_{\min}(\boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}^{-1})} \le c_0^{3/2}.$$

Under the given conditions, we have $\kappa_n\log p = o(n)$. Under the events defined in (B.50)-(B.52), we obtain that

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

This proves (B.49). The proof is hence completed.

B.6 Proof of Lemma A.3

We first prove

$$\max_{i\in[1,\ldots,n]}\max_{j\in[1,\ldots,p]} |X_{i,j}| \le \sqrt{2\omega_0^2\log p}. \tag{B.54}$$

Note that

$$\max_{i,j}\Pr\left(|X_{i,j}|>\sqrt{2\omega_0^2\log p}\right) \le \frac{\mathrm{E}\exp\left(|X_{i,j}|^2/\omega_0^2\right)}{\exp\left(2\omega_0^2\log p/\omega_0^2\right)} \le \frac{2}{p^2},$$

where the first inequality follows from Markov's inequality and the second inequality is due to the definition of the Orlicz norm. Since $p \gg n$, it follows from Bonferroni's inequality that

$$\Pr\left(\max_{i,j}|X_{i,j}| > \sqrt{2\omega_0^2 \log p}\right) \leq np \max_{i,j} \Pr\left(|X_{i,j}| > \sqrt{2\omega_0^2 \log p}\right) \leq \frac{2n}{p} \to 0.$$

This proves (B.54).

Under Condition (A1), we have for any $t \in [s_n, \dots, n-1]$,

$$\begin{aligned}
& \mathrm{E}\left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)^2 X^2_{t+1,j} | \mathcal{F}_t\right\} \\
\leq & \sqrt{\mathrm{E}\left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)^4 | \mathcal{F}_t\right\}} \sqrt{\mathrm{E}\left(X^4_{t+1,j} | \mathcal{F}_t\right)} \\
\leq & \max_{j_0 \notin \mathcal{M}, |\mathcal{M}| \leq \kappa_n} \sqrt{\mathrm{E}\left(X_{0,j_0} - \boldsymbol{\omega}^T_{\mathcal{M},j_0} \boldsymbol{X}_{0,\mathcal{M}}\right)^4} \max_{j \in [1,\dots,p]} \sqrt{\mathrm{E}X^4_{0,j}} \\
\leq & \max_{j_0 \notin \mathcal{M}, |\mathcal{M}| \leq \kappa_n} \sqrt{8\mathrm{E}X^4_{0,j_0} + 8\mathrm{E}|\boldsymbol{\omega}^T_{\mathcal{M},j_0} \boldsymbol{X}_{0,\mathcal{M}}|^4} \max_{j \in [1,\dots,p]} \sqrt{\mathrm{E}X^4_{0,j}}
\end{aligned}$$

where the first inequality follows from Cauchy-Schwarz inequality, and the last inequality is due to that the elementary inequality that $(a+b)^4 \leq 8a^4 + 8b^4, \forall a, b \in \mathbb{R}$.

Under (A3), we have $\max_j \mathrm{E}X^4_{0,j} \leq c_0$, $\mathrm{E}X^4_{0,j_0} \leq c_0$. By (B.46), we have $\mathrm{E}|\boldsymbol{X}^T_{0,\mathcal{M}} \boldsymbol{\omega}_{\mathcal{M},j_0}|^4 \leq c_0^3/\bar{c}^2$. Thus,

$$\mathrm{E}\left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)^2 X^2_{t+1,j} | \mathcal{F}_t\right\} \leq \sqrt{8}\sqrt{1 + \frac{c_0^2}{\bar{c}^2}} c_0.$$

Combining this together with (24), we have almost surely,

$$\mathrm{E}\frac{1}{\sigma^2_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right)^2 X^2_{t+1,j} | \mathcal{F}_t\right\} \leq \sqrt{8 + \frac{8c_0^2}{\bar{c}^2}} \frac{c_0}{\bar{c}}. \tag{B.55}$$

Let

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

Define $Z^*_{t+1,j} = \left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0} \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}^{(t)}_{j_0}}\right) X_{t+1,j}$ and

$$Z^{**}_{t+1,j} = Z^*_{t+1,j} I\left\{|Z^*_{t+1,j}| \leq 2\omega_0^2 \left(1 + \sqrt{\frac{c_0 \kappa_n}{\bar{c}}}\right) \log p\right\}.$$

Under the events $\max_{i,j}|X_{i,j}| \le \sqrt{2\omega_0^2 \log p}$ and $|\widehat{\mathcal{M}}_{j_0}^{(t)}| \le \kappa_n$, it follows from (24) that

$$
\begin{aligned}
|Z_{t+1,j}^*| \le & \left| X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T \boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}} \right| |X_{t+1,j}| \\
\le & \left(|X_{t+1,j_0}| + \|\boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\|_2 \sqrt{\kappa_n} \max_j |X_{t+1,j}| \right) |X_{t+1,j}| \\
\le & \left\{ \sqrt{2\omega_0^2 \log p} \left(1 + \sqrt{\frac{c_0\kappa_n}{\bar{c}}} \right) \right\} \sqrt{2\omega_0^2 \log p} = 2\omega_0^2 \left(1 + \sqrt{\frac{c_0\kappa_n}{\bar{c}}} \right) \log p,
\end{aligned}
$$

and hence $Z_{t+1,j}^{**} = Z_{t+1,j}^*$.

