Online Supplementary Materials for “Optimal averaging estimator of heterogeneous treatment effects for single-index models”

Na Li[1], Maoyuan Wang[1], Yu Fei[2], Xinyu Zhang [3]

[1] *Faculty of Science, Kunming University of Science and Technology*

[2] *School of Statistics and Mathematics, Yunnan University of Finance and Economics*

[3] *Academy of Mathematics and Systems Science, Chinese Academy of Sciences*

April 25, 2025

The online supplemental file contains the proofs of Lemmas 1-3 and detailed proofs of Theorems 1, 2, and 4.

1 Proof of Lemmas 1-2

Proof Lemmas 1: Let $\kappa_n = \sqrt{p_s} n^{-1/2} + \alpha_n \sqrt{p_s}$, and set $\|\boldsymbol{u}\| = C$, where C is a large enough constant. To obtain (8), our aim is to show that, for any given ε and large n, there is a large constant C such that

$$P\left(\min_{\|\boldsymbol{u}\|=C} \|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\hat{\boldsymbol{\xi}}\|^2 > \|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}}\|^2) \right) \geq 1 - \varepsilon. \tag{1}$$

This implies that with probability tending to 1 there is a local maximum $\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}})$ in the ball $\{\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u} : \|\boldsymbol{u}\| \leq C\}$ such that $\|\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}) - \boldsymbol{\beta}_s^*\|_2 = O_p(p_s^{1/2} n^{-1/2} + p_s^{1/2} a_n)$.

We observe that

$$\begin{aligned}
&\|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\hat{\boldsymbol{\xi}}\|^2 - \|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}}\|^2 \\
&= \|\boldsymbol{\xi} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\boldsymbol{\xi}\|^2 - \|\boldsymbol{\xi} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}\|^2 \\
&\quad + \|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\hat{\boldsymbol{\xi}}\|^2 - \|\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}}\|^2 \\
&\quad - [\|\boldsymbol{\xi} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\boldsymbol{\xi}\|^2 - \|\boldsymbol{\xi} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}\|^2] \\
&= H_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u}) - H_{(s)}(\boldsymbol{\beta}_s^*) \\
&\quad - 2[(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\boldsymbol{\xi} - (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}]
\end{aligned}$$

$$
\begin{aligned}
&-2[\boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-\boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})]\\
&-2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})]\\
&+2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}^T(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{\xi}-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}]\\
&+[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}^T(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})]\\
=\;& H_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})-H_{(s)}(\boldsymbol{\beta}_s^*)+\Lambda_1+\Lambda_2+\Lambda_3+\Lambda_4+\Lambda_5,
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_1 &= -2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{\xi}-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}],\\
\Lambda_2 &= -2[\boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-\boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})],\\
\Lambda_3 &= -2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})],\\
\Lambda_4 &= 2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{\xi}-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}],
\end{aligned}
$$

and

$$
\Lambda_5 = [(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}'(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})].
$$

By Taylor expansion, we have

$$
\begin{aligned}
&H_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})-H_{(s)}(\boldsymbol{\beta}_s^*)\\
&=\kappa_n\boldsymbol{u}'\frac{\partial H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s}+\kappa_n^2\boldsymbol{u}'\frac{\partial^2 H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s\partial\boldsymbol{\beta}_s'}\boldsymbol{u}+o_p(\kappa_n^2\|\boldsymbol{u}\|^2).
\end{aligned}
$$

Under Assumption 8, we observe

$$
\begin{aligned}
&\kappa_n^2\boldsymbol{u}'\frac{\partial^2 H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s\partial\boldsymbol{\beta}_s'}\boldsymbol{u}\\
&\geq \lambda_{min}\left\{\frac{\partial^2 H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s\partial\boldsymbol{\beta}_s'}\right\}\kappa_n^2\|\boldsymbol{u}\|^2\\
&\geq c_0 n\kappa_n^2\|\boldsymbol{u}\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\kappa_n\boldsymbol{u}'\frac{\partial H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s}\\
&\leq \kappa_n\|\boldsymbol{u}\|\|\frac{\partial H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s}\|\\
&= O_p(\sqrt{np_s}\kappa_n)\|\boldsymbol{u}\|.
\end{aligned}
$$

First, for Λ_1, it is shown that

$$
\begin{aligned}
&\Lambda_1\\
&= -2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{\xi}-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^T\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}]
\end{aligned}
$$

$$\begin{aligned}
&= -\kappa_n 2(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)\boldsymbol{\xi}}{\partial \boldsymbol{\beta}_s}\boldsymbol{u} \\
&\leq c\kappa_n \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\| \|\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)\boldsymbol{\xi}}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\| \\
&= c\kappa_n \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\| \sqrt{\sum_{i=1}^{n} \boldsymbol{u}' \frac{\hat{g}_{(s)}(\boldsymbol{x}'_{i,s}\boldsymbol{\beta}_s)}{\partial \boldsymbol{\beta}_s} \left(\frac{\hat{g}_{(s)}(\boldsymbol{x}'_{i,s}\boldsymbol{\beta}_s)}{\partial \boldsymbol{\beta}_s}\right)' \boldsymbol{u}} \\
&= O_p(n^{1/2}\alpha_n \kappa_n \sqrt{np_s})\|\boldsymbol{u}\|,
\end{aligned}$$

where $\tilde{\boldsymbol{\beta}}_s$ lies between $\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u}$ and $\boldsymbol{\beta}_s^*$, and the last equation is due to Assumptions 7 and 9.

Next, for Λ_2 and Λ_3, it is seen that

$$\begin{aligned}
&\Lambda_2 \\
&= -2[\boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) - \boldsymbol{\xi}'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})] \\
&= -\kappa_n 2\boldsymbol{\xi}' \frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u} \\
&\leq \kappa_n \|\boldsymbol{\xi}\| \|\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\| \\
&= O_p(n^{1/2}\alpha_n \kappa_n \sqrt{np_s})\|\boldsymbol{u}\|,
\end{aligned}$$

and

$$\begin{aligned}
&\Lambda_3 \\
&= -2[(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) - (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})] \\
&= -\kappa_n 2(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})' \frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u} \\
&\leq \kappa_n \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\| \|\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\| \\
&= O_p(n^{1/2}\alpha_n^2 \kappa_n \sqrt{np_s})\|\boldsymbol{u}\|,
\end{aligned}$$

where the last equations are obtained from Assumptions 2, 7, and 9.

