Supporting Information for "Solving the MAR Problem in Semi-supervised Learning: An Inverse Probability Weighting Method"

Jin Su, Shuyi Zhang and Yong Zhou

KLATASDS-MoE, School of Statistics, Academy of Statistics and Interdisciplinary Sciences,

East China Normal University, Shanghai 200062, China

The supplement is organized as follows. In Section S1, we provide the theoretical proofs of Lemmas 3.1–3.2 and Theorem 3.1. Section S2 contains the additional results for the simulation studies.

S1 Theoretical proofs

S1.1 Useful Lemmas

Now we introduce three important lemmas.

Lemma S1.1. *Let $\{\mathbf{X}_i\}_{i=1}^n$ be independent zero-mean random vectors in $\mathbb{R}^p$, where $p \geq 1$ and $n \geq 2$, such that for some $\alpha > 0$ and some $K_n > 0$, we have*

$$\max_{1\leq i\leq n}\max_{1\leq j\leq p}\left\|\mathbf{X}_{i[j]}\right\|_{\psi_\alpha} \leq K_n, \quad \text{and define} \quad \Gamma_n := \max_{1\leq j\leq q}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left(\mathbf{X}_{i[j]}^2\right).$$

Then, for any $t \geq 0$, with probability at least $1 - 3\exp\{-t\}$, we have

$$\left\|\frac{1}{n}\sum_{i=1}^n \mathbf{X}_i\right\|_\infty \leq 7\sqrt{\frac{\Gamma_n(t+\log p)}{n}} + \frac{C_\alpha K_n(\log n)^{1/\alpha}(t+\log p)^{1/\alpha^*}}{n},$$

where $\alpha^* := \min\{\alpha, 1\}$, and $C_\alpha > 0$ is a constant depending only on α.

Proof. The proof can be found in [1]. ∎

Lemma S1.2. *Denote $k(\cdot) = expit(\cdot)$. Consider the offset logistic model when $p < \infty$. $\pi_M(\mathbf{X}) = k(\log(\pi_M^*) + \bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma})$, $\hat{\pi}_M(\mathbf{X}) = k(\log(\hat{\pi}_M^*) + \bar{\Omega}(\mathbf{X})'\hat{\gamma})$, $p < \infty$. Suppose $\|\boldsymbol{\gamma}\|_2 < C < \infty$, $\|[\mathbb{E}\{\dot{k}(\bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma})\bar{\Omega}(\mathbf{X})\bar{\Omega}(\mathbf{X})'\}]^{-1}\|_2 < C$. As $M\pi_M^* \to \infty$,*

$$\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = M^{-1}\sum_{i=1}^M IF_{\boldsymbol{\gamma}}(\mathbf{Z}_i) + o_p\left((M\pi_M^*)^{-1/2}\right),$$
$$IF_{\boldsymbol{\gamma}}(\mathbf{Z}) := \mathcal{K}^{-1}\left(\boldsymbol{\gamma}, \pi_M^*\right)\left\{D - k(\log(\pi_M^*) + \boldsymbol{\gamma}'\bar{\Omega}(\mathbf{x}))\right\}\bar{\Omega}(\mathbf{x}) - \left(\pi_M^{*-1}D - 1\right)\mathbf{e}_1,$$

where $\mathbf{e}_1 := (1,0,\ldots,0)' \in \mathbb{R}^{Lp+1}, \mathcal{K}(\boldsymbol{\gamma}, \pi_M^*) = \mathbb{E}\left\{\bar{\Omega}(\mathbf{X})\bar{\Omega}(\mathbf{X})'\dot{k}\left(\bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma} + \log(\pi_M^*)\right)\right\}$. *We have*

$$\begin{aligned}
&\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_2 = O_p\left\{(M\pi_M^*)^{-1/2}\right\} \\
&\mathbb{E}\left[\frac{\pi_M^{**}}{\pi_M(\mathbf{X})}\left\{\frac{\pi_M(\mathbf{X})}{\hat{\pi}_M(\mathbf{X})} - 1\right\}^2\right] = O_p\left\{(M\pi_M^*)^{-1}\right\} = o_p(1), \\
&\mathbb{E}\left[\pi_M^{-r}(\mathbf{X})\right]^{1/r} \asymp \pi_M^{*-1} \quad \forall r > 0, \qquad \text{and } \pi_M^{**} \asymp \pi_M^*.
\end{aligned}$$

Proof. The proof can be found in [3]. ∎

Lemma S1.3. *Consider the offset logistic model when $p \to \infty$. $\pi_M(\mathbf{X}) = k(\log \pi_M^* + \bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma})$, $p \to \infty$. Define $\delta\mathcal{L}_M(\boldsymbol{\Delta}; a; \gamma) := \mathcal{L}_M(\boldsymbol{\gamma} + \boldsymbol{\Delta}; a) - \mathcal{L}_M(\boldsymbol{\gamma}; a) - \boldsymbol{\Delta}^T\nabla_{\boldsymbol{\gamma}}\mathcal{L}_M(\boldsymbol{\gamma}; a)$. Let $\bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma}$ be a sub-gaussian random variable, $\bar{\Omega}(\mathbf{X})$ be a marginal sub-gaussian random variable, $\|\bar{\Omega}(\mathbf{X})'\boldsymbol{\gamma}\|_{\psi_2} \leq \sigma_{\boldsymbol{\gamma}} < \infty$, $\max_{1\leq j\leq pL+1}\|\bar{\Omega}(\mathbf{X})_{[j]}\|_{\psi_2} \leq \sigma < \infty$. Suppose the restricted strong convexity (RSC) property holds for $\delta\mathcal{L}_M(\Delta; 1; \boldsymbol{\gamma})$ with parameter κ on a given set $\overline{\mathbb{C}}(S;3) := \left\{\boldsymbol{\Delta} \in \mathbb{R}^{p+1} : \|\boldsymbol{\Delta}_{S^c}\|_1 \leq 3\|\boldsymbol{\Delta}_S\|_1, \|\boldsymbol{\Delta}\|_2 \leq 1\right\}$(i.e. $\delta\mathcal{L}_M(\boldsymbol{\Delta}; 1; \boldsymbol{\gamma}) \geq \kappa\|\boldsymbol{\Delta}\|_2^2$, for all $\boldsymbol{\Delta} \in \overline{\mathbb{C}}(S;3)$) with probability at least $1 - t_M$, where $\boldsymbol{\Delta}_S = \{\boldsymbol{\Delta}_{[j]}\}_{j\in S}$ and $\boldsymbol{\Delta}_{S^c} = \{\boldsymbol{\Delta}_{[j]}\}_{j\notin S}$, $t_M = o(1)$. Assume $\log(p)\log(M) = O(M\pi_M^*)$ and $s_{\boldsymbol{\gamma}}\log(p) = o(M\pi_M^*)$ as $M, p \to \infty$, where $s_{\boldsymbol{\gamma}} := \|\boldsymbol{\gamma}\|_0$. Let*

$$C_M := C_1\sqrt{\frac{\pi_M^*\log(p+1)}{M}} + C_2\frac{\sqrt{\log(2M)}\log(p+1)}{M},$$

with some constants $C_1, C_2 > 0$. For any λ_M satisfying $2(1+c)C_M \leq \lambda_M \leq 9\kappa\pi_M^* s_{\boldsymbol{\gamma}}^{-1/2}$ with $c > 0$, whenever $M\pi_M^* > 9c\log(p+1)$,

$$\|\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\|_2 \leq \frac{1}{9}\lambda_M s_\gamma^{1/2}\pi_M^{*-1}\kappa^{-1}, \qquad \text{with probability at least } 1 - 8(p+1)^{-c} - t_M.$$

Further assume that $\left\|\bar{\Omega}(\mathbf{X})'\mathbf{v}\right\|_{\psi_2} \leq \sigma\|\mathbf{v}\|_2$ for any $\mathbf{v} \in \mathbb{R}^{p+1}$. Then, $\pi_M^* \asymp \pi_M^{**}$ and for any $0 < r \leq 2 + c_1$, with some $\lambda_M \asymp \sqrt{\pi_M^*\log(p)/M}$, and

$$\begin{aligned}
&\left\|\pi_M^{-1}(\mathbf{X})\right\|_{r,\mathbb{P}_{\mathbf{X}}} \asymp \pi_M^{*-1} \quad \forall r > 0, \quad \pi_M^* \asymp \pi_M^{**}, \\
&\left\|1 - \frac{\pi_M(\cdot)}{\hat{\pi}_M(\cdot)}\right\|_{r,\mathbb{P}_{\mathbf{X}}} = O_p\left(\sqrt{\frac{s_{\boldsymbol{\gamma}}\log(p)}{M\pi_M^*}}\right) \quad \forall r > 0, \\
&\mathbb{E}_{\mathbf{X}}\left[\frac{\pi_M^*}{\pi_M(\mathbf{X})}\left\{1 - \frac{\pi_M(\mathbf{X})}{\hat{\pi}_M(\mathbf{X})}\right\}^2\right] = O_p\left\{\frac{s_{\boldsymbol{\gamma}}\log(p)}{M\pi_M^*}\right\} = o_p(1).
\end{aligned}$$

Proof. The proof can be found in [3]. ∎

S1.2 Proof of Lemma 3.1

Proof. Denote

$$\begin{aligned}
\hat{f}_{k,w}(s;\hat{\boldsymbol{\beta}}_k) =& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{x}'\hat{\boldsymbol{\beta}}_k), \\
\hat{f}_{k,w}(s;\boldsymbol{\beta}) =& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}K_h(\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}), \\
f_w(s;\boldsymbol{\beta}) =& \mathbb{E}\left\{\frac{\pi_M(\mathbf{X})}{\pi_M^w(\mathbf{X})}\middle| S = s\right\}f(s;\boldsymbol{\beta}) = w(s)f(s;\boldsymbol{\beta}).
\end{aligned}$$

By the triangle inequality, we have

$$\left|\hat{f}_{k,w}(s;\hat{\boldsymbol{\beta}}_k) - w(s)f(s;\boldsymbol{\beta})\right| \le |\hat{f}_{k,w}(s;\hat{\boldsymbol{\beta}}_k) - \hat{f}_{k,w}(s;\boldsymbol{\beta})| + |\hat{f}_{k,w}(s;\boldsymbol{\beta}) - \mathbb{E}\{\hat{f}_{k,w}(s;\boldsymbol{\beta})\}|$$
$$+ |\mathbb{E}\{\hat{f}_{k,w}(s;\boldsymbol{\beta})\} - w(s)f(s;\boldsymbol{\beta})| := |\widehat{R}_{M,w}(s)| + |\tilde{S}_{M,w}(s)| + |\bar{S}_{M,w}(s)|.$$