By (B.54) and Condition (A1), we have with probably tending to 1,

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

Therefore, it suffices to show with probability tending to 1,

$$
\max_j \left| \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} Z_{t+1,j}^{**} I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)} \right) \right| \le \bar{c}_* \log p,
$$

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

$$
I_{2,j}^* = \sum_{t=s_n}^{n-1} \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \mathrm{E}(Z_{t+1,j}^{**}|\mathcal{F}_t) I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)} \right).
$$

We've shown in the proof of Theorem 2.1 that $\mathrm{E}\{Z_{t+1,j}^* I(j \in \widehat{\mathcal{M}}_{j_0}^{(t)})|\mathcal{F}_t\} = 0$. Therefore, we have

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \mathrm{E}(Z_{t+1,j}^{**}|\mathcal{F}_t) I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)} \right) \right| \\
\le & \left| \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \mathrm{E}\left[Z_{t+1,j}^* I\left\{ |Z_{t+1,j}^*| > 2\omega_0^2 \left(1 + \sqrt{\frac{c_0\kappa_n}{\bar{c}}} \right) \log p \right\} |\mathcal{F}_t \right] \right|.
\end{aligned} \tag{B.56}
$$

It follows from Cauchy-Schwarz inequality that

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \mathrm{E}\left[Z_{t+1,j}^* I\left\{ |Z_{t+1,j}^*| > 2\omega_0^2 \left(1 + \sqrt{\frac{c_0\kappa_n}{\bar{c}}} \right) \log p \right\} |\mathcal{F}_t \right] \right| \\
\le & \left| \frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} \sqrt{\mathrm{E}\{(Z_{t+1,j}^*)^2|\mathcal{F}_t\}} \right| \left| \Pr\left\{ |Z_{t+1,j}^*| > 2\omega_0^2 \left(1 + \sqrt{\frac{c_0\kappa_n}{\bar{c}}} \right) \log p \right\} \right|
\end{aligned}
$$

By (B.55), the first term on the second line is upper bounded by $(8+8c_0^2/\bar{c}^2)^{1/4}(c_0/\bar{c})^{1/2}/\sqrt{n}$ with probability 1. As for the second term, under the event $|\widehat{\mathcal{M}}_{j_0}^{(t)}| \le \kappa_n$, using similar arguments in the proof of Lemma A.2, we can show that

$$\Pr\left\{|Z_{t+1,j}| > 2\omega_0^2\left(1+\sqrt{\frac{c_0\kappa_n}{\bar{c}}}\right)\log p\right\} \le \Pr\left(\max_j |X_{t+1,j}| > \sqrt{2\omega_0 \log p}\right) \le \frac{2}{p}.$$

Hence, we have

$$\left|\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\mathrm{E}\left[Z_{t+1,j}I\left\{|Z_{t+1,j}| > 2\omega_0^2\left(1+\sqrt{\frac{c_0\kappa_n}{\bar{c}}}\right)\log p\right\}|\mathcal{F}_t\right]\right| \le \sqrt{\frac{2c_0}{\bar{c}np}}(8+8c_0^2/\bar{c}^2)^{1/4}.$$

Therefore, it follows from (B.56) that

$$\left|\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\mathrm{E}(Z_{t+1,j}^*|\mathcal{F}_t)I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}\right)\right| \le \sqrt{\frac{2c_0}{\bar{c}np}}(8+8c_0^2/\bar{c}^2)^{1/4}.$$

Since $p \gg n$, we have

$$|I_{2,j}^*| \le \sqrt{\frac{2c_0 n}{\bar{c}p}}(8+8c_0^2/\bar{c}^2)^{1/4} = o(1),$$

where the $o(1)$ term is uniform in j.