Final, we consider Λ_4 and Λ_5. Combined $\lambda_{max}(\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s)) = O(1)$ with Assumptions 2, 7, and 9, we can obtain that

$$\begin{aligned}
&\Lambda_4 \\
&= 2[(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{K}'_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^* + \kappa_n \boldsymbol{u})\boldsymbol{\xi} - (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{K}'_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}] \\
&= 2\kappa_n(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})'\boldsymbol{K}'_{(s)}(\tilde{\boldsymbol{\beta}}_s)\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)\boldsymbol{\xi}}{\partial \boldsymbol{\beta}_s}\boldsymbol{u} \\
&\quad + 2\kappa_n \left[\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\right]' \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)\boldsymbol{\xi}
\end{aligned}$$

$$
\begin{aligned}
&+\kappa_n^2\left[\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\right]^{'}\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)\boldsymbol{\xi}}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\\
=\;& O_p(n^{1/2}\alpha_n\kappa_n\sqrt{np_s})\|\boldsymbol{u}\|+O_p(n^{1/2}\alpha_n\kappa_n\sqrt{np_s})\|\boldsymbol{u}\|+O_p(n\alpha_n\kappa_n^2p_s)\|\boldsymbol{u}\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
&\Lambda_5\\
=\;&2[(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{'}\boldsymbol{K}_{(s)}^{'}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*+\kappa_n\boldsymbol{u})(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})-(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{'}\boldsymbol{K}_{(s)}^{'}(\boldsymbol{\beta}_s^*)\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})]\\
=\;&2\kappa_n(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})^{'}\boldsymbol{K}_{(s)}^{'}(\tilde{\boldsymbol{\beta}}_s)\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\\
&+2\kappa_n\left[\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\right]^{'}\boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})\\
&+\kappa_n^2\left[\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\right]^{'}\frac{\partial \boldsymbol{K}_{(s)}(\tilde{\boldsymbol{\beta}}_s)(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})}{\partial \boldsymbol{\beta}_s}\boldsymbol{u}\\
=\;&O_p(n^{1/2}\alpha_n^2\kappa_n\sqrt{np_s})\|\boldsymbol{u}\|+O_p(n\alpha_n^2\kappa_n^2p_s)\|\boldsymbol{u}\|^2.
\end{aligned}
$$

We put these results together, and allow $\|\boldsymbol{u}\|=C$ to be large enough, then Λ_1, Λ_2, Λ_3, Λ_4 and Λ_5 are dominated by $\kappa_n^2\boldsymbol{u}^{'}\frac{\partial^2 H(\boldsymbol{\beta}_s^*)}{\partial\boldsymbol{\beta}_s\partial\boldsymbol{\beta}_s^{'}}\boldsymbol{u}$, which is positive. Therefore, the proof of (8) is completed.

Proof Lemma 2: Under Assumptions 1-5, and 7-9, it is shown that

$$
\begin{aligned}
&|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}),\hat{e}(\boldsymbol{x}))-\tau(\boldsymbol{x}_i)|\\
\le\;&|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}),\hat{e}(\boldsymbol{x}))-\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}),e(\boldsymbol{x}))|\\
&+|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}),e(\boldsymbol{x}))-\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\boldsymbol{\beta}_s^*,e(\boldsymbol{x}))|+|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\boldsymbol{\beta}_s^*,e(\boldsymbol{x})-\tau(\boldsymbol{x}_i)|\\
\le\;&\|\boldsymbol{e}_i\|_\infty\|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{e}(\boldsymbol{x})))\|_\infty\|\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|_\infty\\
&+C\|\frac{\partial\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\tilde{\boldsymbol{\beta}}_s,e(\boldsymbol{x}))}{\partial\boldsymbol{\beta}_s}\|\|\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}})-\boldsymbol{\beta}_s^*\|+|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\boldsymbol{\beta}_s^*,e(\boldsymbol{x})-\tau(\boldsymbol{x}_i)|\\
\le\;&\|\boldsymbol{e}_i\|_\infty\|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{e}(\boldsymbol{x})))\|_\infty\|\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|_\infty\\
&+C\left\|\frac{\partial\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\tilde{\boldsymbol{\beta}}_s,e(\boldsymbol{x}))}{\partial\boldsymbol{\beta}_s}\right\|\|\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}})-\boldsymbol{\beta}_s^*\|+|\hat{g}_{(s)}(\boldsymbol{x}_{i,s}^{'}\boldsymbol{\beta}_s^*,e(\boldsymbol{x}))-\tau(\boldsymbol{x}_i)|\\
=\;&O_p(p_sn^{-1/2}+p_sa_n+\sqrt{m_n^{(s)}/n}+m_n^{(s)-r}).
\end{aligned}
$$

2 Proof of Theorem 1

Let $\upsilon_n=n^\delta S_n(p^*n^{-1/2}+p^*\alpha_n+\sqrt{m_n^\star/n})$, and set $\|\boldsymbol{u}\|=C$, where C is a large enough constant. Similar to the proof of Lemma 1, we only to show that, for any given ϵ and large n, there is a large constant C such that

$$
P\bigg(\min_{\|\boldsymbol{u}\|=C}\mathrm{CV}_n(\boldsymbol{\omega}^{opt}+\upsilon_n\boldsymbol{u})>\mathrm{CV}_n(\boldsymbol{\omega}^{opt}))\bigg)\ge 1-\epsilon. \tag{2}
$$

To prove Theorem 1, we conduct the following decomposition:

$$\begin{aligned}
&\mathrm{CV}_n(\boldsymbol{\omega}^{opt}+\upsilon_n\boldsymbol{u})-\mathrm{CV}_n(\boldsymbol{\omega}^{opt})\\
&=\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2-2[\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\\
&\quad+\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2+2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})-\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]\\
&\quad+2[\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]-2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]\\
&\quad-2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})-\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]\\
&=\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2+\Xi_1+\Xi_2+\Xi_3+\Xi_4+\Xi_5+\Xi_6,
\end{aligned}$$

where

$$\begin{aligned}
\Xi_1&=-2[\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}),\\
\Xi_2&=\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2,\\
\Xi_3&=2[\hat{\boldsymbol{\xi}}-\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})],\\
\Xi_4&=2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})-\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})],\\
\Xi_5&=-2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})-\tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})],
\end{aligned}$$

and

$$\Xi_6=-2[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]'[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})-\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})].$$

Under Assumption 6, we can obtain that

$$\begin{aligned}
&\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2\\
&=\upsilon_n^2\boldsymbol{u}^T\Upsilon_n\boldsymbol{u}\\
&\geq \upsilon_n^2\|\boldsymbol{u}\|^2\lambda_{min}(\Upsilon_n)\\
&>Cn\upsilon_n^2\|\boldsymbol{u}\|^2.
\end{aligned}$$

where Υ_n is $S_n\times S_n$ matrix with (i,j)th element is $\Upsilon_{i,j}=\hat{\boldsymbol{\Delta}}^{(i)'}\hat{\boldsymbol{\Delta}}^{(j)}$.

