OPTIMAL MODEL AVERAGING FOR SINGLE-INDEX MODELS WITH DIVERGENT DIMENSIONS

Jiahui Zou[a], Wendun Wang[b] and Xinyu Zhang[c] and Guohua Zou[d]

Capital University of Economics and Business[a],

Erasmus University Rotterdam[b],

Chinese Academy of Sciences[c]

and Capital Normal University[d]

Online Supplement

This file contains some explanations of conditions, technical proofs and other results of simulation studies. Specifically, Section S1 provides explanations of conditions in Appendix and conditions needed for Corrolary 2. Section S2 contains detailed simulation setup and additional numerical results. Section S3 presents the proofs of lemmas, theorems and corollaries. Section S4 discusses the related methods.

S1 Conditions

This section provides some detailed explanations for the additional conditions in Appendix.

S1.1 Conditions for Lemma 1

Condition S.1. (i) The kernel function $k(s)$ is a bounded symmetric density with a compact support. (ii) The following quantities are finite: $\int |\tau k'(\tau)|\,\mathrm{d}\tau$, $\int \tau^2|k'(\tau)|\,\mathrm{d}\tau$, $\int k'^2(\tau)\,\mathrm{d}\tau$, $\int |\tau|k'^2(\tau)\,\mathrm{d}\tau$ and $\int \tau^2 k'^2(\tau)\,\mathrm{d}\tau$, where $k'(s)$ is the first-order derivative of $k(s)$.

These are common restrictions on the kernel function in nonparametric statistics, such as Lemmas .2–.4 in Ichimura (1993) and Condition (C.5) in Zhu et al. (2019).

Condition S.2. (i) $\max_{1\le s\le S_n} h_s \to 0$. (ii) $\sum_{s=1}^{S_n} n^{-1}h_s^{-3}p_s = O_P(1)$. (iii) $\max_{1\le s\le S_n}(nh_s^4 + h_s^{-1})/M_n^2 n d_n^2 = O(1)$.

This condition pertains to the bandwidth of the averaging estimator. S_n and p_s appear because we need to solve S_n candidate models simultaneously. The similar conditions are also used in Condition (C5) in Wang et al. (2011) and Condition (C.5) in Zhu et al. (2019).

Condition S.3. (i) There exists a universal constant $\bar{C} > 0$ such that

$\max_{1\le s\le S_n}\max_{1\le i\le n}\|\mathbf{x}_{(s),i}\| \le \sqrt{p_s}\bar{C}$. (ii) $\max_{1\le s\le S_n}\max_{1\le i\le n}|\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*|$ and $\max_{1\le s\le S_n}\max_{1\le i\le n}|x_{(s),ir}\beta_r^*|$ are bounded. (iii) $\max_{1\le s\le S_n}\max_{1\le i\le n}$ $\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right| = O_P(1)$. (iv) There exists a constant $\underline{c}$ such that $\min_{1\le s\le S_n}$ $\min_{r:1\le r\le p_s,\beta_{(s),r}^*\neq 0}|\beta_{(s),r}^*| > \underline{c} > 0$. (v) $\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\|\partial\sum_{i=1}^n$ $\{\mu_i - \sum_{j\neq i}^n K_{(s),ij}(\boldsymbol{\beta}_{(s)}^*)\mu_j\}^2/\partial\boldsymbol{\beta}_{(s)}\| = O_P(1)$.

Condition S.3(i) holds if each element of $\mathbf{x}_i$ is uniformly bounded, an assumption also imposed by Radchenko (2015, Assumption A1). Conditions S.3(ii) and S.3(iii) require that the quasi-true parameter is not abnormal so that the estimator for $\boldsymbol{\beta}_{(s)}$ is well-behaved. Condition S.3(iv) guarantees that the nonzero parameters, $\beta_{(s),r}^*$, have a uniform lower bound. Condition S.3(v) requires that the difference between $\boldsymbol{\mu}$ and the theoretical estimator from the s^{th} candidate model, $\mathbf{K}_{(s)}(\boldsymbol{\beta}_{(s)}^*)\boldsymbol{\mu}$, is smooth enough around $\boldsymbol{\beta}_{(s)}^*$ such that there is sufficient information to estimate the quasi-true parameter $\boldsymbol{\beta}_{(s)}^*$. $\sqrt{p_s}$ and $\sqrt{S_n}$ appear in the left-side denominator in this condition, because $\|\partial\sum_{i=1}^n\{\mu_i - \sum_{j\neq i}^n K_{(s),ij}(\boldsymbol{\beta}_{(s)}^*)\mu_j\}^2/\partial\boldsymbol{\beta}_{(s)}\|$ is of order $\sqrt{p_s}$ and there are S_n candidate models.

Let $\rho_{(s)}(v_1,\ldots,v_{p_s})$ denote the joint density function of $x_{(s),1}\beta_1^*,\ldots,x_{(s),p_s}\beta_{p_s}^*$ for the s^{th} candidate model, where $\mathbf{x}_{(s)} = (x_{(s),1},\ldots,x_{(s),p_s})^{\mathrm{T}}$ and $\boldsymbol{\beta}_{(s)}^* = (\beta_1^*,\ldots,\beta_{p_s}^*)^{\mathrm{T}}$. Let $f_{(s)}(t)$ denote the density of $\mathbf{x}_{(s)}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*$, and denote $f'_{(s)}(t)$ and $f''_{(s)}(t)$ as the first and second-order derivatives of $f_{(s)}(t)$, respectively.

Further let $\phi_{(s)}(t) = g_{(s)}(t) f_{(s)}(t)$ and $\varphi_{(s)}(t) = g^2_{(s)}(t) f_{(s)}(t)$.

Condition S.4. (i) There exists a constant $\bar{C}$ such that

$$\int \rho_{(s)} \left(v_1, \ldots, v_{k-1}, t - \sum_{l \neq k}^{p_s} v_l, v_{k+1}, \ldots, v_{p_s} \right) \mathrm{d}v_1 \ldots \mathrm{d}v_{k-1} \, \mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s} < \bar{C}$$

uniformly for s and t. (ii) There exist some constants $\underline{c}$ and $\bar{C}$ such that $\underline{c} < f_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i} \boldsymbol{\beta}^*_{(s)}) < \bar{C}$ almost surely for $s = 1, \ldots, S_n$; $i = 1, \ldots, n$. (iii) There exists a universal constant $\bar{C}$ such that $|f'_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i} \boldsymbol{\beta}^*_{(s)})| < \bar{C}$, $|f''_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i} \boldsymbol{\beta}^*_{(s)})| < \bar{C}$ almost surely for $s = 1, \ldots, S_n$; $i = 1, \ldots, n$. (iv) There exists a constant $G > 0$ and $\omega_{(s)} \left(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s} \right) > 0$ such that $\left| \rho_{(s)} \left(v_1, \ldots, v_{k-1}, t_1, v_{k+1}, \ldots, v_{p_s} \right) - \rho_{(s)} \left(v_1, \ldots, v_{k-1}, t_2, v_{k+1}, \ldots, v_{p_s} \right) \right|$ $\leq G \omega_{(s)} \left(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s} \right) |t_1 - t_2|$, for any s and k, where $\int \omega_{(s)} \left(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s} \right) \mathrm{d}v_1 \ldots \mathrm{d}v_{k-1} \, \mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s} < \infty$ and $\sum_{r=1, r \neq k}^{p_s}$ $\int |v_l| \omega_{(s)} \left(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s} \right) \mathrm{d}v_1 \ldots \mathrm{d}v_{k-1} \, \mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s} < \infty$ uniformly for any s. (v) $f'_{(s)}(t)$ and $f''_{(s)}(t)$ satisfy the Lipschitz condition, i.e., there exist two constants c_1 and c_2 such that $|f'_{(s)}(t_1) - f'_{(s)}(t_2)| \leq c_1 |t_1 - t_2|$ and $|f''_{(s)}(t_1) - f''_{(s)}(t_2)| \leq c_2 |t_1 - t_2|$; (vi) $\phi'_{(s)}(t)$ and $\varphi'_{(s)}(t)$ satisfy the Lipschitz condition.

This condition imposes restrictions on the joint density of $x_{(s),1} \beta^*_1, \ldots,$ $x_{(s),p_s} \beta^*_{p_s}$ and the density of $\sum_{r=1}^{p_s} x_{(s),r} \beta^*_r$. Especially, when $\{x_{(s),r}\}_{r=1}^{p_s}$ are independent, $f_{(s)}(t) = \int \rho_{(s)}(v_1, \ldots, v_{k-1}, t - \sum_{l \neq k}^{p_s} v_l, v_{k+1}, \ldots, v_{p_s}) \, \mathrm{d}v_1 \ldots$ $\mathrm{d}v_{k-1} \, \mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s}$ because of the convolution product. To illustrate this condition, we can consider the simple case where each $x_{(s),r}$ is i.i.d. $N(0,1)$, then $\mathbf{x}_{(s),r}\beta_r^* \sim N(0, \beta_r^{*2})$ and

$$f_{(s)}(t) = \int \rho_{(s)}\left(v_1, \dots, v_{k-1}, t - \sum_{l\neq k}^{p_s} v_l, v_{k+1}, \dots, v_{p_s}\right) \mathrm{d}v_1 \dots \mathrm{d}v_{k-1}\, \mathrm{d}v_{k+1} \dots \mathrm{d}v_{p_s}$$

is the density of $N(0, \sum_{r=1}^{p_s} \beta_{(s),r}^{*2})$. In Conditions S.4($i$) and S.4($ii$), if $\beta_r^* = 1$ for $r = 1, \dots, p_s$, then $|f_{(s)}(t)| \leq (2\pi)^{-1/2}$ and we can take $\bar{C} = (2\pi)^{-1/2}$. Condition S.4(ii) is also similar to Condition (C.2) in Zhu et al. (2019).

Condition S.4(iii) ensures that $f'_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)$ and $f''_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)$ are both uniformly bounded. From the discussion of Condition S.4(i), we have $|f'_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)| = (2\pi e)^{-1/2}\sum_{r=1}^{p_s}(\beta_{(s),r}^{*2})^{-1} \leq (2\pi e)^{-1/2}$ and $|f''_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)| = (2\pi)^{-1/2}\sum_{r=1}^{p_s}(\beta_{(s),r}^{*2})^{-3/2} \leq (2\pi)^{-1/2}$ uniformly for s and i.

Condition S.4(iv) guarantees that the joint density is Lipschitz continuous so that the data are smooth around $x_{(s),r}^{\mathrm{T}}\beta_{(s),r}^*$. For example, if each $x_{(s),r}$ is i.i.d. from $N(0,1)$,

$$\begin{aligned}
&\left|\rho_{(s)}(v_1, \dots, v_{k-1}, t_1, v_{k+1}, \dots, v_{p_s}) - \rho_{(s)}(v_1, \dots, v_{k-1}, t_2, v_{k+1}, \dots, v_{p_s})\right| \\
&\leq \frac{1}{(2\pi)^{(p_s-1)/2}\prod_{i=1,i\neq k}^{p_s}\beta_i^*}\exp\left(-\sum_{i=1,i\neq k}^{p_s}\frac{v_i^2}{2\beta_i^{*2}}\right)\frac{1}{(2\pi)^{1/2}\beta_k^*}\left|\exp\left(-\frac{t_1^2}{2\beta_k^{*2}}\right) - \exp\left(-\frac{t_2^2}{2\beta_k^{*2}}\right)\right| \\
&\leq \frac{1}{(2\pi)^{(p_s-1)/2}\prod_{i=1,i\neq k}^{p_s}\beta_i^*}\exp\left(-\sum_{i=1,i\neq k}^{p_s}\frac{v_i^2}{2\beta_i^{*2}}\right)\cdot\frac{1}{(2\pi e)^{1/2}\beta_k^{*2}}|t_1 - t_2|.
\end{aligned}$$

Then, we can take $G = (2\pi)^{-1/2}\underline{c}^{*-2}\exp\{-1/2\}$ according to Condition S.3(iv) and

$$\omega_{(s)}(v_1, \dots, v_{(k-1)}, v_{(k+1)}, \dots, v_{p_s}) =$$

$$(2\pi)^{-(p_s-1)/2}\left(\prod_{i=1,i\neq k}^{p_s}\beta_i^*\right)^{-1}\exp\left\{-\sum_{i=1,i\neq k}^{p_s}2^{-1}v_i^2\beta_i^{*-2}\right\}.$$

Further, if $\beta_r^* = 1$ for $r = 1, \ldots, p_s$, we have

$$\int \omega_{(s)}(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s})\,\mathrm{d}v_1 \ldots \mathrm{d}v_{k-1}\,\mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s} = 1 \quad \text{and}$$

$$\sum_{l=1,l\neq k}^{p_s}\int |v_l|\omega_{(s)}(v_1, \ldots, v_{k-1}, v_{k+1}, \ldots, v_{p_s})\,\mathrm{d}v_1 \ldots \mathrm{d}v_{k-1}\,\mathrm{d}v_{k+1} \ldots \mathrm{d}v_{p_s}$$
$$= (2/\pi)^{(p_s-1)/2} < 1.$$

Condition S.4(v) requires that the first and second-order derivatives of the density $f_{(s)}(\cdot)$ have a stronger continuous property so that there is enough information around the quasi-true parameter $\boldsymbol{\beta}^*_{(s)}$. When each $x_{(s),r}$ is i.i.d. from $N(0,1)$, we have $|f'_{(s)}(t_1) - f'_{(s)}(t_2)| \leq (2\pi)^{-1/2}|t_1 - t_2|$ and $|f''_{(s)}(t_1) - f''_{(s)}(t_2)| \leq (2/\pi)^{1/2}|t_1 - t_2|$. Conditions S.4($v$) and S.4($vi$) are also similarly used in Lemmas .2–.4 of Ichimura (1993) and Condition (ii) of Liang et al. (2010). Note that we require uniformity across s in this condition because our averaging allows the number of candidate models to diverge.

Condition S.5. (i)For any $s = 1, \ldots, S_n$, the objective function $H_{(s),n}(\boldsymbol{\beta}_{(s)})$ defined in (2.1) is twice continuously differentiable. (ii) There exists a constant $c_0 > 0$ such that

$$\min\left[\min_{1\le s\le S_n}\lambda_{\min}\left\{\frac{\partial^2 H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\right\},\min_{\substack{1\le s\le S_n\\1\le j\le J_n}}\lambda_{\min}\left\{\frac{\partial^2 H^{[-j]}_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\right\}\right]\ge c_0>0.$$

Condition A1.5 is crucial for the convergence of $\widehat{\boldsymbol{\beta}}_{(s)}$ and $\widehat{\boldsymbol{\beta}}^{[-j]}_{(s)}$. Condition S.5($i$) is the same as the condition of Lemma 5.4 in Ichimura (1993), and it requires the smoothness of the objective function. Condition S.5(ii) holds if there exists a local maximum for every s and can be roughly regarded as a minimum-eigenvalue requirement of the "Fisher information" matrix. When J_n and S_n are divergent, we need to restrict the eigenvalues of "Fisher information" across all blocks and candidate models, and thus $\min_{1\le s\le S_n}$ and $\min_{1\le j\le J_n}$ are needed.

S1.2 Conditions for Corollary 2

The following conditions are required to prove Corollary 2.

Condition S.6. For $s = 1,\ldots,S_n$ and $r = 1,\ldots,p_s$, the r^{th} element of $\widehat{\boldsymbol{\beta}}^{R}_{(s)}$ obtained from (4.13), $\widehat{\beta}^{R}_{(s),r}$, has a limiting value $\beta^{R*}_{(s),r}$. Furthermore,

$$\max_{1\le s\le S_n}\left\|\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right\| = O_P(n^{\alpha-1/2}S_n^{\gamma}),$$

$$\max_{1\le s\le S_n}\max_{1\le j\le J_n}\left\|\widehat{\boldsymbol{\beta}}^{R[-j]}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right\| = O_P\left\{(n-M_n)^{\alpha-1/2}S_n^{\gamma}\right\},$$

where $\alpha\in(0,1/2)$, $\gamma>0$, and $\widehat{\boldsymbol{\beta}}^{R[-j]}_{(s)}$ is the CV estimator obtained from the regularized estimation in (4.13), but excluding observations of the j^{th}

block.

This condition imposes restrictions on the distances between the quasi-true parameter $\boldsymbol{\beta}_{(s)}^{R*}$ and the two regularized estimators, $\widehat{\boldsymbol{\beta}}_{(s)}^{R}$ and $\widehat{\boldsymbol{\beta}}_{(s)}^{R[-j]}$, respectively. Similar results have been shown by Radchenko (2015) in a study of regularized estimators for SIMs, and the author focused on the deviation between the estimator and the true parameter as their fitted model coincides with the DGP. With $\widehat{\boldsymbol{\beta}}_{(s)}^{R[-j]}$ and $\boldsymbol{\beta}_{(s)}^{R*}$ given above, we can similarly define $\boldsymbol{\mu}^{R*}(\mathbf{w})$, $\widetilde{\boldsymbol{\mu}}^{R}(\mathbf{w})$ and $\widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})$ as the averaging estimators of $\boldsymbol{\mu}$ using $\boldsymbol{\beta}_{(s)}^{R*}$, $\widetilde{\boldsymbol{\beta}}_{(s)}^{R}$ and $\widetilde{\boldsymbol{\beta}}_{(s)}^{R*}$, respectively. Furthermore, we denote $L_n^{R*}(\mathbf{w})$ as the squared loss of the regularization-based averaging estimator and $\xi_n^R = \inf_{\mathbf{w}\in\mathcal{W}} L_n^{R*}(\mathbf{w})$.

For $s = 1, \ldots, S_n$, let $\mathcal{C}_s = \left\{ r \middle| \widehat{\beta}_{(s),r}^{R} \neq 0 \text{ or } \beta_{(s),r}^{R*} \neq 0 \right\}$ and denote the cardinality of $\mathcal{C}_s$ as q_s.

Condition S.7. (i) $\xi_n^{R^{-1}} S_n^{\gamma} n q_{\max}^{1/2} (n - M_n)^{\alpha-1/2} = o_P(1)$, where $q_{\max} = \max_{1\le s\le S_n} q_s$. (ii) $\xi_n^{R-1} d_n M_n n = o_P(1)$.

This condition is a regularization version of Condition 6, and provides restrictions on the relative divergent speeds of S_n, ξ_n^R, $q_{\max}$ and d_n. Compared with Condition 6(i), Condition S.7(i) concerns $q_{\max}$ rather than $p_{\max}$, and it is likely to hold if $q_{\max}$ grows at a slow speed. Note that q_s is related with the "variable selection consistency", an important property of Lasso-type estimators (see, e.g., Zou, 2006; Leng, 2010).

Condition S.8. There exists a positive ρ such that, for any $s = 1, \ldots, S_n$ and any $p_s \times 1$ vector $\mathbf{e}_{(s)}$ consisting of 1 or 0, we have

$$\lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{e}_{(s)} \odot \frac{\partial \widehat{g}_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}}\right)\left(\frac{\partial \widehat{g}_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}^{\mathrm{T}}} \odot \mathbf{e}_{(s)}^{\mathrm{T}}\right)\right\} = O_P(\|\mathbf{e}_{(s)}\|_1), \quad \text{(S1.1)}$$

and

$$\lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{e}_{(s)} \odot \frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}}\right)\left(\frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}^{\mathrm{T}}} \odot \mathbf{e}_{(s)}^{\mathrm{T}}\right)\right\} = O_P(\|\mathbf{e}_{(s)}\|_1), \quad \text{(S1.2)}$$

uniformly for any $\boldsymbol{\beta}_{(s)}^{(1)}, \ldots, \boldsymbol{\beta}_{(s)}^{(n)} \in \mathcal{O}(\boldsymbol{\beta}_{(s)}^{*}, \rho)$, where $\odot$ denotes the Hadamard product.

This condition is an extension of Condition 5. It degrades to Condition 5 if $\mathbf{e}_{(s)}^{\mathrm{T}} = (1, \ldots, 1)_{1 \times p_s}$.

S2 Simulation setup and additional results

S2.1 Benchmark experimental designs

To verify the theory in Section 3, we consider two exemplifying nonlinear functions that associate the response variable and covariates. For each nonlinear function, we study two cases that differ in the dimension of covariates. First, we fix the dimension of covariates to be finite. Second, we allow the dimension of covariates and the number of candidate models to be divergent. Furthermore, for each of the cases, we consider whether the correct models are included in the set of candidate models.

Example 1: We follow Naik and Tsai (2001) to consider the following DGP

$$y_i = \mu_i + c\epsilon_i = \sin(\pi \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}/6) + c\epsilon_i \quad (i = 1, \ldots, n),$$

where $\mathbf{x}_i$ is a $p \times 1$ vector generated from a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma} = (0.5^{|i-j|})_{p\times p}$. The settings of p and $\boldsymbol{\beta}$ vary across the four situations as specified below. ϵ_i is i.i.d. and follows a standard normal distribution. c controls the signal-to-noise ratio, and we vary c such that $R^2 = \mathrm{var}(\mu_i)/\mathrm{var}(y_i)$ ranges from 0.1 to 0.9.

Example 2: We follow Kong and Xia (2007) and Ichimura (1993) to consider the Tobit DGP as

$$y_i = (\mu_i + c\epsilon_i) I(\mu_i + c\epsilon_i > 0) = (\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + c\epsilon_i) I(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} + c\epsilon_i > 0) \quad (i = 1, \ldots, n),$$

where $I(\cdot)$ is an indicator function, and the remaining settings are the same as Example 1.

For each nonlinear link function, we consider the following four situations.

(1) **Finite dimension with all candidate models misspecified**

We fix $p = 7$ and set the coefficient vector as $\boldsymbol{\beta} = (1, 1.5, 1, 0, 0.1, -1.5, 1.5)^{\mathrm{T}}$. To construct misspecified candidate models, we include the first covariate but omit the last in all candidate models. The remaining covariates are uncertain, but at least one of them is included, which leads to $S_n = 2^5 - 1 = 31$

candidate models.

(2) **Finite dimension with correct candidate models**

In this case, we set $\boldsymbol{\beta} = (1, 1.5, 0, 1, 0, -1.5, 1.5)^{\mathrm{T}}$, where two of the coefficients are set to zero to increase the number of correct models for demonstration purposes. All candidate models include the first and last covariates but differ in the specification of the remainders, so correct models are contained in the candidate model set.

(3) **Divergent dimension with all candidate models misspecified**

To mimic the cases where the dimension of covariates increases with the sample size n, we set $\boldsymbol{\beta} = (1, (\underline{1.5, 1, 0, 0.1, -1.5}, \underline{1.5, 1, 0, 0.1, -1.5}, \ldots)_{\lceil 1.5n^{1/3} \rceil}, 1, 1.5)^{\mathrm{T}}$, where the subscript $\lceil 1.5n^{1/3} \rceil$ is the speed at which the dimension of $\boldsymbol{\beta}$ increases and $\lceil \cdot \rceil$ ascertains the ceiling of a number. We include the first covariate but omit the last two in all candidate models to construct misspecified candidate models as above. To reduce the computational burden, we employ the pre-screening method based on an ordering of covariates as discussed in Section 4, such that the number of candidate models also increases at a rate of $\lceil 1.5n^{1/3} \rceil$.

(4) **Divergent dimension with correct candidate models**

The setting is similar to (3), except that we include the first and last two covariates in all candidate models and set

$$\boldsymbol{\beta} = (1, (\underline{1.5, 0, 1, 0, 0, -1.5, 0}, \underline{1.5, 0, 1, 0, 0, -1.5, 0}, \ldots)_{\lceil 1.5n^{1/3} \rceil}, 1, 1.5)^{\mathrm{T}}.$$

We consider the sample sizes for estimation as $n = 100$, 200, 300, 400 and 500, and set the testing size as 1,000; all results are reported based on $D = 1000$ replications.

S2.2 Implementation and comparison

To implement the proposed JCVMA, we set the number of observations in each CV block to be $M_n = 50$. Robustness checks suggest that the results are qualitatively similar as long as there are sufficient observations in each block. Following Yu et al. (2014), we suggest to take the bandwidth of order $\kappa n^{-1/5} \log^{-1/6}(n)$ and choose the optimal κ via cross-validation. We follow the convention to use the Gaussian kernel $K(u) = \exp(-u^2/2)/\sqrt{2\pi}$ when estimating each candidate model.