Hence, it suffices to show $\Pr(\max_j |I_{2,j}^{**}| \le \bar{c}_* \log p) \to 1$, for some constant $\bar{c}_* > 0$, where

$$I_{2,j}^{**} = \sum_{t=s_n}^{n-1}\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\{Z_{t+1,j}^* - \mathrm{E}(Z_{t+1,j}^*|\mathcal{F}_t)\}I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}\right).$$

For any $j \in [1,\ldots,p]$, $I_{2,j}^{**}$ is a mean zero martingale with respect to the filtration $\{\sigma(\mathcal{F}_t)\}$. Besides, given $\mathcal{F}_t$, by the definition of $Z_{t+1,j}^*$ and (24), we have that

$$\begin{aligned}
&\left|\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\{Z_{t+1,j}^* - \mathrm{E}(Z_{t+1,j}^*|\mathcal{F}_t)\}I\left(j \in \widehat{\mathcal{M}}_{j_0}^{(t)}\right)\right| \\
\le\ &\left|\frac{1}{\sqrt{n}\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\{Z_{t+1,j}^* - \mathrm{E}(Z_{t+1,j}^*|\mathcal{F}_t)\}\right| \le \frac{4\omega_0^2}{\sqrt{n\bar{c}}}\left(1+\sqrt{\frac{c_0\kappa_n}{\bar{c}}}\right)\log p, \qquad a.s.
\end{aligned}$$

Moreover, it follows from (B.55) that we have, almost surely,

$$\sum_{t=s_n}^{n-1}\mathrm{E}\frac{1}{n\sigma_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^2}\left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}^T\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)^2 X_{t+1,j}^2|\mathcal{F}_t\right\} \le \sqrt{8+\frac{8c_0^2}{\bar{c}^2}}\frac{c_0}{\bar{c}}. \qquad \text{(B.57)}$$

Set $y_0 = \sqrt{8(1+c_0^2/\bar{c}^2)}c_0/\bar{c}$ and

$$c_0^* = \frac{4\omega_0^2}{\sqrt{n\bar{c}}}\left(1+\sqrt{\frac{c_0\kappa_n}{\bar{c}}}\right)\log p.$$

It follows from Theorem 9.12 in de la Peña et al. (2009) that

$$\Pr\left(|I_{2,j}^{**}| > z, \sum_{t=s_n}^{n-1} \mathrm{E}\frac{1}{n\sigma^2_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}\left\{\left(X_{t+1,j_0} - \boldsymbol{\omega}^T_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}\boldsymbol{X}_{t+1,\widehat{\mathcal{M}}_{j_0}^{(t)}}\right)^2 X_{t+1,j}^2|\mathcal{F}_t\right\} \le y_0\right)$$
$$\le 2\exp\left(-\frac{z^2}{2(y_0+c_0^* z)}\right), \quad \forall z > 0.$$

In view of (B.57), we have

$$\Pr\left(|I_{2,j}^{**}| > z\right) \le 2\exp\left(-\frac{z^2}{2(y_0+c_0^* z)}\right) \le 2\exp\left(-\min\left(\frac{z^2}{4y_0}, \frac{z}{2c_0^*}\right)\right).$$

Let $z_0 = 3\max(\sqrt{y_0}\log p, c_0^*\log p)$, we have

$$\Pr\left(|I_{2,j}^{**}| > z_0\right) \le 2\exp\left(-\min\left(\frac{9}{4}\log^2 p, \frac{3}{2}\log p\right)\right) \le 2\exp\left(-\frac{3}{2}\log p\right).$$

It follows from Bonferroni's inequality that

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

Under the given conditions, we have $\kappa_n \log^2 p = o(n)$ and hence $z_0 = 3\sqrt{y_0}\log p$ for sufficiently large n. By taking $\bar{c}_* = 3\sqrt{y_0}$, this shows $\Pr(\max_j |I_{2,j}^{**}| \le \bar{c}_*\log p) \to 1$. The proof is hence completed.

B.7 Proof of Lemma B.1

Assertion (B.1) can be proven in a similar manner as (24). We omit its proof for brevity. To prove (B.2) and (B.3), we first show the following events occur with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\right\|_2 \le \frac{\bar{c}_*\kappa_n\sqrt{\log p}}{\sqrt{n}} + \bar{c}_*\eta_n, \tag{B.58}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\|\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \boldsymbol{\Sigma}_{\mathcal{M},j_0}\right\|_2 \le \frac{\bar{c}_*\sqrt{\kappa_n\log p}}{\sqrt{n}} + \bar{c}_*\eta_n, \tag{B.59}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\|\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right\|_2 \le \frac{\bar{c}_*(\sqrt{\kappa_n\log p}+\kappa_n)}{\sqrt{n}} + \bar{c}_*\eta_n. \tag{B.60}$$

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,

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n} \left\|\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \boldsymbol{\Sigma}_{\mathcal{M},\mathcal{M}}\right\|_2 \le \frac{\bar{c}_{**}\kappa_n\sqrt{\log p}}{\sqrt{n}}, \tag{B.61}$$

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

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

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

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n} \left\|\left(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right\|_2 \le \bar{c}_{***}\eta_n, \tag{B.62}$$