Combined with Assumptions 2,7, and $\lambda_{max}(\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s))=O(1)$, it follows that

$$\begin{aligned}
&\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2\\
&\leq S_n\upsilon_n^2\sum_{s=1}^{S_n}u_s^2\|\hat{\boldsymbol{\Delta}}^{(s)}\|\\
&=S_n\upsilon_n^2\sum_{s=1}^{S_n}u_s^2\hat{\boldsymbol{\xi}}^{'}\boldsymbol{K}_{(s)}'(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))\hat{\boldsymbol{\xi}}\\
&=O_p(nS_n^2\upsilon_n^2\|\boldsymbol{u}\|^2).
\end{aligned}\tag{3}$$

From Assumptions 2, 7, and 9, and Lemma 1, it follows that

$$\|\hat{\boldsymbol{\Delta}}^{(s)}-\boldsymbol{\Delta}^{*(s)}\|^2$$

$$
\begin{aligned}
&= \|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))\hat{\boldsymbol{\xi}} - \boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))\boldsymbol{\xi} + \boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))\boldsymbol{\xi} - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\boldsymbol{\xi}\|^2 \\
&\leq 2\|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}))(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\|^2 \\
&\quad + 2\sum_{j}^{n}[\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}) - \boldsymbol{\beta}_s^*]' \frac{\partial \hat{g}_{(s)}(\boldsymbol{x}_{j,s}'\tilde{\boldsymbol{\beta}}_s, e(\boldsymbol{x}_j))}{\partial \boldsymbol{\beta}_s} \left[\frac{\partial \hat{g}_{(s)}(\boldsymbol{x}_{j,s}'\tilde{\boldsymbol{\beta}}_s, e(\boldsymbol{x}_j))}{\partial \boldsymbol{\beta}_s}\right]' [\hat{\boldsymbol{\beta}}_s(\hat{\boldsymbol{e}}) - \boldsymbol{\beta}_s^*] \\
&= O_p(n\alpha_n^2 + p_s^2 + np_s^2\alpha_n^2),
\end{aligned} \tag{4}
$$

which lead to

$$
\begin{aligned}
&\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \boldsymbol{\Delta}^*(\upsilon_n\boldsymbol{u})\|^2 \\
&\leq \upsilon_n^2 S_n \sum_{s=1}^{S_n} u_s^2 \|\hat{\boldsymbol{\Delta}}^{(s)} - \boldsymbol{\Delta}^{*(s)}\|^2 \\
&= O_p(\upsilon_n^2 S_n^2 n(\frac{p^{*2}}{n} + p^{*2}\alpha_n^2)\|\boldsymbol{u}\|^2).
\end{aligned} \tag{5}
$$

Similar to (5) and, we can obtain that

$$
\begin{aligned}
&\|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega}^{opt})\|^2 \\
&= O_p(S_n^2 n(\frac{p^{*2}}{n} + p^{*2}\alpha_n^2)),
\end{aligned} \tag{6}
$$

and

$$
\begin{aligned}
&\|\tilde{\boldsymbol{\Delta}}^{(s)} - \boldsymbol{\Delta}^{*(s)}\|^2 \\
&\leq 2\|\tilde{\boldsymbol{\Delta}}^{(s)} - \tilde{\boldsymbol{\Delta}}^{*(s)}\|^2 + 2\|\tilde{\boldsymbol{\Delta}}^{*(s)} - E[\hat{\boldsymbol{\xi}}|\boldsymbol{x}_s'\boldsymbol{\beta}_s^*] - \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}} + E[\hat{\boldsymbol{\xi}}|\boldsymbol{x}_s'\boldsymbol{\beta}_s^*] + \boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}} - \boldsymbol{\Delta}^{*(s)}\|^2 \\
&= 2\|\tilde{\boldsymbol{\Delta}}^{(s)} - \tilde{\boldsymbol{\Delta}}^{*(s)}\|^2 + C\|\tilde{\boldsymbol{\Delta}}^{*(s)} - E[\hat{\boldsymbol{\xi}}|\boldsymbol{x}_s'\boldsymbol{\beta}_s^*]\|^2 \\
&\quad + C\|\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}} - E[\hat{\boldsymbol{\xi}}|\boldsymbol{x}_s'\boldsymbol{\beta}_s^*]\|^2 + C\|\boldsymbol{K}_{(s)}(\boldsymbol{\beta}_s^*)\hat{\boldsymbol{\xi}} - \boldsymbol{\Delta}^{*(s)}\|^2 \\
&= O_p(p_s^2 + np_s^2\alpha_n^2 + m_n^{(s)} + nm_n^{(s)-2r}).
\end{aligned} \tag{7}
$$

First, for Ξ_1, we observe that

$$
\begin{aligned}
&\Xi_1 \\
&= -2[\boldsymbol{\xi} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - 2[\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}]^T\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) \\
&= -2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - 2[\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}]^T\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - 2\boldsymbol{\varepsilon}^T\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}).
\end{aligned}
$$

Combined with (3), (5) and Assumptions 2 and 7 we can obtain that

$$
\begin{aligned}
&-2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) \\
&= O_p(\varphi_n^{1/2}\upsilon_n S_n n^{1/2}\|\boldsymbol{u}\|),
\end{aligned}
$$

$$
-2[\boldsymbol{\xi} - \hat{\boldsymbol{\xi}}]'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})
$$

$$= O_p(n^{1/2}\alpha_n \upsilon_n S_n n^{1/2}\|\boldsymbol{u}\|),$$

and

$$\begin{aligned}
&-2\boldsymbol{\varepsilon}'\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\\
&= -2\boldsymbol{\varepsilon}'\boldsymbol{\Delta}^*(\upsilon_n\boldsymbol{u}) - 2\boldsymbol{\varepsilon}'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \boldsymbol{\Delta}^*(\upsilon_n\boldsymbol{u})]\\
&= O_p(\upsilon_n S_n n^{1/2}\|\boldsymbol{u}\|) + O_p(\upsilon_n S_n n(\frac{p^*}{n^{1/2}} + p^*\alpha_n)\|\boldsymbol{u}\|).
\end{aligned}$$

Second, we consider $\Xi_2 - \Xi_6$. Form (3), (5), and Assumptions 2 and 7, we have

$$\begin{aligned}
&\Xi_2\\
&= \|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2\\
&\leq c[\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \boldsymbol{\Delta}^*(\upsilon_n\boldsymbol{u})\|^2 + \|\boldsymbol{\Delta}^*(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2]\\
&= O_p(\upsilon_n^2 S_n^2 n(\frac{p^{*2}}{n} + p^{*2}\alpha_n^2 + m_n^\star/n)\|\boldsymbol{u}\|^2),
\end{aligned}$$

$$\begin{aligned}
&|\Xi_5|\\
&\leq \|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|\\
&= O_p(nS_n^2\upsilon_n^2(\frac{p^*}{n^{1/2}} + p^*\alpha_n + \sqrt{m_n^\star/n})\|\boldsymbol{u}\|^2),
\end{aligned}$$