We compare JCVMA with three information criteria: AIC,BIC, and a variant of AIC, which is designed especially for SIMs. The AIC and BIC scores of the s^{th} candidate model are given by

$$\mathrm{AIC}_s = n \log(\widehat{\sigma}_s^2) + 2\mathrm{trace}\{\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\},$$

$$\mathrm{BIC}_s = n \log(\widehat{\sigma}_s^2) + \log(n)\mathrm{trace}\{\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\},$$

where $\widehat{\sigma}_s^2 = n^{-1}\|\mathbf{y} - \widehat{\boldsymbol{\mu}}_{(s)}\|^2$. Naik and Tsai (2001) proposed a variant of AIC based on the Kullback-Leibler distance as

$$\mathrm{AICC}_s = \log(\widehat{\sigma}_s^2) + \frac{n + \mathrm{trace}\left\{\widehat{\mathbf{H}}_{(s)} + \mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)}) - \widehat{\mathbf{H}}_{(s)}\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\right\}}{n - 2 - \mathrm{trace}\left\{\widehat{\mathbf{H}}_{(s)} + \mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)}) - \widehat{\mathbf{H}}_{(s)}\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\right\}},$$

where $\widehat{\mathbf{H}}_{(s)} = \widehat{\mathbf{V}}_{(s)}(\widehat{\mathbf{V}}_{(s)}^{\mathrm{T}}\widehat{\mathbf{V}}_{(s)})^{-1}\widehat{\mathbf{V}}_{(s)}^{\mathrm{T}}$ with

$$\begin{aligned}\widehat{\mathbf{V}}_{(s)} &= \left\{\partial\widehat{g}_{(s)}(\mathbf{x}_{(s),1}^{\mathrm{T}}\boldsymbol{\beta}_{(s)})/\partial\boldsymbol{\beta}_{(s)}, \ldots, \partial\widehat{g}_{(s)}(\mathbf{x}_{(s),n}^{\mathrm{T}}\boldsymbol{\beta}_{(s)})/\partial\boldsymbol{\beta}_{(s)}\right\}^{\mathrm{T}}\Big|_{\boldsymbol{\beta}_{(s)}=\widehat{\boldsymbol{\beta}}_{(s)}} \\ &= \left\{\widehat{g}'_{(s)}(\mathbf{x}_{(s),1}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{(s)})\mathbf{x}_{(s)}, \ldots, \widehat{g}'_{(s)}(\mathbf{x}_{(s),n}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{(s)})\mathbf{x}_{(s)}\right\}^{\mathrm{T}},\end{aligned}$$

and $\widehat{g}'_{(s)}(\cdot)$ denotes the derivative of $\widehat{g}_{(s)}(\cdot)$.

We also compare the smoothed versions of the three information criteria, which use the values of the criteria for each candidate model as weights to construct the averaging estimators, namely, SAIC, SBIC and SAICC, e.g., $\mathrm{SAICC}_s = \exp(-\mathrm{AICC}_s/2)/\sum_{l=1}^{S_n}\exp(-\mathrm{AICC}_l/2)$.

We evaluate the performance of the methods from three perspectives. First, since our theory shows the asymptotic optimality of the JCVMA, we report the relative squared loss of each method with respect to the best possible averaging estimator, namely

$$D^{-1}\sum_{d=1}^{D} L_n^{(d)}/\inf_{\mathbf{w}\in\mathcal{W}} L_n^{(d)}(\mathbf{w}),$$

where $\boldsymbol{\mu}^{(d)}$ is the true value of the testing set, $\widehat{\boldsymbol{\mu}}^{(d)}$ is the predicted value produced by each method, $L_n^{(d)} = \|\widehat{\boldsymbol{\mu}}^{(d)} - \boldsymbol{\mu}^{(d)}\|^2$ is the loss, $\inf_{\mathbf{w}\in\mathcal{W}} L_n^{(d)}(\mathbf{w})$

is the minimum squared loss over all possible averaging estimators, and the superscript (d) denotes the d^{th} replication.

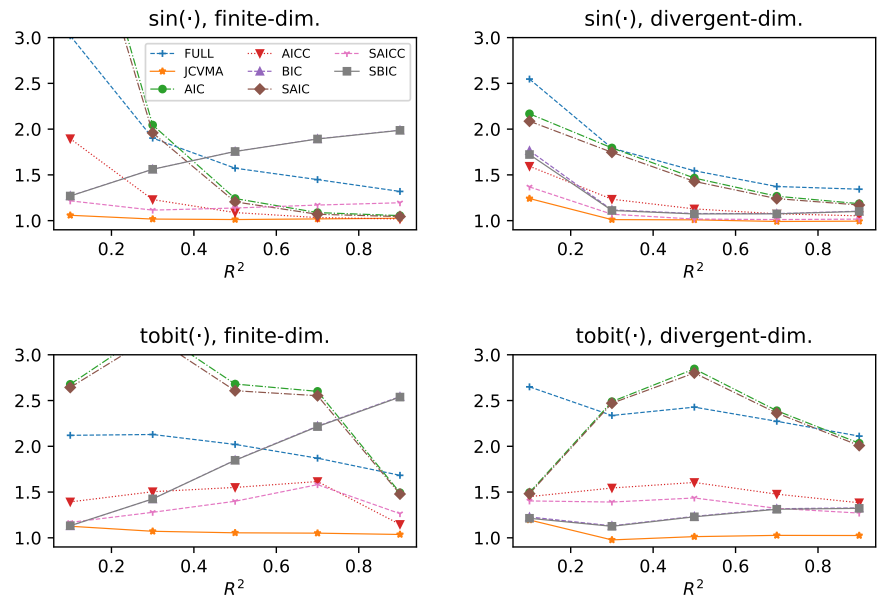
Second, we compare the prediction performance of various methods using the normalized mean squared prediction error (NMSPE), which is defined by NMSPE $= D^{-1}\sum_{d=1}^{D} L_n^{(d)}/L_{\min}^{(d)}$, where $L_{\min}^{(d)}$ is the minimum squared loss over all candidate models.

Finally, to verify the consistency of weights when correct models exist in the candidate model set as shown in Theorem 2, we plot the weights assigned to the correct models when n increases. We also examine the validity of Corollary 1 by reporting the relative squared loss of JCVMA with respect to the best possible averaging estimators using only misspecified models, namely $D^{-1}\sum_{d=1}^{D} L_n^{(d)}(\widehat{\mathbf{w}})/\inf_{\mathbf{w}\in\mathcal{W}_F} L_n^{(d)}(\mathbf{w})$, where $\widehat{\mathbf{w}}$ is the JCVMA weight vector obtained by minimizing $\text{CV}_{J_n}(\mathbf{w})$ as in (2.3).

S2.3 Simulation results

The upper four diagrams in Figure S.1 compare the NMSPEs of the eight methods when all candidate models are misspecified. We report the results of $n = 300$, while those of other sample sizes are highly similar and are thus provided in the Online Supplement. Again, we find that JCVMA produces the lowest NMSPE in almost all cases, followed by AICC or SAICC. The

All candidate models are misspecified



Correct models are included in the candidate set

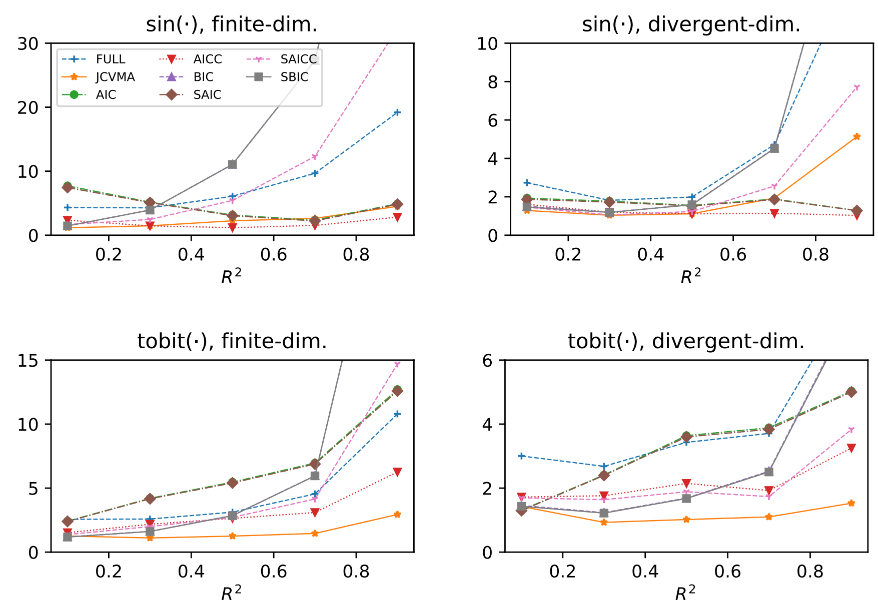


Figure S.1: Normalized mean squared prediction error ($n = 300$)

Note: The upper four figures plot the NMSPE when all candidate models are misspecified, and the bottom four figures consider the cases in which the candidate set includes correct models.

discrepancy between JCVMA and other methods seems larger in finite-dimensional cases than in divergent cases and appears to increase with n.

The bottom four diagrams of Figure S.1 present the related NMSPEs when $n = 300$ and the candidate set includes correct models. In this case, the difference between the methods appears to be relatively small, especially when R^2 is small. Nevertheless, JCVMA continues to perform well with low NMSPEs and is always ranked among the top three methods, if not the best in a small proportion of cases. When R^2 is moderate or large, its improvement over other methods is particularly large for the Tobit model.

Figures S.2 and S.3 present the relative squared losses of competing methods under different levels of R^2, and Figures S.4–S.7 present the normalized mean squared prediction errors under different sample sizes.

S2.4 Simulation under $p > n$

Thus far, we have studied the finite-sample performance of JCVMA when $p < n$. Now we consider situations in which p is larger than n. We set $n = 100$ and $p = 200$. The coefficient vector is set as $\boldsymbol{\beta} = (1, 2, 0.1, 3, 0.08, 4, 0.06, 5, 0.04, 6, 0.02, 0, 0, 0, 0, 0, \ldots, 0, 4)^{\mathrm{T}}$, which is characterized by sparsity. We consider the misspecified scenario, where the last covariate is omitted by all fitted models.

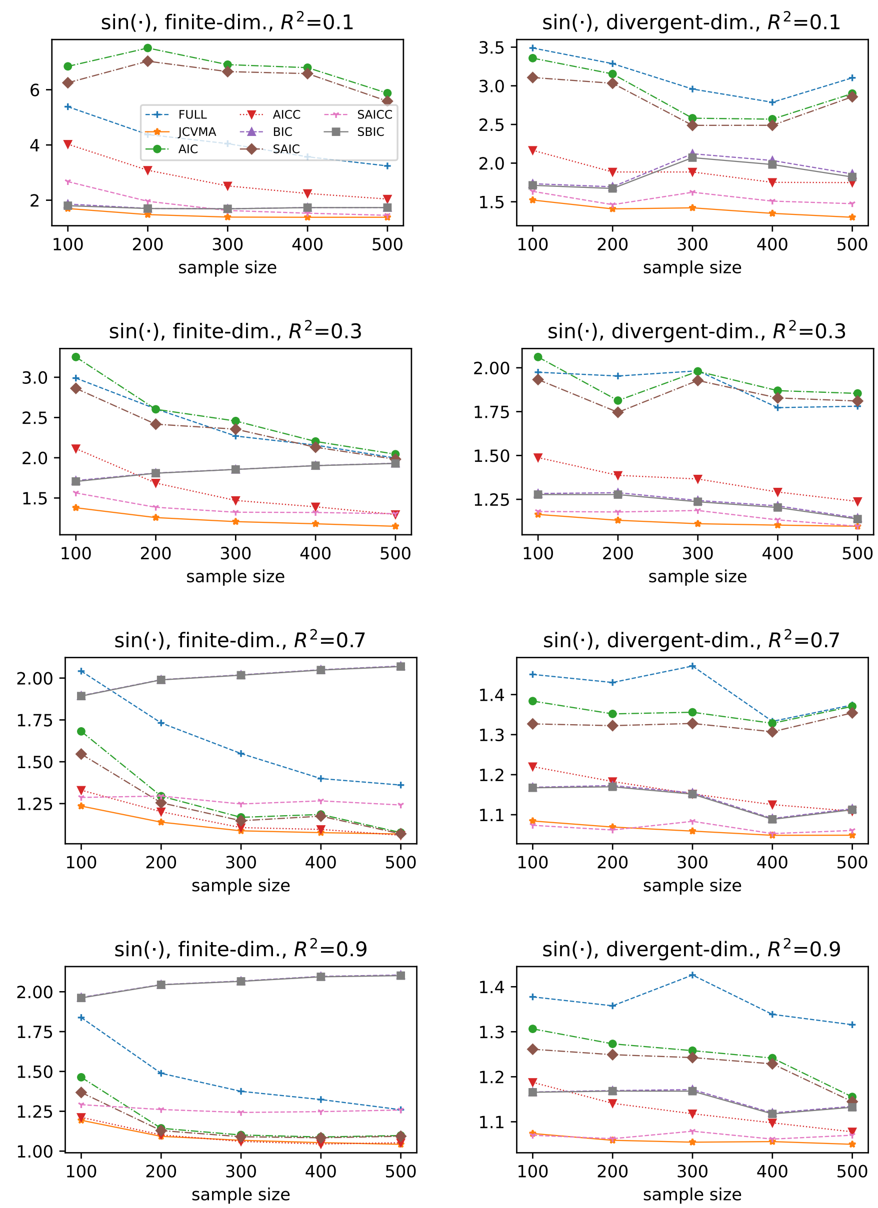


Figure S.2: Relative squared losses when all candidate models are misspecified (sine fun.)

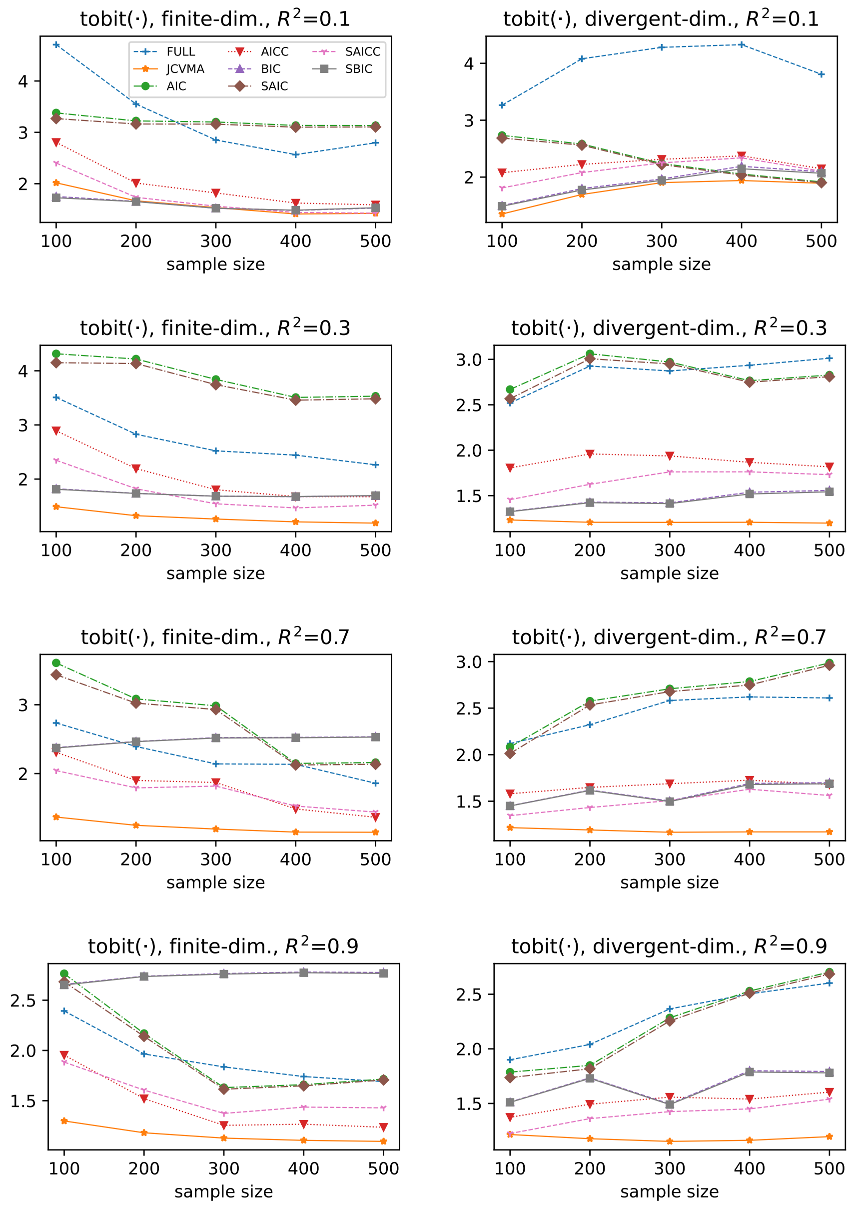
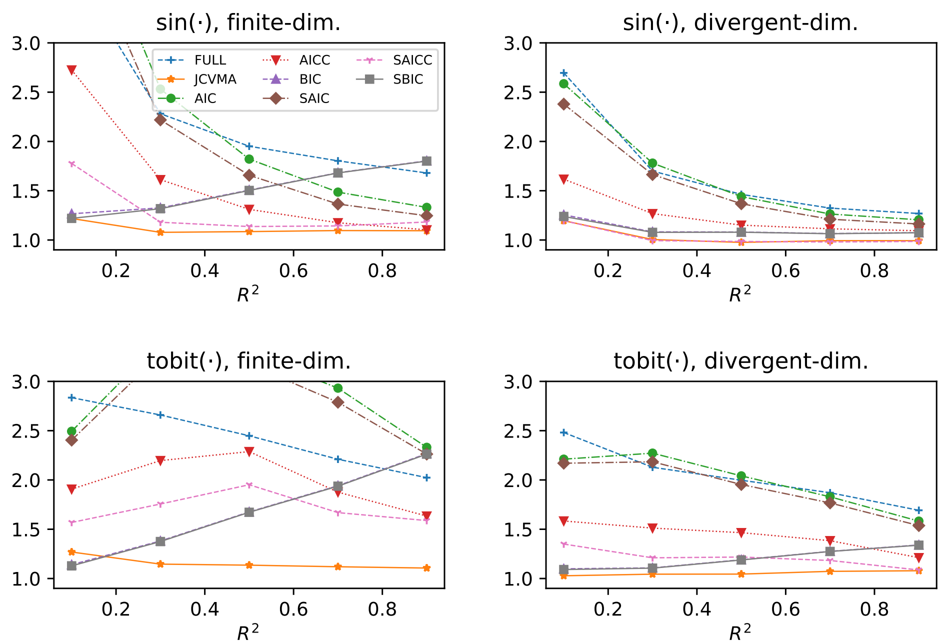


Figure S.3: Relative squared losses when all candidate models are misspecified (Tobit fun.)

All candidate models are misspecified



Correct models are included in the candidate set

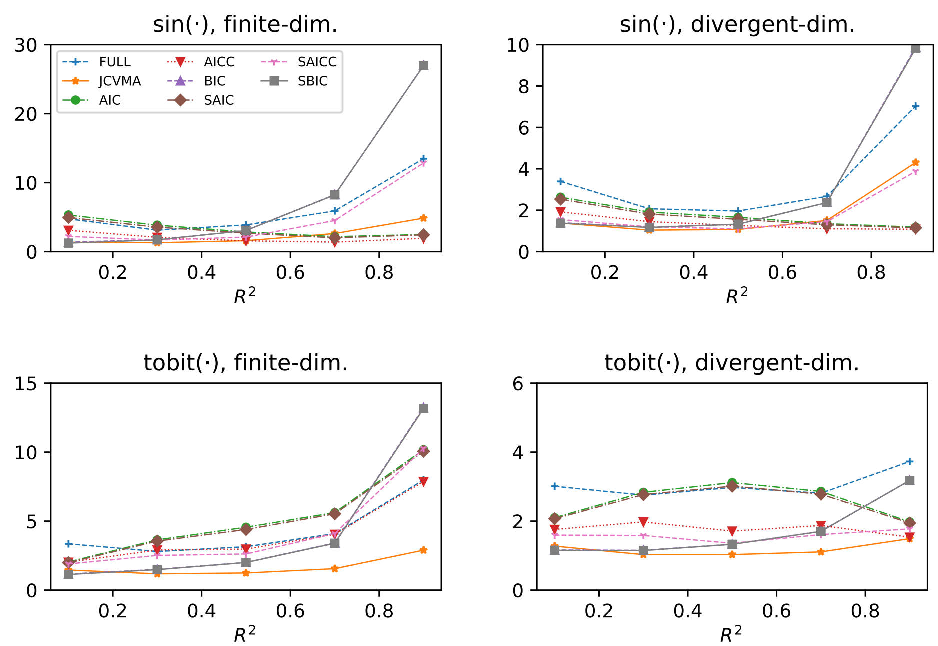
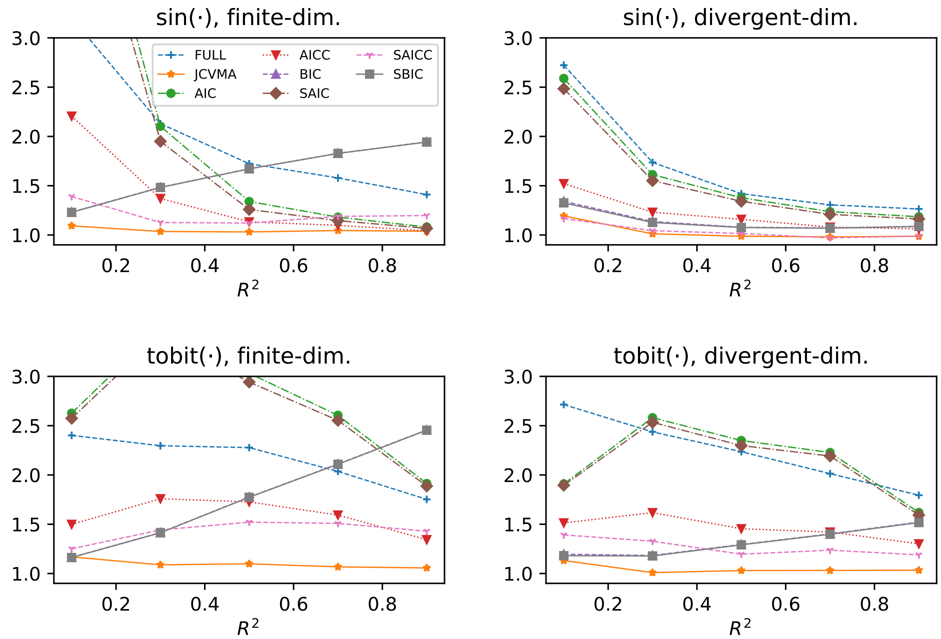


Figure S.4: Normalized mean squared prediction error ($n = 100$)

Notes: The upper four figures plot the NMSPEs when all candidate models are misspecified, and the bottom four figures consider the cases where the candidate set includes correct models.