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n} \left\|\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right\|_2 \le \bar{c}_{***}\eta_n, \tag{B.63}$$

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n} \left\|\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0}\right\|_2 \le \bar{c}_{***}\eta_n. \tag{B.64}$$

Using similar arguments in (B.15), we can show that

$$\max_{i\in[1,\dots,n]} |b''(\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}) - b''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)| \le \bar{c}_{****}\omega_0\eta_n,$$

for some constant $\bar{c}_{****} > 0$. By Condition (A4*), we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\|\left(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right\|_2 \qquad (B.65)\\
\le\;& \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\le1}} \left|\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}} - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right)\boldsymbol{\omega}_{\mathcal{M},j_0}\right|\\
\le\;& \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\le1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}||b''(\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}) - b''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)|\\
\le\;& \bar{c}_{****}\omega_0\eta_n \max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2\le1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}|.
\end{aligned}$$

Note that

$$
\begin{aligned}
& \max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\leq \kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}| \\
\leq & \underbrace{\max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\leq \kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \frac{1}{n}\sum_{i=1}^n \left(|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{i,\mathcal{M}}| - \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|\right)}_{\eta_1} \\
+ & \underbrace{\max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\leq \kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|}_{\eta_2}.
\end{aligned}
$$

Using similar arguments in bounding (B.40), we can show that

$$
|\eta_1| = O\left(\frac{\sqrt{\kappa_n \log p}}{\sqrt{n}} + \frac{\kappa_n}{\sqrt{n}}\right), \tag{B.66}
$$

with probability tending to 1. Under the given conditions, the RHS of (B.66) is $o(1)$. Besides, by Condition (A3*), (B.1) and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
|\eta_2| \leq & \max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\leq \kappa_n} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \left(\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2\right)^{1/2} \left(\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^2\right)^{1/2} \\
\leq & \max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\leq \kappa_n} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \left(\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^4\right)^{1/4} \left(\mathrm{E}|\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}|^4\right)^{1/4} \leq \frac{c_0}{\sqrt{\bar{c}}}.
\end{aligned}
$$

Combining this together with (B.66) gives

$$
\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\leq \kappa_n} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}||\boldsymbol{\omega}_{\mathcal{M},j_0}^T\boldsymbol{X}_{0,\mathcal{M}}| \leq \frac{2c_0}{\sqrt{\bar{c}}}, \tag{B.67}
$$

with probability tending to 1. Assertion (B.62) thus follows by combining (B.65) together with (B.67).

As for (B.63), we have with probability tending to 1,

$$
\begin{aligned}
\max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\leq \kappa_n} \left\|\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}^* - \widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right\|_2 \leq & \max_{\substack{\mathcal{M}\subseteq \mathbb{I}_{j_0} \\ |\mathcal{M}|\leq \kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2\leq 1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2 |b''(\boldsymbol{X}_i^T\widetilde{\boldsymbol{\beta}}) - b''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)| \\
\leq & \bar{c}_{****}\omega_0\eta_n \max_{\mathcal{M}\subseteq \mathbb{I}_{j_0}, |\mathcal{M}|\leq \kappa_n} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|} \\ \|\boldsymbol{a}\|_2\leq 1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2. \quad \text{(B.68)}
\end{aligned}
$$

Besides,

$$
\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2\\
\le\ &\underbrace{\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \frac{1}{n}\sum_{i=1}^{n}\left(|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2-\mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2\right)}_{\eta_1^*}+\underbrace{\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \mathrm{E}|\boldsymbol{a}^T\boldsymbol{X}_{0,\mathcal{M}}|^2}_{\eta_2^*}.
\end{aligned}
$$

Using similar arguments in bounding (B.40), we have with probability tending to 1,

$$
|\eta_1^*| = O\left(\frac{\sqrt{\kappa_n\log p}}{\sqrt{n}}+\frac{\kappa_n^{3/2}}{\sqrt{n}}\right).
$$

Moreover, by (A3*) and Cauchy-Schwarz inequality, we have $\eta_2^*\le c_0^{1/2}$. Combining these together with (B.68), we obtain

$$
\Pr\left(\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}\right\|_2\le 2c_0^{1/2}\bar{c}_{****}\omega_0\eta_n\right)\to 1.
$$

This proves (B.63). Similarly, we can show (B.64) holds. This proves (B.58)-(B.60). Based on these results, following the arguments in the proof of Lemma A.2, we can show (B.2) and (B.3) hold. Besides, based on (B.62)-(B.64), we can similarly show (B.5) holds.