Firstly, we foucus on the second term $|\tilde{S}_{M,w}(s)| = |\hat{f}_{k,w}(s;\boldsymbol{\beta}) - \mathbb{E}\{\hat{f}_{k,w}(s;\boldsymbol{\beta})\}|$. Denote $T_h(\mathbf{Z};\mathbf{x},\beta) = \frac{D}{\pi_M^w(\mathbf{X})}K_h(\mathbf{X}'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta})$, then $|T_h(\mathbf{Z};\mathbf{x},\beta)| \le C_{w,2}C_sM_K/\pi_M^*h$. Thus

$$\begin{aligned}\|T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\|_{\psi_2} \le& \frac{C_sC_{w,2}M_K}{\pi_M^*h},\\ \|T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta}) - \mathbb{E}\{T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\}\|_{\psi_2} \le& \frac{3C_sC_{w,2}M_K}{\pi_M^*h}, \quad \text{holds uniformly for all} \quad \mathbf{x}\in\mathcal{X}.\end{aligned} \tag{S1.1}$$

For the variance term, we have

$$\begin{aligned}\mathrm{Var}\{T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\} \le \mathbb{E}\{T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\}^2 &= \mathbb{E}_S[\mathbb{E}\{T_h^2(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\}|S]\\ &= h^{-2}\int_{\mathbb{R}}\mathbb{E}\left(\frac{D_j^2}{\pi_M^{w2}(\mathbf{X})}\Big|S=s\right)K^2((s-s_{\mathbf{x}})/h)f(s;\boldsymbol{\beta})ds\\ &= h^{-2}\int_{\mathbb{R}}\mathbb{E}\left(\frac{\pi_M(\mathbf{X})}{\pi_M^{w2}(\mathbf{X})}\Big|S=s\right)K^2((s-s_{\mathbf{x}})/h)f(s;\boldsymbol{\beta})ds\\ &\le \frac{C_{w,2}^2}{h}\int_{\mathbb{R}}\mathbb{E}\left(\frac{1}{\pi_M(\mathbf{X})}\Big|S=s_{\mathbf{x}}+hu\right)K^2(u)f(s_{\mathbf{x}}+hu;\boldsymbol{\beta})du\\ &\le \frac{C_{w,2}^2C_sC_fM_K}{\pi_M^*h}.\end{aligned} \tag{S1.2}$$

Acccording to Lemma S1.1, for any $t\ge 0$, with probability at least $1-3\exp(-t^2)$, we have

$$\begin{aligned}&\left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta}) - \mathbb{E}\{T_h(\mathbf{Z};\mathbf{x},\boldsymbol{\beta})\}\right\|_\infty\\ \le& 7t\sqrt{\frac{C_{w,2}^2C_sC_fM_K}{|\mathcal{Z}_{-k(i)}|\pi_M^*h}} + \frac{3C_\alpha C_sC_{w,2}M_K(\log|\mathcal{Z}_{-k(i)}|)^{1/2}t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^*h}.\end{aligned} \tag{S1.3}$$

Thus for any $t\ge 0$, with probability at least $1-3\exp(-t^2)$, we have

$$|\tilde{S}_{M,w}(s)| \le 7t\sqrt{\frac{C_{w,2}^2C_sC_fM_K}{|\mathcal{Z}_{-k(i)}|\pi_M^*h}} + \frac{3C_\alpha C_sC_{w,2}M_K(\log|\mathcal{Z}_{-k(i)}|)^{1/2}t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^*h}. \tag{S1.4}$$

For the variance term $\bar{S}_{M,w}(s)$, we have

$$\begin{aligned}
\bar{S}_{M,w}(s) &= \mathbb{E}\{\hat{f}_{k,w}(s;\boldsymbol{\beta})\} - w(s)f(s;\boldsymbol{\beta}) \\
&= \mathbb{E}\left\{\frac{D_j}{\pi_M^w(\mathbf{X}_j)} K_h(\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta})\right\} - w(s)f(s;\boldsymbol{\beta}) \\
&= h^{-1}\int \mathbb{E}\left\{\frac{\pi_M(\mathbf{X})}{\pi_M^w(\mathbf{X})}\middle| S = s\right\} K\left\{(s - s_{\mathbf{x}})/h\right\} f(s;\boldsymbol{\beta})ds - w(s)f(s;\boldsymbol{\beta}) \\
&= \int \mathbb{E}\left\{\frac{\pi_M(\mathbf{X})}{\pi_M^w(\mathbf{X})}\middle| S = s_{\mathbf{x}} + hu\right\} K(u)f(s_{\mathbf{x}} + hu;\boldsymbol{\beta})du - w(s)f(s;\boldsymbol{\beta}) \\
&= h\left\{w(s)\dot{f}(s_{\mathbf{x}};\boldsymbol{\beta}) + \dot{w}(s_{\mathbf{x}})f(s_{\mathbf{x}};\boldsymbol{\beta})\right\}\underbrace{\int uK(u)du}_{=0} + h^2R^*(\mathbf{x}) \\
&:= h^2\int_{\mathbb{R}}\left\{w(s^*_{\mathbf{x},u})\ddot{f}(s^*_{\mathbf{x},u};\boldsymbol{\beta}) + \ddot{w}(s^*_{\mathbf{x},u})f(s^*_{\mathbf{x},u};\boldsymbol{\beta}) + \dot{w}(s^*_{\mathbf{x},u})\dot{f}(s^*_{\mathbf{x},u};\boldsymbol{\beta})\right\}u^2K(u)du,
\end{aligned}$$

where $s^*_{\mathbf{x},u}$ is some "immediate"point, satisfying $\left|s^*_{\mathbf{x},u} - s_{\mathbf{x}}\right| \le h|u|$. Then we get

$$\left|\bar{S}_{M,w}(s)\right| \le h^2C_{w,f}\int_{\mathbb{R}} u^2|K(u)|du \le C_{w,f}\kappa_2h^2 \tag{S1.5}$$

holds uniformly for all $\mathbf{x} \in \mathcal{X}$. Lastly, let us focus on the first term $|\widehat{R}_{M,w}(s)| = |\hat{f}_{k,w}(s;\hat{\boldsymbol{\beta}}_k) - \hat{f}_{k,w}(s;\boldsymbol{\beta})|$. According to the first order Taylor expansion, we have

$$\widehat{R}_{M,w}(s) := (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})'\left\{\frac{1}{|\mathcal{Z}_{-k(i)}|h}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\frac{(\mathbf{X}_j - \mathbf{x})}{h}\dot{K}\left(\frac{S_j^* - s_{\mathbf{x}}^*}{h}\right)\right\},$$

where $\left\{S_j^*\right\}_{j\in\mathcal{Z}_{-k(i)}}$ and $s^*_{\mathbf{x}}$ are some "immediate" points satisfying for each $j \in \mathcal{Z}_{-k(i)}$, $\left|\left(S_j^* - s_{\mathbf{x}}^*\right) - (S_j - s_{\mathbf{x}})\right| \le \left|\left(\hat{S}_j - \hat{s}_{\mathbf{x}}\right) - (S_j - s_{\mathbf{x}})\right| \equiv \left|(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})'(\mathbf{X}_j - \mathbf{x})\right|$. Now we rewrite $\widehat{R}_{M,w}(s) = \widehat{R}_{M,1}(s) + \widehat{R}_{M,2}(s)$, where

$$\begin{aligned}
\widehat{R}_{M,1}(s) &:= (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})'\left\{\frac{1}{|\mathcal{Z}_{-k(i)}|h}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\frac{(\mathbf{X}_j - \mathbf{x})}{h}\dot{K}\left(\frac{S_j - s_{\mathbf{x}}}{h}\right)\right\} \\
&:= (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})'\hat{\mathbf{T}}_M(\mathbf{x}), \quad \widehat{R}_{M,2}(s) := \widehat{R}_{M,w}(s) - \widehat{R}_{M,1}(s).
\end{aligned} \tag{S1.6}$$

To control $\widehat{R}_{M,1}(s) = (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})'\hat{\mathbf{T}}_M(\mathbf{x})$, note that

$$|\widehat{R}_{M,1}(s)| \le \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\|_1[\|\hat{\mathbf{T}}_M(\mathbf{x}) - \mathbb{E}\{\hat{\mathbf{T}}_M(\mathbf{x})\}\|_\infty + \|\mathbb{E}\{\hat{\mathbf{T}}_M(\mathbf{x})\}\|_\infty]. \tag{S1.7}$$

Now we only need to bound $\|\hat{\mathbf{T}}_M(\mathbf{x}) - \mathbb{E}\{\hat{\mathbf{T}}_M(\mathbf{x})\}\|_\infty$ and $\|\mathbb{E}\{\hat{\mathbf{T}}_M(\mathbf{x})\}\|_\infty$. For each $\mathbf{x} \in \mathcal{X}$, define $\mathbf{T}_h^*(\mathbf{Z};\mathbf{x}) := \frac{D}{\pi_M^w(\mathbf{X})h^2}(\mathbf{X} - \mathbf{x})\dot{K}(\frac{S - s_{\mathbf{x}}}{h})$, $\hat{\mathbf{T}}_M(\mathbf{x}) = \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathbf{T}_h^*(\mathbf{Z}_j;\mathbf{x})$. Similarly, we have

$$\begin{aligned}
\max_{1\le d\le p}\left\|\mathbf{T}^*_{h[d]}(\mathbf{Z};\mathbf{x})\right\|_{\psi_2} &\le 2h^{-2}\pi_M^{*-1}C_sC_{w,2}M_{\mathbf{X}}M_{\dot{K}}, \\
\max_{1\le d\le p}\left\|\mathbf{T}^*_{h[d]}(\mathbf{Z};\mathbf{x}) - \mathbb{E}\left\{\mathbf{T}_h^*(\mathbf{Z};\mathbf{x})\right\}\right\|_{\psi_2} &\le 6h^{-2}\pi_M^{*-1}C_sC_{w,2}M_{\mathbf{X}}M_{\dot{K}},
\end{aligned}$$