All candidate models are misspecified



Correct models are included in the candidate set

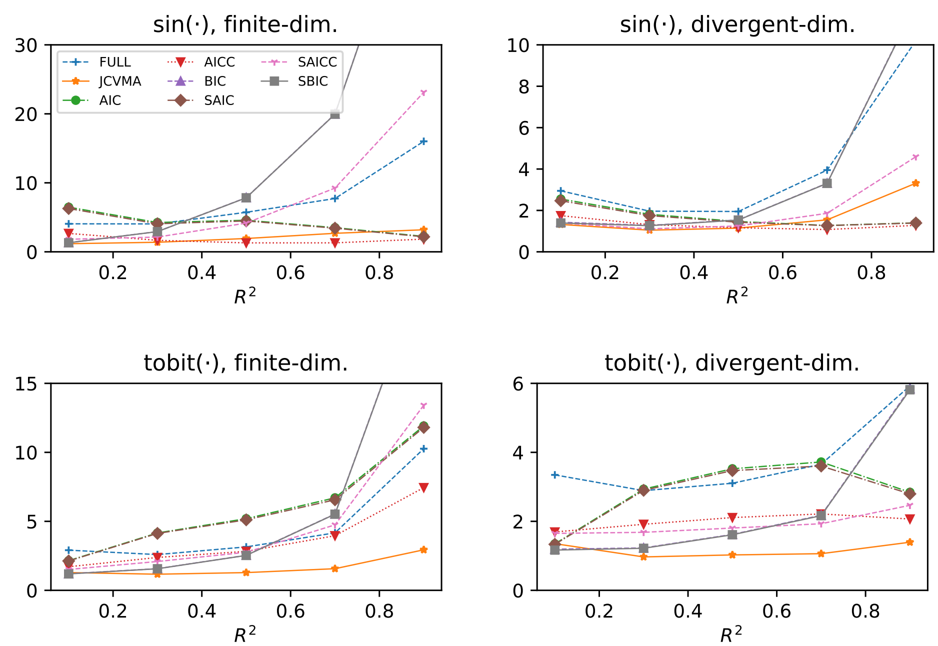
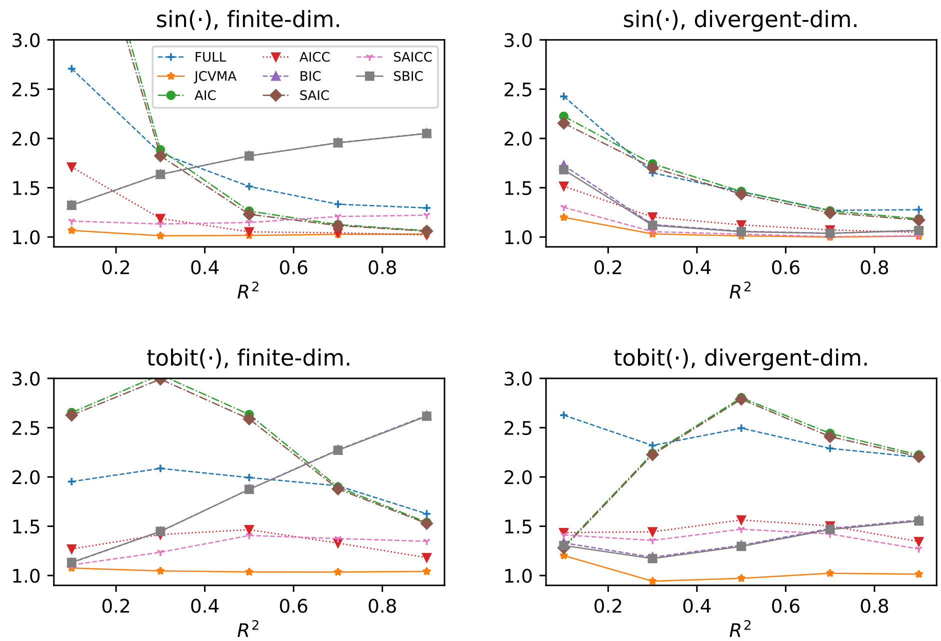


Figure S.5: Normalized mean squared prediction error ($n = 200$)

Notes: The upper four figures plot the NMSPEs when all candidate models are misspecified, and the bottom four figures consider the cases where the candidate set includes correct models.

All candidate models are misspecified



Correct models are included in the candidate set

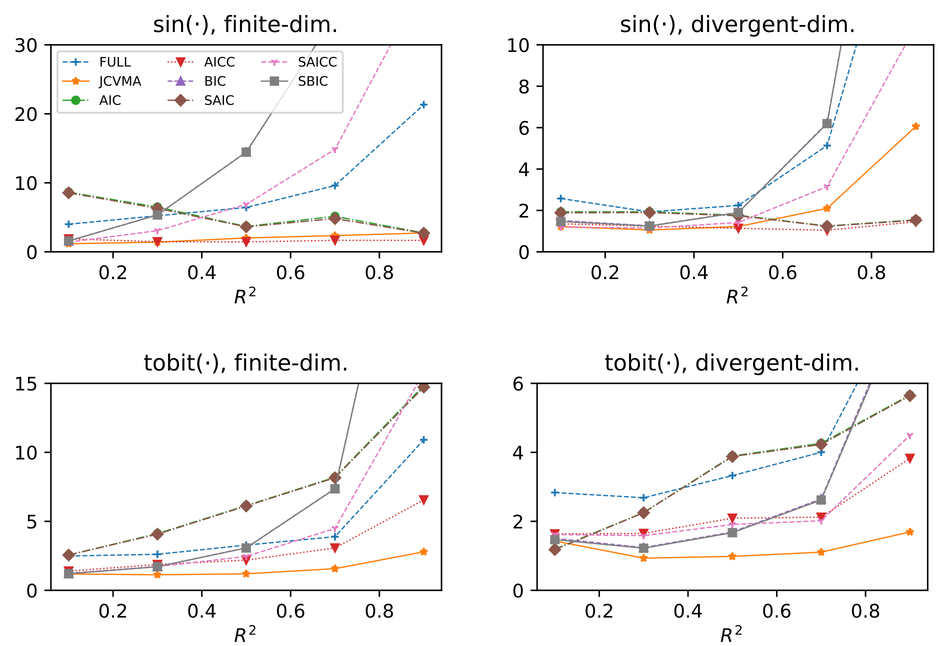


Figure S.6: Normalized mean squared prediction error ($n = 400$)

Notes: The upper four figures plot the NMSPEs when all candidate models are misspecified, and the bottom four figures consider the cases where the candidate set includes correct models.

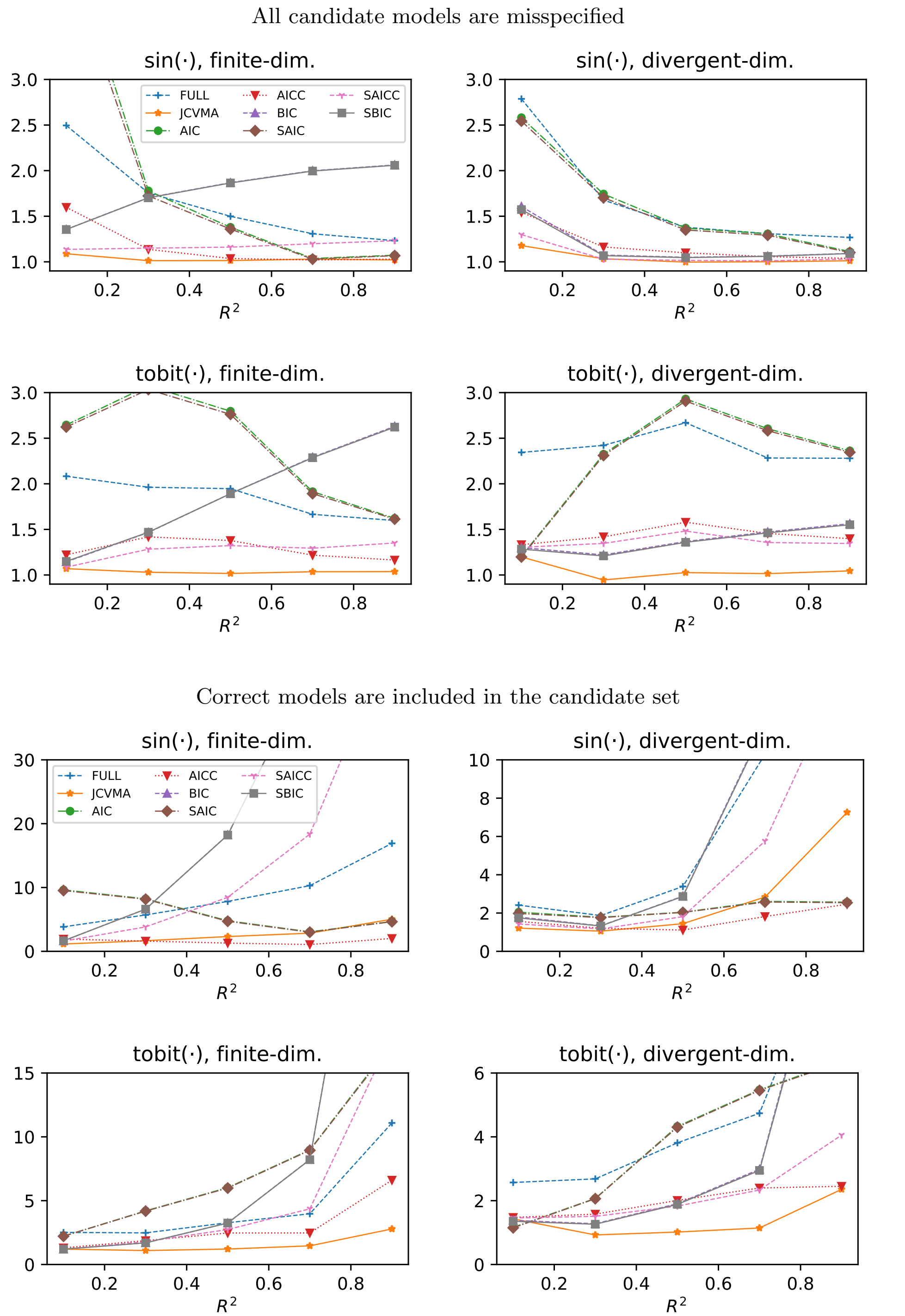


Figure S.7: Normalized mean squared prediction error ($n = 500$)

Notes: The upper four figures plot the NMSPEs when all candidate models are misspecified, and the bottom four figures consider the cases where the candidate set includes correct models.

Since the full model contains more parameters than the number of observations, it cannot be estimated by standard NLS. An excessively large p also implies that there are a formidable number of candidate models if we consider all possible combinations of covariates. Hence, we employ the regularization method with an L_1 penalty to estimate the full SIM, and pre-screen candidate models using the second approach (regularization-based screening) discussed in Section 4. Particularly, we vary the tuning parameter λ by taking 10 evenly spaced points between 0.001 (which yields on average 150 non-zero coefficient estimates across replications) and 0.02 (which forces all coefficient estimates to be zero across replications). Such a variation of λ leads to 10 candidate models, over which all selection and averaging methods are applied to predict the response variable.

Figure S.8 presents the NMSPEs of the competing methods when $p > n$ and all candidate models are misspecified. It shows that JCVMA based on regularized estimation and pre-screening outperforms other selection and averaging methods, and again its advantage is particularly prominent when R^2 is small. When R^2 is large, the squared losses of all methods are asymptotically identical to that of the best single model because all of the candidate models perform similarly after this pre-screening.

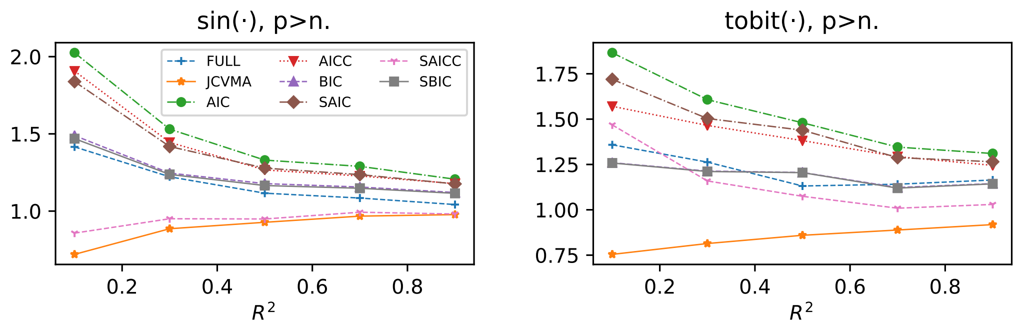


Figure S.8: NMSPE when $p > n$ and all candidate models are misspecified

Note: We set $p = 200$ and $n = 100$. The NMSPE is defined as $D^{-1}\sum_{d=1}^{D} L_n^{(d)}/L_{\min}^{(d)}$.

S3 Lemmas and proofs

For convenience, two new notations should be introduced first. If a function $g_{(s),ij} = O_u(a_n)$, then $g_{(s),ij}/a_n$ is bounded uniformly for any s, i and j. Besides, if a function $g_{(s),ij} = O_{uP}(a_n)$, then $g_{(s),ij}/a_n$ is bounded in probability uniformly for any s, i and j.

S3.1 Proof of Lemma 1

To prove Lemma 1, we need several additional lemmas. We first introduce some quantities which will be useful in the following proof. Denote

$$A_{(s),ni} = \frac{1}{n-1}\sum_{j\neq i}^{n} \mu_j k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}),$$

$$B_{(s),ni} = \frac{1}{n-1}\sum_{j\neq i}^{n} k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}),$$

and

$$C_{(s),ni} = \frac{1}{n-1}\sum_{j\neq i}^{n} \epsilon_j k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}).$$

S3.2 Additional lemmas and proofs

Lemma S.1. *Under Conditions 1, 2 and A1.1– A1.4, the following equalities hold:*

$$\max_{1\leq s\leq S_n} \frac{1}{np_s}\sum_{i=1}^{n} \|A'_{(s),ni}\|^2 = O_P\left(1 + \sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s\right), \tag{S3.3}$$

$$B_{(s),ni}|\mathbf{x}_{(s),i} = f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + O_P(h_s^2) + O_P(h_s^{-1/2}n^{-1/2}), \tag{S3.4}$$

$$\min_{1\leq s\leq S_n}\min_{1\leq i\leq n} B_{(s),ni}|\mathbf{x}_{(s),i} \geq \underline{c} + o_P(1), \tag{S3.5}$$

$$\max_{1\leq s\leq S_n}\sum_{i=1}^{n}\frac{1}{np_s}\|B'_{(s),ni}\|^2 = O_P\left(1 + \sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s\right), \tag{S3.6}$$

$$\max_{1\leq i\leq n}\max_{1\leq i\leq n} C_{(s),ni}|\mathbf{x}_{(s),i} = O_P\left(\max_{1\leq s\leq S_n} h_s^{-1/2} n^{-1/2}\right), \tag{S3.7}$$

$$\max_{1\leq s\leq S_n}\sum_{i=1}^{n}\frac{1}{np_s}\|C'_{(s),ni}\|^2 = O_P\left(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s\right), \tag{S3.8}$$

and

$$\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\left|\frac{A_{(s),ni}}{B_{(s),ni}}\right| = O_P(1). \tag{S3.9}$$

where the superscript "′" means the derivate with respect to $\boldsymbol{\beta}_{(s)}^{*}$ *and* $\cdot|\mathbf{x}_{(s),i}$ *means conditional on* $\mathbf{x}_{(s),i}$.

Proof of Lemma S.1. First, we consider (S3.3). We can show that

$$\begin{aligned}&\mathrm{E}\left\{\mu_j k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)})\Big|\mathbf{x}_{(s),i}\right\}\\&=\mathrm{E}\left\{\frac{\mu_j(\mathbf{x}_{(s),i}-\mathbf{x}_{(s),j})}{h_s^2}k'\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}}{h_s}\right)\Big|\mathbf{x}_{(s),i}\right\}. \qquad \text{(S3.10)}\end{aligned}$$

Let $v_{(s),jr}=x_{(s),jr}\beta^*_{(s),r}$, $r=1,...,p_s$ and $t_{(s),i}=\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}$, $i=1,...,n$. On the one hand, we consider the case that $\beta^*_{(s),k}\neq 0$. For the k^{th} element of (S3.10),

$$\begin{aligned}&\mathrm{E}\left\{\frac{\mu_j(x_{(s),ik}-x_{(s),jk})}{h_s^2}k'\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}}{h_s}\right)\Big|\mathbf{x}_{(s),i}\right\}\\
&=\int\frac{\mu_j(x_{(s),ik}-x_{(s),jk})}{h_s^2}k'\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)\rho_{(s)}(v_{(s),j1},...,v_{(s),jp_s})\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),jp_s}\\
&=\int\mu_j h_s^{-2}\beta^{*-1}_{(s),k}(v_{(s),ik}-v_{(s),jk})k'\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)\rho_{(s)}(v_{(s),j1},...,v_{(s),jp_s})\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),jp_s}\\
&=-\int\mu_j h_s^{-1}\beta^{*-1}_{(s),k}\left(\sum_{l\neq k}^{p_s}v_{(s),il}-\sum_{l\neq k}^{p_s}v_{(s),jl}-\tau h_s\right)k'(\tau)\\
&\quad\times\rho_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s}v_{(s),jl}-\tau h_s,v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\\
&\quad\times\,\mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}\\
&=-\int\mu_j h_s^{-1}\beta^{*-1}_{(s),k}\left(\sum_{l\neq k}^{p_s}v_{(s),il}-\sum_{l\neq k}^{p_s}v_{(s),jl}-\tau h_s\right)k'(\tau)\\
&\quad\times\rho_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s}v_{(s),jl},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\\
&\quad\times\,\mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}\\
&\quad-\int\mu_j h_s^{-1}\beta^{*-1}_{(s),k}\left(\sum_{l\neq k}^{p_s}v_{(s),il}-\sum_{l\neq k}^{p_s}v_{(s),jl}-\tau h_s\right)k'(\tau)\\
&\quad\times\left\{\rho_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s}v_{(s),jl}-\tau h_s,v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\right.\\
&\quad\left.-\rho_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s}v_{(s),jl},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\right\}\\
&\quad\times\,\mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}\end{aligned}$$

$$\equiv -\Delta_1 - \Delta_2. \tag{S3.11}$$

Considering that $\int k'(u)\,\mathrm{d}u = 0$, $\int uk'(u)\,\mathrm{d}u = -1$ from Condition A1.1,

$$\begin{aligned}
\Delta_1 = &\int \mu_j h_s^{-1} \beta_{(s),k}^{*-1} \left(\sum_{l\neq k}^{p_s} v_{(s),il} - \sum_{l\neq k}^{p_s} v_{(s),jl} \right) k'(\tau) \\
&\times \rho_{(s)} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
&\times \mathrm{d}\tau\, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
&- \int \mu_j \beta_{(s),k}^{*-1} \tau k'(\tau) \rho_{(s)} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
&\times \mathrm{d}\tau\, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
= &\, 0 + \int \mu_j \beta_{(s),k}^{*-1} \rho_{(s)} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
&\times \mathrm{d}\tau\, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
= &\, \mu_j \beta_{(s),k}^{*-1} \int \rho_{(s)} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
&\times \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
= &\, O_u(1),
\end{aligned} \tag{S3.12}$$

where the last inequality is due to Conditions 2(ii), S.3(iv) and S.4(i). Next,

$$\begin{aligned}
\Delta_2 = &\int \mu_j h_s^{-1} \beta_{(s),k}^{*-1} \left(\sum_{l\neq k}^{p_s} v_{(s),il} - \sum_{l\neq k}^{p_s} v_{(s),jl} - \tau h_s \right) k'(\tau) \\
&\times \Bigg\{ \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl} - \tau h_s, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
&- \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l\neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \Bigg\} \\
&\times \mathrm{d}\tau\, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s}
\end{aligned}$$

$$
\begin{aligned}
&= \int \mu_j h_s^{-1} \beta_{(s),k}^{*-1} \left(\sum_{l \neq k}^{p_s} v_{(s),il} \right) k'(\tau) \Bigg\{ \rho_{(s),k} \left(v_{(s),j1}, ..., t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl} - \tau h_s, , ..., v_{(s),jp_s} \right) \\
&\quad - \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \Bigg\} \\
&\quad - \int \mu_j h_s^{-1} \beta_{(s),k}^{*-1} \left(\sum_{l \neq k}^{p_s} v_{(s),jl} \right) k'(\tau) \Bigg\{ \rho_{(s),k} \left(v_{(s),j1}, ..., t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl} - \tau h_s, ..., v_{(s),jp_s} \right) \\
&\quad - \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \Bigg\} \\
&\quad \times \, \mathrm{d}\tau \, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)} \, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
&\quad - \int \mu_j \tau \beta_{(s),k}^{*-1} k'(\tau) \Bigg\{ \rho_{(s),k} \left(v_{(s),j1}, ..., t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl} - \tau h_s, ..., v_{(s),jp_s} \right) \\
&\quad - \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \Bigg\} \\
&\quad \times \, \mathrm{d}\tau \, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)} \, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s} \\
&\equiv \Omega_1 - \Omega_2 - \Omega_3 \\
&= O_u(1), \qquad \text{(S3.13)}
\end{aligned}
$$

where the last equality is based on the following equalities (S3.14), (S3.15) and (S3.16):

$$
\begin{aligned}
|\Omega_1| \leq & \int \mu_j h_s^{-1} \beta_{(s),k}^{*-1} \left| \sum_{l \neq k}^{p_s} v_{(s),il} \right| |k'(\tau)| \\
& \times \Bigg| \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl} - \tau h_s, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \\
& - \rho_{(s),k} \left(v_{(s),j1}, ..., v_{(s),j(k-1)}, t_{(s),i} - \sum_{l \neq k}^{p_s} v_{(s),jl}, v_{(s),j(k+1)}, ..., v_{(s),jp_s} \right) \Bigg| \\
\leq & \int G |\mu_j| \beta_{(s),k}^{*-1} \left| \sum_{l \neq k}^{p_s} v_{(s),il} \right| |\tau k'(\tau)| \omega_{(s)}(v_{(s),j1}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),jp_s}) \\
& \times \, \mathrm{d}\tau \, \mathrm{d}v_{(s),j1} ... \mathrm{d}v_{(s),j(k-1)} \, \mathrm{d}v_{(s),j(k+1)} ... \mathrm{d}v_{(s),jp_s}
\end{aligned}
$$

$$= G|\mu_j|\beta^{*-1}_{(s),k}\left|\sum_{l\neq k}^{p_s} v_{(s),il}\right|\int |\tau k'(\tau)|\,\mathrm{d}\tau \int \omega_{(s)}(v_{(s),j1},...,v_{(s),j(k-1)},v_{(s),j(k+1)},...,v_{(s),jp_s})$$

$$\times\ \mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}$$

$$= O_u(1), \tag{S3.14}$$

where the last inequality is from Condition S.4(iv), the last quality is due to Conditions 2(ii), A1.1, S.3(ii) and S.3(iv).

$$|\Omega_2| \leq \int |\mu_j| h_s^{-1} |\beta^{*-1}_{(s),k}| \left(\sum_{l\neq k}^{p_s} |v_{(s),jl}|\right) |k'(\tau)| \Bigg| \rho_{(s),k}\left(v_{(s),j1},...,t_{(s),i}-\sum_{l\neq k}^{p_s} v_{(s),jl}-\tau h_s,...,v_{(s),jp_s}\right)$$

$$- \rho_{(s),k}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s} v_{(s),jl},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\Bigg|$$

$$\times\ \mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}$$

$$\leq G\int |\mu_j||\beta^{*-1}_{(s),k}| \left(\sum_{l\neq k}^{p_s} |v_{(s),jl}|\right) |\tau k'(\tau)| \omega_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\Bigg|$$

$$\times\ \mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}$$

$$= G|\mu_j||\beta^{*-1}_{(s),k}|\int |\tau k'(\tau)|\,\mathrm{d}\tau \sum_{l\neq k}^{p_s}\int |v_{(s),jl}|\omega_{(s)}\left(v_{(s),j1},...,v_{(s),j(k-1)},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)$$

$$\times\ \mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}$$

$$= O_u(1), \tag{S3.15}$$

where the last equality is due to Conditions 2(ii), A1.1, S.3(iv) and S.4(iv).

Finally,

$$|\Omega_3| \leq \int |\mu_j||\beta^{*-1}_{(s),k}||\tau k'(\tau)| \Bigg| \rho_{(s),k}\left(v_{(s),j1},...,t_{(s),i}-\sum_{l\neq k}^{p_s} v_{(s),jl}-\tau h_s,...,v_{(s),jp_s}\right)$$

$$- \rho_{(s),k}\left(v_{(s),j1},...,v_{(s),j(k-1)},t_{(s),i}-\sum_{l\neq k}^{p_s} v_{(s),jl},v_{(s),j(k+1)},...,v_{(s),jp_s}\right)\Bigg|$$

$$\times\ \mathrm{d}\tau\,\mathrm{d}v_{(s),j1}...\,\mathrm{d}v_{(s),j(k-1)}\,\mathrm{d}v_{(s),j(k+1)}...\,\mathrm{d}v_{(s),jp_s}$$

$$
\begin{aligned}
&\leq G|\mu_j||\beta^{*-1}_{(s),k}|h_s \int \tau^2 |k'(\tau)| \,\mathrm{d}\tau \int \omega_{(s)}\big(v_{(s),j1}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)},\\
&\quad ..., v_{(s),jp_s}\big) \,\mathrm{d}v_{(s),j1} ... \,\mathrm{d}v_{(s),j(k-1)} \,\mathrm{d}v_{(s),j(k+1)} ... \,\mathrm{d}v_{(s),jp_s}\\
&= O_u(h_s), \qquad\qquad \text{(S3.16)}
\end{aligned}
$$

where the last equality is due to Conditions 2(ii), A1.1, S.3(iv) and S.4(iv).