Now, we focus on proving (B.4). We first show

$$
\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\right\|_2\le\bar{c}_0^*\eta_n, \tag{B.69}
$$

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

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

By the condition $\kappa_n^3\log p=o(n)$, (A2*), (A5*), (B.58) and (B.61), we can show the following events occur with probability tending to 1,

$$
\min_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}})\ge\frac{\bar{c}}{2},\qquad\min_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\lambda_{\min}(\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}})\ge\frac{\bar{c}}{2}, \tag{B.70}
$$

for sufficiently large n. Hence, we have

$$\Pr\left(\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\mathbf{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}\right\|_2\left\|\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\right\|_2\le\frac{4}{\bar{c}^2}\right)\to 1.$$

Combining this together with (B.63) yields

$$\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\left\|\widehat{\mathbf{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}}-\widehat{\mathbf{\Sigma}}^{-1}_{\mathcal{M},\mathcal{M}}\right\|_2\le\frac{4\bar{c}_{***}\eta_n}{\bar{c}^2}.$$

This proves (B.69).

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

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

By (B.64) and (B.69), 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. Besides, similar to (B.14), we can show

$$\Pr\left(\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}\|\widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}\|_2\le 2\sqrt{c_0/\bar{c}}\right)\to 1. \tag{B.71}$$

This together with (B.63) and (B.69) yields that

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

Recall that

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

Hence, in order to prove

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\omega}}_{\mathcal{M},j_0}-\widetilde{\boldsymbol{\omega}}_{\mathcal{M},j_0}\|_2\le\bar{c}_0\eta_n^2, \tag{B.72}$$

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

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\| I_1^* - \sum_{j=1}^p \widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}} \widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}(\tilde{\beta}_j - \beta_{0,j}) \right\|_2 \le \bar{c}_0\eta_n^2, \tag{B.73}$$

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \left\| I_4^* - \sum_{j=1}^p \widehat{\boldsymbol{\Sigma}}^{*-1}_{\mathcal{M},\mathcal{M}} \widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}} \widehat{\boldsymbol{\omega}}^*_{\mathcal{M},j_0}(\tilde{\beta}_j - \beta_{0,j}) \right\|_2 \le \bar{c}_0\eta_n^2. \tag{B.74}$$

We first prove (B.73). By (B.70) and the definition of $\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}$, it suffices to show

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

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

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

for some $\boldsymbol{\beta}_a^*$ lying on the line segment joining $\boldsymbol{\beta}_0$ and $\widetilde{\boldsymbol{\beta}}$. By (A5*), we have

$$\Pr\left(\sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \|\boldsymbol{\beta}_0 - \boldsymbol{\beta}_a^*\|_1 \le \eta_n \right) \to 1. \tag{B.75}$$

The function b''' is Lipschitz continuous. By (A4*), under the event defined in (B.75), we have

$$\sup_{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}} \max_{i=1,\ldots,n} |b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_a^*) - b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)| \le L_0 \max_{i=1,\ldots,n} \|\boldsymbol{X}_i\|_\infty \|\boldsymbol{\beta}_0 - \boldsymbol{\beta}_a^*\|_1 \le L_0\omega_0\eta_n, \tag{B.76}$$

for some constant $L_0 > 0$. Therefore, we have

$$\begin{aligned}
&\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}} \left| \boldsymbol{a}^T \left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},j_0} - \widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},j_0} - \frac{1}{n}\sum_{i=1}^n \boldsymbol{X}_{i,\mathcal{M}} b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0) X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\} \right) \right| \\
&\le \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}} \left| \frac{1}{n}\sum_{i=1}^n \boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}} \{b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0) - b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_a^*)\} X_{i,j_0}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\} \right| \\
&\le L_0\omega_0^3\eta_n^2 \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}} \sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}} \frac{1}{n}\sum_{i=1}^n |\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|,
\end{aligned} \tag{B.77}$$

with probability tending to 1. Similar to (B.67), we can show

$$\Pr\left(\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|\le c_2^*\right)\to 1,$$

for some constant $c_2^*>0$. This proves (B.73).

Similarly, to prove (B.74), it suffices to show with probability tending to 1, we have for any $\mathcal{M}\subseteq\mathbb{I}_{j_0}$ such that $|\mathcal{M}|\le\kappa_n$ and any $\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}$ such that $\|\boldsymbol{a}\|_2=1$,

$$\left|\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}}-\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_{i,\mathcal{M}}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)\boldsymbol{X}_{i,\mathcal{M}}\{\boldsymbol{X}_i^T(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\}\right)\boldsymbol{a}\right|\le c_3^*\eta_n^2,$$

for some constant $c_3^*>0$. By Taylor's theorem, we have for any $\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}$,