For the variance term, we have

$$
\begin{aligned}
&\max_{1\le d\le p} \mathbb{E}\left[\left\{\mathbf{T}^*_{h[d]}(\mathbf{Z};\mathbf{x})\right\}^2\right]\\
\le&\frac{4}{h^4}M_{\mathbf{X}}^2\int_{\mathbb{R}}\mathbb{E}\left(\frac{\pi_M(\mathbf{X})}{\pi_M^{w*2}(\mathbf{X})}\middle|S=s\right)\left[\dot{K}\left\{(s-s_{\mathbf{x}})/h\right\}\right]^2 f(s;\boldsymbol{\beta})ds\\
=&\frac{4}{h^3}C_{w,2}^2M_{\mathbf{X}}^2\int_{\mathbb{R}}\mathbb{E}\left(\frac{1}{\pi_M(\mathbf{X})}\middle|S=s_{\mathbf{x}}+hu\right)f(s_{\mathbf{x}}+hu;\boldsymbol{\beta})\left\{\dot{K}(u)\right\}^2du\\
\le&\frac{4C_sC_fC_{w,2}^2M_{\mathbf{X}}^2M_{\dot{K}}C_{\dot{K}}}{\pi_M^*h^3}.
\end{aligned}
$$

By Lemma S1.1, for any $\mathbf{x}\in\mathcal{X}$ and $t\ge 0$, with probability at least $1-3\exp(-t^2)$,

$$
\begin{aligned}
&\left\|\hat{\mathbf{T}}_M(\mathbf{x})-\mathbb{E}\left\{\hat{\mathbf{T}}_M(\mathbf{x})\right\}\right\|_\infty=\left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathbf{T}^*_h\left(\mathbf{Z}_i;\mathbf{x}\right)-\mathbb{E}\left\{\mathbf{T}^*_h(\mathbf{Z};\mathbf{x})\right\}\right\|_\infty\\
\le&7\sqrt{\frac{4C_sC_fC_{w,2}^2M_{\mathbf{X}}^2M_{\dot{K}}C_{\dot{K}}(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^3}}+\frac{6C_sC_\alpha C_{w,2}M_{\mathbf{X}}M_{\dot{K}}(\log|\mathcal{Z}_{-k(i)}|)^{1/2}(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^2}.
\end{aligned}
\tag{S1.8}
$$

Next, we control $\|\mathbb{E}\left\{\mathbf{T}^*_h(\mathbf{Z};\mathbf{x})\right\}\|_\infty$,

$$
\begin{aligned}
\mathbb{E}\left\{\mathbf{T}^*_h(\mathbf{Z};\mathbf{x})\right\}=&\mathbb{E}\left\{\frac{D}{\pi_M^w(\mathbf{x})h^2}(\mathbf{X}-\mathbf{x})\dot{K}\left(\frac{S-s_{\mathbf{x}}}{h}\right)\right\}\\
=&\frac{1}{h^2}\mathbb{E}\left\{\frac{\pi_M(\mathbf{X})}{\pi_M^w(\mathbf{X})}(\mathbf{X}-\mathbf{x})\dot{K}\left(\frac{S-s_{\mathbf{x}}}{h}\right)\right\}\\
\le&\frac{C_{w,2}}{h^2}\int_{\mathbb{R}}\left\{\boldsymbol{\eta}^{(1)}_{\boldsymbol{\beta}}(s)-\mathbf{x}f(s;\boldsymbol{\beta})\right\}\dot{K}\left\{(s-s_{\mathbf{x}})/h\right\}ds\\
=&\frac{C_{w,2}}{h}\int_{\mathbb{R}}\left\{\boldsymbol{\eta}^{(1)}_{\boldsymbol{\beta}}\left(s_{\mathbf{x}}+hu\right)-\mathbf{x}f(s_{\mathbf{x}}+hu;\boldsymbol{\beta})\right\}\dot{K}(u)du\\
=&C_{w,2}\int_{\mathbb{R}}\left\{\dot{\boldsymbol{\eta}}^{(1)}_{\boldsymbol{\beta}}\left(s_{\mathbf{x}}+hu\right)-\mathbf{x}\dot{f}(s_{\mathbf{x}}+hu;\boldsymbol{\beta})\right\}K(u)du,
\end{aligned}
$$

$$
\begin{aligned}
\|\mathbb{E}\left\{\mathbf{T}^*_h(\mathbf{Z};\mathbf{x})\right\}\|_\infty&\le C_{w,2}\{\max_{1\le d\le p}\|\dot{\boldsymbol{\eta}}^{(1)}_{\boldsymbol{\beta}[d]}(\cdot)\|_\infty+\|\mathbf{x}\|_\infty\|\dot{f}(\cdot;\boldsymbol{\beta})\|_\infty\}\int_{\mathbb{R}}|K(u)|du\\
&\le C_{w,2}\left(C_{\eta,1,1}+M_{\mathbf{X}}C_{f,1}\right)
\end{aligned}
\tag{S1.9}
$$

It holds uniformly for all $\mathbf{x}\in\mathcal{X}$. Therefore for any fixed $\mathbf{x}\in\mathcal{X}$ and $t\ge 0$, with probability at least $1-3\exp(-t^2)-q_M$,

$$
\begin{aligned}
\left|\widehat{R}_{M,1}(s)\right|\le&b_M\left\{C_{w,2}\left(C_{\eta,1,1}+M_{\mathbf{X}}C_{f,1}\right)+7\sqrt{\frac{4C_sC_fC_{w,2}^2M_{\mathbf{X}}^2M_{\dot{K}}C_{\dot{K}}(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^3}}\right.\\
&\left.+\frac{6C_sC_\alpha C_{w,2}M_{\mathbf{X}}M_{\dot{K}}(\log|\mathcal{Z}_{-k(i)}|)^{1/2}(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^2}\right\}.
\end{aligned}
\tag{S1.10}
$$

For $\widehat{R}_{M,2}(s)$, we have

$$\begin{aligned}\left|\widehat{R}_{M,2}(s)\right| =& \left|\frac{(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})'}{|\mathcal{Z}_{-k(i)}|h^2}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}(\mathbf{X}_j-\mathbf{x})\left\{\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)-\dot{K}\left(\frac{S_j-s_{\mathbf{x}}}{h}\right)\right\}\right| \\ \leq& \frac{\|\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1}{|\mathcal{Z}_{-k(i)}|h^2}\sum_{j\in\mathcal{Z}_{-k(i)}}\|\mathbf{X}_i-\mathbf{x}\|_\infty\left|\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right|\left|\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)-\dot{K}\left(\frac{S_j-s_{\mathbf{x}}}{h}\right)\right| \\ \leq& 2M_{\mathbf{X}}C_{w,2}\|\widehat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1 \\ &\times\left\{\frac{1}{|\mathcal{Z}_{-k(i)}|h^2}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\frac{D_j}{\pi_M(\mathbf{X}_j)}\right|\left|\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)-\dot{K}\left(\frac{S_j-s_{\mathbf{x}}}{h}\right)\right|\right\}.\end{aligned} \tag{S1.11}$$

Let $\mathcal{B}_M$ represent event $\|\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1\leq b_M$. On event $\mathcal{B}_M$, we have

$$\begin{aligned}\left|\dot{K}\left(\frac{S_j-s_{\mathbf{x}}}{h}\right)-\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)\right| \leq& \frac{1}{h}\left|(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})'(\mathbf{X}_i-\mathbf{x})\right|\varphi\left(\frac{S_j-s_{\mathbf{x}}}{h}\right) \\ \leq& \frac{1}{h}\|(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1\,\|(\mathbf{X}_i-\mathbf{x})\|_\infty\,\varphi\left(\frac{S_j-s_{\mathbf{x}}}{h}\right) \\ \leq& \frac{2M_{\mathbf{X}}b_M}{h}\varphi\left(\frac{S_j-s_{\mathbf{x}}}{h}\right).\end{aligned} \tag{S1.12}$$

Thus

$$\left|\widehat{R}_{M,2}(s)\right| \leq \frac{4C_{w,2}M_{\mathbf{X}}^2b_M^2}{|\mathcal{Z}_{-k(i)}|h^3}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\frac{D_j}{\pi_M(\mathbf{X}_j)}\right|\varphi\left(\frac{S_j-s_{\mathbf{x}}}{h}\right), \quad \forall\mathbf{x}\in\mathcal{X}. \tag{S1.13}$$

Define $\mathcal{T}_h(\mathbf{Z};\mathbf{x})=\mathcal{T}_h(\mathbf{Z};\mathbf{x},\beta):=h^{-3}|D/\pi_M(\mathbf{X})|\varphi\left\{(S-s_{\mathbf{x}})/h\right\}$, we have

$$\begin{aligned}\mathbb{E}\left\{\mathcal{T}_h^2(\mathbf{Z};\mathbf{x})\right\} =& \frac{1}{h^6}\int\frac{D_j^2}{\pi_M^2(\mathbf{X}_j)}\varphi^2\left(\frac{S-s_{\mathbf{x}}}{h}\right)f(s;\boldsymbol{\beta})ds \\ =& \frac{1}{h^5}\int\mathbb{E}\left\{\frac{1}{\pi_M(\mathbf{X}_j)}\middle|S=s_{\mathbf{x}}+hu\right\}f(s_{\mathbf{x}}+hu;\boldsymbol{\beta})\varphi^2(u)du \\ \leq& \frac{C_sC_fM_\varphi C_\varphi}{\pi_M^*h^5}, \\ \mathbb{E}\left\{\mathcal{T}_h(\mathbf{Z};\mathbf{x})\right\} =& \frac{1}{h^3}\int\varphi\left(\frac{S-s_{\mathbf{x}}}{h}\right)f(s;\boldsymbol{\beta})ds=\frac{1}{h^2}\int_{\mathbb{R}}\varphi(u)f(s_{\mathbf{x}}+hu;\boldsymbol{\beta})du\leq\frac{M_\varphi C_f}{h^2}.\end{aligned}$$

Furthermore, we have

$$\|\mathcal{T}_h(\mathbf{Z};\mathbf{x})-\mathbb{E}\left\{\mathcal{T}_h(\mathbf{Z};\mathbf{x})\right\}\|_{\psi_2}\leq 3\,\|\mathcal{T}_h(\mathbf{Z};\mathbf{x})\|_{\psi_2}\leq 3C_s\pi_M^{*-1}h^{-3}M_\varphi, \quad \forall\mathbf{x}\in\mathcal{X}. \tag{S1.14}$$

For any $\mathbf{x}\in\mathcal{X}$ and any $t\geq 0$, with probability at least $1-3\exp\left(-t^2\right)$, we have

$$\left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathcal{T}_h\left(\mathbf{Z}_i;\mathbf{x}\right)\right\|_\infty \leq 7t\sqrt{\frac{C_sC_fM_\varphi C_\varphi}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^5}}+\frac{3C_sC_\alpha M_\varphi(\log|\mathcal{Z}_{-k(i)}|)^{1/2}t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^3}+\frac{M_\varphi C_f}{h^2},$$