On the other hand, we consider the case that $\beta^*_{(s),k} = 0$. Let $t_{(s),i} = \sum_{r=1,r\neq k}^{p_s} x_{(s),ir}\beta^*_{(s),r}$, $i = 1, ..., n$. For the k^{th} element of (S3.10),

$$
\begin{aligned}
&\mathrm{E}\left\{ \frac{\mu_j(x_{(s),ik} - x_{(s),jk})}{h_s^2} k'\left(\frac{\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j}\boldsymbol{\beta}^*_{(s)}}{h_s} \right) \Big| \mathbf{x}_{(s),i} \right\}\\
&= \frac{\mu_j\{x_{(s),ik} - \mathrm{E}(x_{(s),jk})\}}{h_s^2} \int k'\left(\frac{t_{(s),i} - t_{(s),j}}{h_s} \right) \rho_{(s)}(v_{(s),j1}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ...v_{(s),jp_s})\\
&\quad \times \,\mathrm{d}v_{(s),j1} ... \,\mathrm{d}v_{(s),j(k-1)} \,\mathrm{d}v_{(s),j(k+1)} ... \,\mathrm{d}v_{(s),jp_s}\\
&= \mu_j h_s^{-1}\{x_{(s),ik} - \mathrm{E}(x_{(s),jk})\} \int k'(\tau)\rho_{(s)}\Bigg(t_{(s),i} - \sum_{l\neq 1,k}^{p_s} v_{(s),jl} - \tau h_s, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)},\\
&\quad ..., v_{(s),p_s} \Bigg) \times \,\mathrm{d}\tau \,\mathrm{d}v_{(s),j2} ... \,\mathrm{d}v_{(s),j(k-1)} \,\mathrm{d}v_{(s),j(k+1)} ... \,\mathrm{d}v_{(s),jp_s}\\
&= \mu_j h_s^{-1}\{x_{(s),ik} - \mathrm{E}(x_{(s),jk})\} \int k'(\tau)\rho_{(s)} \left(t_{(s),i} - \sum_{l\neq 1,k}^{p_s} v_{(s),jl}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s} \right)\\
&\quad \times \,\mathrm{d}\tau \,\mathrm{d}v_{(s),j2} ... \,\mathrm{d}v_{(s),j(k-1)} \,\mathrm{d}v_{(s),j(k+1)} ... \,\mathrm{d}v_{(s),jp_s}\\
&\quad + \mu_j h_s^{-1}\{x_{(s),ik} - \mathrm{E}(x_{(s),jk})\} \int k'(\tau) \Bigg\{ \rho_{(s)}\Bigg(t_{(s),i} - \sum_{l\neq 1,k}^{p_s} v_{(s),jl} - \tau h_s, ..., v_{(s),j(k-1)},\\
&\quad v_{(s),j(k+1)}, ..., v_{(s),p_s} \Bigg) - \rho_{(s)} \left(t_{(s),i} - \sum_{l\neq 1,k}^{p_s} v_{(s),jl}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s} \right) \Bigg\}\\
&\quad \times \,\mathrm{d}\tau \,\mathrm{d}v_{(s),j2} ... \,\mathrm{d}v_{(s),j(k-1)} \,\mathrm{d}v_{(s),j(k+1)} ... \,\mathrm{d}v_{(s),jp_s}\\
&= 0 + O_u(1), \qquad\qquad \text{(S3.17)}
\end{aligned}
$$

where 0 in the last equality is due to $\int k'(\tau) \,\mathrm{d}\tau = 0$ and the $O_u(1)$ in the last equality is due to Conditions 2(ii), S.3(i) and the following inequality:

$$
\begin{aligned}
& h_s^{-1}\bigg|\int k'(\tau)\bigg\{\rho_{(s)}\bigg(t_{(s),i}-\sum_{l\neq 1,k}^{p_s} v_{(s),jl}-\tau h_s, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\bigg) \\
& -\rho_{(s)}\bigg(t_{(s),i}-\sum_{l\neq 1,k}^{p_s} v_{(s),jl}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\bigg)\bigg\} \\
& \times\, \mathrm{d}\tau\, \mathrm{d}v_{(s),j2} ...\, \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ...\, \mathrm{d}v_{(s),jp_s}\bigg| \\
\leq\, & h_s^{-1}\int |k'(\tau)|\bigg|\rho_{(s)}\bigg(t_{(s),i}-\sum_{l\neq 1,k}^{p_s} v_{(s),jl}-\tau h_s, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\bigg) \\
& -\rho_{(s)}\bigg(t_{(s),i}-\sum_{l\neq 1,k}^{p_s} v_{(s),jl}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\bigg)\bigg| \\
& \times\, \mathrm{d}\tau\, \mathrm{d}v_{(s),j2} ...\, \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ...\, \mathrm{d}v_{(s),jp_s} \\
\leq\, & G h_s^{-1}\int |k'(\tau)||\tau| h_s \omega_{(s)}\left(v_{(s),j2}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\right) \\
& \times\, \mathrm{d}\tau\, \mathrm{d}v_{(s),j2} ...\, \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ...\, \mathrm{d}v_{(s),jp_s} \\
=\, & G\int |\tau||k'(\tau)|\, \mathrm{d}\tau \int \omega_{(s)}\left(v_{(s),j2}, ..., v_{(s),j(k-1)}, v_{(s),j(k+1)}, ..., v_{(s),p_s}\right) \\
& \times\, \mathrm{d}\tau\, \mathrm{d}v_{(s),j2} ...\, \mathrm{d}v_{(s),j(k-1)}\, \mathrm{d}v_{(s),j(k+1)} ...\, \mathrm{d}v_{(s),jp_s} \\
=\, & O_u(1), \qquad \text{(S3.18)}
\end{aligned}
$$

where the last equality is due to Conditions A1.1 and S.4(iv).

Based on (S3.11), (S3.12), (S3.13) and (S3.17), for the k^{th} element of

(S3.10), we obtain that

$$\mathrm{E}\left\{\frac{\mu_j(x_{(s),ik}-x_{(s),jk})}{h_s^2}k'\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*}{h_s}\right)\Big|\mathbf{x}_{(s),i}\right\}=O_{uP}(1), \tag{S3.19}$$

which indicates, considering the dimension of (S3.10) is p_s,

$$\max_{1\le j\le n}\max_{1\le i\le n}\left\|\mathrm{E}\left\{\mu_j k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\Big|\mathbf{x}_{(s),i}\right\}\right\|=O_{uP}(p_s^{1/2}). \tag{S3.20}$$

Hence, the norm of the conditional expectation of $A'_{(s),ni}$ is

$$\begin{aligned}
&\max_{1\le i\le n}\|\mathrm{E}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|\\
&=\frac{1}{n-1}\max_{1\le i\le n}\left\|\sum_{j\ne i}^{n}\mathrm{E}\left\{\mu_j k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\Big|\mathbf{x}_{(s),i}\right\}\right\|\\
&\le\max_{1\le j\le n}\max_{1\le i\le n}\left\|\mathrm{E}\left\{\mu_j k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\Big|\mathbf{x}_{(s),i}\right\}\right\|\\
&=O_{uP}(p_s^{1/2}).
\end{aligned} \tag{S3.21}$$

Next, from Conditions 2(ii) and S.3(i), we see that the norm of the conditional variance of $A'_{(s),ni}$ is given by

$$\begin{aligned}
&\|\mathbf{V}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|\\
&\le\frac{1}{(n-1)^2}\sum_{j\ne i}^{n}\left\|\mathbf{V}\left\{\mu_j k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\Big|\mathbf{x}_{(s),i}\right\}\right\|\\
&\le\frac{\max_{1\le j\le n}|\mu_j|^2}{n-1}\max_{1\le j\le n}\mathrm{E}\left\{\frac{\|\mathbf{x}_{(s),i}-\mathbf{x}_{(s),j}\|^2}{h_s^4}k'^2\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*}{h_s}\right)\Big|\mathbf{x}_{(s),i}\right\}\\
&\le\frac{2\max_{1\le j\le n}|\mu_j|^2\max_{1\le s\le S_n}\max_{1\le i\le n}\|\mathbf{x}_{(s),i}\|^2}{(n-1)h_s^4}\max_{1\le j\le n}\int k'^2\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j}\\
&=O_u(h_s^{-3}n^{-1}p_s)\int k'^2(\tau)f_{(s)}(t_{(s),i}-\tau h_s)\,\mathrm{d}\tau\\
&=O_u(h_s^{-3}n^{-1}p_s)\int k'^2(\tau)\left\{f_{(s)}(t_{(s),i})-\tau h_s f'_{(s)}(\tilde{t})\right\}\,\mathrm{d}\tau
\end{aligned}$$

$$= O_u(h_s^{-3}n^{-1}p_s)\left\{f_{(s)}(t_{(s),i})\int k^{'2}(\tau)\,\mathrm{d}\tau + O_u(h_s)\right\}$$
$$= f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})O_{uP}(h_s^{-3}n^{-1}p_s)$$
$$= O_{uP}(h_s^{-3}n^{-1}p_s), \tag{S3.22}$$

where $\tilde{t}$ lies between $t_{(s),i}$ and $t_{(s),i}-\tau h_s$, $O_u(h_s)$ in the third-to-last equality is due to (S3.23):

$$\begin{aligned}\left|\int f'_{(s)}(\tilde{t})k^{'2}(\tau)\tau\,\mathrm{d}\tau\right| &\le \int \left|f'_{(s)}(t_{(s),i})\right| k^{'2}(\tau)\tau\,\mathrm{d}\tau + \int \left|f'_{(s)}(\tilde{t}) - f'_{(s)}(t_{(s),i})\right| k^{'2}(\tau)\tau\,\mathrm{d}\tau \\ &\le \bar{C}\int |\tau|k^{'2}(\tau)\,\mathrm{d}\tau + c_1 h_s \int \tau^2 k^{'2}(\tau)\,\mathrm{d}\tau \\ &= O_u(1+h_s),\end{aligned} \tag{S3.23}$$

where the last inequality is from S.4(v) and the last equality is due to Condition A1.1. Further, from Conditions A1.1 and S.4(ii),

$$\begin{aligned}\mathrm{E}\{\|\mathbf{V}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|\} &\le O_u(h_s^{-3}n^{-1}p_s)\mathrm{E}\left\{f_{(s)}(t_{(s),i})\int k^{'2}(\tau)\,\mathrm{d}\tau + O_u(h_s)\right\} \\ &= O_u(h_s^{-3}n^{-1}p_s)\left\{\int f_{(s)}^2(t_{(s),i})\,\mathrm{d}t_{(s),i}\int k^{'2}(\tau)\,\mathrm{d}\tau + O_u(h_s)\right\} \\ &= O_u(h_s^{-3}n^{-1}p_s).\end{aligned} \tag{S3.24}$$

Let $a_n = \sum_{s=1}^{S_n} h_s^{-3}n^{-1}p_s$. Combining with (S3.22)–(S3.24) and according to Markov's Inequality, we know for some large $\delta > 0$, when n is large enough,

$$\mathrm{Pr}\left(\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni} - \mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\|^2 \ge \delta a_n\right)$$
$$\le \sum_{s=1}^{S_n}\mathrm{Pr}\left(\frac{1}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni} - \mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\|^2 \ge \delta a_n\right)$$
$$\le (np_s\delta a_n)^{-1}\sum_{s=1}^{S_n}\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{E}\left\{\|A'_{(s),ni} - \mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\|^2\left|\mathbf{x}_{(s),i}\right.\right\}\right]$$

$$
\begin{aligned}
&= (np_s\delta a_n)^{-1}\sum_{s=1}^{S_n}\sum_{i=1}^{n}\mathrm{E}\Bigg\{\mathrm{trace}\Big(\mathrm{E}\Big[\left\{A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\right\} \\
&\quad \times \left\{A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\right\}^{\mathrm{T}}\left|\mathbf{x}_{(s),i}\right.\Big]\Big)\Bigg\} \\
&\le (np_s\delta a_n)^{-1}\sum_{s=1}^{S_n}\sum_{i=1}^{n}p_s\mathrm{E}\Bigg\{\lambda_{\max}\Big(\mathrm{E}\Big[\left\{A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\right\} \\
&\quad \times \left\{A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\right\}^{\mathrm{T}}\left|\mathbf{x}_{(s),i}\right.\Big]\Big)\Bigg\} \\
&= (n\delta a_n)^{-1}\sum_{s=1}^{S_n}\sum_{i=1}^{n}\mathrm{E}\left\{\|\mathbf{V}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\|\right\} \\
&\le (\delta a_n)^{-1}\sum_{s=1}^{S_n}\max_{1\le i\le n}\mathrm{E}\left\{\|\mathbf{V}(A'_{(s),ni}\left|\mathbf{x}_{(s),i}\right.)\|\right\} \\
&= O_{up}(\delta^{-1}) \to 0 \text{ as } \delta\to\infty, \qquad\qquad (\text{S3.25})
\end{aligned}
$$

which implies, according to van de Geer (2000),

$$
\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni}-\mathrm{E}(A_{(s),ni}|\mathbf{x}_{(s),i})\|^2 = O_P(a_n) = O_P\left(\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right). \qquad (\text{S3.26})
$$

Hence,

$$
\begin{aligned}
&\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni}\|^2 \\
&\le \max_{1\le s\le S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|\mathrm{E}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|^2 + \max_{1\le s\le S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|^2 \\
&\le \max_{1\le s\le S_n}\max_{1\le i\le n}p_s^{-1}\left\|\mathrm{E}(A'_{(s),ni}|\mathbf{x}_{(s),i})\right\|^2 + \max_{1\le s\le S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|A'_{(s),ni}-\mathrm{E}(A'_{(s),ni}|\mathbf{x}_{(s),i})\|^2 \\
&= O_P\left(1+\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right). \qquad\qquad (\text{S3.27})
\end{aligned}
$$

Now the proof of (S3.3) is completed.

Second, we consider (S3.4) and (S3.5). Under Conditions A1.1, S.4(*iii*) and S.4(v), we have

$$
\begin{aligned}
&\mathrm{E}(B_{(s),ni}|\mathbf{x}_{(s),i})\\
&= \mathrm{E}\left\{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\\
&= \int \frac{1}{h_s} k\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right) f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j}\\
&= \int k(\tau) f_{(s)}(t_{(s),i}-h_s\tau)\,\mathrm{d}\tau\\
&= \int k(\tau)\left[f_{(s)}(t_{(s),i}) - h_s\tau f_{(s)}'(t_{(s),j}) + \frac{h_s\tau^2}{2} f_{(s)}''(t) + \frac{h_s\tau^2}{2}\{f_{(s)}''(\tilde{t}) - f_{(s)}''(t_{(s),j})\}\right]\mathrm{d}\tau\\
&= f_{(s)}(t_{(s),i}) + \frac{1}{2}h_s^2 f_{(s)}''(t_{(s),i})\int \tau^2 k(\tau)\,\mathrm{d}\tau + O_u(h_s^2)\\
&= f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + O_u(h_s^2), \qquad\qquad \text{(S3.28)}
\end{aligned}
$$

where $\tilde{t}$ lies between $t_{(s),j}$ and $t_{(s),j} - \tau h_s$, and $O_u(h_s^2)$ in second-to-last equality can be obtained in a similar way as (S3.23). Based on Conditions A1.1, S.4(iii) and S.4(v), we reach that

$$
\begin{aligned}
&\mathbf{V}(B_{(s),ni}|\mathbf{x}_{(s),i})\\
&= \frac{1}{n-1}\mathbf{V}\left\{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\\
&\le \frac{1}{n-1}\mathrm{E}\left\{k_{h_s}^2(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\\
&= \frac{1}{(n-1)h_s^2}\int k^2\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right) f(t_{(s),j})\,\mathrm{d}t_{(s),j}\\
&= \frac{1}{(n-1)h_s}\int k^2(\tau) f_{(s)}(t_{(s),i}-h_s\tau)\,\mathrm{d}\tau\\
&= \frac{1}{(n-1)h_s}\int k^2(\tau)\left[f_{(s)}(t_{(s),i}) - h_s\tau f_{(s)}'(t_{(s),i}) + h_s\tau\{f_{(s)}'(t_{(s),i}) - f_{(s)}'(\tilde{t})\}\right]\mathrm{d}\tau\\
&\le \frac{1}{(n-1)h_s}\left\{f_{(s)}(t_{(s),i})\int k^2(\tau)\,\mathrm{d}\tau + h_s|f_{(s)}'(t_{(s),i})|\int \tau k^2(\tau)\,\mathrm{d}\tau + O_u(h_s^2)\right\}\\
&= O_u(h_s^{-1}n^{-1}), \qquad\qquad \text{(S3.29)}
\end{aligned}
$$

where $\tilde{t}$ lies between $t_{(s),i}$ and $t_{(s),i} - \tau h_s$, the $O_u(h_s^2)$ in the second-to-last equality is obtained like (S3.23). Combining with (S3.28), (S3.29), given

$\mathbf{x}_{(s),i}$,

$$B_{(s),ni} = \mathrm{E}(B_{(s),ni}|\mathbf{x}_{(s),i}) + O_P\left[\{\mathbf{V}(B_{(s),ni}|\mathbf{x}_{(s),i})\}^{1/2}\right]$$
$$= f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + O_P(h_s^2) + O_P(h_s^{-1/2}n^{-1/2}). \qquad \text{(S3.30)}$$

Further, combining with (S3.28), (S3.29) and based on Conditions A1.2 and S.4(ii) , given $\mathbf{x}_{(s),i}$, we have

$$\min_{1\le s\le S_n}\min_{1\le i\le n} B_{(s),ni}$$
$$= \min_{1\le s\le S_n}\min_{1\le i\le n}\left[\mathrm{E}(B_{(s),ni}|\mathbf{x}_{(s),i}) + \{B_{(s),ni} - \mathrm{E}(B_{(s),ni}|\mathbf{x}_{(s),i})\}\right]$$
$$\ge \min_{1\le s\le S_n}\min_{1\le i\le n}\mathrm{E}(B_{(s),ni}|\mathbf{x}_{(s),i}) - \max_{1\le s\le S_n}\max_{1\le i\le n} O_{uP}\left[\{\mathbf{V}(B_{(s),ni}|\mathbf{x}_{(s),i})\}^{1/2}\right]$$
$$\ge \min_{1\le s\le S_n}\min_{1\le i\le n} f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) - O\left(\max_{1\le s\le S_n} h_s^2\right) - O_P\left(\max_{1\le s\le S_n} h_s^{-1/2}n^{-1/2}\right)$$
$$\ge \underline{c} + o_P(1). \qquad \text{(S3.31)}$$

Now the proof of (S3.4) and (S3.5) is completed.

Third, we consider (S3.6). Following (S3.20), we have

$$\max_{1\le i\le n}\|\mathrm{E}(B'_{(s),ni}|\mathbf{x}_{(s),i})\| = \max_{1\le i\le n}\left\|\frac{1}{n-1}\sum_{j\ne i}^{n}\mathrm{E}\left\{k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\right\|$$
$$\le \max_{1\le j\le n}\max_{1\le i\le n}\left\|\mathrm{E}\left\{k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\right\|$$
$$= O_{uP}(p_s^{1/2}). \qquad \text{(S3.32)}$$

According to the process of (S3.22), based on Conditions A1.1, A1.3 and A1.4, we note that

$$\|\mathbf{V}(B'_{(s),ni}|\mathbf{x}_{(s),i})\|$$

$$\leq \frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\left\|\mathbf{V}\left\{k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\big|\mathbf{x}_{(s),i}\right\}\right\|$$

$$\leq \frac{1}{n-1}\max_{1\leq j\leq n}\mathrm{E}\left\{\frac{\|\mathbf{x}_{(s),i}-\mathbf{x}_{(s),j}\|^2}{h_s^4}k'^2\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}}{h_s}\right)\bigg|\mathbf{x}_{(s),i}\right\}$$

$$= O_{uP}(h_s^{-3}n^{-1}p_s). \tag{S3.33}$$

Thus, combining (S3.32) and (S3.33), resembling (S3.27), we obtain

$$\begin{aligned}
&\max_{1\leq s\leq S_n}\frac{1}{np_s}\sum_{i=1}^{n}\|B'_{(s),ni}\|^2\\
&\leq \max_{1\leq s\leq S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|\mathrm{E}(B'_{(s),ni}|\mathbf{x}_{(s),i})\|^2+\max_{1\leq s\leq S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|B'_{(s),ni}-\mathrm{E}(B'_{(s),ni}|\mathbf{x}_{(s),i})\|^2\\
&\leq \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}2p_s^{-1}\left\|\mathrm{E}(B'_{(s),ni}|\mathbf{x}_{(s),i})\right\|^2+\max_{1\leq s\leq S_n}\frac{2}{np_s}\sum_{i=1}^{n}\|B'_{(s),ni}-\mathrm{E}(B'_{(s),ni}|\mathbf{x}_{(s),i})\|^2\\
&= O_P\left(1+\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right).
\end{aligned}\tag{S3.34}$$

Now the proof of (S3.6) is completed.

Fourth, we consider (S3.7). It is readily shown that

$$\begin{aligned}
&\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\mathrm{E}(C_{(s),ni}|\mathbf{x}_{(s),i})\\
&=\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\frac{1}{n-1}\sum_{j\neq i}^{n}\mathrm{E}(\epsilon_j)\mathrm{E}\left\{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}\\
&=0.
\end{aligned}\tag{S3.35}$$

Under Conditions $2(i)$, S.1(ii) and A1.4, we can obtain its conditional variance as

$$\begin{aligned}
&\mathbf{V}(C_{(s),ni}|\mathbf{x}_{(s),i})\\
&=\frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\mathbf{V}\left\{\epsilon_j k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\Big|\mathbf{x}_{(s),i}\right\}
\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\mathrm{E}\left[\left\{\epsilon_j^2 k_{h_s}^2(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}\Big|\mathbf{x}_{(s),i}\right]\\
&=\frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\mathrm{E}(\epsilon_j^2)\mathrm{E}\left[\left\{k_{h_s}^2(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}\Big|\mathbf{x}_{(s),i}\right]\\
&=\frac{1}{(n-1)^2}\sum_{j\neq i}^{n}\sigma_j^2\int\frac{1}{h_s^2}k^2\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j}\\
&\leq\frac{\sigma_{\max}^2}{h_s^2(n-1)}\int k^2\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j}\\
&=\frac{\sigma_{\max}^2}{h_s(n-1)}\int k^2(\tau)\left[f_{(s)}(t_{(s),i})-h_s\tau f'_{(s)}(t_{(s),i})+h_s\tau\{f'_{(s)}(t_{(s),i})-f'_{(s)}(\tilde{t})\}\right]\,\mathrm{d}\tau\\
&=\frac{\sigma_{\max}^2}{h_s(n-1)}\left\{f_{(s)}(t_{(s),i})\int k^2(\tau)\,\mathrm{d}\tau+|f'_{(s)}(t_{(s),i})|h_s\int\tau k^2(\tau)\,\mathrm{d}\tau+O_u(h_s^2)\right\}\\
&=f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})O_{uP}(h_s^{-1}n^{-1})\\
&=O_{uP}(h_s^{-1}n^{-1}),
\end{aligned}
\tag{S3.36}
$$

where $\tilde{t}$ lies between $t_{(s),i}$ and $t_{(s),i}-\tau h_s$, the last equality is from Condition S.4(ii) and $O_u(h_s^2)$ in the third-to-last equality is obtained similarly as (S3.23). Like (S3.25), we can combine (S3.35) and (S3.36) to show that, given $\mathbf{x}_{(s),i}$,

$$
\begin{aligned}
\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}C_{(s),ni}&\leq\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\mathrm{E}(C_{(s),ni}|\mathbf{x}_{(s),i})+\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}O_{uP}\left[\{\mathbf{V}(B_{(s),ni}|\mathbf{x}_{(s),i})\}^{1/2}\right]\\
&=O_P\left(\max_{1\leq s\leq S_n}h_s^{-1/2}n^{-1/2}\right).
\end{aligned}
$$

Now the proof of (S3.7) is completed.