$$\begin{aligned}
&|\Xi_6|\\
&\leq \|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|\|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})\|\\
&= O_p(nS_n^2\upsilon_n(\frac{p^*}{n^{1/2}} + p^*\alpha_n + \sqrt{m_n^\star/n})\|\boldsymbol{u}\|),
\end{aligned}$$

$$\begin{aligned}
&\Xi_4\\
&= \|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|\|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})\|\\
&= O_p(\upsilon_n S_n^2 n(\frac{p^{*2}}{n} + p^{*2}\alpha_n^2 + m_n^\star/n)\|\boldsymbol{u}\|),
\end{aligned}$$

and

$$\begin{aligned}
\Xi_3 &= 2[\hat{\boldsymbol{\xi}} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]\\
&= 2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}^{opt})]'[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})] + 2[\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}]^T[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]\\
&\quad + 2\boldsymbol{\varepsilon}^T[\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u}) - \tilde{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})]\\
&= O_p(\varphi_n^{1/2}\upsilon_n S_n n^{1/2}(\frac{p^*}{n^{1/2}} + p^*\alpha_n + \sqrt{m_n^\star/n})\|\boldsymbol{u}\|)\\
&\quad + O_p(\alpha_n\upsilon_n S_n n(\frac{p^*}{n^{1/2}} + p^*\alpha_n + \sqrt{m_n^\star/n})\|\boldsymbol{u}\|).
\end{aligned}$$

Based on Assumptions 11 and 12, it is shown that Ξ_1, Ξ_2, Ξ_3, Ξ_4, Ξ_5, and Ξ_6 are dominated by $\|\hat{\boldsymbol{\Delta}}(\upsilon_n\boldsymbol{u})\|^2$ which is positive. That completes the proof of Theorem 1.

3 Proof Theorem 2

We conduct the following decomposition

$$\begin{aligned}
&\mathrm{CV}_n(\boldsymbol{\omega}) \\
&= \|\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 \\
&= \|\boldsymbol{\xi} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 \\
&= \|\boldsymbol{\Delta} + \boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 \\
&= \|\boldsymbol{\Delta} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2 + 2[\boldsymbol{\Delta} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\
&= \|\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 + \|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 + 2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})] \\
&\quad + 2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T\boldsymbol{\varepsilon} + 2[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \\
&\quad + 2[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T\boldsymbol{\varepsilon} + 2[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]^T(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2 \\
&= L_n(\boldsymbol{\omega}) + \Phi_1(\boldsymbol{\omega}) + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2.
\end{aligned}$$

Theorem 2 is valid if the following holds:

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{\Phi_1(\boldsymbol{\omega})}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1) \tag{8}$$

and

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{L_n^*(\boldsymbol{\omega}) - L_n(\boldsymbol{\omega})}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1). \tag{9}$$

In order to prove (8), we need only to verify that

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{\|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{10}$$

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{11}$$

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'\boldsymbol{\varepsilon}}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{12}$$

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'\boldsymbol{\varepsilon}}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{13}$$

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{14}$$

and

$$\sup_{\boldsymbol{\omega}\in\mathcal{W}}\left|\frac{[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})}\right| = o_p(1), \tag{15}$$

First, the proof of (9)-(13) are similar to Zou et al. (2021), so we only need to prove (14) and (15).

Second, from (4), we have

$$\|\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\|^2$$

$$\begin{aligned}&\leq S_n \sum_{s=1}^{S_n} \omega_s^2 \|\hat{\boldsymbol{\Delta}}^{(s)} - \boldsymbol{\Delta}^{*(s)}\|^2 \\ &= O_p(n\alpha_n^2 S_n^2) + O_n(p^{*2} S_n^2 + np^{*2} S_n^2 \alpha_n^2). \qquad (16)\end{aligned}$$

From (7), we can obtain that

$$\|\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\|^2 = O_n(p^{*2} S_n^2 + np^{*2} S_n^2 \alpha_n^2 + S_n^2 m_n^\star). \qquad (17)$$

Combined with (16) and (17), and Assumption 7, it is can be seen that

$$\begin{aligned}&\sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\boldsymbol{\Delta} - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| \\ &\leq \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| + \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\boldsymbol{\Delta}^*(\boldsymbol{\omega}) - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| \\ &\leq \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{\|\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\| \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|}{L_n^*(\boldsymbol{\omega})} \right| + \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{\|\boldsymbol{\Delta}^*(\boldsymbol{\omega}) - \hat{\boldsymbol{\Delta}}(\boldsymbol{\omega})\| \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|}{L_n^*(\boldsymbol{\omega})} \right| \\ &= O_p(n^{1/2}\alpha_n \eta_n^{-1/2}) + O_n(\frac{n\alpha_n^2 S_n + n^{1/2}\alpha_n p^* S_n + np^* S_n \alpha_n^2}{\eta_n}) \\ &= o_p(1),\end{aligned}$$

and

$$\begin{aligned}&\sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| \\ &\leq \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\hat{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| + \sup_{\boldsymbol{\omega}\in\mathcal{W}} \left| \frac{[\boldsymbol{\Delta}^*(\boldsymbol{\omega}) - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})]'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{L_n^*(\boldsymbol{\omega})} \right| \\ &= O_n(\frac{n\alpha_n^2 S_n + n^{1/2}\alpha_n p^* S_n + np^* S_n \alpha_n^2 + n^{1/2}\alpha_n S_n \sqrt{m_n^\star}}{\eta_n}) \\ &= o_p(1),\end{aligned}$$

where the last equation is due to Assumptions 4 and 10.

4 Proof of Lemma 3

Note that

$$\begin{aligned}&\mathrm{CV}_n(\boldsymbol{\omega}) \\ &= \|\hat{\boldsymbol{\xi}} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 \\ &= \|\boldsymbol{\Delta} + \boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 \\ &= \|\boldsymbol{\Delta} - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})\|^2 + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2 + 2[\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}]'[\Delta - \tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega})] \\ &= \|\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\|^2 + \|\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\|^2 - 2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'[\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]\end{aligned}$$

$$
\begin{aligned}
&+ 2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + 2[\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2 \\
&\equiv L_n^*(\boldsymbol{\omega}) + \Phi_1(\boldsymbol{\omega}) + \Phi_2(\boldsymbol{\omega}) + \Phi_3(\boldsymbol{\omega}) + \Phi_4(\boldsymbol{\omega}) + \|\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|^2, \qquad (18)
\end{aligned}
$$

where

$$
\begin{aligned}
\Phi_1(\boldsymbol{\omega}) &= \|\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})\|^2, \\
\Phi_2(\boldsymbol{\omega}) &= -2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'[\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})], \\
\Phi_3(\boldsymbol{\omega}) &= 2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}),
\end{aligned}
$$

and

$$\Phi_4(\boldsymbol{\omega}) = 2[\tilde{\boldsymbol{\Delta}}(\boldsymbol{\omega}) - \boldsymbol{\Delta}^*(\boldsymbol{\omega})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}).$$