Therefore by Lemma S1.1, for any $t\geq 0$, with probability at least $1-3\exp(-t^2)-q_M$,

$$\left|\widehat{R}_{M,2}(s)\right|\leq 4C_{w,2}M_{\mathbf{X}}^2b_M^2\left(7t\sqrt{\frac{C_sM_\varphi C_\varphi}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^5}}+\frac{3C_sC_\alpha M_\varphi(\log|\mathcal{Z}_{-k(i)}|)^{1/2}t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^*h^3}+\frac{M_\varphi C_f}{h^2}\right). \tag{S1.15}$$

Combined with (S1.7), for any $t \geq 0$, with probability at least $1 - 6\exp(-t^2) - 2q_M$,

$$\begin{aligned}\left|\widehat{R}_{M,w}(s)\right| \leq & b_M \left\{ C_{w,2}\left(C_{\eta,1,1} + M_{\mathbf{X}} C_{f,1}\right) + 7\sqrt{\frac{4C_s C_f C_{w,2}^2 M_{\mathbf{X}}^2 M_{\dot{K}} C_{\dot{K}} (t^2 + \log p)}{|\mathcal{Z}_{-k(i)}| \pi_M^* h^3}} \right. \\ & \left. + \frac{6C_s C_\alpha C_{w,2} M_{\mathbf{X}} M_{\dot{K}} (\log|\mathcal{Z}_{-k(i)}|)^{1/2} (t^2 + \log p)}{|\mathcal{Z}_{-k(i)}| \pi_M^* h^2} \right\} \\ & + 4C_{w,2} M_{\mathbf{X}}^2 b_M^2 \left(7t\sqrt{\frac{C_s C_f M_\varphi C_\varphi}{|\mathcal{Z}_{-k(i)}| \pi_M^* h^5}} + \frac{3C_s C_\alpha M_\varphi (\log|\mathcal{Z}_{-k(i)}|)^{1/2} t^2}{|\mathcal{Z}_{-k(i)}| \pi_M^* h^3} + \frac{M_\varphi C_f}{h^2} \right).\end{aligned} \tag{S1.16}$$

Define

$$\begin{aligned}\epsilon_{M,1}(t) \equiv & D_1 \frac{t}{\sqrt{M\pi_M^* h}} + D_2 \frac{t^2\sqrt{\log M}}{M\pi_M^* h} + D_3 h^2, \\ \epsilon_{M,2}(t) \equiv & D_4 \left(b_M + \frac{b_M^2}{h^2} \right) + D_5 \left\{ \frac{b_M t}{\sqrt{M\pi_M^* h^3}} \left(1 + \frac{b_M}{h}\right) + \frac{b_M \sqrt{\log p}}{\sqrt{M\pi_M^* h^3}} \right\} \\ & + D_6 \left\{ \frac{b_M t^2 \sqrt{\log M}}{M\pi_M^* h^2} \left(1 + \frac{b_M}{h}\right) + \frac{b_M \log p \sqrt{\log M}}{M\pi_M^* h^2} \right\}.\end{aligned}$$

Combined with (S1.4) and (S1.5), with probability at least $1 - 9\exp(-t^2) - 2q_M$, we have

$$|\hat{f}_{k,w}(s; \hat{\boldsymbol{\beta}}_k) - w(s) f(s; \boldsymbol{\beta})| \leq \epsilon_{M,1}(t) + \epsilon_{M,2}(t). \tag{S1.17}$$

With assumption $b_M/h = o(1)$, $(b_M\sqrt{\log p})/h = o(1)$ and $\log M \log p = O(M\pi_M^* h)$, we have

$$\begin{aligned}& \sup_{\mathbf{x}\in\mathcal{X}} |\hat{f}_{k,w}(s; \hat{\boldsymbol{\beta}}_k) - w(s) f(s; \boldsymbol{\beta})| \\ = & O_p \left\{ \frac{1}{\sqrt{M\pi_M^* h}} + \frac{\sqrt{\log M}}{M\pi_M^* h} + h^2 + b_M + \frac{b_M^2}{h^2} + \frac{\sqrt{\log M}\log p}{M\pi_M^* h} \right\}.\end{aligned} \tag{S1.18}$$

Next we consider $|\hat{f}_{k,\hat{w}}(s; \hat{\boldsymbol{\beta}}_k) - \hat{f}_{k,w}(s; \hat{\boldsymbol{\beta}}_k)|$. We have

$$\hat{f}_{k,\hat{w}}(s; \hat{\boldsymbol{\beta}}_k) - \hat{f}_{k,w}(s; \hat{\boldsymbol{\beta}}_k) = \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j\in\mathcal{Z}_{-k(i)}} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{x}'\hat{\boldsymbol{\beta}}_k).$$

Denote $L_1 = \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j\in\mathcal{Z}_{-k(i)}} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{x}'\hat{\boldsymbol{\beta}}_k)$. By the first-order Taylor expansion, we have

$$\begin{aligned}L_1 = & \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j\in\mathcal{Z}_{-k(i)}} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{x}'\hat{\boldsymbol{\beta}}_k) \\ = & \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j\in\mathcal{Z}_{-k(i)}} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} \frac{1}{h} K\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h} \right) \\ & + \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j\in\mathcal{Z}_{-k(i)}} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} \frac{1}{h} \dot{K}\left(\frac{S_j^* - s_{\mathbf{x}}^*}{h} \right) \frac{\mathbf{X}_j - \mathbf{x}}{h} (\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}) \\ := & L_{11} + L_{12},\end{aligned}$$

where $\left\{S_j^*\right\}_{j\in\mathcal{Z}_{-k(i)}}$ and $s_{\mathbf{x}}^*$ are some "immediate" points satisfying for any $j \in \mathcal{Z}_{-k(i)}$, $\left|\left(S_j^* - s_{\mathbf{x}}^*\right) - \left(S_j - s_{\mathbf{x}}\right)\right| \leq$

$\left|\left(\hat{S}_j - \hat{s}_{\mathbf{x}}\right) - (S_j - s_{\mathbf{x}})\right| \equiv \left|(\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta})' (\mathbf{X}_j - \mathbf{x})\right|$. For term L_{11}, we have

$$
\begin{aligned}
L_{11} =& \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \frac{1}{h} \left\{ \frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)} \right\} K\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h} \right) \\
=& \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \frac{1}{h} \left\{ \frac{\pi_M^w(\mathbf{X}_j)}{\hat{\pi}_M^w(\mathbf{X}_j)} - 1 \right\} \frac{D_j}{\pi_M^w(\mathbf{X}_j)} K\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h} \right) \\
=& \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{F}_{2,j}' (\hat{\boldsymbol{\xi}}^w - \boldsymbol{\xi}^w)\{1 + o_p(1)\},
\end{aligned}
$$

where $\mathbf{F}_{2,j} := \frac{1}{h}\{\pi_M^w(\mathbf{X}_j) - 1\} \frac{D_j}{\pi_M^w(\mathbf{X}_j)} K\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h}\right) \bar{\Omega}^w(\mathbf{X}_j)$. By Hölder inequality, we have

$$
\frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{F}_{2,j}' (\hat{\boldsymbol{\xi}}^w - \boldsymbol{\xi}^w) \leq \left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{F}_{2,j} \right\|_\infty \times \|\hat{\boldsymbol{\xi}}^w - \boldsymbol{\xi}^w\|_1.
$$

After simple calculations, we have

$$
\begin{aligned}
&\max_{1 \leq d \leq pL+1} \|\mathbf{F}_{2,j,[d]}\|_{\psi_2} \leq \frac{C_{w,2} M_{\mathbf{X},L} M_K}{\pi_M^* h}, \\
&\max_{1 \leq d \leq pL+1} \|\mathbf{F}_{2,j,[d]} - \mathbb{E}\mathbf{F}_{2,j,[d]}\|_{\psi_2} \leq \frac{3C_{w,2} M_{\mathbf{X},L} M_K}{\pi_M^* h}, \\
&\max_{1 \leq d \leq pL+1} \mathbb{E}\mathbf{F}_{2,j,[d]}^2 \leq \frac{C_{w,2}^2 M_K M_{\mathbf{X},L}^2}{\pi_M^* h}, \\
&\|\mathbb{E}\mathbf{F}_{2,j}\|_\infty \leq C_{w,2} M_{\mathbf{X},L} C_f.
\end{aligned}
$$

By Lemma S1.1, for any $t \geq 0$, with probability at least $1 - 3\exp(-t^2)$,

$$
\begin{aligned}
\left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{F}_{2,j} \right\|_\infty =& \left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} (\mathbf{F}_{2,j} - \mathbb{E}\mathbf{F}_{2,j} + \mathbb{E}\mathbf{F}_{2,j}) \right\|_\infty \\
\leq& \left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} (\mathbf{F}_{2,j} - \mathbb{E}\mathbf{F}_{2,j}) \right\|_\infty + \|\mathbb{E}\mathbf{F}_{2,j}\|_\infty \\
\leq& 7\sqrt{\frac{C_{w,2}^2 M_K M_{\mathbf{X},L}^2 (t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}| \pi_M^* h}} + C_{w,2} M_{\mathbf{X},L} C_f \\
&+ \frac{3C_\alpha C_{w,2} M_{\mathbf{X},L} M_K (\log|\mathcal{Z}_{-k(i)}|)^{1/2} (t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}| \pi_M^* h}.
\end{aligned}
$$

Thus for any $t \geq 0$, with probability at least $1 - 3\exp(-t^2) - p_M$,

$$
\begin{aligned}
L_{11} \leq& r_M \left(7\sqrt{\frac{C_{w,2}^2 M_K M_{\mathbf{X},L}^2 (t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}| \pi_M^* h}} \right. \\
&\left. + \frac{3C_\alpha C_{w,2} M_{\mathbf{X},L} M_K (\log|\mathcal{Z}_{-k(i)}|)^{1/2} (t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}| \pi_M^* h} + C_{w,2} M_{\mathbf{X},L} C_f \right).
\end{aligned}
\tag{S1.19}
$$

We have

$$
L_{11} = O_p\left(r_M \left\{ \sqrt{\frac{\log p}{M\pi_M^* h}} + \frac{\sqrt{\log M}\log p}{M\pi_M^* h} + 1 \right\} \right). \tag{S1.20}
$$