Fifth, we consider (S3.8). It is straightforward to shown that

$$
\mathrm{E}(C'_{(s),ni}|\mathbf{x}_{(s),i})=\frac{1}{n-1}\sum_{j\neq i}^{n}\mathrm{E}(\epsilon_j)\mathrm{E}\left\{k'_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})|\mathbf{x}_{(s),i}\right\}=\mathbf{0}.
\tag{S3.37}
$$

Based on Conditions 2(i), A1.1, S.3(i) and A1.4, we see that

$$
\begin{aligned}
&\left\|\mathbf{V}(C'_{(s),ni}|\mathbf{x}_{(s),i})\right\| \\
&\le \frac{\sigma^2_{\max}}{n-1}\left\|\mathrm{E}\left\{\frac{\partial k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}^{\mathrm{T}}_{(s),j}\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}^*_{(s)}}\frac{\partial k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}^{\mathrm{T}}_{(s),j}\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}^{*T}_{(s)}}\Big|\mathbf{x}_{(s),i}\right\}\right\| \\
&\le \frac{\sigma^2_{\max}}{n-1}\int\frac{\|\mathbf{x}_{(s),i}-\mathbf{x}_{(s),j}\|^2}{h_s^4}k'^2_{h_s}\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j} \\
&\le \frac{2\sigma^2_{\max}\max_{1\le s\le S_n}\max_{1\le i\le n}\|\mathbf{x}_{(s),i}\|^2}{n-1} \\
&\quad\times\int\frac{1}{h_s^4}k'^2_{h_s}\left(\frac{t_{(s),i}-t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j} \\
&= \frac{2\sigma^2_{\max}\max_{1\le s\le S_n}\max_{1\le i\le n}\|\mathbf{x}_{(s),i}\|^2}{h_s^3(n-1)} \\
&\quad\times\int\left[f_{(s)}(t_{(s),i})-h_s\tau f'_{(s)}(t_{(s),i})+h_s\tau\{f'_{(s)}(t_{(s),i})-f'_{(s)}(\tilde{t})\}\right]k'^2(\tau)\,\mathrm{d}\tau \\
&\le \frac{O_{uP}(p_s)}{h_s^3(n-1)}\left\{f_{(s)}(t_{(s),i})\int k'^2(\tau)\,\mathrm{d}\tau+h_s|f'_{(s)}(t_{(s),i})|\int|\tau|k'^2(\tau)\,\mathrm{d}\tau+O(h_s^2)\right\} \\
&= O_{uP}(h_s^{-3}n^{-1}p_s), \qquad\qquad (S3.38)
\end{aligned}
$$

where $\tilde{t}$ lies between $t_{(s),i}$ and $t_{(s),i}-\tau h_s$, $O_u(h_s^2)$ in the second-to-last equality is obtained similarly as (S3.23). Meanwhile, based on Conditions A1.1, S.4(ii) and S.4(iii), we obtain that

$$
\begin{aligned}
\mathrm{E}\left\{\|\mathbf{V}(C'_{(s),ni}|\mathbf{x}_{(s),i})\|\right\} &= \frac{O_{uP}(p_s)}{h_s^3(n-1)}\left\{\int f^2_{(s)}(t_{(s),i})\,\mathrm{d}t_{(s),i}\int k'^2(\tau)\,\mathrm{d}\tau\right. \\
&\quad\left.+h_s\int|f'_{(s)}(t_{(s),i})|f_{(s)}(t_{(s),i})\,\mathrm{d}t_{(s),i}\int|\tau|k'^2(\tau)\,\mathrm{d}\tau+O_u(h_s^2)\right\} \\
&= O_{uP}(h_s^{-3}n^{-1}p_s). \qquad\qquad (S3.39)
\end{aligned}
$$

Again, like (S3.25), we can show, based on (S3.37)–(S3.39), that

$$
\begin{aligned}
\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{i=1}^n\left\|C'_{(s),ni}\right\|^2 &\le \max_{1\le s\le S_n}\frac{2}{np_s}\sum_{i=1}^n\|\mathrm{E}(C'_{(s),ni}|\mathbf{x}_{(s),i})\|^2 \\
&\quad+\max_{1\le s\le S_n}\frac{2}{np_s}\sum_{i=1}^n\|C'_{(s),ni}-\mathrm{E}(C'_{(s),ni}|\mathbf{x}_{(s),i})\|^2
\end{aligned}
$$

$$= O_P\left(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s\right). \tag{S3.40}$$

This completes the proof of (S3.8).

Finally, as for (S3.9), from Condition 2(ii), we have that

$$\begin{aligned}
\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\frac{A_{(s),ni}}{B_{(s),ni}}\right| &= \max_{1\le s\le S_n}\max_{1\le i\le n}\left|\frac{\sum_{j\ne i}^n \mu_j k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right| \\
&\le \max_{1\le i\le n}|\mu_i| \max_{1\le s\le S_n}\left|\frac{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right| \\
&= \max_{1\le i\le n}|\mu_i| = O(1).
\end{aligned}$$

This completes the proof. □

Lemma S.2. *Under Conditions A1.1, A1.2, S.3(iii), S.4(ii) and S.4(vi), we have*

$$\max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) - \frac{1}{B_{(s),ni}(n-1)}\sum_{j\ne i}^n\{g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) + \epsilon_j\}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|$$
$$= o_P(1)$$

Proof of Lemma S.2. Under Conditions A1.1, S.3(iii) and S.4(vi), we observe that

$$\begin{aligned}
&\mathrm{E}\left\{\frac{1}{n-1}\sum_{j\ne i}^n\{g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) + \epsilon_j\}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\middle|\mathbf{x}_{(s),i}\right\} \\
&= \mathrm{E}\left\{g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\frac{1}{h_s}k\left(\frac{\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*}{h_s}\right)\middle|\mathbf{x}_{(s),i}\right\} \\
&= \frac{1}{h_s}\int g_{(s)}(t_{(s),j})k\left(\frac{t_{(s),i} - t_{(s),j}}{h_s}\right)f_{(s)}(t_{(s),j})\,\mathrm{d}t_{(s),j}
\end{aligned}$$

$$
\begin{aligned}
&= \int g_{(s)}(t_{(s),i} - \tau h_s) k(\tau) f_{(s)}(t_{(s),i} - \tau h_s) \,\mathrm{d}\tau \\
&= \int \left[\phi_{(s)}(t_{(s),i}) - \tau h_s \phi'_{(s)}(t_{(s),j}) + \tau h_s \{\phi'_{(s)}(t_{(s),j}) - \phi'_{(s)}(\tilde{t})\}\right] k(\tau) \,\mathrm{d}\tau \\
&= g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}} \boldsymbol{\beta}_{(s)}^{*}) f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}} \boldsymbol{\beta}_{(s)}^{*}) + O_u(h_s^2), \qquad \text{(S3.41)}
\end{aligned}
$$

where $\tilde{t}$ lies between $t_{(s),j}$ and $t_{(s),j} - \tau h_s$, and $O_u(h_s^2)$ in the last quality is obtained similarly as (S3.23). On the other hand, we have

$$
\begin{aligned}
&\mathbf{V}\left[\frac{1}{n-1}\sum_{j\neq i}^{n}\{g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + \epsilon_j\} k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \middle| \mathbf{x}_{(s),i}\right] \\
&= \frac{1}{n-1}\mathbf{V}\left[\{g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + \epsilon_j\} k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \middle| \mathbf{x}_{(s),i}\right] \\
&\leq \frac{1}{n-1}\mathrm{E}\left[\{g_{(s)}^2(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) + \sigma_j^2\} k_{h_s}^2(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \middle| \mathbf{x}_{(s),i}\right] \\
&= \frac{1}{h_s^2(n-1)}\int \{g_{(s)}^2(t_{(s),j}) + \sigma_j^2\} k^2\left(\frac{t_{(s),i} - t_{(s),j}}{h_s}\right) f_{(s)}(t_{(s),j}) \,\mathrm{d}t_{(s),j} \\
&\leq \frac{1}{h_s(n-1)}\int \varphi_{(s)}(t_{(s),i} - \tau h_s) k^2(\tau) \,\mathrm{d}\tau \\
&\quad + \frac{\sigma_{\max}^2}{h_s(n-1)}\int f_{(s)}(t_{(s),i} - \tau h_s) k^2(\tau) \,\mathrm{d}\tau \\
&= \frac{1}{h_s(n-1)}\int \left[\varphi_{(s)}(t_{(s),i}) - \tau h_s \varphi'_{(s)}(t_{(s),j}) + \tau h_s\{\varphi'_{(s)}(t_{(s),j}) - \varphi'_{(s)}(\tilde{t})\}\right] k^2(\tau) \,\mathrm{d}\tau \\
&\quad + \frac{\sigma_{\max}^2}{h_s(n-1)}\int \left[f_{(s)}(t_{(s),i}) - \tau h_s f'_{(s)}(t_{(s),j}) + \tau h_s\{f'_{(s)}(t_{(s),j}) - f'_{(s)}(\tilde{t})\}\right] k^2(\tau) \,\mathrm{d}\tau \\
&= \frac{1}{h_s(n-1)}\left\{\varphi_{(s)}(t_{(s),i})\int k^2(\tau) \,\mathrm{d}\tau + O_u(h_s^2)\right\} \\
&\quad + \frac{\sigma_{\max}^2}{h_s(n-1)}\left\{f_{(s)}(t_{(s),i})\int k^2(\tau) \,\mathrm{d}\tau + O_u(h_s^2)\right\}
\end{aligned}
$$

$$= \{g_{(s)}^2(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) + \sigma_{\max}^2\}O_{uP}\left(h_s^{-1}n^{-1}\right) + O_u(h_s n^{-1}), \tag{S3.42}$$

where $\tilde{t}$ lies between $t_{(s),j}$ and $t_{(s),j} - \tau h_s$, and $O_u(h_s^2)$ in the second-to-last equality is calculated in the same manner as (S3.23). Now, from Lemma S.1 and combining (S3.41) and (S3.42), we obtain

$$\begin{aligned}
&\max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) - \frac{1}{B_{(s),ni}(n-1)}\sum_{j\neq i}^{n} g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right| \\
&= \max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) - \frac{g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) + O_{uP}(h_s^2) + O_{uP}(h_s^{-1/2}n^{-1/2})}{f_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*) + O_P(h_s^2) + O_P(h_s^{-1/2}n^{-1/2})}\right| \\
&= O_P\left(\max_{1\le s\le S_n} h_s^2\right) + O_P\left(\max_{1\le s\le S_n} h_s^{-1/2}n^{-1/2}\right) \\
&= o_P(1), \qquad\qquad (S3.43)
\end{aligned}$$

where the last second equality is due to Condition S.4(ii) and the last equality is from Condition A1.2.

□

Lemma S.3. *If Conditions 1– 4 and A1.1–A1.4 hold, then*

$$\max_{1\le s\le S_n}\left\|\sqrt{\frac{n}{S_n p_s}}\frac{\partial H_{(s),n}(\boldsymbol{\beta}_{(s)}^*)}{\partial\boldsymbol{\beta}_{(s)}}\right\| = O_P(1). \tag{S3.44}$$

Proof of Lemma S.3. We first note that

$$\begin{aligned}
&\max_{1\le s\le S_n}\left\|\sqrt{\frac{n}{S_n p_s}}\frac{\partial H_{(s),n}(\boldsymbol{\beta}_{(s)}^*)}{\partial\boldsymbol{\beta}_{(s)}}\right\| \\
&= \max_{1\le s\le S_n}\left\|\frac{2}{(nS_np_s)^{1/2}}\sum_{i=1}^{n}\left\{y_i - \widehat{g}_{(s),i}(\boldsymbol{\beta}_{(s)}^*)\right\}\frac{\partial\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\partial\boldsymbol{\beta}_{(s)}}\right\| \\
&= \max_{1\le s\le S_n}\left\|\frac{2}{(nS_np_s)^{1/2}}\sum_{i=1}^{n}\left\{y_i - \widehat{g}_{(s),i}(\boldsymbol{\beta}_{(s)}^*)\right\}\left(\sum_{j\neq i}^{n}\mu_j\mathbf{m}_{(s),ij} + \sum_{j\neq i}^{n}\epsilon_j\mathbf{m}_{(s),ij}\right)\right\|
\end{aligned}$$

$$\leq \max_{1\leq s\leq S_n} \frac{2}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \epsilon_i \mathbf{z}_{(s),i}\right\| + \max_{1\leq s\leq S_n} \frac{2}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \sum_{j\neq i}^n \epsilon_i\epsilon_j \mathbf{m}_{(s),ij}\right\|$$
$$+ \max_{1\leq s\leq S_n} \frac{2}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \left\{\mu_i - \widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\} \mathbf{z}_{(s),i}\right\|$$
$$+ \max_{1\leq s\leq S_n} \frac{2}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \left\{\mu_i - \widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\} \sum_{j\neq i}^n \epsilon_j \mathbf{m}_{(s),ij}\right\|$$
$$= \Pi_1 + \Pi_2 + \Pi_3 + \Pi_4, \tag{S3.45}$$

where

$$\mathbf{m}_{(s),ij} = \frac{\partial}{\partial \boldsymbol{\beta}_{(s)}} \frac{k_{h_s,ij}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\neq i}^n k_{h_s,ij}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}, \tag{S3.46}$$
$$\mathbf{z}_{(s),i} = \sum_{j\neq i}^n \mu_j \mathbf{m}_{(s),ij} = \frac{A'_{(s),ni}}{B_{(s),ni}} - \frac{A_{(s),ni}}{B_{(s),ni}} \frac{B'_{(s),ni}}{B_{(s),ni}}. \tag{S3.47}$$

Next, we will examine each term in (S3.45). First, we calculate the volume of Π_1. For $\delta > 0$, when n is large enough,

$$\Pr\left(\max_{1\leq s\leq S_n} \frac{1}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}}\right\| > \delta \Big| \mathbf{x}_i, i=1,...,n.\right)$$
$$\leq \sum_{s=1}^{S_n} \Pr\left(\frac{1}{(nS_np_s)^{1/2}} \left\|\sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}}\right\| > \delta \Big| \mathbf{x}_i, i=1,...,n.\right)$$
$$. \leq n^{-1}S_n^{-1}\delta^{-2} \sum_{s=1}^{S_n} p_s^{-1} \mathrm{E}\left(\left\|\sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}}\right\|^2 \Big| \mathbf{x}_i, i=1,...,n.\right)$$
$$\leq \delta^{-2} n^{-1} \max_{1\leq s\leq S_n} p_s^{-1} \sum_{i=1}^n \sigma_i^2 \left\|\frac{A'_{(s),ni}}{B_{(s),ni}}\right\|^2$$
$$\leq \frac{\sigma_{\max}^2 n^{-1} \max_{1\leq s\leq S_n} \sum_{i=1}^n p_s^{-1} \|A'_{(s),ni}\|^2}{\delta^2 \min_{1\leq s\leq S_n} \min_{1\leq i\leq n} B_{(s),ni}^2}$$
$$\leq \frac{\sigma_{\max}^2 \{O_P(1) + O_P(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s)\}}{\delta^2 \{\underline{c} + o_P(1)\}^2}$$

$$= O_P(\delta^{-2}) \to 0 \text{ as } \delta \to \infty, \tag{S3.48}$$

where the second-to-last equality is based on Lemma S.1. According to van de Geer (2000), (S3.48) further implies that

$$\lim_{\delta\to\infty} \limsup_{n\to\infty} \Pr\left(\max_{1\le s\le S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}} \right\| > \delta \middle| \mathbf{x}_i, i = 1, ..., n. \right) = 0, \tag{S3.49}$$

in probability. Furthermore, according to Dominated Convergence Theorem and Theorem 6.5.6 of (Ash, 1972), we have

$$\begin{aligned} &\lim_{\delta\to\infty} \limsup_{n\to\infty} \Pr\left(\max_{1\le s\le S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}} \right\| > \delta \right) \\ &\le \mathrm{E}\left\{ \lim_{\delta\to\infty} \limsup_{n\to\infty} \Pr\left((nS_np_s)^{-1/2} \left\| \sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}} \right\| > \delta \middle| \mathbf{x}_i, i = 1, ..., n. \right) \right\} \\ &= 0, \end{aligned} \tag{S3.50}$$

which indicates

$$\max_{1\le s\le S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^n \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}} \right\| = O_P(1). \tag{S3.51}$$

From LemmaS.1, we also have

$$\begin{aligned} &\max_{1\le s\le S_n} \frac{1}{np_s} \sum_{i=1}^n \left\| \frac{A_{(s),ni}}{B_{(s),ni}} \frac{B'_{(s),ni}}{B_{(s),ni}} \right\|^2 \\ &\le \max_{1\le s\le S_n} \max_{1\le i\le n} \left| \frac{A_{(s),ni}}{B_{(s),ni}} \right|^2 \max_{1\le s\le S_n} \frac{1}{np_s} \sum_{i=1}^n \left\| \frac{B'_{(s),ni}}{B_{(s),ni}} \right\|^2 \\ &= O(1) \frac{\max_{1\le s\le S_n} n^{-1} \sum_{i=1}^n p_s^{-1} \|B'_{(s),ni}\|^2}{\min_{1\le s\le S_n} \min_{1\le i\le n} B_{(s),ni}} \end{aligned}$$

$$\leq \frac{O_P(1) + O_P\left(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s\right)}{\underline{c} + o_P(1)}$$
$$= O_P(1). \tag{S3.52}$$

So in the similar manner as (S3.51),

$$\max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \epsilon_i \frac{A_{(s),ni}}{B_{(s),ni}} \frac{B'_{(s),ni}}{B_{(s),ni}} \right\| = O_P(1) \tag{S3.53}$$

Putting (S3.51) and (S3.53) together, we obtain

$$\max_{1\leq s\leq S_n} \left\| (nS_np_s)^{-1/2} \sum_{i=1}^{n} \epsilon_i \mathbf{z}_{(s),i} \right\|$$
$$= \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \epsilon_i \left(\frac{A'_{(s),ni}}{B_{(s),ni}} - \frac{A_{(s),ni}}{B_{(s),ni}} \frac{B'_{(s),ni}}{B_{(s),ni}} \right) \right\|$$
$$\leq \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \epsilon_i \frac{A'_{(s),ni}}{B_{(s),ni}} \right\| + \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \epsilon_i \frac{A_{(s),ni}}{B_{(s),ni}} \frac{B'_{(s),ni}}{B_{(s),ni}} \right\|$$
$$= O_P(1). \tag{S3.54}$$

We then consider Π_2. Making use of Lemma S.1 and Condition A1.2, it is shown that

$$\max_{1\leq s\leq S_n} \frac{1}{np_s} \sum_{i=1}^{n} \left\| \frac{C'_{(s),ni}}{B_{(s),ni}} \right\|^2 \leq \frac{\max_{1\leq s\leq S_n} n^{-1} p_s^{-1} \sum_{i=1}^{n} \|C'_{(s),ni}\|^2}{\min_{1\leq s\leq S_n} \min_{1\leq i\leq n} B^2_{(s),ni}}$$
$$\leq \frac{O_P(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s)}{\underline{c}^2 + o_P(1)}$$
$$= O_P\left(\sum_{s=1}^{S_n} h_s^{-3} n^{-1} p_s \right)$$
$$= O_P(1), \tag{S3.55}$$

and

$$\max_{1\le s\le S_n}\sum_{i=1}^{n}\frac{1}{np_s}\left\|\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^2$$
$$\le\frac{\max_{1\le s\le S_n}\max_{1\le i\le n}C^2_{(s),ni}\max_{1\le s\le S_n}n^{-1}p_s^{-1}\sum_{i=1}^{n}\|B'_{(s),ni}\|^2}{\min_{1\le s\le S_n}\min_{1\le i\le n}B^4_{(s),ni}}$$
$$\le\frac{O_P\left(\max_{1\le s\le S_n}h_s^{-1}n^{-1}\right)\left\{O_P(1)+O_P(\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s)\right\}}{\underline{c}^4+o_P(1)}$$
$$=O_P\left(1\right). \tag{S3.56}$$

Then resembling (S3.51) and based on (S3.55) and (S3.56), we also have

$$\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\epsilon_i\frac{C'_{(s),ni}}{B_{(s),ni}}\right\|=O_P\left(1\right), \tag{S3.57}$$
$$\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\epsilon_i\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|=O_P\left(1\right). \tag{S3.58}$$

Put (S3.57) and (S3.58) together,

$$\max_{1\le s\le S_n}\frac{1}{\sqrt{nS_np_s}}\left\|\sum_{i=1}^{n}\sum_{j\ne i}^{n}\epsilon_i\epsilon_j\mathbf{m}_{(s),ij}\right\|$$
$$=\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\epsilon_i\left(\frac{C'_{(s),ni}}{B_{(s),ni}}-\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right)\right\|$$
$$\le\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\epsilon_i\frac{C'_{(s),ni}}{B_{(s),ni}}\right\|+\max_{1\le s\le S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\epsilon_i\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|$$
$$=O_P(1). \tag{S3.59}$$

We now consider Π_3. Based on Lemma S.1 and Conditions 3 – 4, by Cauchy's Inequation,

$$\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{j=1}^{n}\left\|\frac{1}{n-1}\sum_{i\ne j}^{n}k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)}-\mathbf{x}^{\mathrm{T}}_{(s),j}\boldsymbol{\beta}^*_{(s)})\left(\frac{A'_{(s),ni}}{B^2_{(s),ni}}-\frac{A_{(s),ni}}{B^2_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right)\right\|^2$$

$$
\begin{aligned}
&= \max_{1\le s\le S_n} \frac{1}{np_s}\sum_{j=1}^{n}\left\|\sum_{i\neq j}^{n}\frac{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})}{(n-1)B_{(s),ni}}\left(\frac{A'_{(s),ni}}{B_{(s),ni}}-\frac{A_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right)\right\|^{2}\\
&\le \max_{1\le s\le S_n} \frac{1}{np_s}\sum_{j=1}^{n}\left\{\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right\}\sum_{i\neq j}^{n}\left\|\frac{A'_{(s),ni}}{B_{(s),ni}}-\frac{A_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^{2}\\
&\le \max_{1\le s\le S_n} \frac{1}{p_s}\left\{\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right\}\sum_{i=1}^{n}\left\|\frac{A'_{(s),ni}}{B_{(s),ni}}-\frac{A_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^{2}\\
&= O_P\left(n\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right)\\
&= O_P(1), \qquad \text{(S3.60)}
\end{aligned}
$$

where $K_{(s),ij}(\boldsymbol{\beta}_{(s)}^{*}) = k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})/\sum_{j\neq i}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})$ and the penultimate equality is due to the following equalities (S3.61) and (S3.62):

$$
\begin{aligned}
&\max_{1\le s\le S_n}\max_{1\le j\le n}\frac{1}{p_s}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\sum_{i=1}^{n}\left\|\frac{A'_{(s),ni}}{B_{(s),ni}}\right\|^{2}\\
&\le \max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\frac{\max_{1\le s\le S_n}\sum_{i=1}^{n}p_s^{-1}\|A'_{(s),ni}\|^{2}}{\min_{1\le s\le S_n}\min_{1\le i\le n}B_{(s),ni}^{2}}\\
&\le n\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\frac{O_P(1)+O_P(\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s)}{\underline{c}^{2}+o_P(1)}\\
&= O_P\left(n\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right) \qquad \text{(S3.61)}
\end{aligned}
$$

and

$$
\begin{aligned}
&\max_{1\le s\le S_n}\max_{1\le j\le n}\frac{1}{p_s}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\sum_{i=1}^{n}\left\|\frac{A_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^{2}\\
&\le \left(\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right)\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\frac{A_{(s),ni}}{B_{(s),ni}}\right|^{2}\frac{\max_{1\le s\le S_n}\sum_{i=1}^{n}p_s^{-1}\|B'_{(s),ni}\|^{2}}{\min_{1\le s\le S_n}\min_{1\le i\le n}B_{(s),ni}^{2}}\\
&= \left(n\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right)O(1)\frac{O_P(1)+O_P(\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s)}{\underline{c}^{2}+o_P(1)}\\
&= O_P\left(n\max_{1\le s\le S_n}\max_{1\le j\le n}\sum_{i\neq j}^{n}K_{(s),ij}^{2}(\boldsymbol{\beta}_{(s)}^{*})\right). \qquad \text{(S3.62)}
\end{aligned}
$$

Then resembling (S3.60) and by Chebyshev's Inequation, we know

$$\max_{1\leq s\leq S_n} \frac{1}{\sqrt{nS_np_s}} \left\| \frac{1}{n-1}\sum_{j=1}^{n}\epsilon_j \sum_{i\neq j}^{n} k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \left(\frac{A'_{(s),ni}}{B^2_{(s),ni}} - \frac{A_{(s),ni}}{B^2_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}} \right) \right\| = O_P(1). \tag{S3.63}$$

Further, from Condition S.3(v), we know

$$\begin{aligned}
&\max_{1\leq s\leq S_n} \frac{2}{\sqrt{nS_np_s}} \left\| \sum_{i=1}^{n} \left(\mu_i - \frac{A_{(s),ni}}{B_{(s),ni}} \right) \mathbf{z}_{(s),i} \right\| \\
&= \max_{1\leq s\leq S_n} \frac{2}{\sqrt{nS_np_s}} \left\| \sum_{i=1}^{n} \left(\mu_i - \frac{A_{(s),ni}}{B_{(s),ni}} \right) \frac{\partial}{\partial \boldsymbol{\beta}_{(s)}} \frac{A_{(s),ni}}{B_{(s),ni}} \right\| \\
&= \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \frac{\partial}{\partial \boldsymbol{\beta}_{(s)}} \sum_{i=1}^{n} \left\{ \mu_i - \sum_{j\neq i}^{n} K_{(s),ij}(\boldsymbol{\beta}_{(s)}^{*})\mu_j \right\}^2 \right\| \\
&= O_P(1).
\end{aligned} \tag{S3.64}$$