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

for some $\boldsymbol{\beta}_a^*$ lying on the line segment joining $\boldsymbol{\beta}_0$ and $\widetilde{\boldsymbol{\beta}}$. Hence, similar to (B.76) and (B.77), we can show that with probability tending to 1,

$$\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\boldsymbol{a}^T\left(\widehat{\boldsymbol{\Sigma}}_{\mathcal{M},\mathcal{M}}-\widehat{\boldsymbol{\Sigma}}^*_{\mathcal{M},\mathcal{M}}-\frac{1}{n}\sum_{j=1}^{p}\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}(\widetilde{\beta}_j-\beta_{0,j})\right)\boldsymbol{a}\right|$$
$$\le c_4^*\eta_n^2\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2, \tag{B.78}$$

for some constant $c_4^*>0$. Using similar arguments in (B.40), we can show

$$\Pr\left(\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\frac{1}{n}\sum_{i=1}^{n}|\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}}|^2\le c_5^*\right)\to 1, \tag{B.79}$$

for some $c_5^*>0$. This together with (B.78) proves (B.74). Hence, (B.72) is 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 (B.72) proves (B.4).

Finally, we show

$$\sum_{t=0}^{n-1} \frac{\widetilde{Z}_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}} \left(\frac{1}{\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} - \frac{\sum_j \hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}(\tilde{\beta}_j - \beta_{0,j})}{\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(-s_n)},j_0}} \right) = \sum_{t=0}^{n-1} \frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}}{\sqrt{n}\hat{\sigma}^*_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}} + o_p(1). \quad \text{(B.80)}$$

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

The function $b'''(\cdot)$ is Lipschitz continuous. Hence, there exists some constant L_0 such that

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

By Condition (A4*), we obtain

$$\max_{1 \le i \le n} |b'''(\boldsymbol{X}_i^T \boldsymbol{\beta}_0)| \le c_6^*, \quad \text{(B.81)}$$

for some constant $c_6^* > 0$. By Condition (A4*), we have

$$\max_{1 \le j \le p} |\widehat{\boldsymbol{\Psi}}^{(j)}_{j_0,j_0}| = \max_{1 \le j \le p} \left| \frac{1}{n} \sum_i X_{i,j}^3 b'''(\boldsymbol{X}_i^T \boldsymbol{\beta}_0) \right| \le \omega_0^3 c_6^*. \quad \text{(B.82)}$$

Besides, by (B.79), (B.81) and Condition (A4*), we have

$$\begin{aligned}
\max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\|_2 &= \max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}|\boldsymbol{a}^T\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},\mathcal{M}}\boldsymbol{a}| \qquad\text{(B.83)}\\
&= \max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^n(\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}})^2X_{i,j}b'''(\boldsymbol{X}_i^T\boldsymbol{\beta}_0)\right| \le c_6^*\omega_0\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\sup_{\substack{\boldsymbol{a}\in\mathbb{R}^{|\mathcal{M}|}\\\|\boldsymbol{a}\|_2=1}}\left|\frac{1}{n}\sum_{i=1}^n(\boldsymbol{a}^T\boldsymbol{X}_{i,\mathcal{M}})^2\right|\\
&\le c_5^*c_6^*\omega_0,
\end{aligned}$$

with probability tending to 1. Similarly, we can show

$$\max_{1\le j\le p}\max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\le\kappa_n}}\|\widehat{\boldsymbol{\Psi}}^{(j)}_{\mathcal{M},j_0}\|_2 \le c_2^*c_6^*\omega_0^2, \qquad\text{(B.84)}$$

with probability tending to 1. This together with (B.71), (B.82) and (B.83) yields

$$\Pr\left(\max_{1\le j\le p}\max_{\mathcal{M}\subseteq\mathbb{I}_{j_0},|\mathcal{M}|\le\kappa_n}|\hat{\xi}^{(j)}_{\mathcal{M},j_0}|\le c_7^*\right)\to 1, \qquad\text{(B.85)}$$

for some constant $c_7^*>0$.

Besides, similar to (B.12), we can show

$$\Pr\left\{\max_{t\in[0,\dots,n-1]}\left|\widehat{Z}^*_{t+1}\right|\le\omega_0\left(1+2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right)\right\}\to 1. \qquad\text{(B.86)}$$

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

$$\varepsilon^*_t=\varepsilon_t I(|\varepsilon_t|\le n^{-1/3}c_n),$$

for some sequence diverging c_n which will be specified later. By Bonferrni's inequality and Markov's inequality, we have

$$\Pr\left\{\bigcup_{t=1}^n(\varepsilon^*_t\ne\varepsilon_t)\right\}\le n\Pr\left(\varepsilon^*_t\ne\varepsilon_t\right)\le n\Pr\left(|\varepsilon_0|>n^{-1/3}c_n\right)\le n\frac{\mathrm{E}|\varepsilon_0|^3}{nc_n^3}\to 0.$$