For term L_{12},

$$\begin{aligned}
L_{12} =& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left\{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\frac{1}{h}\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)\frac{\mathbf{X}_j-\mathbf{x}}{h}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})\\
=& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left\{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\frac{1}{h}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta}-\mathbf{x}'\boldsymbol{\beta}}{h}\right)\frac{\mathbf{X}_j-\mathbf{x}}{h}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})\\
&+\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left\{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\frac{1}{h}\left\{\dot{K}\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)-\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta}-\mathbf{x}'\boldsymbol{\beta}}{h}\right)\right\}\\
&\times\frac{\mathbf{X}_j-\mathbf{x}}{h}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})\\
:=& L_{12,1}+L_{12,2}.
\end{aligned}$$

Next we analyze $L_{12,1}$,

$$\begin{aligned}
L_{12,1} =& \frac{1}{|\mathcal{Z}_{-k(i)}|h}\sum_{j\in\mathcal{Z}_{-k(i)}}\left\{\frac{\pi_M^w(\mathbf{X}_j)}{\hat{\pi}_M^w(\mathbf{X}_j)}-1\right\}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta}-\mathbf{x}'\boldsymbol{\beta}}{h}\right)\frac{\mathbf{X}_j-\mathbf{x}}{h}(\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta})\\
\le& \left\|\frac{1}{|\mathcal{Z}_{-k(i)}|h}\sum_{j\in\mathcal{Z}_{-k(i)}}\left\{\frac{\pi_M^w(\mathbf{X}_j)}{\hat{\pi}_M^w(\mathbf{X}_j)}-1\right\}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta}-\mathbf{x}'\boldsymbol{\beta}}{h}\right)\frac{\mathbf{X}_j-\mathbf{x}}{h}\right\|_\infty\|\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1\\
\le& \left\|\frac{1}{|\mathcal{Z}_{-k(i)}|h}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\left\{\frac{\pi_M^w(\mathbf{X}_j)}{\hat{\pi}_M^w(\mathbf{X}_j)}-1\right\}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta}-\mathbf{x}'\boldsymbol{\beta}}{h}\right)\right|\frac{2M_{\mathbf{X}}}{h}\right\|_\infty\|\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1\\
=& \left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\mathbf{L}_{12,1,j}'(\hat{\boldsymbol{\xi}}^w-\boldsymbol{\xi}^w)\right|2M_{\mathbf{X}}\{1+o_p(1)\}\right\|_\infty\|\hat{\boldsymbol{\beta}}_k-\boldsymbol{\beta}\|_1.
\end{aligned}$$

On event $\mathcal{B}_M$, we have $L_{12,1}\le\frac{2M_{\mathbf{X}}b_M}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\mathbf{L}_{12,1,j}'(\hat{\boldsymbol{\xi}}^w-\boldsymbol{\xi}^w)\right|$.

$$\begin{aligned}
\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\left|\mathbf{L}_{12,1,j}'(\hat{\boldsymbol{\xi}}^w-\boldsymbol{\xi}^w)\right| \le& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\|\mathbf{L}_{12,1,j}\|_\infty\|(\hat{\boldsymbol{\xi}}^w-\boldsymbol{\xi}^w)\|_1\\
\le& \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}|L_{12,1,j}^1|M_{\mathbf{X},L}\|(\hat{\boldsymbol{\xi}}^w-\boldsymbol{\xi}^w)\|_1,
\end{aligned}$$

where $L^1_{12,1,j} = \{\pi^w_M(\mathbf{X}_j) - 1\} \frac{D_j}{\pi^w_M(\mathbf{X}_j)} \frac{1}{h^2} \dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h}\right)$. We have

$$
\begin{aligned}
|L^1_{12,1,j}| \le & \frac{C_{w,2} M_{\dot{K}}}{\pi^*_M h^2}, \\
||L^1_{12,1,j}| - \mathbb{E}(|L^1_{12,1,j}|)| \le & \frac{3 C_{w,2} M_{\dot{K}}}{\pi^*_M h^2}, \\
\mathbb{E}(L^1_{12,1,j})^2 = & \mathbb{E}\left[\{\pi^w_M(\mathbf{X}_j) - 1\} \frac{D_j}{\pi^w_M(\mathbf{X}_j)} \frac{1}{h^2} \dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h}\right)\right]^2 \\
\le & \frac{C^2_{w,2}}{h^4} \mathbb{E}\left\{\frac{1}{\pi_M(\mathbf{X}_j)} \dot{K}^2\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta}}{h}\right)\right\} \\
= & \frac{C^2_{w,2}}{h^4} \int \mathbb{E}\left\{\frac{1}{\pi_M(\mathbf{X})}\Big| S = s\right\} \dot{K}^2\left(\frac{s - s_{\mathbf{x}}}{h}\right) f(s; \boldsymbol{\beta}) ds \\
= & \frac{C^2_{w,2}}{h^3} \int \mathbb{E}\left\{\frac{1}{\pi_M(\mathbf{X})}\Big| S = s_{\mathbf{x}} + th\right\} \dot{K}^2(t) f(s_{\mathbf{x}} + th; \boldsymbol{\beta}) dt \\
\le & \frac{C^2_{w,2} C_s C_f M_{\dot{K}} C_{\dot{K}}}{\pi^*_M h^3}, \\
\mathbb{E}(|L^1_{12,1,j}|) \le & \frac{C_{w,2} C_{\dot{K}}}{h^2}.
\end{aligned}
$$

With probability at least $1 - 3\exp(-t^2)$,

$$
\begin{aligned}
\frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} |L^1_{12,1,j}| \le & 7t\sqrt{\frac{C^2_{w,2} C_s C_f M_{\dot{K}} C_{\dot{K}}}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^3}} + \frac{C_\alpha C_{w,2} M_{\dot{K}} (\log|\mathcal{Z}_{-k(i)}|)^{1/2} t^2}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^2} \\
& + \frac{C_{w,2} C_{\dot{K}}}{h}.
\end{aligned}
$$

Thus with probability at least $1 - 3\exp(-t^2) - p_M - q_M$,

$$
L_{12,1} \le 2 r_M b_M M_{\mathbf{X}} M_{\mathbf{X},L} \left(7t\sqrt{\frac{C^2_{w,2} C_s C_f M_{\dot{K}} C_{\dot{K}}}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^3}} + \frac{C_\alpha C_{w,2} M_{\dot{K}} (\log|\mathcal{Z}_{-k(i)}|)^{1/2} t^2}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^2} + \frac{C_{w,2} C_{\dot{K}}}{h^2}\right). \quad \text{(S1.21)}
$$

Similarly, with probability at least $1 - 3\exp(-t^2) - p_M - 2q_M$,

$$
L_{12,2} \le 4 r_M b^2_M M^2_{\mathbf{X}} M_{\mathbf{X},L} \left(7t\sqrt{\frac{C^2_{w,2} C_s C_f M_\varphi C_\varphi}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^5}} + \frac{3 C_s C_{w,2} C_\varphi C_\alpha \sqrt{\log M} t^2}{|\mathcal{Z}_{-k(i)}| \pi^*_M h^3} + \frac{C_{w,2} C_f C_\varphi}{h^3}\right). \quad \text{(S1.22)}
$$

Combine all the above results, with probability at least $1 - 9\exp(-t^2) - 3p_M - 3q_M$, we have

$$
\begin{aligned}
&|\hat{f}_{k,\hat{w}}(s;\hat{\boldsymbol{\beta}}_k) - \hat{f}_{k,w}(s;\hat{\boldsymbol{\beta}}_k)| \\
\leq & r_M \left(7\sqrt{\frac{C_{w,2}^2 M_K M_{\mathbf{X},L}^2 (t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}|\pi_M^* h}} \right. \\
& \left. + \frac{C_\alpha C_{w,2} M_{\mathbf{X},L} M_K (\log|\mathcal{Z}_{-k(i)}|)^{1/2}(t^2 + \log(pL+1))}{|\mathcal{Z}_{-k(i)}|\pi_M^* h} + C_{w,2} M_{\mathbf{X},L} C_f \right) \\
& + 2 r_M b_M M_{\mathbf{X}} M_{\mathbf{X},L} \left(7t\sqrt{\frac{C_{w,2}^2 C_s C_f M_{\dot{K}} C_{\dot{K}}}{|\mathcal{Z}_{-k(i)}|\pi_M^* h^3}} + \frac{C_\alpha C_{w,2} M_{\dot{K}} (\log|\mathcal{Z}_{-k(i)}|)^{1/2} t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^* h^2} + \frac{C_{w,2} C_{\dot{K}}}{h^2} \right) \\
& + 4 r_M b_M^2 M_{\mathbf{X}}^2 M_{\mathbf{X},L} \left(\frac{C_{w,2} C_f C_\varphi}{h^3} + 7t\sqrt{\frac{C_{w,2}^2 C_s C_f M_\varphi C_\varphi}{|\mathcal{Z}_{-k(i)}|\pi_M^* h^5}} + \frac{3 C_s C_{w,2} C_\varphi C_\alpha \sqrt{\log M} t^2}{|\mathcal{Z}_{-k(i)}|\pi_M^* h^3} \right) \equiv \epsilon_{M,3}(t).
\end{aligned} \tag{S1.23}
$$

Combined with (S1.18), we have

$$
|\hat{f}_{k,\hat{w}}(s;\hat{\boldsymbol{\beta}}_k) - w(s) f(s;\boldsymbol{\beta})| \leq \epsilon_M(t). \tag{S1.24}
$$

∎

S1.3 Proof of Lemma 3.2

Proof. Denote

$$
\begin{aligned}
\Delta_{\mathcal{S}}^* &= \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{D_j}{\hat{\pi}_M(\mathbf{X}_j)} Y_j K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{D_j}{\hat{\pi}_M(\mathbf{X}_j)} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)} - \frac{1}{M}\sum_{i=1}^M g(S_i) \\
&= \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{D_j}{\hat{\pi}_M(\mathbf{X}_j)} \{Y_j - g(S_i)\} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{D_j}{\hat{\pi}_M(\mathbf{X}_j)} K_H(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}
\end{aligned}
$$