Therefore, combining with (S3.63) and (S3.64), we have

$$\begin{aligned}
&\max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \left\{ \mu_i - \widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \right\} \mathbf{z}_{(s),i} \right\| \\
&= \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \left(\mu_i - \frac{A_{(s),ni}}{B_{(s),ni}} - \frac{C_{(s),ni}}{B_{(s),ni}} \right) \mathbf{z}_{(s),i} \right\| \\
&= O_P(1) + \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \frac{C_{(s),ni}}{B_{(s),ni}} \mathbf{z}_{(s),i} \right\| \\
&= O_P(1) + \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \left\| \sum_{i=1}^{n} \frac{C_{(s),ni}}{B_{(s),ni}} \left(\frac{A'_{(s),ni}}{B_{(s),ni}} - \frac{A_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}} \right) \right\| \\
&= O_P(1) + \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \\
&\quad \times \left\| \sum_{i=1}^{n} \frac{1}{n-1} \sum_{j\neq i}^{n} \epsilon_i k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \left(\frac{A'_{(s),ni}}{B^2_{(s),ni}} - \frac{A_{(s),ni}}{B^2_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}} \right) \right\| \\
&= O_P(1) + \max_{1\leq s\leq S_n} (nS_np_s)^{-1/2} \\
&\quad \times \left\| \frac{1}{n-1} \sum_{j=1}^{n} \epsilon_j \sum_{i\neq j}^{n} k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*} - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}) \left(\frac{A'_{(s),ni}}{B^2_{(s),ni}} - \frac{A_{(s),ni}}{B^2_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}} \right) \right\|
\end{aligned}$$

$$= O_P(1). \tag{S3.65}$$

Finally, we deal with Π_4. According to Lemma S.2 and Conditions 2(ii) and S.3(iii), we have

$$\begin{aligned}
&\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&=\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-g_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)+g_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&=\max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)-\frac{\sum_{j\ne i}^n y_j k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right|\\
&\quad+\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&\le\max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)-\frac{\sum_{j\ne i}^n\{g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)+\epsilon_j\}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right|\\
&\quad+\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\frac{\sum_{j\ne i}^n\{\mu_j-g_{(s)}(\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right|\\
&\quad+\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&\le\max_{1\le s\le S_n}\max_{1\le i\le n}\left|g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)-\frac{\sum_{j\ne i}^n\{g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)+\epsilon_j\}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j\ne i}^n k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right|\\
&\quad+2\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&=o_P(1)+2\max_{1\le s\le S_n}\max_{1\le i\le n}\left|\mu_i-g_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|\\
&=O_P(1).
\end{aligned} \tag{S3.66}$$

Resembling (S3.27), we can obtain

$$\max_{1\le s\le S_n}\frac{1}{np_s}\sum_{j=1}^n\left(\frac{1}{n-1}\sum_{i\ne j}^n k_{h_s}^{\prime 2}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right)=O_P\left(1+\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right). \tag{S3.67}$$

Then combining with (S3.66), (S3.67) and Lemma S.1, we obtain that

$$\begin{aligned}
&\frac{1}{np_s}\sum_{j=1}^{n}\left\|\frac{1}{n-1}\sum_{i\neq j}^{n}\frac{\partial k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})}{\partial\boldsymbol{\beta}_{(s)}^{*}}\frac{\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}}{B_{(s),ni}}\right\|^2\\
&\leq\frac{\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right|^2}{\min_{1\leq s\leq S_n}\min_{1\leq i\leq n}B_{(s),ni}^2}\\
&\quad\times\max_{1\leq s\leq S_n}\frac{1}{np_s}\sum_{j=1}^{n}\left(\frac{1}{n-1}\sum_{i\neq j}^{n}k_{h_s}^{\prime 2}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right)\\
&=\frac{O_P(1)}{\underline{c}^2+o_P(1)}O_P\left(1+\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right)\\
&=O_P(1), \qquad\qquad (S3.68)
\end{aligned}$$

where the second-to-last equality is due to Conditions 2(ii) and S.3(iii), and the last equality is due to A1.2. So based on (S3.68), resembling (S3.51), we obtain that

$$\begin{aligned}
&\max_{1\leq s\leq S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}\frac{C_{(s),ni}^{\prime}}{B_{(s),ni}}\right\|\\
&=\max_{1\leq s\leq S_n}(nS_np_s)^{-1/2}\left\|\frac{1}{n-1}\sum_{j=1}^{n}\epsilon_j\sum_{i\neq j}^{n}\frac{\partial k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})}{\partial\boldsymbol{\beta}_{(s)}^{*}}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}\frac{1}{B_{(s),ni}}\right\|\\
&=O_P\left(1\right). \qquad\qquad (S3.69)
\end{aligned}$$

On the other hand, based on Lemma S.1 and Condition 3, by Cauchy's inequality, we have

$$\begin{aligned}
&\max_{1\leq s\leq S_n}\frac{1}{np_s}\sum_{j=1}^{n}\left\|\frac{1}{n-1}\sum_{i\neq j}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right\}\frac{B_{(s),ni}^{\prime}}{B_{(s),ni}^2}\right\|^2\\
&\leq\max_{1\leq i\leq n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})\right|^2\max_{1\leq s\leq S_n}\frac{1}{np_s}\left\{\sum_{i=1}^{n}\frac{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*}-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{*})}{B_{(s),ni}}\left\|\frac{B_{(s),ni}^{\prime}}{B_{(s),ni}}\right\|\right\}^2
\end{aligned}$$

$$
\begin{aligned}
&\leq \max_{1\leq i\leq n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|^2 \max_{1\leq s\leq S_n}\frac{1}{np_s}\left\{\max_{1\leq j\leq n}\sum_{i\neq j}^{n}K_{(s),ij}^2(\boldsymbol{\beta}_{(s)}^*)\sum_{i=1}^{n}\left\|\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^2\right\}\\
&\leq \max_{1\leq i\leq n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|^2\left(\max_{1\leq s\leq S_n}\max_{1\leq j\leq n}\sum_{i\neq j}^{n}K_{(s),ij}^2(\boldsymbol{\beta}_{(s)}^*)\right)\max_{1\leq s\leq S_n}\sum_{i=1}^{n}\frac{1}{np_s}\left\|\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|^2\\
&\leq \max_{1\leq i\leq n}\left|\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right|^2\left(\max_{1\leq s\leq S_n}\max_{1\leq j\leq n}\sum_{i\neq j}^{n}K_{(s),ij}^2(\boldsymbol{\beta}_{(s)}^*)\right)\frac{\sum_{i=1}^{n}n^{-1}p_s^{-1}\|B'_{(s),ni}\|^2}{\min_{1\leq s\leq S_n}\min_{1\leq i\leq n}B_{(s),ni}^2}\\
&=O_P(1)\left(\max_{1\leq s\leq S_n}\max_{1\leq j\leq n}\sum_{i\neq j}^{n}K_{(s),ij}^2(\boldsymbol{\beta}_{(s)}^*)\right)\frac{O_P\left(1+\sum_{s=1}^{S_n}h_s^{-3}n^{-1}p_s\right)}{\underline{c}^2+o_P(1)}\\
&=O_P(1). \qquad (S3.70)
\end{aligned}
$$

Based on (S3.70), resembling (S3.51), we also have

$$
\begin{aligned}
&\max_{1\leq s\leq S_n}(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|\\
&=\max_{1\leq s\leq S_n}(nS_np_s)^{-1/2}\left\|\sum_{j=1}^{n}\epsilon_j\frac{1}{n-1}\sum_{i\neq j}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*-\mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\frac{B'_{(s),ni}}{B_{(s),ni}^2}\right\|\\
&=O_P(1), \qquad (S3.71)
\end{aligned}
$$

where the last equality is due to Condition A1.2.

Putting (S3.69) and (S3.71) together, we obtain

$$
\begin{aligned}
&\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}2(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\sum_{j\neq i}^{n}\epsilon_j\mathbf{m}_{(s),ij}\right\|\\
&=\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}2(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\left(\frac{C'_{(s),ni}}{B_{(s),ni}}-\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right)\right\|\\
&\leq \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}2(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\frac{C'_{(s),ni}}{B_{(s),ni}}\right\|\\
&\quad+\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}2(nS_np_s)^{-1/2}\left\|\sum_{i=1}^{n}\left\{\mu_i-\widehat{g}_{(s),i}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)\right\}\frac{C_{(s),ni}}{B_{(s),ni}}\frac{B'_{(s),ni}}{B_{(s),ni}}\right\|\\
&=O_P(1). \qquad (S3.72)
\end{aligned}
$$

Finally, combining (S3.45) with (S3.54), (S3.59), (S3.65) and (S3.72), the proof of (S3.44) is completed.

□

S3.3 Proof of Lemma 1

Proof of Lemma 1. From Lemma S.3, we have

$$\max_{1\le s\le S_n}\left\|\sqrt{\frac{n}{S_np_s}}\frac{\partial H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\right\|=O_P(1). \tag{S3.73}$$

Also, noting that $\widehat{\boldsymbol{\beta}}_{(s)}=\arg\min_{\boldsymbol{\beta}_{(s)}}H_{(s),n}(\boldsymbol{\beta}_{(s)})$, we have

$$\max_{1\le s\le S_n}\{H_{(s),n}(\widehat{\boldsymbol{\beta}}_{(s)})-H_{(s),n}(\boldsymbol{\beta}^*_{(s)})\}\le 0.$$

From Condition S.5(i), we obtain

$$\begin{aligned}0&\ge\max_{1\le s\le S_n}\{H_{(s),n}(\widehat{\boldsymbol{\beta}}_{(s)})-H_{(s),n}(\boldsymbol{\beta}^*_{(s)})\}\\&=\max_{1\le s\le S_n}\Bigg\{(\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)})^{\mathrm{T}}\frac{\partial H_{(s),n}(\partial\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\\&\quad+\frac{1}{2}(\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)})^{\mathrm{T}}\frac{\partial^2 H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}(\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)})+o_P(\|\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)}\|^2)\Bigg\}.\end{aligned}$$

Multiplying both sides by $np_s^{-1}S_n^{-1}(1+\max_{1\le s\le S_n}n^{1/2}p_s^{-1/2}S_n^{-1/2}\|\boldsymbol{\beta}_{(s)}-\boldsymbol{\beta}^*_{(s)}\|)^{-2}$, we have

$$\begin{aligned}0&\ge\max_{1\le s\le S_n}\Bigg[\mathbf{c}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{(s)})\sqrt{\frac{n}{S_np_s}}\frac{\partial H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\left\{1+\max_{1\le s\le S_n}\sqrt{\frac{n}{S_np_s}}\left\|\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)}\right\|\right\}^{-1}\\&\quad+\frac{1}{2}\mathbf{c}^{\mathrm{T}}(\widehat{\boldsymbol{\beta}}_{(s)})\frac{\partial^2 H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\mathbf{c}(\widehat{\boldsymbol{\beta}}_{(s)})\\&\quad+o_P\left\{\frac{n}{S_np_s}\left\|\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)}\right\|^2\left(1+\max_{1\le s\le S_n}\sqrt{\frac{n}{S_np_s}}\left\|\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)}\right\|\right)^{-2}\right\}\Bigg]\\&\ge-\max_{1\le s\le S_n}\|\mathbf{c}(\widehat{\boldsymbol{\beta}}_{(s)})\|\max_{1\le s\le S_n}\left\|\sqrt{\frac{n}{S_np_s}}\frac{\partial H_{(s),n}(\boldsymbol{\beta}^*_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\right\|\\&\quad\times\left\{1+\max_{1\le s\le S_n}\sqrt{\frac{n}{S_np_s}}\left\|\widehat{\boldsymbol{\beta}}_{(s)}-\boldsymbol{\beta}^*_{(s)}\right\|\right\}^{-1}+\left\{\frac{c_0}{2}+o_P(1)\right\}\max_{1\le s\le S_n}\|\mathbf{c}(\widehat{\boldsymbol{\beta}}_{(s)})\|^2,\end{aligned} \tag{S3.74}$$

where the last inequality is due to S.5(ii) and

$$\mathbf{c}_n(\widehat{\boldsymbol{\beta}}_{(s)}) = \left(1 + \max_{1\le s\le S_n} n^{1/2}p_s^{-1/2}S_n^{-1/2}\|\widehat{\boldsymbol{\beta}}_{(s)} - \boldsymbol{\beta}^*_{(s)}\|\right)^{-1} n^{1/2}p_s^{-1/2}S_n^{-1/2}\left(\widehat{\boldsymbol{\beta}}_{(s)} - \boldsymbol{\beta}^*_{(s)}\right). \tag{S3.75}$$

Finally, combining with (S3.73), if $\max_{1\le s\le S_n} n^{1/2}p_s^{-1/2}S_n^{-1/2}\|\widehat{\boldsymbol{\beta}}_{(s)} - \boldsymbol{\beta}^*_{(s)}\| \to \infty$, then inequality (S3.74) implies $\max_{1\le s\le S_n}\|\mathbf{c}(\widehat{\boldsymbol{\beta}}_{(s)})\| = o_P(1)$. However, from $\max_{1\le s\le S_n}\|\mathbf{c}(\widehat{\boldsymbol{\beta}}_{(s)})\| = o_P(1)$ and (S3.75), we can obtain $\max_{1\le s\le S_n} n^{1/2} p_s^{-1/2}S_n^{-1/2}\|\widehat{\boldsymbol{\beta}}_{(s)} - \boldsymbol{\beta}^*_{(s)}\| \to 0$. This leads to a contradiction, and therefore we can conclude that

$$\max_{1\le s\le S_n} n^{1/2}p_s^{-1/2}S_n^{-1/2}\|\widehat{\boldsymbol{\beta}}_{(s)} - \boldsymbol{\beta}^*_{(s)}\| = O_P(1).$$

Similarly, we can show (3.8) by contradiction in a similar way. □

S3.4 Proof of Theorem 1

To prove Theorem 1, we first introduce some lemmas.

Lemma S.4. *Denote*

$$\mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)}) = \begin{pmatrix} B_{1(s)}(\boldsymbol{\beta}^*_{(s)}) & & & \\ & B_{2(s)}(\boldsymbol{\beta}^*_{(s)}) & & \\ & & \ddots & \\ & & & B_{n(s)}(\boldsymbol{\beta}^*_{(s)}) \end{pmatrix}$$

and

$$\mathbf{D}(\boldsymbol{\beta}^*_{(s)}) = \begin{pmatrix} \mathbf{D}_1(\boldsymbol{\beta}^*_{(s)}) & & & \\ & \mathbf{D}_2(\boldsymbol{\beta}^*_{(s)}) & & \\ & & \ddots & \\ & & & \mathbf{D}_{J_n}(\boldsymbol{\beta}^*_{(s)}) \end{pmatrix},$$

where

$$B_{i(s)}(\boldsymbol{\beta}^*_{(s)}) = \frac{\sum_{j^*=1}^{n} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})}{\sum_{j^*\in\mathcal{A}(i)} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})}, \quad i = 1, \ldots, n,$$

and

$$\mathbf{D}_l(\boldsymbol{\beta}^*_{(s)}) = \left\{ \frac{k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),m}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j}\boldsymbol{\beta}^*_{(s)})}{\sum_{j^*=1}^{n} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),m}\boldsymbol{\beta}^*_{s} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{s})} \right\}_{M_n\times M_n}, \quad m, j = (l-1)M_n + 1, \ldots, lM_n.$$

Then we have that

$$\widetilde{\mathbf{K}}_{(s)}(\boldsymbol{\beta}^*_{(s)}) = \mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)}) \left\{ \mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)}) - \mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)}) \right\}.$$

Lemma S.4 can be obtained with some algebra, and we omit the proof here.

Lemma S.5. *Under Conditions 3 and 4, we have that (i)*$\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} = O_P(1)$. *(ii)*$\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} = O_P(M_n d_n)$. *(iii)* $\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} = 1 + O_P(M_n d_n)$. *(iv)* $\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)}) - I_n\} = O_P(M_n d_n)$.

Proof of Lemma S.5. First, we prove Lemma S.5*(i)*. Using Reisz inequality (see. e.g., Hardy et al. (1952); Speckman (1988)) and Condition 4, we have

$$\begin{aligned}
&\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} \\
&\le \max_{1\le s\le S_n} \left\{ \max_{1\le i\le n} \sum_{j=1}^{n} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \max_{1\le j\le n} \sum_{i=1}^{n} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \right\}^{1/2} \\
&= O_P(1). \qquad \text{(S3.76)}
\end{aligned}$$

Second, we prove Lemma S.5*(ii)*. From Condition 3, we have that

$$\begin{aligned}
&\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} \\
&\le \max_{1\le s\le S_n} \left\{ \max_{1\le i\le n} \sum_{j=1}^{n} |D_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \max_{1\le j\le n} \sum_{i=1}^{n} |D_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \right\}^{1/2} \\
&= \max_{1\le s\le S_n} \left\{ \max_{1\le i\le n} \sum_{j\in\mathcal{B}(i)} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \max_{1\le j\le n} \sum_{i\in\mathcal{B}(j)} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \right\}^{1/2} \\
&\le \max_{1\le s\le S_n} \left\{ M_n \max_{1\le i\le n} \max_{1\le j\le n} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \cdot M_n \max_{1\le i\le n} \max_{1\le j\le n} |K_{(s),ij}(\boldsymbol{\beta}^*_{(s)})| \right\}^{1/2} \\
&= M_n \max_{1\le s\le S_n} \max_{1\le i\le n} \max_{1\le j\le n} K_{(s),ij}(\boldsymbol{\beta}^*_{(s)}) \\
&= O_P(M_n d_n),
\end{aligned}$$

Next, Lemma S.5*(iii)* can be obtained from Condition 3 as

$$\begin{aligned}
&\max_{1\le s\le S_n} \lambda_{\max}\{\mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\} \\
&= \max_{1\le s\le S_n} \max_{1\le i\le n} \frac{\sum_{j^*=1}^{n} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})}{\sum_{j^*\in\mathcal{A}(i)} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})} \\
&= \max_{1\le s\le S_n} \max_{1\le i\le n} \left\{ 1 - \frac{\sum_{j^*\in\mathcal{B}(i)} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})}{\sum_{j^*=1}^{n} k_{h_s}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^*_{(s)} - \mathbf{x}^{\mathrm{T}}_{(s),j^*}\boldsymbol{\beta}^*_{(s)})} \right\}^{-1}
\end{aligned}$$

$$\leq \left\{1 - \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\frac{\sum_{j^*\in\mathcal{B}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right\}^{-1}$$

$$\leq \left\{1 - M_n\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\max_{1\leq j\leq n}\frac{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right\}^{-1}$$

$$= 1 + O_P(M_n d_n). \tag{S3.77}$$

Finally, we can show Lemma S.5*(iv)*. Using Condition 3, we observe that

$$\begin{aligned}
&\max_{1\leq s\leq S_n}\lambda_{\max}\{B_{(s)}(\boldsymbol{\beta}_{(s)}^*) - I_n\}\\
&= \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\frac{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*\in\mathcal{A}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)} - 1\\
&= \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\frac{\sum_{j^*\in\mathcal{B}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*\in\mathcal{A}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\\
&= \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\left[\frac{\sum_{j^*\in\mathcal{B}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\frac{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*\in\mathcal{A}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right]\\
&= \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\left[\frac{\sum_{j^*\in\mathcal{B}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right.\\
&\quad\left.\times\left\{1 - \frac{\sum_{j^*\in\mathcal{B}(i)}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right\}^{-1}\right]\\
&\leq \max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\max_{1\leq j\leq n}\left[M_n\frac{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right.\\
&\quad\left.\times\left\{1 - M_n\max_{1\leq s\leq S_n}\max_{1\leq i\leq n}\max_{1\leq j\leq n}\frac{k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}{\sum_{j^*=1}^{n}k_{h_s}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^* - \mathbf{x}_{(s),j^*}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^*)}\right\}^{-1}\right]\\
&= O_P(M_n d_n)O_P(1 + M_n d_n) = O_P(M_n d_n).
\end{aligned}$$

□

Proof of Theorem 1. For the exposition purpose, we assume non-random $\{\mathbf{x}_i\}$ in this proof by conditioning on $\{\mathbf{x}_i\}$. We can extend the arguments to random $\{\mathbf{x}_i\}$ in a straightforward manner; See, for example, Zhu et al. (2019). Define $\boldsymbol{\epsilon} = (\epsilon_1, ..., \epsilon_n)^{\mathrm{T}}$, then we can write $\mathrm{CV}_{J_n}(\mathbf{w})$ as

$$\mathrm{CV}_{J_n}(\mathbf{w})$$

$$
\begin{aligned}
&= \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \mathbf{y}\|^2 \\
&= \big\| \{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\} + \{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\} - \{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\} - \{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\} - \boldsymbol{\epsilon}\big\|^2 \\
&= \underbrace{\|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2}_{L_n(\mathbf{w})} + \|\boldsymbol{\epsilon}\|^2 + \underbrace{\|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2}_{①} + \underbrace{\|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2}_{②} + \underbrace{\|\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2}_{③} \\
&\quad + \underbrace{2\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}}_{④} + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}}_{⑤} + \underbrace{2\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}}_{⑥} \\
&\quad + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon}}_{⑦} + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}}_{⑧} + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}}_{⑨} \\
&\quad + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}}_{⑩} + \underbrace{2\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}}_{⑪} \\
&\quad + \underbrace{2\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}}_{⑫} + \underbrace{2\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}}_{⑬} \\
&\equiv L_n(\mathbf{w}) + \|\boldsymbol{\epsilon}\|^2 + \Xi_{1n}(\mathbf{w}), \qquad \text{(S3.78)}
\end{aligned}
$$

where $\Xi_{1n}(\mathbf{w})$ collects the terms ①–⑬. We can also write $L_n(\mathbf{w})$ as

$$
\begin{aligned}
L_n(\mathbf{w}) &= \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|^2 \\
&= \|\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu} + \widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2 \\
&= \underbrace{\|\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu}\|^2}_{L_n^*(\mathbf{w})} + \underbrace{\|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2}_{①} + \underbrace{2\{\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}}_{⑭} \\
&\equiv L_n^*(\mathbf{w}) + \Xi_{2n}(\mathbf{w}), \qquad \text{(S3.79)}
\end{aligned}
$$

where $\Xi_{2n}(\mathbf{w})$ collects the terms ① and ⑭. Similar to the proof of Theorem 1' in Wan et al. (2010) (see also Li (1987) and Gao et al. (2019)), Theorem

1 holds if we can prove the following two equalities:

$$\sup_{\mathbf{w}\in\mathcal{W}} |\Xi_{1n}(\mathbf{w})|/L_n^*(\mathbf{w}) = o_P(1), \quad \text{and} \quad \sup_{\mathbf{w}\in\mathcal{W}} |\Xi_{2n}(\mathbf{w})|/L_n^*(\mathbf{w}) = o_P(1). \tag{S3.80}$$

To show (S3.80), we need to examine each term in Ξ_{1n} and Ξ_{2n}. We first prove the following equations regarding ①–⑥ and related terms:

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.81}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.82}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.83}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.84}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.85}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) = o_P(1). \tag{S3.86}$$

We start with proving (S3.81). Based on Lemma 1, we observe that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2 \\
&= \sup_{\mathbf{w}\in\mathcal{W}} \left\| \sum_{s=1}^{S_n} w_s \mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\mathbf{y} - \sum_{s=1}^{S_n} w_s \mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\mathbf{y} \right\|^2 \\
&\leq \max_{1\leq s\leq S_n} \left\| \mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}_{(s)})\mathbf{y} - \mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\mathbf{y} \right\|^2 \\
&= \max_{1\leq s\leq S_n} \sum_{i=1}^{n} \left\{ \widehat{g}_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{(s)}) - \widehat{g}_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}^*_{(s)}) \right\}^2
\end{aligned}$$

$$
\begin{aligned}
&= \max_{1\le s\le S_n} \sum_{i=1}^{n} \left(\widehat{\boldsymbol\beta}_{(s)} - \boldsymbol\beta^*_{(s)}\right)^{\mathrm T} \frac{\partial \widehat g_{(s)}(\mathbf x_{(s),i}^{\mathrm T}\widetilde{\boldsymbol\beta}_{(s)}^{(i)})}{\partial \boldsymbol\beta_{(s)}} \frac{\partial \widehat g_{(s)}(\mathbf x_{(s),i}^{\mathrm T}\widetilde{\boldsymbol\beta}_{(s)}^{(i)})}{\partial \boldsymbol\beta_{(s)}^{\mathrm T}} \left(\widehat{\boldsymbol\beta}_{(s)} - \boldsymbol\beta^*_{(s)}\right)\\
&\le n \max_{1\le s\le S_n} \lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \widehat g_{(s)}(\mathbf x_{(s),i}^{\mathrm T}\widetilde{\boldsymbol\beta}_{(s)}^{(i)})}{\partial \boldsymbol\beta_{(s)}} \frac{\partial \widehat g_{(s)}(\mathbf x_{(s),i}^{\mathrm T}\widetilde{\boldsymbol\beta}_{(s)}^{(i)})}{\partial \boldsymbol\beta_{(s)}^{\mathrm T}}\right\} \max_{1\le s\le S_n}\left\|\widehat{\boldsymbol\beta}_{(s)} - \boldsymbol\beta^*_{(s)}\right\|^2\\
&= O_P(p_{\max}) n \max_{1\le s\le S_n}\left\|\widehat{\boldsymbol\beta}_{(s)} - \boldsymbol\beta^*_{(s)}\right\|^2\\
&= O_P(p_{\max}) S_n p_{\max} \max_{1\le s\le S_n} \frac{n}{S_n p_s}\left\|\widehat{\boldsymbol\beta}_{(s)} - \boldsymbol\beta^*_{(s)}\right\|^2 = O_P(S_n p_{\max}^2), \qquad \text{(S3.87)}
\end{aligned}
$$

where $\widetilde{\boldsymbol\beta}_{(s)}^{(i)}$ lies between $\widehat{\boldsymbol\beta}_{(s)}$ and $\boldsymbol\beta^*_{(s)}$ and thus also in the neighborhood $\mathcal O(\boldsymbol\beta^*_{(s)}, \rho)$ for some constant ρ according to (3.7), when n is sufficiently large. The third-to-last equality is due to (3.9) in Condition 5. Moreover, we note that

$$
\begin{aligned}
\frac{S_n p_{\max}^2}{\xi_n} &= \frac{n S_n^{1/2} p_{\max} S_n^{1/2} p_{\max}}{n\xi_n}\\
&\le \frac{S_n^{1/2} n p_{\max}}{\xi_n (n-M_n)^{1/2}} \frac{S_n^{1/2} p_{\max}}{(n-M_n)^{1/2}}\\
&= \frac{S_n^{1/2} n p_{\max}}{\xi_n (n-M_n)^{1/2}} \frac{S_n^{1/2} n p_{\max}}{\xi_n (n-M_n)^{1/2}} \frac{\xi_n}{n} = o_P(1), \qquad \text{(S3.88)}
\end{aligned}
$$

where the last equality is from Condition 6(i) and the fact that $\xi_n = O_P(n)$. Combining with (S3.87) and (S3.88), we obtain (S3.81).