When $\|\tilde{\boldsymbol{\omega}}_F\| \equiv 0$, Lemma 3 is obviously true. Therefore, we only consider $\|\tilde{\boldsymbol{\omega}}_F\| > 0$. Let $\hat{\boldsymbol{\omega}}_T = (\hat{\omega}_1, \hat{\omega}_2, \ldots, 1 - \sum_{k=1}^{S_0-1} \hat{\omega}_k, 0, \ldots, 0)' \in \mathcal{W}$. By the definition of $\hat{\boldsymbol{\omega}}$ and (8), we have

$$
\begin{aligned}
0 &\geq \big\{\mathrm{CV}_n(\hat{\boldsymbol{\omega}}) - \mathrm{CV}_n(\hat{\boldsymbol{\omega}}_T)\big\}/n \\
&= \big\{L_n^*(\hat{\boldsymbol{\omega}}) - L_n^*(\hat{\boldsymbol{\omega}}_T) + \Phi_1(\hat{\boldsymbol{\omega}}) - \Phi_1(\hat{\boldsymbol{\omega}}_T) + \Phi_2(\hat{\boldsymbol{\omega}}) - \Phi_2(\hat{\boldsymbol{\omega}}_T) \\
&\quad + \Phi_3(\hat{\boldsymbol{\omega}}) - \Phi_3(\hat{\boldsymbol{\omega}}_T) + \Phi_4(\hat{\boldsymbol{\omega}}) - \Phi_4(\hat{\boldsymbol{\omega}}_T)\big\}/n.
\end{aligned}
$$

First,

$$
\begin{aligned}
&\frac{\big\{L_n^*(\hat{\boldsymbol{\omega}}) - L_n^*(\hat{\boldsymbol{\omega}}_T)}{n} \\
&= \frac{\|\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}) - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T)\|^2}{n} + \frac{2[\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}) - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T)]'[\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta}]}{n} \\
&= \tilde{\boldsymbol{\omega}}_F'\boldsymbol{\Psi}_n\tilde{\boldsymbol{\omega}}_F + \tilde{\boldsymbol{\omega}}_F'\Gamma_{n,1},
\end{aligned}
$$

where

$$
\begin{aligned}
&\boldsymbol{\Gamma}_{n,1} \\
&= \frac{2}{n}\big\{(\boldsymbol{\Delta}^{(S_0+1)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta}), \ldots, (\boldsymbol{\Delta}^{(S_n)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta})\big\}'.
\end{aligned}
$$

Second, we can rewrite $1/n[\Phi_1(\hat{\boldsymbol{\omega}}) - \Phi_1(\hat{\boldsymbol{\omega}}_T)]$ as

$$
\begin{aligned}
&\frac{\Phi_1(\hat{\boldsymbol{\omega}}) - \Phi_1(\hat{\boldsymbol{\omega}}_T)}{n} \\
&= \hat{\boldsymbol{\omega}}'\boldsymbol{A}_{n,1}\hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}_T'\boldsymbol{A}_{n,1}\hat{\boldsymbol{\omega}}_T \\
&= (\hat{\boldsymbol{\omega}}_T + \hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}_T)'\boldsymbol{A}_{n,1}(\hat{\boldsymbol{\omega}}_T + \hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}_T) - \hat{\boldsymbol{\omega}}_T'\boldsymbol{A}_{n,1}\hat{\boldsymbol{\omega}}_T \\
&= \tilde{\boldsymbol{\omega}}_F'\boldsymbol{D}_{n,1} + 2\tilde{\boldsymbol{\omega}}_F'\boldsymbol{\Gamma}_{n,2},
\end{aligned}
$$

where $\boldsymbol{A}_{n,1}$ is a $S_n \times S_n$ matrix with (i,j)th element is

$$\frac{1}{n}[\tilde{\boldsymbol{\Delta}}^{(i)} - \boldsymbol{\Delta}^{(i)*}]'[\tilde{\boldsymbol{\Delta}}^{(j)} - \boldsymbol{\Delta}^{(j)*}],$$

$$\boldsymbol{D}_{n,1} = \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix}' \boldsymbol{A}_{n,1} \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix},$$

and

$$\boldsymbol{\Gamma}_{n,2} = \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix}^T \boldsymbol{A}_{n,1} \hat{\boldsymbol{\omega}}_T$$

Third, we observe that

$$\begin{aligned} &\frac{\Phi_2(\hat{\boldsymbol{\omega}}) - \Phi_2(\hat{\boldsymbol{\omega}}_T)}{n} \\ &= \hat{\boldsymbol{\omega}}' \boldsymbol{A}_{n,2} \hat{\boldsymbol{\omega}} - \hat{\boldsymbol{\omega}}_T' \boldsymbol{A}_{n,2} \hat{\boldsymbol{\omega}}_T \\ &= \tilde{\boldsymbol{\omega}}_F' \boldsymbol{D}_{n,2} \tilde{\boldsymbol{\omega}}_F + \tilde{\boldsymbol{\omega}}_F' \boldsymbol{\Gamma}_{n,3}, \end{aligned}$$

where $\boldsymbol{A}_{n,2}$ is a $S_n \times S_n$ matrix with (i,j)th element is

$$\frac{2}{n} [\boldsymbol{\Delta} - \boldsymbol{\Delta}^{(i)*}]' [\boldsymbol{\Delta}^{(j)*} - \tilde{\boldsymbol{\Delta}}^{(j)}],$$

$$\boldsymbol{D}_{n,2} = \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix}' \boldsymbol{A}_{n,2} \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix},$$

and

$$\boldsymbol{\Gamma}_{n,3} = \begin{pmatrix} \mathbf{0}_{S_0 \times (K_n - S_0)} \\ -\mathbf{1}'_{K_n - S_0} \\ \boldsymbol{I}_{K_n - S_0} \end{pmatrix}' \boldsymbol{A}_{n,2} \hat{\boldsymbol{\omega}}_T.$$

Forth,

$$\begin{aligned} &\frac{\Phi_3(\hat{\boldsymbol{\omega}}) - \Phi_3(\hat{\boldsymbol{\omega}}_T)}{n} \\ &= \frac{2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{n} - \frac{2[\boldsymbol{\Delta} - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T)]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{n} \\ &= \frac{2[\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})}{n} \\ &= -\tilde{\boldsymbol{\omega}}_F' \boldsymbol{\Gamma}_{n,4}, \end{aligned}$$

where

$$\boldsymbol{\Gamma}_{n,4} = \frac{2}{n} \left\{ (\boldsymbol{\Delta}^{(S_0+1)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}), \ldots, (\boldsymbol{\Delta}^{(S_n)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) \right\}'.$$