Since $\hat{f}_{k,\hat{w}}(s;\hat{\boldsymbol{\beta}}_k) = w(s)f(s;\boldsymbol{\beta}) + o_p(1)$, we can focus on

$$
\begin{aligned}
\Delta_{\mathcal{S},1}^* =& \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)}\{Y_j - g(S_i)\} K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{w(S_i) f(S_i;\boldsymbol{\beta})} \\
:=& \Delta_{\mathcal{S},1,1}^* + \Delta_{\mathcal{S},1,2}^* + \Delta_{\mathcal{S},1,3}^* + \Delta_{\mathcal{S},1,4}^*,
\end{aligned} \tag{S1.25}
$$

where

$$
\begin{aligned}
\Delta_{\mathcal{S},1,1}^* =& \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\{Y_j - g(S_j)\} K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{w(S_i) f(S_i;\boldsymbol{\beta})}, \\
\Delta_{\mathcal{S},1,2}^* =& \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{\left\{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\{Y_j - g(S_j)\} K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{w(S_i) f(S_i;\boldsymbol{\beta})}, \\
\Delta_{\mathcal{S},1,3}^* =& \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\{g(S_j) - g(S_i)\} K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{w(S_i) f(S_i;\boldsymbol{\beta})}, \\
\Delta_{\mathcal{S},1,4}^* =& \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}} \frac{\left\{\frac{D_j}{\hat{\pi}_M^w(\mathbf{X}_j)} - \frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\{g(S_j) - g(S_i)\} K_h(\mathbf{X}_j'\hat{\boldsymbol{\beta}}_k - \mathbf{X}_i'\hat{\boldsymbol{\beta}}_k)}{w(S_i) f(S_i;\boldsymbol{\beta})}.
\end{aligned}
$$

Now we focus on $\Delta^*_{\mathcal{S},1,1}$,

$$\begin{aligned}\Delta^*_{\mathcal{S},1,1} =&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\hat{\boldsymbol\beta}_k-\mathbf{X}_i'\hat{\boldsymbol\beta}_k)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ :=&\Delta^*_{\mathcal{S},1,1,1}+\Delta^*_{\mathcal{S},1,1,2},\end{aligned}$$

where

$$\begin{aligned}\Delta^*_{\mathcal{S},1,1,1} =&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)},\\ \Delta^*_{\mathcal{S},1,1,2} =&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}\frac{1}{h}\dot{K}\left(\frac{S_j^*-s^*_{\mathbf{x}}}{h}\right)}{w(S_i)f(S_i;\boldsymbol\beta)}\left(\frac{\mathbf{X}_j-\mathbf{X}_i}{h}\right)'\\ &\times(\hat{\boldsymbol\beta}_k-\boldsymbol\beta).\end{aligned}$$

Firstly we analyse $\Delta^*_{\mathcal{S},1,1,1}$. Denote $\Delta^*_{\mathcal{S},1,1,1}(\mathbf{Z}_j,\mathbf{X}_i)=\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)}$. We have

$$\begin{aligned}\mathbb{E}\{\Delta^*_{\mathcal{S},1,1,1}(\mathbf{Z}_j,\mathbf{X}_i)\mid\mathbf{Z}_j\} &= \frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}\mathbb{E}\left\{\frac{K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)}\right\}\\ &= \frac{D_j}{\pi^w_M(\mathbf{X}_j)w(S_j)}\{Y_j-g(S_j)\},\\ \mathbb{E}\{\Delta^*_{\mathcal{S},1,1,1}(\mathbf{Z}_j,\mathbf{X}_i)\mid\mathbf{X}_i\} &= \frac{1}{w(S_i)f(S_i;\boldsymbol\beta)}\mathbb{E}\left\{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)\right\}\\ &= \frac{w_y(S_i)}{w(S_i)}-g(S_i),\end{aligned}$$

where $w_y(S_i)=\mathbb{E}\left\{\frac{\pi_M(\mathbf{X}_i)}{\pi^w_M(\mathbf{X}_i)}Y_i|S_i\right\}$. By Hoeffding decomposition, we have

$$\begin{aligned}\Delta^*_{\mathcal{S},1,1,1} &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ &= \frac{1}{M}\sum_{i=1}^{M}\frac{D_i\{Y_i-g(S_i)\}}{\pi^w_M(\mathbf{X}_i)w(S_i)}+\frac{1}{M}\sum_{i=1}^{M}\frac{w_y(S_i)}{w(S_i)}-g(S_i)+O_p(d_M),\end{aligned}\tag{S1.26}$$

where $d_M=(M^2\pi^*_M h)^{-1/2}$. Next we analyse $\Delta^*_{\mathcal{S},1,1,2}$. Using Taylor expansion, we have

$$\Delta^*_{\mathcal{S},1,1,2} =\Delta^*_{\mathcal{S},1,1,2,1}+\Delta^*_{\mathcal{S},1,1,2,2},$$

where

$$\begin{aligned}\Delta^*_{\mathcal{S},1,1,2,1} =&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}\frac{1}{h}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta}{h}\right)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ &\times\left(\frac{\mathbf{X}_j-\mathbf{X}_i}{h}\right)'(\hat{\boldsymbol\beta}_k-\boldsymbol\beta),\\ \Delta^*_{\mathcal{S},1,1,2,2} =&\frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j-g(S_j)\}\frac{1}{h}\left\{\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta}{h}\right)-\dot{K}\left(\frac{S_j^*-s^*_{\mathbf{x}}}{h}\right)\right\}}{w(S_i)f(S_i;\boldsymbol\beta)}\\ &\times\left(\frac{\mathbf{X}_j-\mathbf{X}_i}{h}\right)'(\hat{\boldsymbol\beta}_k-\boldsymbol\beta).\end{aligned}$$

We can focus on $\Delta^*_{\mathcal{S},1,1,2,1}$ since $\Delta^*_{\mathcal{S},1,1,2,2} = o_p(\Delta^*_{\mathcal{S},1,1,2,1})$. By Hölder inequality, we have

$$\begin{aligned}
\Delta^*_{\mathcal{S},1,1,2,1} \leq & \left\| \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \frac{\frac{D_j}{\pi^w_M(\mathbf{X}_j)}\{Y_j - g(S_j)\}\frac{1}{h}\dot{K}\left(\frac{\mathbf{X}_j'\boldsymbol{\beta} - \mathbf{X}_i'\boldsymbol{\beta}}{h}\right)}{w(S_i)f(S_i;\boldsymbol{\beta})} \left(\frac{\mathbf{X}_j - \mathbf{X}_i}{h}\right)' \right\|_\infty \\
& \times \max_k \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\|_1 \qquad \text{(S1.27)} \\
:= & \frac{1}{K}\sum_{k=1}^K \left\| \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) \right\|_\infty \max_k \|\hat{\boldsymbol{\beta}}_k - \boldsymbol{\beta}\|_1
\end{aligned}$$

Next we analyse $\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i)$. Denote $\mathbf{q}_2(S_j) = \mathbb{E}(\mathbf{X}_j|S_j)$, $\mathbf{w}_3(S_j) = \mathbb{E}\left\{\frac{\pi_M(\mathbf{X}_j)}{\pi^w_M(\mathbf{X}_j)} Y_j \mathbf{X}_j \big| S_j\right\}$, $\mathbf{w}_4(S_j) = \mathbb{E}\left\{\frac{\pi_M(\mathbf{X}_j)}{\pi^w_M(\mathbf{X}_j)} \mathbf{X}_j \big| S_j\right\}$, $\mathbf{w}_5(S_j) = \mathbf{w}_3(S_j) - \mathbf{w}_4(S_j)g(S_j)$, $w_6(S_j) = w_y(S_j) - w(S_j)g(S_j)$. After simple calculations, we have

$$\begin{aligned}
\mathbb{E}\{\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i)\} =& \mathbb{E}\left(\frac{\nabla_s[\{w_6(S_i)\mathbf{X}_i - \mathbf{w}_5(S_i)\}f(S_i)]}{w(S_i)f(S_i)}\right) := \boldsymbol{\Delta}_{\boldsymbol{\beta}}, \\
\mathbb{E}\{\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) \mid \mathbf{Z}_j\} =& \frac{D_j}{\pi^*_M(\mathbf{X}_j)w(S_j)}\{Y_j - g(S_j)\}\left\{\frac{\dot{w}(S_j)}{w(S_j)}\mathbf{q}_2(S_j) - \dot{\mathbf{q}}_2(S_j) - \mathbf{X}_j\frac{\dot{w}(S_j)}{w(S_j)}\right\}, \\
\mathbb{E}\{\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) \mid \mathbf{X}_i\} =& \frac{\{w_6(S_i)\mathbf{X}_i - \mathbf{w}_5(S_i)\}\dot{f}(S_i) + \{\dot{w}_6(S_i)\mathbf{X}_i - \dot{\mathbf{w}}_5(S_i)\}f(S_i)}{w(S_i)f(S_i)}.
\end{aligned}$$

Let $\mathbf{h}^*(\mathbf{Z}_j, \mathbf{X}_i) = \boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) - \boldsymbol{\Delta}_{\boldsymbol{\beta}}$, $\mathbf{h}_1(\mathbf{Z}_j) = \mathbb{E}\{\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i)|\mathbf{Z}_j\} - \boldsymbol{\Delta}_{\boldsymbol{\beta}}$, $\mathbf{h}_2(\mathbf{X}_i) = \mathbb{E}\{\boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i)|\mathbf{X}_i\} - \boldsymbol{\Delta}_{\boldsymbol{\beta}}$. We have

$$\begin{aligned}
& \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) \\
=& \boldsymbol{\Delta}_{\boldsymbol{\beta}} + \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{h}_1(\mathbf{Z}_j) + \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \mathbf{h}_2(\mathbf{X}_i) + \mathbf{R}_M,
\end{aligned}$$

where $\mathbf{R}_M$ is a reminder term satisfying $\mathbb{E}(\mathbf{R}_M) = \mathbf{0}$ and $\|\mathbf{R}_M\|_\infty = O_p\left(\sqrt{\frac{\log p}{|\mathcal{L}||\mathcal{U}|}}\right)$.