The similar arguments can be used to prove

$$
\sup_{\mathbf w\in\mathcal W} \|\widetilde{\boldsymbol\mu}(\mathbf w) - \widetilde{\boldsymbol\mu}^*(\mathbf w)\|^2 = O_P\left\{S_n n p_{\max}^2 (n-M_n)^{-1}\right\}, \qquad \text{(S3.89)}
$$

and also note that $S_n n p_{\max}^2/\left\{\xi_n(n-M_n)\right\} = o_P(1)$ for the similar reasons as (S3.88). Then we can prove (S3.82).

In order to prove (S3.83). We note that for $\delta > 0$, we have

$$\Pr(\|\boldsymbol{\epsilon}\|^2 > n\delta) \leq \frac{\mathrm{E}(\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon})}{n\delta} = \frac{\sum_{i=1}^{n}\sigma_i^2}{n\delta} = O(\delta^{-1}) \to 0 \quad \text{as} \quad \delta \to \infty, \tag{S3.90}$$

where the last equality is from Condition 2(i). This implies that $\|\boldsymbol{\epsilon}\|^2 = O_P(n)$. Further, it can be shown that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \|\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2 \\
&= \sup_{\mathbf{w}\in\mathcal{W}} \left\|\mathbf{K}(\mathbf{w},\boldsymbol{\beta}^*_{(s)})\mathbf{y} - \widetilde{\mathbf{K}}(\mathbf{w},\boldsymbol{\beta}^*_{(s)})\mathbf{y}\right\|^2 \\
&\leq \max_{1\leq s\leq S_n} \left\|\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\mathbf{y} - \widetilde{\mathbf{K}}_{(s)}(\boldsymbol{\beta}^*_{(s)})\mathbf{y}\right\|^2 \\
&\leq \max_{1\leq s\leq S_n} \left\|\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)}) - \mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\{\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)}) - \mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)})\}\right\|^2 \|\mathbf{y}\|^2 \\
&\leq \max_{1\leq s\leq S_n} \left\|\left\{\mathbf{I}_n - \mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\right\}\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)}) + \mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)})\right\|^2 \|\mathbf{y}\|^2 \\
&\leq \max_{1\leq s\leq S_n} \left\{\|\mathbf{I}_n - \mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\|\|\mathbf{K}_{(s)}(\boldsymbol{\beta}^*_{(s)})\| + \|\mathbf{B}_{(s)}(\boldsymbol{\beta}^*_{(s)})\|\|\mathbf{D}_{(s)}(\boldsymbol{\beta}^*_{(s)})\|\right\}^2 (\|\boldsymbol{\mu}\|^2 + \|\boldsymbol{\epsilon}\|^2) \\
&= [O_P(M_n d_n)O_P(1) + \{1 + O_P(M_n d_n)\}O_P(M_n d_n)]^2 O_P(n) = O_P(M_n^2 n d_n^2),
\end{aligned} \tag{S3.91}$$

where the third step holds due to Lemma S.4, and the second-to-last equality is obtained from Lemma S.5 and Condition 2(ii). By combining (S3.91) and Condition 6(ii), we can obtain (S3.83).

For (S3.84), using (S3.87) and (S3.90), we can show that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) \\
&\leq \xi_n^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|\|\boldsymbol{\epsilon}\| \\
&= O_P(\xi_n^{-1} S_n^{1/2} p_{\max} n^{1/2}) = o_P(1),
\end{aligned} \tag{S3.92}$$

where the last equality holds due to Condition 6(i). Thus, (S3.84) holds.

Finally, using (S3.89), (S3.90) and Conditions 6(i) and 6(ii), we have

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) \\
&\leq \xi_n^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|\|\boldsymbol{\epsilon}\| \\
&= \xi_n^{-1} O_P\left\{S_n^{1/2} n^{1/2} p_{\max}(n-M_n)^{-1/2}\right\} O_P(n^{1/2}) = o_P(1),
\end{aligned} \tag{S3.93}$$

and

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) \leq \xi_n^{-1} \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|\|\boldsymbol{\epsilon}\| = o_P(1). \tag{S3.94}$$

In the following, we treat the remaining terms ⑦–⑭ in $\Xi_{1n}(\mathbf{w})$ and $\Xi_{2n}(\mathbf{w})$. Using (S3.81)–(S3.84) and Condition 6, we can obtain the following results:

For term ⑦,

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^*(\mathbf{w}) \\
&\leq \sup_{\mathbf{w}\in\mathcal{W}} \|\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}\|/L_n^*(\mathbf{w}) + \sup_{\mathbf{w}\in\mathcal{W}} \|\{\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon}\|/L_n^*(\mathbf{w}) = o_P(1).
\end{aligned} \tag{S3.95}$$

For term ⑧,

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) \\
&\leq \sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\|/L_n^{*\,1/2}(\mathbf{w}) \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|/L_n^{*\,1/2}(\mathbf{w}) \\
&\leq \sup_{\mathbf{w}\in\mathcal{W}} \{\|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\| + \|\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu}\|\}/L_n^{*\,1/2}(\mathbf{w}) \times \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|/L_n^{*\,1/2}(\mathbf{w})
\end{aligned}$$

$$= \{O_P(\xi_n^{-1} S_n^{1/2} p_{\max}) + 1\} o_P(1) = o_P(1). \tag{S3.96}$$

Similarly, for terms ⑨ and ⑩, it can be readily shown that

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) = o_P(1) \tag{S3.97}$$

and

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) = o_P(1). \tag{S3.98}$$

For term ⑪, it is clear that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) \\
&\leq \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|/L_n^{*1/2}(\mathbf{w}) \times \sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|/L_n^{*1/2}(\mathbf{w}) \\
&= o_P(1).
\end{aligned} \tag{S3.99}$$

Similarly, for terms ⑫, ⑬ and ⑭, we can show, respectively, that

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.100}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) = o_P(1), \tag{S3.101}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\boldsymbol{\mu}^*(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\}|/L_n^*(\mathbf{w}) = o_P(1). \tag{S3.102}$$

Then, the first equality in (S3.80) follows from (S3.81)–(S3.86) and (S3.95)–(S3.101). Combining (S3.79), (S3.81) and (S3.102), we obtain the second equality in (S3.80). This completes the proof of Theorem 1. □

S3.5 Proof of Theorem 2

Proof of Theorem 2. We first calculate the value of our criterion $\mathrm{CV}_{J_n}(\mathbf{w})$ at the special weight

$$\mathbf{w}^0 = (w_1^0, ..., w_{S_0}^0, 0, ..., 0)^{\mathrm{T}} \in \mathcal{W}.$$

It is clear that

$$\begin{aligned}
&\mathrm{CV}_{J_n}(\mathbf{w}^0) \\
&= \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \mathbf{y}\|^2 \\
&= \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0) + \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0) - \boldsymbol{\mu} - \boldsymbol{\epsilon}\|^2 \\
&= \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0)\|^2 + \|\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0) - \boldsymbol{\mu}\|^2 + \|\boldsymbol{\epsilon}\|^2 + 2\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0)\}^{\mathrm{T}}\boldsymbol{\epsilon} \\
&\quad + 2\{\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon} + 2\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0)\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0) - \boldsymbol{\mu}\}. \qquad \text{(S3.103)}
\end{aligned}$$

We treat each term in (S3.103) below. From (S3.89), we observe that

$$\begin{aligned}
\|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w}^0)\|^2 &= \left\|\sum_{s=1}^{S_0} w_s^0 \left(\widetilde{\boldsymbol{\mu}}_{(s)} - \widetilde{\boldsymbol{\mu}}^*_{(s)}\right)\right\|^2 \\
&\le \sup_{\mathbf{w}\in\mathcal{W}\setminus\mathcal{W}_F} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2 \\
&\le \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2 = O_P\left\{S_n n p_{\max}^2 (n - M_n)^{-1}\right\}.
\end{aligned} \qquad \text{(S3.104)}$$

Recall that $\|\boldsymbol{\mu}_{(s)}^{*}-\boldsymbol{\mu}\|=O_P(nh_s^4+h_s^{-1})$ when the s^{th} candidate model is correct. Then from (S3.91), we have that

$$\begin{aligned}
\|\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)-\boldsymbol{\mu}\|^2 &= \left\|\sum_{s=1}^{S_0} w_s^0(\widetilde{\boldsymbol{\mu}}_{(s)}^{*}-\boldsymbol{\mu}_{(s)}^{*})+\sum_{s=1}^{S_0} w_s^0(\boldsymbol{\mu}_{(s)}^{*}-\boldsymbol{\mu})\right\|^2 \\
&\le 2\sup_{\mathbf{w}\in\mathcal{W}}\|\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w})-\boldsymbol{\mu}^{*}(\mathbf{w})\|^2+2\max_{1\le s\le S_0}\left\|\boldsymbol{\mu}_{(s)}^{*}-\boldsymbol{\mu}\right\|^2 \\
&= O_P(M_n^2nd_n^2)+O_P\left\{\max_{1\le s\le S_0}(nh_s^4+h_s^{-1})\right\} \\
&= O_P(M_n^2nd_n^2), \qquad \text{(S3.105)}
\end{aligned}$$

where the last equality is due to Condition S.2(iii). Thus, from (S3.90), (S3.104) and (S3.105), we obtain

$$\begin{aligned}
|\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0)-\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)\}^{\mathrm{T}}\boldsymbol{\epsilon}| &\le \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0)-\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)\|\|\boldsymbol{\epsilon}\| \\
&= O_P\left\{S_n^{1/2}n^{1/2}p_{\max}(n-M_n)^{-1/2}\right\}O_P(n^{1/2}) \\
&= O_P\left\{S_n^{1/2}np_{\max}(n-M_n)^{-1/2}\right\}, \qquad \text{(S3.106)}
\end{aligned}$$

$$|\{\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)-\boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon}| \le \|\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)-\boldsymbol{\mu}\|\|\boldsymbol{\epsilon}\| = O_P(M_nnd_n), \qquad \text{(S3.107)}$$

$$\begin{aligned}
&|\{\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0)-\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)-\boldsymbol{\mu}\}| \\
&\le \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}^0)-\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)\|\|\widetilde{\boldsymbol{\mu}}^{*}(\mathbf{w}^0)-\boldsymbol{\mu}\| \\
&= O_P\left\{S_n^{1/2}p_{\max}M_nnd_n(n-M_n)^{-1/2}\right\}. \qquad \text{(S3.108)}
\end{aligned}$$

Combining (S3.103)–(S3.108) and considering that $M_n d_n = o(1)$ and $S_n^{1/2} p_{\max}(n-M_n)^{-1/2} = o(1)$ implied by Condition 7, we can show

$$\mathrm{CV}_{J_n}(\mathbf{w}^0) - \|\boldsymbol{\epsilon}\|^2 = O_P\left\{S_n^{1/2} n p_{\max}(n-M_n)^{-1/2}\right\} + O_P(M_n n d_n). \tag{S3.109}$$

We now calculate the value of our criterion $\mathrm{CV}_{J_n}(\mathbf{w})$ at the weight estimator $\widehat{\mathbf{w}}$. It is obvious that

$$\begin{aligned}
&\mathrm{CV}_{J_n}(\widehat{\mathbf{w}}) \\
&= \|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) + \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}}) + \boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu} - \boldsymbol{\epsilon}\|^2 \\
&= \|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\|^2 + \|\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\|^2 + \|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\|^2 + \|\boldsymbol{\epsilon}\|^2 \\
&\quad + 2\{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\boldsymbol{\epsilon} + 2\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\boldsymbol{\epsilon} + 2\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon} \\
&\quad + 2\{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\} + \{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\} \\
&\quad + 2\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\}.
\end{aligned} \tag{S3.110}$$

In the following, we consider each term of (S3.110). From (S3.89) and (S3.91),

$$\|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\|^2 \le \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^*(\mathbf{w})\|^2 = O_P\left\{S_n n p_{\max}^2 (n-M_n)^{-1}\right\}, \tag{S3.111}$$

and

$$\|\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\|^2 \le \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}^*(\mathbf{w}) - \boldsymbol{\mu}^*(\mathbf{w})\|^2 = O_P(M_n^2 n d_n^2). \tag{S3.112}$$

Combining (S3.90), (S3.111) and (S3.112), we obtain

$$
\begin{aligned}
|\{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\boldsymbol{\epsilon}| &\leq \|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\|\|\boldsymbol{\epsilon}\| \\
&= O_P\left\{S_n^{1/2}(n-M_n)^{-1/2}n^{1/2}p_{\max}\right\}O_P(n^{1/2}) \\
&= O_P\left\{S_n^{1/2}(n-M_n)^{-1/2}np_{\max}\right\}, \qquad \text{(S3.113)}
\end{aligned}
$$

$$
|\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\boldsymbol{\epsilon}| \leq \|\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\|\|\boldsymbol{\epsilon}\| = O_P(M_n n d_n), \qquad \text{(S3.114)}
$$

$$
\begin{aligned}
&|\{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\}| \\
&\leq \|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\|\|\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\| \\
&= O_P\left\{S_n^{1/2}d_n M_n n p_{\max}(n-M_n)^{-1/2}\right\}, \qquad \text{(S3.115)}
\end{aligned}
$$

$$
\begin{aligned}
|\{\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\}| &\leq \|\widetilde{\boldsymbol{\mu}}(\widehat{\mathbf{w}}) - \widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}})\|\|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\| \\
&= O_P\{S_n^{1/2}n^{1/2}p_{\max}(n-M_n)^{-1/2}L_n^{*\,1/2}(\widehat{\mathbf{w}})\}, \qquad \text{(S3.116)}
\end{aligned}
$$

$$
\begin{aligned}
|\{\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\}^{\mathrm{T}}\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\}| &\leq \|\widetilde{\boldsymbol{\mu}}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}^*(\widehat{\mathbf{w}})\|\|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\| \\
&= O_P\{M_n n^{1/2} d_n L_n^{*\,1/2}(\widehat{\mathbf{w}})\}, \qquad \text{(S3.117)}
\end{aligned}
$$

and

$$
\begin{aligned}
|\{\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\}^{\mathrm{T}}\boldsymbol{\epsilon}| &\leq |\boldsymbol{\mu}^*(\widehat{\mathbf{w}})^{\mathrm{T}}\boldsymbol{\epsilon}| + |\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\epsilon}| \\
&\leq \max_{1\leq s\leq S_n}|\boldsymbol{\mu}^{*\mathrm{T}}_{(s)}\boldsymbol{\epsilon}| + |\boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\epsilon}| = O_P(S_n^{1/2}n^{1/2}). \qquad \text{(S3.118)}
\end{aligned}
$$

Considering the index of correct model is $s = 1, ..., S_0$, then we have

$$\begin{aligned}
L_n^*(\widehat{\mathbf{w}}) &= \|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}) - \boldsymbol{\mu}\|^2 = \left\|\sum_{s=1}^{S_n} \widehat{w}_s \boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu}\right\|^2 \\
&= \left\|\sum_{s=1}^{S_0} \widehat{w}_s \boldsymbol{\mu}^*_{(s)} + \sum_{s=S_0+1}^{S_n} \widehat{w}_s \boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu}\right\|^2 \\
&= \left\|\widehat{w}_\Delta \sum_{s=1}^{S_0} \frac{\widehat{w}_s}{\widehat{w}_\Delta}(\boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu}) + (1-\widehat{w}_\Delta) \sum_{s=S_0+1}^{S_n} \frac{\widehat{w}_s}{1-\widehat{w}_\Delta}(\boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu})\right\|^2 \\
&= (1-\widehat{w}_\Delta)^2 \left\|\sum_{s=S_0+1}^{S_n} \frac{\widehat{w}_s}{1-\widehat{w}_\Delta} \boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu}\right\|^2 \\
&= (1-\widehat{w}_\Delta)^2 \left\|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}_F) - \boldsymbol{\mu}\right\|^2, \qquad \text{(S3.119)}
\end{aligned}$$

where $\widehat{\mathbf{w}}_F = (0, 0, ..., \widehat{w}_{S_0+1}/(1-\widehat{w}_\Delta), ..., \widehat{w}_{S_n}/(1-\widehat{w}_\Delta))^{\mathrm{T}}$. Considering $\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}\in\mathcal{W}} \mathrm{CV}_{J_n}(\mathbf{w})$, we see that $\mathrm{CV}_{J_n}(\widehat{\mathbf{w}}) - \|\boldsymbol{\epsilon}\|^2 \le \mathrm{CV}_{J_n}(\mathbf{w}^0) - \|\boldsymbol{\epsilon}\|^2$. Combining (S3.110)–(S3.118) leads to

$$\begin{aligned}
& L_n^*(\widehat{\mathbf{w}}) \\
&= O_P\{S_n^{1/2} n p_{\max}(n-M_n)^{-1/2}\} + O_P(M_n n d_n) + O_P\{S_n n p_{\max}^2 (n-M_n)^{-1}\} \\
&\quad + O_P(M_n^2 n d_n^2) + O_P(S_n^{1/2} n^{1/2}) + O_P\{S_n^{1/2} d_n M_n n p_{\max}(n-M_n)^{-1/2}\} \\
&\quad + O_P\{S_n^{1/2} n^{1/2} p_{\max}(n-M_n)^{-1/2} L_n^{*\,1/2}(\widehat{\mathbf{w}})\} \\
&\quad + O_P\{M_n n^{1/2} d_n L_n^{*\,1/2}(\widehat{\mathbf{w}})\} \\
&= O_P\{S_n^{1/2} n p_{\max}(n-M_n)^{-1/2}\} + O_P(M_n n d_n) + O_P(S_n^{1/2} n^{1/2}) \\
&\quad + O_P\{S_n^{1/2} n^{1/2} p_{\max}(n-M_n)^{-1/2} L_n^{*\,1/2}(\widehat{\mathbf{w}})\} \\
&\quad + O_P\{M_n n^{1/2} d_n L_n^{*\,1/2}(\widehat{\mathbf{w}})\}, \qquad \text{(S3.120)}
\end{aligned}$$

since $M_n d_n = o(1)$ and $S_n^{1/2} p_{\max}(n - M_n)^{-1/2} = o(1)$ implied by Condition 7.

which further leads to

$$\begin{aligned} L_n^*(\widehat{\mathbf{w}}) &= O_P(S_n^{1/2} n^{1/2}) + O_P\{S_n n p_{\max}^2 (n - M_n)^{-1}\} + O_P(M_n^2 n d_n^2) \\ &\quad + O_P\left(M_n n d_n\right) + O_P\left\{S_n^{1/2} n p_{\max}(n - M_n)^{-1/2}\right\} \\ &= O_P\left(M_n n d_n\right) + O_P\left\{S_n^{1/2} n p_{\max}(n - M_n)^{-1/2}\right\}. \end{aligned} \tag{S3.121}$$

Thus, by (S3.119) and (S3.121), we achieve

$$(1 - \widehat{w}_\Delta)^2 = \|\boldsymbol{\mu}^*(\widehat{\mathbf{w}}_F) - \boldsymbol{\mu}\|^{-2} O_P\left\{M_n n d_n + S_n^{1/2} n p_{\max}(n - M_n)^{-1/2}\right\}. \tag{S3.122}$$

Now combining (S3.122) and Conditions 7(i) and 7(ii), it is clear that $\widehat{w}_\Delta \to 1$ in probability. This completes the proof of Theorem 2. □

S3.6 Proof of Corollary 1

Proof of Corollary 1. First, we consider the squared loss of the averaging estimator with $\widehat{\mathbf{w}}$. Based on Lemma 1 and Condition 7, we can obtain (S3.122). Moreover, we observe that

$$\begin{aligned} & L_n(\widehat{\mathbf{w}}) \\ &= \left\|\sum_{s=1}^{S_n} \widehat{w}_s \widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}\right\|^2 \\ &\leq 2\left\|\sum_{s=1}^{S_0} \widehat{w}_s (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu})\right\|^2 + 2\left\|\sum_{s=S_0+1}^{S_n} \widehat{w}_s (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu})\right\|^2 \end{aligned}$$

$$
\begin{aligned}
&\leq 2 \max_{1\leq s\leq S_0} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}\right\|^2 + 2(1-\widehat{w}_\Delta) \max_{S_0<s\leq S_n} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}\right\|^2 \\
&\leq 2 \max_{1\leq s\leq S_0} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)}\right\|^2 + 2(1-\widehat{w}_\Delta) \max_{S_0<s\leq S_n} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}\right\|^2 \\
&\leq 2 \max_{1\leq s\leq S_0} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)}\right\|^2 + 4(1-\widehat{w}_\Delta) \max_{S_0<s\leq S_n} \left(\left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)}\right\|^2 + \left\|\boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu}_{(s)}\right\|^2\right) \\
&= O_P(S_0 p_{\max}^2) + \xi_F^{-1/2} O_P\left\{(M_n n d_n)^{1/2} + S_n^{1/4} n^{1/2} p_{\max}^{1/2} (n-M_n)^{-1/4}\right\} \\
&\quad \times \left[O_P\{(S_n - S_0)p_{\max}^2\} + O(n)\right], \qquad \text{(S3.123)}
\end{aligned}
$$

where the last equality holds due to (S3.87), (S3.122) and Condition 2.