Final, we have

$$\frac{\Phi_4(\hat{\boldsymbol{\omega}}) - \Phi_4(\hat{\boldsymbol{\omega}}_T)}{n}$$

$$
\begin{aligned}
&= \frac{2}{n}\{[\tilde{\boldsymbol{\Delta}}(\hat{\boldsymbol{\omega}}) - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}})]^{'}(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}) - [\tilde{\boldsymbol{\Delta}}(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T)]^{'}(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\} \\
&= \tilde{\boldsymbol{\omega}}_F^{'}\boldsymbol{\Gamma}_{n,5},
\end{aligned}
$$

where

$$
\begin{aligned}
&\boldsymbol{\Gamma}_{n,5} \\
&= \frac{2}{n}\big\{[(\tilde{\boldsymbol{\Delta}}^{(S_0+1)} - \boldsymbol{\Delta}^{(S_0+1)*}) - (\tilde{\boldsymbol{\Delta}}^{(S_0)} - \boldsymbol{\Delta}^{(S_0)*})]^{'}(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}), \ldots, \\
&\quad [(\tilde{\boldsymbol{\Delta}}^{(S_n)} - \boldsymbol{\Delta}^{(S_n)*}) - (\tilde{\boldsymbol{\Delta}}^{(S_0)} - \boldsymbol{\Delta}^{(S_0)*})]^{'}(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\big\}^{'}.
\end{aligned}
$$

Thus, we can rewrite (18) as

$$
\begin{aligned}
0 &\geq \big\{\mathrm{CV}_n(\hat{\boldsymbol{\omega}}) - \mathrm{CV}_n(\hat{\boldsymbol{\omega}}_T)\big\}/n \\
&= \tilde{\boldsymbol{\omega}}_F^{'}(\boldsymbol{\Psi}_n + \boldsymbol{D}_{n,1} + \boldsymbol{D}_{n,2})\tilde{\boldsymbol{\omega}}_F + \tilde{\boldsymbol{\omega}}_F^{'}(\boldsymbol{\Gamma}_{n,1} + 2\boldsymbol{\Gamma}_{n,2} + \boldsymbol{\Gamma}_{n,3} - \boldsymbol{\Gamma}_{n,4} + \boldsymbol{\Gamma}_{n,5}) \\
&= \tilde{\boldsymbol{\omega}}_F^{'}\boldsymbol{\Phi}_n\tilde{\boldsymbol{\omega}}_F + \tilde{\boldsymbol{\omega}}_F^{'}\boldsymbol{c}_n, \qquad (19)
\end{aligned}
$$

where $\boldsymbol{\Phi}_n = \boldsymbol{\Psi}_n + \boldsymbol{D}_{n,1} + \boldsymbol{D}_{n,2}$ and $\boldsymbol{c}_n = \boldsymbol{\Gamma}_{n,1} + 2\boldsymbol{\Gamma}_{n,2} + \boldsymbol{\Gamma}_{n,3} - \boldsymbol{\Gamma}_{n,4} + \boldsymbol{\Gamma}_{n,5}$.

Next, we will examine each term in (19) in turns. From the definition of Ψ_n, if follows from Assumption 2 that

$$
\begin{aligned}
&\|\boldsymbol{\Psi}_n\| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\zeta}_i\|^2 \\
&\leq \frac{\sum_{i=1}^{n}\sum_{s=S_0+1}^{S_n}(\Delta_i^{(s)*} - \Delta_i^{(S_0)*})^2}{n} \\
&= \frac{\sum_{s=S_0+1}^{S_n}\sum_{i=1}^{n}(\Delta_i^{(s)*} - \Delta_i^{(S_0)*})^2}{n} \\
&= \frac{\sum_{s=S_0+1}^{S_n}\|\boldsymbol{\Delta}^{(s)*} - \boldsymbol{\Delta}^{(S_0)*}\|^2}{n} \\
&= O_p(S_n - S_0). \qquad (20)
\end{aligned}
$$

By Cauchy inequality, we can obtain that

$$
\begin{aligned}
&\|\boldsymbol{\Gamma}_{n,1}\|^2 \\
&= \frac{4}{n^2}\sum_{k=S_0+1}^{S_n}[(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})^T(\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta})]^2 \\
&\leq \frac{1}{n^2}\sum_{k=S_0+1}^{S_n}\|\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*}\|^2\|\boldsymbol{\Delta}^*(\hat{\boldsymbol{\omega}}_T) - \boldsymbol{\Delta})\|^2 \\
&\leq \frac{1}{n^2}\sum_{k=S_0+1}^{S_n}\|\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*}\|^2 \times \left\|\sum_{s=1}^{S_0}\hat{\omega}_s(\Delta_i - \Delta_i^{*(s)}) + (1 - \sum_{s=1}^{S_0}\hat{\omega}_s)(\Delta_{S_0} - \Delta_i^{*(S_0)})\right\|^2
\end{aligned}
$$

$$= O_p(S_0^2(S_n - S_0)(\sqrt{m_n^\star/n} + (m_n^\star)^{-r})^2),$$

which implies

$$\|\boldsymbol{\Gamma}_{n,1}\| = O_p(S_0\sqrt{S_n - S_0}(\sqrt{m_n^\star/n} + (m_n^\star)^{-r})). \tag{21}$$

From (4) and (17), we obtain

$$\begin{aligned}
&\|\boldsymbol{A}_{n,1}\|^2 \\
&\le \sum_{i=1}^{S_n}\sum_{j=1}^{S_n} \frac{\{[\tilde{\boldsymbol{\Delta}}^{(i)} - \boldsymbol{\Delta}^{(i)*}]'[\tilde{\boldsymbol{\Delta}}^{(j)} - \boldsymbol{\Delta}^{(j)*}]\}^2}{n^2} \\
&\le \sum_{i=1}^{S_n}\sum_{j=1}^{S_n} \frac{\|\tilde{\boldsymbol{\Delta}}^{(i)} - \boldsymbol{\Delta}^{(i)*}\|^2\|\tilde{\boldsymbol{\Delta}}^{(j)} - \boldsymbol{\Delta}^{(j)*}\|^2}{n^2} \\
&= O_p(\frac{S_n^2(n\alpha_n^2 + p^{*2} + np^{*2}\alpha_n^2 + m_n^\star)^2}{n^2}),
\end{aligned} \tag{22}$$

Thus, we have

$$\begin{aligned}
&\|\boldsymbol{D}_{n,1}\| \\
&\le \left\| \begin{pmatrix} \mathbf{0}_{S_0\times(K_n-S_0)} \\ -\mathbf{1}'_{S_n-S_0} \\ \boldsymbol{I}_{S_n-S_0} \end{pmatrix} \right\|^2 \|\boldsymbol{A}_{n,1}\| \\
&= O_p(S_n(S_n - S_0)(\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n)),
\end{aligned} \tag{23}$$

and

$$\begin{aligned}
&\|\boldsymbol{\Gamma}_{n,2}\| \\
&\le \left\| \begin{pmatrix} \mathbf{0}_{S_0\times(K_n-S_0)} \\ -\mathbf{1}'_{S_n-S_0} \\ \boldsymbol{I}_{S_n-S_0} \end{pmatrix} \right\| \|\boldsymbol{A}_{n,1}\|\|\hat{\boldsymbol{\omega}}_T\| \\
&= O_p(S_nS_0\sqrt{S_n - S_0}(\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n)),
\end{aligned} \tag{24}$$

where the last equation follows from

$$\left\| \begin{pmatrix} \mathbf{0}_{S_0\times(S_n-S_0)} \\ -\mathbf{1}'_{S_n-S_0} \\ \boldsymbol{I}_{S_n-S_0} \end{pmatrix}' \right\| = O_p(\sqrt{K_n - S_0}). \tag{25}$$