By triangle inequality, we have

$$\begin{aligned}
& \left\| \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \boldsymbol{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j, \mathbf{X}_i) - \boldsymbol{\Delta}_{\boldsymbol{\beta}} \right\|_\infty \\
\leq & \left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{h}_1(\mathbf{Z}_j) \right\|_\infty + \left\| \frac{1}{|\mathcal{Z}_{k(i)}|} \sum_{i \in \mathcal{Z}_{k(i)}} \mathbf{h}_2(\mathbf{X}_i) \right\|_\infty + \|\mathbf{R}_M\|_\infty .
\end{aligned}$$

For $\left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{h}_1(\mathbf{Z}_j) \right\|_\infty$,

$$\max_{1 \leq d \leq p} \|\mathbf{h}_1(\mathbf{Z})_{[d]}\|_{\psi_2} \leq C\pi^{*-1}_M, \max_{1 \leq d \leq p} \|\mathbb{E}\mathbf{h}_1(\mathbf{Z})^2_{[d]}\| \leq C\pi^{*-1}_M,$$

By Lemma S1.1, for any $t \geq 0$, with probability at least $1 - 3\exp(-t^2)$, we have

$$\left\| \frac{1}{|\mathcal{Z}_{-k(i)}|} \sum_{j \in \mathcal{Z}_{-k(i)}} \mathbf{h}_1(\mathbf{Z}_j) \right\|_\infty \leq 7C\sqrt{\frac{(t^2 + \log p)}{|\mathcal{Z}_{-k(i)}|\pi^*_M}} + \frac{C_\alpha K_n (\log|\mathcal{Z}_{-k(i)}|)^{1/2}(t^2 + \log p)}{|\mathcal{Z}_{-k(i)}|\pi^*_M}. \qquad \text{(S1.28)}$$

Similarly, for term $\left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathbf{h}_2(\mathbf{X}_i)\right\|_\infty$, for any $t\geq 0$, with probability at least $1-3\exp(-t^2)$,

$$\left\|\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathbf{h}_2(\mathbf{X}_i)\right\|_\infty \leq 7C\sqrt{\frac{(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|}}+\frac{C_\alpha K_n(\log|\mathcal{Z}_{-k(i)}|)^{1/2}(t^2+\log p)}{|\mathcal{Z}_{-k(i)}|\pi_M^*}. \tag{S1.29}$$

Combined with (S1.28) and (S1.29), we have

$$\begin{aligned}&\left\|\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\mathbf{\Delta}^*_{\mathcal{S},1,1,2,1}(\mathbf{Z}_j,\mathbf{X}_i)-\mathbf{\Delta}_{\boldsymbol\beta}\right\|_\infty\\ =&O_p\left(\sqrt{\frac{\log p}{|\mathcal{Z}_{k(i)}|\pi_M^*}}+\frac{(\log|\mathcal{Z}_{k(i)}|)^{1/2}\log p}{|\mathcal{Z}_{k(i)}|\pi_M^*}+\sqrt{\frac{\log p}{|\mathcal{Z}_{-k(i)}|}}+\frac{(\log|\mathcal{Z}_{-k(i)}|)^{1/2}\log p}{|\mathcal{Z}_{-k(i)}|}+\sqrt{\frac{\log p}{|\mathcal{Z}_{k(i)}||\mathcal{Z}_{-k(i)}|}}\right)\end{aligned} \tag{S1.30}$$

Combined with (S1.26) and (S1.27),

$$\Delta^*_{\mathcal{S},1,1}=\frac{1}{M}\sum_{i=1}^M\frac{D_i\{Y_i-g(S_i)\}}{\pi_M^w(\mathbf{X}_i)w(S_i)}+\frac{1}{M}\sum_{i=1}^M\frac{w_y(S_i)}{w(S_i)}-g(S_i)+O_p\left(d_M+b_M\|\mathbf{\Delta}_{\boldsymbol\beta}\|_\infty+b_M\sqrt{\frac{\log p}{|\mathcal{Z}_{k(i)}|\pi_M^*}}\right). \tag{S1.31}$$

For $\Delta^*_{\mathcal{S},1,2}$, we have

$$\Delta^*_{\mathcal{S},1,2}=\Delta^*_{\mathcal{S},1,2,1}+\Delta^*_{\mathcal{S},1,2,2},$$

where

$$\begin{aligned}\Delta^*_{\mathcal{S},1,2,1}=&\frac{1}{K}\sum_{k=1}^K\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\left\{\frac{D_j}{\hat\pi_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)},\\ \Delta^*_{\mathcal{S},1,2,2}=&\frac{1}{K}\sum_{k=1}^K\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\left\{\frac{D_j}{\hat\pi_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\{Y_j-g(S_j)\}\frac{1}{h}\dot K\left(\frac{S_j^*-s_{\mathbf{x}}^*}{h}\right)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ &\times\left(\frac{\mathbf{X}_j-\mathbf{X}_i}{h}\right)'(\hat{\boldsymbol\beta}_k-\boldsymbol\beta).\end{aligned}$$

For $\Delta^*_{\mathcal{S},1,2,1}$, we have

$$\begin{aligned}\Delta^*_{\mathcal{S},1,2,1}=&\frac{1}{K}\sum_{k=1}^K\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\left\{\frac{D_j}{\hat\pi_M^w(\mathbf{X}_j)}-\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\right\}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ =&\frac{1}{K}\sum_{k=1}^K\frac{1}{|\mathcal{Z}_{k(i)}|}\sum_{i\in\mathcal{Z}_{k(i)}}\frac{1}{|\mathcal{Z}_{-k(i)}|}\sum_{j\in\mathcal{Z}_{-k(i)}}\frac{\{\pi_M^w(\mathbf{X}_j)-1\}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)}{w(S_i)f(S_i;\boldsymbol\beta)}\\ &\times\bar\Omega^w(\mathbf{X}_j)'(\hat{\boldsymbol\xi}^w-\boldsymbol\xi^w)\{1+o_p(1)\}.\end{aligned}$$

Similarly, define $\mathbf{\Delta}_\xi=\mathbb{E}\left\{\frac{\{\pi_M^w(\mathbf{X}_j)-1\}\frac{D_j}{\pi_M^w(\mathbf{X}_j)}\{Y_j-g(S_j)\}K_h(\mathbf{X}_j'\boldsymbol\beta-\mathbf{X}_i'\boldsymbol\beta)\bar\Omega^w(\mathbf{X}_j)}{w(S_i)f(S_i;\boldsymbol\beta)}\right\}$,
$\Delta^*_{\mathcal{S},1,2,1}=O_p\left(r_M\|\mathbf{\Delta}_\xi\|_\infty+r_M\sqrt{\frac{\log p}{|\mathcal{Z}_{k(i)}|\pi_M^*}}\right)$, $\Delta^*_{\mathcal{S},1,2,2}=o_p(\Delta^*_{\mathcal{S},1,2,1})$.

Similarly, we have $\Delta^*_{\mathcal{S},1,3} = o_p\left(\frac{1}{\sqrt{M\pi^*_M}}\right)$, $\Delta^*_{\mathcal{S},1,4} = o_p\left(\frac{1}{\sqrt{M\pi^*_M}}\right)$. Thus

$$\begin{aligned}\hat{\theta}^{\mathrm{sp,w}} - \theta_0 =& \frac{1}{M}\sum_{i=1}^{M}\frac{D_i\{Y_i - g(S_i)\}}{\pi^w_M(\mathbf{X}_i)w(S_i)} + \frac{1}{M}\sum_{i=1}^{M}\frac{w_y(S_i)}{w(S_i)} - \theta_0 \\ &+ O_p\left\{\|\boldsymbol{\Delta}_{\boldsymbol{\beta}}\|_\infty b_M + \|\boldsymbol{\Delta}_\xi\|_\infty r_M + d_M + (b_M + r_M)\sqrt{\frac{\log p}{|\mathcal{Z}_{k(i)}|\pi^*_M}}\right\}.\end{aligned} \tag{S1.32}$$

∎

S1.4 Proof of Theorem 3.1

Proof. Denote $\Psi(\mathbf{Z}) = \frac{D\{Y-g(S)\}}{\pi^w_M(\mathbf{X})w(S)} + g(S) - \theta_0$, $V_M = \mathrm{Var}\{\Psi(\mathbf{Z})\}$. We have

$$\hat{\theta}^{\mathrm{sp,w}} - \theta_0 = \frac{1}{M}\sum_{i=1}^{M}\Psi(\mathbf{Z}_i) + o_p\left(\frac{1}{\sqrt{M\pi^*_M}}\right).$$

It is obvious that $\mathbb{E}\{\Psi(\mathbf{Z})\} = 0$. If for any $\epsilon > 0$, $\pi^*_M\mathbb{E}[\Psi^2(\mathbf{Z})1\{|\Psi(\mathbf{Z})| > \epsilon\sqrt{M/\pi^*_M}\}] \to 0$ as $M \to \infty$. By the Proposition 2.27(Lindeberg-Feller theorem) in [2], we have

$$\frac{\frac{1}{M}\sum_{i=1}^{M}\Psi(\mathbf{Z}_i)}{(\frac{\mathrm{Var}\{\Psi(\mathbf{Z})\}}{M})^{1/2}} \xrightarrow{d} \mathcal{N}(0,1), \quad \frac{1}{M}\sum_{i=1}^{M}\Psi(\mathbf{Z}_i) = O_p\{(M\pi^*_M)^{-1/2}\}$$

as $M\pi^*_M \to \infty$. Thus

$$(M\pi^*_M)^{1/2}\left(\hat{\theta}^{\mathrm{sp,w}} - \theta_0\right) = O_p(1), \quad M^{1/2}V_M^{-1/2}\left(\hat{\theta}^{\mathrm{sp,w}} - \theta_0\right) \xrightarrow{d} \mathcal{N}(0,1). \tag{S1.33}$$

∎

S2 Additional simulation results

Table S1: The results of Simulation (Y1) on 1000 simulation runs under $p = 10, M = 1000, 2000, 5000$.