Let $\widetilde{\mathbf{w}}_F = \arg\min_{\mathbf{w}\in\mathcal{W}_F} \mathrm{CV}_{J_n}(\mathbf{w})$ and $\widetilde{\mathbf{w}}_F = (0, ..., 0, \widetilde{w}^F_{S_0+1}, \widetilde{w}^F_{S_0+2}, ..., \widetilde{w}^F_{S_n})^{\mathrm{T}}$. Next, we consider the squared loss of the averaging estimator using $\widetilde{\mathbf{w}}_F$. Then, from (S3.87), we have

$$
\begin{aligned}
\left\|\sum_{s=S_0+1}^{S_n} \widetilde{w}^F_s (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)})\right\| &\leq \sum_{s=S_0+1}^{S_n} \widetilde{w}^F_s \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)}\right\| \\
&\leq \max_{S_0<s\leq S_n} \left\|\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}^*_{(s)}\right\| \\
&= O_P\{(S_n - S_0)^{1/2} p_{\max}\}. \qquad \text{(S3.124)}
\end{aligned}
$$

On the other hand, it is clear that

$$
\left\|\sum_{s=S_0+1}^{S_n} \widetilde{w}^F_s (\boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu})\right\| \geq \inf_{\mathbf{w}\in\mathcal{W}_F} \left\|\sum_{s=S_0+1}^{S_n} w_s (\boldsymbol{\mu}^*_{(s)} - \boldsymbol{\mu})\right\| = \xi_F^{1/2}. \qquad \text{(S3.125)}
$$

So combining (S3.124), (S3.125) and Condition 8, we can write $L_n(\widetilde{\mathbf{w}}_F)$ as

$$L_n(\widetilde{\mathbf{w}}_F)$$

$$
\begin{aligned}
&= \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}_{(s)}^* + \boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\|^2 \\
&\geq \left| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}_{(s)}^*) \right\| - \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\| \right|^2 \\
&= \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\|^2 \left| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}_{(s)}^*) \right\| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\|^{-1} - 1 \right|^2 \\
&\geq \xi_F \left| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}_{(s)}^*) \right\| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\|^{-1} - 1 \right|^2 \\
&= \xi_F (a_n - 1)^2, \qquad \text{(S3.126)}
\end{aligned}
$$

where

$$
a_n = \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\widehat{\boldsymbol{\mu}}_{(s)} - \boldsymbol{\mu}_{(s)}^*) \right\| \left\| \sum_{s=S_0+1}^{S_n} \widetilde{w}_s^F (\boldsymbol{\mu}_{(s)}^* - \boldsymbol{\mu}) \right\|^{-1} \leq \frac{(S_n - S_0)^{1/2} p_{\max}}{\xi_F^{1/2}} = o_P(1), \qquad \text{(S3.127)}
$$

and the last equality in (S3.127) is due to (S3.124), (S3.125) and Condition 8(i). Thus, following the similar proof strategy of Theorem 1, we can use Conditions 5 and 8(ii) to obtain that

$$
\frac{L_n(\widetilde{\mathbf{w}}_F)}{\inf_{\mathbf{w} \in \mathcal{W}_F} L_n(\mathbf{w})} \to 1 \quad \text{in probability.} \qquad \text{(S3.128)}
$$

Finally, combining (S3.123), (S3.126) and (S3.128), we see that

$$
\begin{aligned}
&\frac{L_n(\widehat{\mathbf{w}})}{\inf_{\mathbf{w} \in \mathcal{W}_F} L_n(\mathbf{w})} \\
&= \frac{L_n(\widehat{\mathbf{w}})}{L_n(\widetilde{\mathbf{w}}_F)} \frac{L_n(\widetilde{\mathbf{w}}_F)}{\inf_{\mathbf{w} \in \mathcal{W}_F} L_n(\mathbf{w})} \\
&\leq \frac{L_n(\widehat{\mathbf{w}})}{\xi_F (a_n - 1)^2} \frac{L_n(\widetilde{\mathbf{w}}_F)}{\inf_{\mathbf{w} \in \mathcal{W}_F} L_n(\mathbf{w})} \\
&= \Bigg(\xi_F^{-1} O_P(S_0 p_{\max}^2) + \xi_F^{-3/2} O_P \left\{ (M_n n d_n)^{1/2} + S_n^{1/4} n^{1/2} p_{\max}^{1/2} (n - M_n)^{-1/4} \right\}
\end{aligned}
$$

$$\times \left[O_P\{(S_n - S_0)p_{\max}^2\} + O(n)\right] \Big) \{1 + o_P(1)\}$$

$$= \xi_F^{-3/2} O_P\left\{n^{3/2}(M_n d_n)^{1/2} + S_n^{1/4} n^{3/2} p_{\max}^{1/2} (n - M_n)^{-1/4}\right\}$$

$$= o_P(1), \tag{S3.129}$$

where the last equality holds due to Condition 8. This completes the proof of Corollary 1. □

S3.7 Proof of Corollary 2

Proof of Corollary 2. This corollary can be shown in the similar way of proving Theorem 1. First, we need prove the following equations:

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}^{R*}(\mathbf{w})\|^2 / L_n^{R*}(\mathbf{w}) = o_P(1), \tag{S3.130}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\|^2 / L_n^{R*}(\mathbf{w}) = o_P(1), \tag{S3.131}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} \|\boldsymbol{\mu}^{R*}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\|^2 / L_n^{R*}(\mathbf{w}) = o_P(1), \tag{S3.132}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}^{R*}(\mathbf{w})\}^{\mathrm{T}} \boldsymbol{\epsilon}| / L_n^{R*}(\mathbf{w}) = o_P(1), \tag{S3.133}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\}^{\mathrm{T}} \boldsymbol{\epsilon}| / L_n^{R*}(\mathbf{w}) = o_P(1), \tag{S3.134}$$

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\boldsymbol{\mu}^{R*}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\}^{\mathrm{T}} \boldsymbol{\epsilon}| / L_n^{R*}(\mathbf{w}) = o_P(1). \tag{S3.135}$$

Here (S3.132) and (S3.135) can be obtained similarly to (S3.91) and (S3.94), respectively. Thus we examine the remaining equations in turn.

We start with showing (S3.130). It can be shown that

$$
\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}}\left\|\widehat{\boldsymbol{\mu}}^{R}(\mathbf{w})-\boldsymbol{\mu}^{R*}(\mathbf{w})\right\|^2\\
&=\sup_{\mathbf{w}\in\mathcal{W}}\left\|\sum_{s=1}^{S_n}w_s\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}^{R}_{(s)})\mathbf{y}-\sum_{s=1}^{S_n}w_s\mathbf{K}_{(s)}(\boldsymbol{\beta}^{R*}_{(s)})\mathbf{y}\right\|^2\\
&\le\max_{1\le s\le S_n}\left\|\mathbf{K}_{(s)}(\widehat{\boldsymbol{\beta}}^{R}_{(s)})\mathbf{y}-\mathbf{K}_{(s)}(\boldsymbol{\beta}^{R*}_{(s)})\mathbf{y}\right\|^2\\
&=\max_{1\le s\le S_n}\sum_{i=1}^{n}\left\{\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widehat{\boldsymbol{\beta}}^{R}_{(s)})-\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\boldsymbol{\beta}^{R*}_{(s)})\right\}^2\\
&=\max_{1\le s\le S_n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right)^{\mathrm{T}}\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\left(\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right)\\
&=\max_{1\le s\le S_n}\sum_{i=1}^{n}\left(\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right)^{\mathrm{T}}\left(\mathbf{e}_{(s)}\odot\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\right)\left(\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\odot\mathbf{e}^{\mathrm{T}}_{(s)}\right)\\
&\quad\times\left(\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right)\\
&\le n\max_{1\le s\le S_n}\lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{e}_{(s)}\odot\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}_{(s)}}\right)\left(\frac{\partial\widehat{g}_{(s)}(\mathbf{x}^{\mathrm{T}}_{(s),i}\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)})}{\partial\boldsymbol{\beta}^{\mathrm{T}}_{(s)}}\odot\mathbf{e}^{\mathrm{T}}_{(s)}\right)\right\}\\
&\quad\times\max_{1\le s\le S_n}\left\|\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right\|^2\\
&=O_P(q_{\max})n\max_{1\le s\le S_n}\left\|\widehat{\boldsymbol{\beta}}^{R}_{(s)}-\boldsymbol{\beta}^{R*}_{(s)}\right\|^2\\
&=O_P(q_{\max})n^{2\alpha}S_n^{2\gamma}=O_P(n^{2\alpha}S_n^{2\gamma}q_{\max}),
\end{aligned}\tag{S3.136}
$$

where $\widetilde{\boldsymbol{\beta}}^{(i)}_{(s)}$ lies between $\widehat{\boldsymbol{\beta}}^{R}_{(s)}$ and $\boldsymbol{\beta}^{R*}_{(s)}$ and thus also in the neighborhood $\mathcal{O}(\boldsymbol{\beta}^{R*}_{(s)},\rho)$ for some constant ρ according to Condition S.6 when n is sufficiently large, $\mathbf{e}_{(s)}$ is a $p_s\times 1$ vector whose i^{th} element $e_{(s),i}=I(\widehat{\beta}^{R}_{(s),i}\neq 0$ or $\beta^{R*}_{(s),i}\neq 0)$. The third-to-last equality is due to (S1.1) in Condition S.8. Now, noting that

$$
\frac{n^{2\alpha}S_n^{2\gamma}q_{\max}}{\xi_n^R}=\frac{S_n^{\gamma}nq_{\max}^{1/2}(n-M_n)^{\alpha-1/2}}{\xi_n^R}\frac{S_n^{\gamma}nq_{\max}^{1/2}(n-M_n)^{\alpha-1/2}}{\xi_n^R}\frac{\xi_n^R n^{2\alpha}}{n^2(n-M_n)^{2\alpha-1}}
$$

$$= o_P(1), \tag{S3.137}$$

where the last equality is from Condition S.7(i) and the fact that $\xi_n = O_P(n)$.

Similar to (S3.136), it can be seen that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\|^2 \\
&= \max_{1\le s\le S_n} \sum_{i=1}^n \left\{\widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{(s)}^{H[-\lceil i/M_n\rceil]}) - \widehat{g}_{(s)}(\mathbf{x}_{(s),i}^{\mathrm{T}}\boldsymbol{\beta}_{(s)}^{R*})\right\}^2 \\
&\le \max_{1\le s\le S_n}\max_{1\le i\le n}\left(\widehat{\boldsymbol{\beta}}_{(s)}^{R[-\lceil i/M_n\rceil]} - \boldsymbol{\beta}_{(s)}^{R*}\right)^{\mathrm{T}} \left\{\sum_{i=1}^n \frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}} \frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}^{\mathrm{T}}}\right\} \\
&\quad \times \max_{1\le i\le n}\left(\widehat{\boldsymbol{\beta}}_{(s)}^{R[-\lceil i/M_n\rceil]} - \boldsymbol{\beta}_{(s)}^{R*}\right) \\
&\le n \max_{1\le s\le S_n} \lambda_{\max}\left\{\frac{1}{n}\sum_{i=1}^n \left(\mathbf{e}_{(s)} \odot \frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}}\right)\left(\frac{\partial \widehat{g}_{(s)}^{[-\mathcal{B}(i)]}(\mathbf{x}_{(s),i}^{\mathrm{T}}\widetilde{\boldsymbol{\beta}}_{(s)}^{(i)})}{\partial \boldsymbol{\beta}_{(s)}^{\mathrm{T}}} \odot \mathbf{e}_{(s)}^{\mathrm{T}}\right)\right\} \\
&\quad \times \max_{1\le i\le n}\left\|\widehat{\boldsymbol{\beta}}_{(s)}^{R[-\lceil i/M_n\rceil]} - \boldsymbol{\beta}_{(s)}^{R*}\right\|^2 \\
&= O_P(q_{\max}) n \max_{1\le s\le S_n}\max_{1\le i\le n}\left\|\widehat{\boldsymbol{\beta}}_{(s)}^{R[-\lceil i/M_n\rceil]} - \boldsymbol{\beta}_{(s)}^{R*}\right\|^2 \\
&= O_P\left\{n(n-M_n)^{2\alpha-1}S_n^{2\gamma}q_{\max}\right\},
\end{aligned} \tag{S3.138}$$

where $\widetilde{\boldsymbol{\beta}}_{(s)}^{(i)}$ lies between $\widehat{\boldsymbol{\beta}}_{(s)}^{R[-\lceil i/M_n\rceil]}$ and $\boldsymbol{\beta}_{(s)}^{R*}$ in the neighborhood $\mathcal{O}(\boldsymbol{\beta}_{(s)}^{R*}, \rho)$ with some constant ρ according to Condition S.6 for sufficiently large n and $\mathbf{e}_{(s)}$ is a $p_s \times 1$ vector whose i^{th} element $e_{(s),i} = I(\widehat{\beta}_{(s),i}^{R[-\lceil i/M_n\rceil]} \neq \beta_{(s),i}^{R*})$, and the second-to-last equality is based on (3.10) in Condition 5. Also note that

$$\begin{aligned}
\frac{n(n-M_n)^{2\alpha-1}S_n^{2\gamma}q_{\max}}{\xi_n^R} &= \frac{S_n^{\gamma} n q_{\max}^{1/2}(n-M_n)^{\alpha-1/2}}{\xi_n^R}\frac{S_n^{\gamma} n q_{\max}^{1/2}(n-M_n)^{\alpha-1/2}}{\xi_n^R}\frac{\xi_n^R}{n} \\
&= o(1),
\end{aligned} \tag{S3.139}$$

where the last equality is based on Condition S.7(i) and the fact that $\xi_n = O_P(n)$.

From (S3.136), we have

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}^{R*}(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^{R*}(\mathbf{w}) \\
&\leq \xi_n^{R-1} \sup_{\mathbf{w}\in\mathcal{W}} \|\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}^{R*}(\mathbf{w})\|\|\boldsymbol{\epsilon}\| \\
&= O_P(\xi_n^{R-1} n^{1/2+\alpha} S_n^{\gamma} q_{\max}^{1/2}) \\
&= o(1). \qquad \text{(S3.140)}
\end{aligned}$$

Further, by (S3.138), it is clear that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} |\{\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\}^{\mathrm{T}}\boldsymbol{\epsilon}|/L_n^{R*}(\mathbf{w}) \\
&\leq \xi_n^{R-1} \sup_{\mathbf{w}\in\mathcal{W}} \|\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\|\|\boldsymbol{\epsilon}\| \\
&= O_P\left\{\xi_n^{R-1} n(n-M_n)^{\alpha-1/2} S_n^{\gamma} q_{\max}^{1/2}\right\} \\
&= o_P(1), \qquad \text{(S3.141)}
\end{aligned}$$

Next, we deal with every term in the high-dimensional versions of (S3.78) and (S3.79), where all $\boldsymbol{\mu}$'s contain the superscript R. The treatments of terms ⑦ and ⑪–⑭ remains the same as in the proof of Theorem 1, but the other terms should be treated differently.

For term ⑧, we see that

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widetilde{\boldsymbol{\mu}}^R(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\}|/L_n^{R*}(\mathbf{w})$$

$$= \{O_P(\xi_n^{-1} n^{\alpha} S_n^{\gamma} q_{\max}^{1/2}) + 1\} o_P(1)$$

$$= o_P(1). \tag{S3.142}$$

For term ⑨, we can show that

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}^{R*}(\mathbf{w})\}|/L_n^{R*}(\mathbf{w})$$

$$= \{O_P(\xi_n^{-1} n^{\alpha} S_n^{\gamma} q_{\max}^{1/2}) + 1\} o_P(1)$$

$$= o_P(1). \tag{S3.143}$$

For term ⑩, we can show

$$\sup_{\mathbf{w}\in\mathcal{W}} |\{\widehat{\boldsymbol{\mu}}^R(\mathbf{w}) - \boldsymbol{\mu}\}^{\mathrm{T}}\{\boldsymbol{\mu}^{R*}(\mathbf{w}) - \widetilde{\boldsymbol{\mu}}^{R*}(\mathbf{w})\}|/L_n^{R*}(\mathbf{w})$$

$$= \{O_P(\xi_n^{-1} n^{\alpha} S_n^{\gamma} q_{\max}^{1/2}) + 1\} o_P(1)$$

$$= o_P(1). \tag{S3.144}$$

Finally, from (S3.95), (S3.99)–(S3.101), (S3.130)–(S3.135) and (S3.142)–(S3.144), we obtain

$$\sup_{\mathbf{w}\in\mathcal{W}} |\Xi_{1n}(\mathbf{w})|/L_n^{R*}(\mathbf{w}) = o_P(1). \tag{S3.145}$$

Note that by (S3.79), (S3.102) and (S3.130), we have

$$\sup_{\mathbf{w}\in\mathcal{W}} |\Xi_{2n}(\mathbf{w})|/L_n^{R*}(\mathbf{w}) = o_P(1). \tag{S3.146}$$

Therefore, similar to the proof of Theorem 1, we see that Corollary 2 holds.

□

S4 Comparison with averaging weighted semiparametric least squares

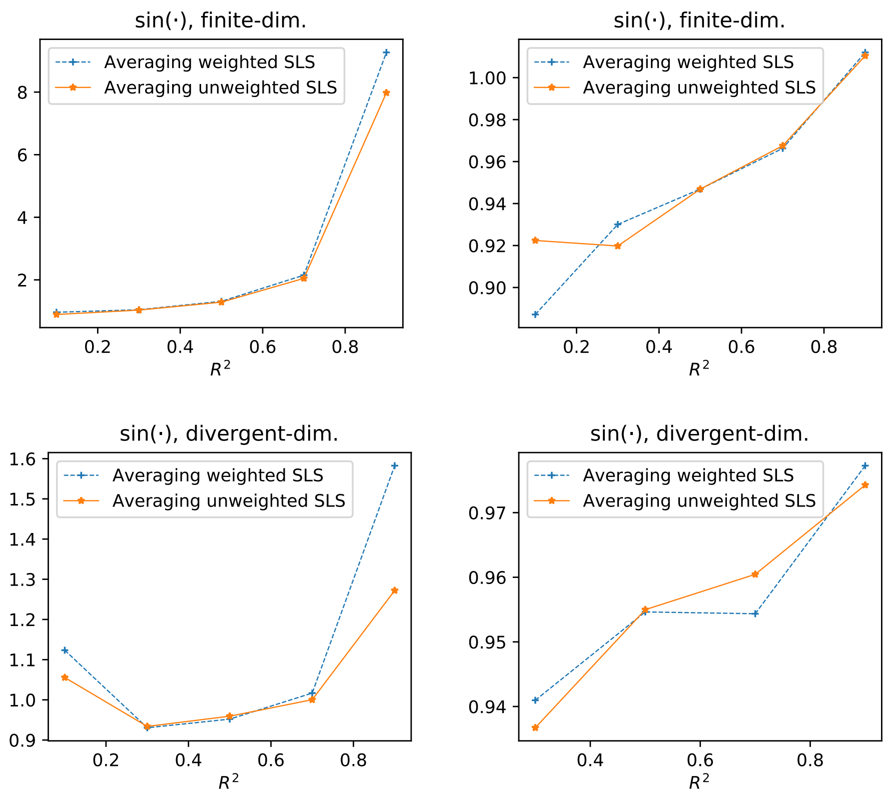
To incorporate heteroscedastic errors, one may consider employing weighted semiparametric least squares (SLS) to estimate each candidate model. Particularly, one can modify the objective function for each candidate model s to incorporate a weighting scheme as

$$H_{(s),n}(\boldsymbol{\beta}_{(s)}) = \frac{1}{n}\sum_{i=1}^{n} w(\mathbf{x}_i)\left(y_i - \sum_{j=1}^{n} K_{(s),ij}(\boldsymbol{\beta}_{(s)})y_j\right)^2$$
$$= n^{-1}\left(\mathbf{y} - \mathbf{K}_{(s)}(\boldsymbol{\beta}_{(s)})\mathbf{y}\right)^{\mathrm{T}}\mathbf{W}(\mathbf{X})\left(\mathbf{y} - \mathbf{K}_{(s)}(\boldsymbol{\beta}_{(s)})\mathbf{y}\right),$$

where $\mathbf{W}(\mathbf{X}) = \mathrm{diag}\{w(\mathbf{x}_1), w(\mathbf{x}_2), ..., w(\mathbf{x}_n)\}$, and $w(\mathbf{x}_i) = \sigma^{-2}(\mathbf{x}_i)$ is the inverse of conditional variance of y_i given $\mathbf{x}_i$. If $\mathbf{W}(\mathbf{X})$ is known, it is straightforward to extend the theoretical analysis of the proposed averaging estimator. However, $\mathbf{W}(\mathbf{X})$ is usually unknown and needs to be estimated in practice. Estimating unknown $\mathbf{W}(\mathbf{X})$ causes great theoretical challenges because all of the candidate models are likely to be misspecified in our framework and thus may lead to an inconsistent estimator of $\mathbf{W}(\mathbf{X})$. Plugging-in an inconsistent estimator $\widehat{\mathbf{W}}(\mathbf{X})$ complicates the whole theoretical analysis. Practically, it is also not clear whether the use of an estimated $\mathbf{W}(\mathbf{X})$ can improve the estimation and prediction for the same reason of possible model misspecification. Moreover, estimating unknown

$\mathbf{W}(\mathbf{X})$ introduces extra errors which may in turn inflate the variance of prediction.

To evaluate the prediction performance of averaging weighted and unweighted semiparametric least squares estimators, we consider finite- and divergent-dimensional cases with and without correct models in the model space, using simulation designs in the paper with the variance of each observation chosen from a uniform distribution $\mathcal{U}(1, 5)$. Again, only the results of $n = 300$ with the sine link function are reported here since those under other settings are largely similar. Figure S.9 presents the NMSPEs of the two estimators. We find that the two methods of estimating candidate models (weighted vs. unweighted) generally lead to similar prediction performance. The averaging estimator based on unweighted SLS even slightly outperforms the weighted version in many cases, possibly due to less estimation errors. For the reasons discussed above, we employ the unweighted SLS to estimate each candidate model in the paper. Similar treatments can be found in, e.g. Hansen and Racine (2012) and Liu and Okui (2013), who averaged ordinary least squares estimators (rather than generalized least squares estimators) of all candidate models, even though heteroscedasticity is assumed there.



Correct candidate models exist — All candidate models are misspecified

Figure S.9: NMSPEs of averaging weighted and unweighted SLS estimators ($n = 300$)

References

Ash, R. B. (1972). *Real Analysis and Probability.* New York: Academic press.

Gao, Y., X. Zhang, S. Wang, T. T. l. Chong, and G. Zou (2019). Frequentist model averaging for threshold models. *Annals of the Institute of Statistical Mathematics 71*, 275–306.

Hansen, B. E. and J. Racine (2012). Jackknife model averaging. *Journal of Econometrics 167*, 38–46.

Hardy, G. H., J. E. Littlewood, and G. Polya (1952). *Inequalities.* Cambridge university press.

Ichimura, H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics 58*, 71–120.

Kong, E. and Y. Xia (2007). Variable selection for the single-index model. *Biometrika 94*, 217–229.

Leng, C. (2010). Variable selection and coefficient estimation via regularized rank regression. *Statistica Sinica 20*, 167–181.

Li, K. (1987). Asymptotic optimality for c_p, c_l, cross-validation and generalized cross-validation: Discrete index set. *The Annals of Statistics 15*, 958–975.

Liang, H., X. Liu, R. Li, and C.-L. Tsai (2010). Estimation and testing for partially linear single-index models. *The Annals of Statistics 38*, 3811–3836.

Liu, Q. and R. Okui (2013). Heteroskedasticity-robust C_p model averaging. *The Econometrics Journal 16*, 463–472.

Naik, P. A. and C.-L. Tsai (2001). Single-index model selections. *Biometrika 88*, 821–832.

Radchenko, P. (2015). High dimensional single index models. *Journal of Multivariate Analysis 139*, 266–282.

Speckman, P. L. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological) 50*, 413–436.

van de Geer, S. (2000). *Empirical Processes in M-Estimation.* New York: Cambridge Uniersity Press.

Wan, A. T. K., X. Zhang, and G. Zou (2010). Least squares model averaging by Mallows criterion. *Journal of Econometrics 156*, 277–283.

Wang, J., L. Xue, L. Zhu, and Y. S. Chong (2011). Estimation for a partial-linear single-index model. *The Annals of Statistics 38*, 246–274.

Yu, Z., B. He, and M. Chen (2014). Empirical likelihood for generalized partially linear single-index models. *Communication in Statistics-Theory and Methods 43*, 4156–4163.

Zhu, R., A. T. K. Wan, X. Zhang, and G. Zou (2019). A Mallows-type model averaging estimator for the varying-coefficient partially linear model. *Journal of the American Statistical Association 114*, 882–892.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*, 1418–1429.