Combining with Assumption 2, (4) and (17), it is seen that

$$\begin{aligned}
&\|\boldsymbol{A}_{n,2}\|^2 \\
&\le \frac{4}{n^2}\sum_{i=1}^{S_n}\sum_{j=1}^{S_n}\big\{[\boldsymbol{\Delta} - \boldsymbol{\Delta}^{(i)*}]^T[\boldsymbol{\Delta}^{(j)*} - \tilde{\boldsymbol{\Delta}}^{(j)}]\big\}^2
\end{aligned}$$

$$= O_p(S_n^2(\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n)),$$

which compels

$$\|\boldsymbol{A}_{n,2}\| = O_p(S_n\sqrt{\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n}). \tag{26}$$

Thus, it is straightforward to show that

$$\begin{aligned}
&\|\boldsymbol{D}_{n,2}\| \\
&\leq \left\| \begin{pmatrix} \mathbf{0}_{S_0\times(S_n-S_0)} \\ -\mathbf{1}'_{S_n-S_0} \\ \boldsymbol{I}_{S_n-S_0} \end{pmatrix} \right\|^2 \|\boldsymbol{A}_{n,2}\| \\
&= O_p(S_n(S_n-S_0)\sqrt{\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n}),
\end{aligned} \tag{27}$$

and

$$\begin{aligned}
\|\boldsymbol{\Gamma}_{n,3}\| &= \left\| \begin{pmatrix} \mathbf{0}_{S_0\times(K_n-S_0)} \\ -\mathbf{1}'_{S_n-S_0} \\ \boldsymbol{I}_{S_n-S_0} \end{pmatrix} \right\| \|\boldsymbol{A}_{n,2}\| \|\hat{\boldsymbol{\omega}}_T\| \\
&= O_p(S_n S_0\sqrt{S_n-S_0}\sqrt{\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^\star/n}).
\end{aligned} \tag{28}$$

By Assumption 2 and $\|\boldsymbol{\varepsilon}\| = O_p(n^{1/2})$, we can obtain that

$$\begin{aligned}
&E[(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})] \\
&\leq C(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})] \\
&= O(n).
\end{aligned}$$

Therefore, we reach that

$$\begin{aligned}
&\|\boldsymbol{\Gamma}_{n,4}\|^2 \\
&= \frac{4}{n^2}\sum_{k=S_0+1}^{S_n}\left\{(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right\}^2 \\
&\leq \frac{8}{n^2}\sum_{k=S_0+1}^{S_n}\left\{(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})'\boldsymbol{\varepsilon}\right\}^2 + \frac{8}{n^2}\sum_{k=S_0+1}^{S_n}\left\{(\boldsymbol{\Delta}^{(k)*} - \boldsymbol{\Delta}^{(S_0)*})'(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right\}^2 \\
&= O_p(\frac{S_n - S_0}{n} + (S_n - S_0)\alpha_n^2),
\end{aligned} \tag{29}$$

and

$$\begin{aligned}
&\|\boldsymbol{\Gamma}_{n,5}\|^2 \\
&= \frac{4}{n^2}\sum_{k=S_0+1}^{S_n}\left\{[(\tilde{\boldsymbol{\Delta}}^{(k)} - \boldsymbol{\Delta}^{(k)*}) - (\tilde{\boldsymbol{\Delta}}^{(S_0)} - \boldsymbol{\Delta}^{(S_0)*})]'(\boldsymbol{\varepsilon} + \hat{\boldsymbol{\xi}} - \boldsymbol{\xi})\right\}^2
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{8}{n^2}\sum_{k=S_0+1}^{S_n}\big\{[\tilde{\boldsymbol{\Delta}}^{(k)}-\boldsymbol{\Delta}^{(k)*}]'(\boldsymbol{\varepsilon}+\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})\big\}^2+\frac{8}{n^2}\sum_{k=S_0+1}^{S_n}\big\{[\tilde{\boldsymbol{\Delta}}^{(S_0)}-\boldsymbol{\Delta}^{(S_0)*}]'(\boldsymbol{\varepsilon}+\hat{\boldsymbol{\xi}}-\boldsymbol{\xi})\big\}^2\\
&\le \frac{C}{n^2}\sum_{k=S_0+1}^{S_n}\|\tilde{\boldsymbol{\Delta}}^{(k)}-\boldsymbol{\Delta}^{(k)*}\|^2\|\boldsymbol{\varepsilon}\|^2+\frac{C}{n^2}\sum_{k=S_0+1}^{S_n}\|\tilde{\boldsymbol{\Delta}}^{(k)}-\boldsymbol{\Delta}^{(k)*}\|^2\|\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2\\
&\quad+\frac{C}{n^2}\sum_{k=S_0+1}^{S_n}\|\tilde{\boldsymbol{\Delta}}^{(S_0)}-\boldsymbol{\Delta}^{(S_0)*}\|^2\|\boldsymbol{\varepsilon}\|^2+\frac{C}{n^2}\sum_{k=S_0+1}^{S_n}\|\tilde{\boldsymbol{\Delta}}^{(S_0)}-\boldsymbol{\Delta}^{(S_0)*}\|^2\|\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}\|^2\\
&= O_p((S_n-S_0)(p^{*2}/n+p^{*2}\alpha_n^2+m_n^\star/n)). \qquad (30)
\end{aligned}
$$

In addition, by (19), we obtain that

$$
\tilde{\boldsymbol{\omega}}_F'(\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n')\tilde{\boldsymbol{\omega}}_F \le 2\tilde{\boldsymbol{\omega}}_F'\boldsymbol{c}_n.
$$

Thus

$$
\begin{aligned}
&\|\tilde{\boldsymbol{\omega}}_F\|\\
&\le \underline{\lambda}^{-1}\big(\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n'\big)2\|c_n\|.
\end{aligned}
$$

Together with (21), (24) and (28)-(30), we know that

$$
\|\boldsymbol{c}_n\| = O_p(S_0\sqrt{S_n-S_0}S_n\sqrt{p^{*2}/n+p^{*2}\alpha_n^2+m_n^\star/n}). \qquad (31)
$$