This implies

$$\eta_{1,j}^{***} = \sum_{t=s_n}^{n-1} \frac{\widehat{Z}_{t+1,j_0}^{*}\varepsilon_{t+1}^{*}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}_{j_0}^{(t)},j_0}}, \tag{B.87}$$

with probability tending to 1. Note that

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

Since $\mathrm{E}(\varepsilon_{t+1}|\boldsymbol{X}_{t+1}) = 0$, we have

$$|\mathrm{E}(\varepsilon_{t+1}^{*}|\boldsymbol{X}_{t+1})| = \left|\mathrm{E}\left(\varepsilon_{t+1}I(|\varepsilon_{t+1}| > n^{1/3}c_n)|\boldsymbol{X}_{t+1}\right)\right| \leq \frac{\mathrm{E}(|\varepsilon_{t+1}|^3|\boldsymbol{X}_{t+1})}{n^{2/3}c_n^2}.$$

By Condition (A6*), there exists some constant $c_8^* > 0$ such that

$$\max_{0\leq t\leq n-1} |\mathrm{E}(\varepsilon_{t+1}^{*}|\boldsymbol{X}_{t+1})| \leq c_8^* n^{-2/3}c_n^{-2}.$$

Combining this together with (B.9), (B.85) and (B.86) yields

$$\max_j |\eta_{1,j}^{*****}| \leq \sqrt{n}c_7^*c_8^*\omega_0\left(1 + 2\sqrt{\frac{\kappa_n c_0}{\bar{c}}}\right)n^{-2/3}c_n^{-2}\frac{8}{\bar{c}^{3/2}},$$

with probability tending to 1. Under the given conditions, we have $\kappa_n^3 = o(n)$. Hence, we've shown

$$\max_j |\eta_{1,j}^{*****}| = o_p(1). \tag{B.88}$$

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

$$|\widehat{Z}_{t+1,j_0}^{*}| \leq \omega_0 \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} (1 + \sqrt{\kappa}_n\|\widehat{\omega}_{\mathcal{M},j_0}^{*}\|_2). \tag{B.89}$$

Let

$$\bar{c}_n^{(j)} \equiv c_n\omega_0 n^{-1/6} \max_{1\leq j\leq p} \max_{\substack{\mathcal{M}\subseteq\mathbb{I}_{j_0}\\|\mathcal{M}|\leq\kappa_n}} \frac{|(1 + \sqrt{\kappa}_n\|\widehat{\omega}_{\mathcal{M},j_0}^{*}\|_2)|}{|\hat{\sigma}_{\mathcal{M},j_0}^{*}|}|\hat{\xi}^{(j)}_{\mathcal{M},j_0}|.$$

By Condition (A1*), (B.89) and the definition of ε_t^*, we have for any t and $1 \le j \le p$,

$$\Pr\left(\left| \frac{\widehat{Z}^*_{t+1,j_0}\varepsilon^*_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right| \le \bar{c}_n^{(j)} \middle| \mathcal{F}^*_t \right) = 1. \tag{B.90}$$

Hence, by Jensen's inequality, we have

$$\Pr\left(\left| \frac{\widehat{Z}^*_{t+1,j_0}\mathrm{E}(\varepsilon^*_{t+1}|\boldsymbol{X}_{t+1})\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right| \le \bar{c}_n^{(j)} \middle| \mathcal{F}^*_t \right) = 1.$$

This together with (B.90) gives

$$\Pr\left(\left| \frac{\widehat{Z}^*_{t+1,j_0}\{\varepsilon^*_{t+1} - \mathrm{E}(\varepsilon^*_{t+1}|\boldsymbol{X}_{t+1})\}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right| \le 2\bar{c}_n^{(j)} \middle| \mathcal{F}^*_t \right) = 1.$$

Besides, it follows from Hölder's inequality that

$$\begin{aligned}
&\sum_{t=s_n}^{n-1} \mathrm{E}\left\{ \left(\frac{\widehat{Z}^*_{t+1,j_0}\{\varepsilon^*_{t+1} - \mathrm{E}(\varepsilon^*_{t+1}|\boldsymbol{X}_{t+1})\}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right)^2 \middle| \mathcal{F}^*_t \right\} \\
\le\; & \sum_{t=s_n}^{n-1} \mathrm{E}\left\{ \left(\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon^*_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right)^2 \middle| \mathcal{F}^*_t \right\} \le \sum_{t=s_n}^{n-1} \mathrm{E}\left\{ \left(\frac{\widehat{Z}^*_{t+1,j_0}\varepsilon_{t+1}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}}{\sqrt{n}\hat{\sigma}^{*3}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \right)^2 \middle| \mathcal{F}^*_t \right\} \\
\le\; & \sum_{t=s_n}^{n-1} \frac{(\widehat{Z}^*_{t+1,j_0}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0})^2}{n\hat{\sigma}^{*6}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \mathrm{E}(\varepsilon^2_{t+1}|\boldsymbol{X}_{t+1}) \le \underbrace{\sum_{t=s_n}^{n-1} \frac{(\widehat{Z}^*_{t+1,j_0}\hat{\xi}^{(j)}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0})^2}{n\hat{\sigma}^{*6}_{\widehat{\mathcal{M}}^{(t)}_{j_0},j_0}} \{\mathrm{E}(|\varepsilon_{t+1}|^3|\boldsymbol{X}_{t+1})\}^{2/3}}_{V_n^{(j)}}.
\end{aligned}$$