Furthermore, for any nonzero vector $\varsigma\in\mathbb{R}^{S_n-S_0}$, we observe that

$$
\begin{aligned}
&\big\|(\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n')\varsigma\big\|^2\\
&\big\|\big[(\boldsymbol{\Psi}_n+\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2})+(\boldsymbol{\Psi}_n+\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2})'\big]\varsigma\big\|^2\\
&=\big\|\{2\boldsymbol{\Psi}_n-[-(2\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2}+\boldsymbol{D}_{n,2}')]\}\varsigma\big\|^2\\
&\ge\bigg|\|2\boldsymbol{\Psi}_{\boldsymbol{n}}\varsigma\|-\|(2\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2}+\boldsymbol{D}_{n,2}')\varsigma\|\bigg|^2\\
&=\|2\boldsymbol{\Psi}_n\varsigma\|^2-2\|2\boldsymbol{\Psi}_n\varsigma\|\|(2\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2}+\boldsymbol{D}_{n,2}')\varsigma\|+\|(2\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2}+\boldsymbol{D}_{n,2}')\varsigma\|^2\\
&\ge\|\varsigma\|^2\Big[\underline{\lambda}(2\boldsymbol{\Psi}_n)-2\|2\boldsymbol{\Psi}_n\|\|2\boldsymbol{D}_{n,1}+\boldsymbol{D}_{n,2}+\boldsymbol{D}_{n,2}'\|+\underline{\lambda}\big(\boldsymbol{D}_{n,3}\big)\Big]\\
&\ge\|\varsigma\|^2\big[2\underline{\lambda}(\boldsymbol{\Psi}_n)-8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,1}\|-8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,2}\|+\underline{\lambda}\big(\boldsymbol{D}_{n,3}\big)\big],
\end{aligned}
$$

which lead to

$$
\begin{aligned}
&\underline{\lambda}\bigg((\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n')'(\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n')\bigg)\\
&=\inf_{\varsigma\in\mathbb{R}^{S_n-S_0},\|\varsigma\|\neq 0}\frac{\big\|(\boldsymbol{\Phi}_n+\boldsymbol{\Phi}_n')\varsigma\big\|^2}{\|\varsigma\|^2}\\
&\ge 2\underline{\lambda}(\boldsymbol{\Psi}_n)-8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,1}\|-8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,2}\|+\underline{\lambda}\big(\boldsymbol{D}_{n,3}\big), \qquad (32)
\end{aligned}
$$

where $\boldsymbol{D}_{n,3} = 2\boldsymbol{D}_{n,1} + \boldsymbol{D}_{n,2} + \boldsymbol{D}_{n,2}^{'}$. By (19) and (32), it is straightforward to show that

$$\begin{aligned}
&\|\tilde{\boldsymbol{\omega}}_F\| \\
&\le \underline{\lambda}^{-1}\big((\boldsymbol{\Phi}_n + \boldsymbol{\Phi}_n^{'})\big)2\|\tilde{c}_n\| \\
&\le \frac{2\|\tilde{c}_n\|}{\big\{2\underline{\lambda}(\boldsymbol{\Psi}_n) - 8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,1}\| - 8\|\boldsymbol{\Psi}_n\|\|\boldsymbol{D}_{n,2}\| + \underline{\lambda}\big(\boldsymbol{D}_{n,3}\big)\big\}^{1/2}}.
\end{aligned}$$

Combining this with Assumptions 3 and 13, (20), (23), (27) and (31), we obtain that

$$\begin{aligned}
&\|\tilde{\boldsymbol{\omega}}_F\| \\
&= \frac{O_p(S_n\sqrt{S_n - S_0}S_n\sqrt{p^{*2}/n + p^{*2}\alpha_n^2 + m_n^{\star}/n})}{[C_2 + O_p((S_n - S_0)^2 S_n\sqrt{\alpha_n^2 + p^{*2}/n + p^{*2}\alpha_n^2 + m_n^{\star}/n}]^{1/2}} \\
&= O_p(S_n\sqrt{S_n - S_0}S_0\sqrt{p^{*2}/n + p^{*2}\alpha_n^2 + m_n^{\star}/n}).
\end{aligned}$$

This completes the proof of Lemma 3.

5 Proof of Theorem 4

First, for $s \in \{S_0 + 1, \dots, S_n\}$, it follows from that

$$\begin{aligned}
&\hat{\Delta}_i^{(k)} \\
&= \boldsymbol{e}_i^T \boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{e}(\boldsymbol{x})))\hat{\boldsymbol{\xi}} \\
&\le \|\boldsymbol{e}_i\|_\infty \|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{e}(\boldsymbol{x})))\|_\infty \|\boldsymbol{\xi}\|_\infty + \|\boldsymbol{e}_i\|_\infty \|\boldsymbol{K}_{(s)}(\hat{\boldsymbol{\beta}}_s(\hat{e}(\boldsymbol{x})))\|_\infty \|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\|_\infty \\
&= O_p(1).
\end{aligned} \tag{33}$$

Combining with Lemmas 2-3, and (33), we get

$$\begin{aligned}
&|\hat{\Delta}_i(\hat{\boldsymbol{\omega}}) - \tau(\boldsymbol{x}_i)| \\
&= \Big|\sum_{k=1}^{S_0} \hat{\omega}_k(\hat{\Delta}_i^{(k)} - \tau(\boldsymbol{x}_i)) + \sum_{k=S_0+1}^{S_n} \hat{\omega}_k(\hat{\Delta}_i^{(k)} - \tau(\boldsymbol{x}_i))\Big| \\
&\le \sum_{k=1}^{S_0} \hat{\omega}_k|\hat{\Delta}_i^{(k)} - \tau(\boldsymbol{x}_i)| + \sum_{k=S_0+1}^{S_n} \hat{\omega}_k|\hat{\Delta}_i^{(k)} - \tau(\boldsymbol{x}_i)| \\
&\le \sum_{k=1}^{S_0} \hat{\omega}_k|\hat{\Delta}_i^{(k)} - \tau(\boldsymbol{x}_i)| + \|\tilde{\boldsymbol{\omega}}_F\| \sum_{k=S_0+1}^{S_n} |\hat{\Delta}_i^{(k)} - \Delta_i| \\
&= O_p(S_0 p^* n^{-1/2} + S_0 p^* \alpha_n + S_0(\sqrt{m_n^{\star}/n} + m_n^{\star -r})) \\
&\quad + O_p(S_n(S_n - S_0)^{3/2} S_0\sqrt{p^{*2}/n + p^{*2}\alpha_n^2 + m_n^{\star}/n}) \\
&= O_p(S_n(S_n - S_0)^{3/2} S_0\sqrt{p^{*2}/n + p^{*2}\alpha_n^2 + m_n^{\star}/n}).
\end{aligned}$$

This completes the proof.

References

Zou, J., Wang, W., Zhang, X., Zou, G., 2021. Optimal model averaging for single-index models with divergent dimensions. arXiv preprint arXiv:2112.15100 .