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

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

Take $z_0^{(j)} = 6\{(V_n^{(j)}\log p)^{1/2} + (\bar{c}_n^{(j)})\log p\}$, we have

$$\begin{aligned}
\Pr(|\eta^{****}_{1,j}| > z_0^{(j)}) \;&\le\; 2\exp\left(-\frac{36V_n^{(j)}\log p + 36(\bar{c}_n^{(j)})^2\log^2 p}{2V_n^{(j)} + 12\bar{c}_n^{(j)}(V_n^{(j)}\log p)^{1/2} + 12(\bar{c}_n^{(j)})^2\log p} \right) \\
&\le\; 2\exp\left(-\frac{36V_n^{(j)}\log p + 36(\bar{c}_n^{(j)})^2\log^2 p}{8V_n^{(j)} + 18(\bar{c}_n^{(j)})^2\log p} \right) \le 2\exp(-2\log p) = \frac{2}{p^2}.
\end{aligned}$$

It follows from Bonferroni's inequality that

$$\Pr\left(\bigcap_{j=1}^{p}\left\{|\eta_{1,j}^{****}|>z_0^{(j)}\right\}\right)\le\sum_{j=1}^{p}\Pr(|\eta_{1,j}^{****}|>z_0^{(j)})=\frac{2}{p}.$$

This implies that

$$\max_{j=1}^{p}|\eta_{1,j}^{****}|\le\max_{j=1}^{p}z_0^{(j)}. \tag{B.91}$$

Set $c_n=\log^{1/3}n$. Under the given conditions, we have $n^{-1/6}\sqrt{\kappa_n}c_n=o(1)$. By Condition (A1*), (A6*), (B.9), (B.71), (B.85) and (B.86), we can show

$$\max_{j=1}^{p}\bar{c}_n^{(j)}=o_p(1)\quad\text{and}\quad\max_{j=1}^{p}V_n^{(j)}=O_p(\kappa_n).$$

In view of (B.91), we've shown $\max_j|\eta_{1,j}^{****}|=O_p(\sqrt{\kappa_n\log p}+\log p)$. This further implies together with (B.87) and (B.88) yields

$$\max_j|\eta_{1,j}^{***}|=O_p(\sqrt{\kappa_n\log p}+\log p). \tag{B.92}$$

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 $\max_j|\eta_{1,j}^{**}|=O_p(\sqrt{\kappa_n\log p}+\log p)$. This together with (B.92) gives $\max_j|\eta_{1,j}^{*}|=O_p(\sqrt{\kappa_n\log p}+\log p)$. 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

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

By Condition (A1*), (A4*), (A5*), (B.70), (B.71), (B.83) and (B.84), we can show

$$\Pr\left(\max_{0\le t\le n-1}|\widehat{Z}^*_{t+1,j_0}-\widetilde{Z}_{t+1,j_0}|\le c_9^*\sqrt{\kappa_n}\eta_n\right)\to 1,$$

for some constant $c_9^*>0$. Combining this together with Condition (A1*), (B.9) and (B.85) yields

$$\max_j|\eta^*_{3,j}|\le c^*_{10}\sqrt{n\kappa_n}\eta_n\left(\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}|\right). \tag{B.93}$$

By Condition (A6*) and Hölder's inequality, we have

$$\mathrm{E}\left(\frac{1}{n}\sum_{t=0}^{n-1}|\varepsilon_{t+1}|\right)=\mathrm{E}|\varepsilon_0|\le(\mathrm{E}|\varepsilon_0|^3)^{1/3}=O(1).$$

Hence, it follows from Markov's inequality that $\sum_{t=0}^{n-1}|\varepsilon_{t+1}|/n=O_p(1)$. This together with (B.93) implies that $\max_j|\eta^*_{3,j}|=O_p(\sqrt{n\kappa_n}\eta_n)$. Therefore, we have $\eta_3^*=O_p(\sqrt{n\kappa_n}\eta_n^2)$. By Condition (A5*), we obtain $\eta_3^*=o_p(1)$. The proof is hence completed.

B.8 Technical lemmas

Lemma B.2 *For any positive definite matrix*

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

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

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

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

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

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.

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.

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