	Bias	SE	SD	CP	Eff(M)	Eff(S)	Bias	SE	SD	CP	Eff(M)	Eff(S)
	MCAR(K1)						MAR(K2)					
	$M = 1000$											
$\overline{Y}$	0.001	0.033	0.033	0.946	-	-	0.016	0.034	0.033	0.927	-	-
$\overline{Y}_{\mathrm{IPW}}$	0.001	0.033	0.033	0.946	0.000	0.000	0.000	0.049	0.045	0.926	-0.725	-0.448
$\overline{Y}_{\mathrm{cc}}$	-0.000	0.010	0.010	0.946	0.900	0.684	-0.000	0.010	0.010	0.950	0.920	0.689
$\hat{\theta}^{\mathrm{sp,T}}_{\mathcal{L}}$	-0.000	0.015	0.015	0.944	0.796	0.548	0.001	0.018	0.020	0.948	0.756	0.456
$\hat{\theta}^{\mathrm{sp,T}}_{\mathrm{IPW}}$	-0.000	0.015	0.015	0.944	0.796	0.548	0.001	0.019	0.020	0.934	0.738	0.436
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathcal{L}}$	-0.000	0.015	0.016	0.956	0.795	0.548	0.001	0.018	0.021	0.959	0.755	0.454
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathrm{IPW}}$	-0.000	0.015	0.016	0.956	0.794	0.546	0.001	0.018	0.022	0.954	0.738	0.436
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathcal{L}}$	-0.000	0.015	0.015	0.946	0.796	0.548	0.001	0.018	0.015	0.903	0.767	0.468
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathrm{IPW}}$	-0.000	0.015	0.015	0.946	0.796	0.548	0.001	0.018	0.015	0.904	0.767	0.468
	$M = 2000$											
$\overline{Y}$	0.000	0.024	0.023	0.944	-	-	0.017	0.024	0.024	0.876	-	-
$\overline{Y}_{\mathrm{IPW}}$	0.000	0.024	0.023	0.944	0.000	0.000	0.001	0.035	0.033	0.930	-0.392	-0.441
$\overline{Y}_{\mathrm{cc}}$	-0.000	0.007	0.007	0.944	0.907	0.695	-0.000	0.007	0.007	0.944	0.940	0.700
$\hat{\theta}^{\mathrm{sp,T}}_{\mathcal{L}}$	-0.000	0.010	0.010	0.957	0.830	0.588	0.000	0.012	0.014	0.961	0.828	0.493
$\hat{\theta}^{\mathrm{sp,T}}_{\mathrm{IPW}}$	-0.000	0.010	0.010	0.957	0.830	0.588	0.000	0.013	0.014	0.947	0.819	0.480
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathcal{L}}$	-0.000	0.010	0.011	0.957	0.830	0.588	0.000	0.012	0.014	0.962	0.828	0.493
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathrm{IPW}}$	-0.000	0.010	0.011	0.959	0.830	0.588	0.000	0.013	0.014	0.956	0.819	0.480
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathcal{L}}$	-0.000	0.010	0.010	0.955	0.830	0.588	0.000	0.012	0.010	0.919	0.842	0.514
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathrm{IPW}}$	-0.000	0.010	0.010	0.955	0.830	0.588	0.000	0.012	0.010	0.919	0.842	0.514
	$M = 5000$											
$\overline{Y}$	-0.000	0.015	0.015	0.939	-	-	0.016	0.016	0.015	0.784	-	-
$\overline{Y}_{\mathrm{IPW}}$	-0.000	0.015	0.015	0.939	0.000	0.000	0.001	0.023	0.022	0.935	-0.060	-0.495
$\overline{Y}_{\mathrm{cc}}$	-0.000	0.005	0.005	0.947	0.897	0.680	-0.000	0.005	0.005	0.948	0.954	0.689
$\hat{\theta}^{\mathrm{sp,T}}_{\mathcal{L}}$	-0.000	0.007	0.006	0.949	0.809	0.563	0.000	0.008	0.008	0.945	0.877	0.490
$\hat{\theta}^{\mathrm{sp,T}}_{\mathrm{IPW}}$	-0.000	0.007	0.006	0.949	0.809	0.563	0.000	0.008	0.008	0.949	0.874	0.484
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathcal{L}}$	-0.000	0.007	0.007	0.952	0.809	0.563	0.000	0.008	0.008	0.950	0.877	0.490
$\hat{\theta}^{\mathrm{sp,MAR}}_{\mathrm{IPW}}$	-0.000	0.007	0.007	0.951	0.809	0.563	0.000	0.008	0.008	0.952	0.882	0.503
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathcal{L}}$	-0.000	0.007	0.007	0.949	0.809	0.562	0.000	0.008	0.006	0.897	0.882	0.503
$\hat{\theta}^{\mathrm{sp,MCAR}}_{\mathrm{IPW}}$	-0.000	0.007	0.007	0.949	0.809	0.562	0.000	0.008	0.006	0.898	0.882	0.503

Table S2: The results of Simulation (Y1) on 1000 simulation runs under $p = 100, M = 1000, 2000, 5000$.

	Bias	SE	SD	CP	Eff(M)	Eff(S)	Bias	SE	SD	CP	Eff(M)	Eff(S)
	MCAR(K1)						MAR(K2)					
	$M = 1000$											
$\overline{Y}$	-0.000	-0.000	0.033	0.948	-	-	0.015	0.033	0.033	0.918	-	-
$\overline{Y}_{\text{IPW}}$	-0.000	-0.000	0.033	0.948	0.000	0.000	0.001	0.048	0.046	0.950	-0.728	-0.443
$\overline{Y}_{\text{cc}}$	0.000	0.000	0.011	0.936	0.892	0.672	0.000	0.011	0.011	0.934	0.913	0.677
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.000	-0.000	0.017	0.959	0.768	0.518	0.002	0.020	0.020	0.948	0.712	0.414
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.000	-0.000	0.017	0.958	0.766	0.516	0.003	0.020	0.021	0.943	0.687	0.390
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.000	-0.000	0.017	0.955	0.767	0.518	0.002	0.020	0.016	0.866	0.709	0.411
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.000	-0.000	0.017	0.955	0.767	0.518		0.020	0.016	0.861	0.692	0.395
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.000	-0.000	0.017	0.958	0.768	0.518	0.002	0.020	0.017	0.892	0.699	0.402
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.000	-0.000	0.017	0.962	0.769	0.519	0.002	0.020	0.017	0.890	0.697	0.401
	$M = 2000$											
$\overline{Y}$	-0.001	0.023	0.023	0.954	-	-	0.016	0.023	0.023	0.912	-	-
$\overline{Y}_{\text{IPW}}$	-0.001	0.023	0.023	0.954	0.000	0.000	0.000	0.034	0.033	0.932	-0.509	-0.497
$\overline{Y}_{\text{cc}}$	0.000	0.007	0.007	0.954	0.897	0.678	0.000	0.008	0.007	0.954	0.927	0.432
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.000	0.010	0.011	0.955	0.798	0.550	0.000	0.013	0.015	0.961	0.782	0.417
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.000	0.010	0.011	0.958	0.799	0.551	0.000	0.013	0.014	0.953	0.771	0.437
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.000	0.010	0.011	0.960	0.798	0.550	0.000	0.013	0.011	0.894	0.786	0.420
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.000	0.010	0.011	0.958	0.798	0.551	0.000	0.013	0.011	0.887	0.773	0.431
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.000	0.010	0.011	0.960	0.798	0.550	0.000	0.013	0.011	0.887	0.785	0.421
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.000	0.010	0.011	0.960	0.798	0.551	0.001	0.013	0.011	0.888	0.782	0.432
	$M = 5000$											
$\overline{Y}$	0.000	0.000	0.015	0.923	-	-	0.017	0.015	0.015	0.794	-	-
$\overline{Y}_{\text{IPW}}$	0.000	0.000	0.015	0.923	0.000	0.000	-0.000	0.022	0.022	0.933	0.047	-0.463
$\overline{Y}_{\text{cc}}$	0.000	0.000	0.005	0.945	0.909	0.699	0.000	0.005	0.005	0.945	0.956	0.687
$\hat{\theta}_{\mathcal{S},K,\text{L}}^{\text{sp,1,T}}$	0.000	0.000	0.007	0.955	0.830	0.588	0.000	0.008	0.008	0.961	0.880	0.481
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	0.000	0.000	0.007	0.954	0.830	0.587	0.000	0.008	0.008	0.949	0.876	0.472
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	0.000	0.000	0.007	0.954	0.830	0.587	-0.000	0.008	0.007	0.922	0.884	0.489
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	0.000	0.000	0.007	0.953	0.830	0.587	0.000	0.008	0.007	0.910	0.884	0.489
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	0.000	0.000	0.007	0.954	0.830	0.587	0.000	0.008	0.007	0.897	0.884	0.490
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	0.000	0.000	0.007	0.955	0.830	0.587	-0.000	0.008	0.007	0.895	0.884	0.489

Table S3: The results of Simulation (Y3) on 1000 simulation runs under $p = 10, M = 1000, 2000, 5000$.

	Bias	SE	SD	CP	Eff(M)	Eff(S)	Bias	SE	SD	CP	Eff(M)	Eff(S)
	MCAR(K1)						MAR(K2)					
	$M = 1000$											
$\overline{Y}$	0.001	0.088	0.088	0.943	-	-	0.042	0.090	0.089	0.940	-	-
$\overline{Y}_{\text{IPW}}$	0.001	0.088	0.088	0.943	0.000	0.000	0.001	0.127	0.119	0.925	-0.651	-0.421
$\overline{Y}_{\text{cc}}$	-0.000	0.027	0.028	0.950	0.902	0.686	0.000	0.027	0.028	0.951	0.923	0.693
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.004	0.031	0.031	0.950	0.875	0.650	-0.003	0.033	0.035	0.953	0.890	0.635
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.004	0.031	0.031	0.950	0.875	0.650	-0.002	0.034	0.036	0.952	0.885	0.625
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.004	0.030	0.032	0.957	0.878	0.654	-0.003	0.033	0.037	0.963	0.892	0.637
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.004	0.030	0.032	0.959	0.878	0.653	-0.001	0.033	0.038	0.967	0.886	0.627
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.004	0.030	0.032	0.954	0.875	0.650	-0.002	0.034	0.032	0.926	0.885	0.625
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.004	0.030	0.032	0.954	0.875	0.650	-0.002	0.034	0.032	0.926	0.885	0.625
	$M = 2000$											
$\overline{Y}$	0.002	0.064	0.062	0.943	-	-	0.048	0.065	0.063	0.878	-	-
$\overline{Y}_{\text{IPW}}$	0.002	0.064	0.062	0.943	0.000	0.000	0.003	0.091	0.088	0.938	-0.257	-0.394
$\overline{Y}_{\text{cc}}$	-0.000	0.019	0.020	0.949	0.909	0.699	0.000	0.019	0.020	0.944	0.944	0.705
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.002	0.022	0.024	0.958	0.926	0.664
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.024	0.955	0.923	0.656
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.002	0.020	0.022	0.954	0.898	0.682	-0.002	0.022	0.024	0.958	0.927	0.665
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.002	0.020	0.022	0.955	0.898	0.682	-0.001	0.022	0.024	0.957	0.925	0.659
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.022	0.932	0.920	0.648
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.022	0.932	0.920	0.648
	$M = 5000$											
$\overline{Y}$	-0.001	0.040	0.039	0.939	-	-	0.045	0.042	0.040	0.793	-	-
$\overline{Y}_{\text{IPW}}$	-0.001	0.040	0.039	0.939	0.000	0.000	0.001	0.061	0.057	0.931	0.002	-0.466
$\overline{Y}_{\text{cc}}$	-0.000	0.013	0.012	0.944	0.897	0.679	0.000	0.013	0.012	0.946	0.956	0.692
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.002	0.014	0.013	0.935	0.882	0.659	-0.001	0.014	0.014	0.949	0.944	0.653
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.002	0.014	0.013	0.935	0.882	0.659	-0.001	0.015	0.015	0.946	0.943	0.651
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.001	0.014	0.013	0.936	0.882	0.659	-0.001	0.014	0.015	0.949	0.944	0.654
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.001	0.014	0.013	0.936	0.882	0.659	-0.001	0.015	0.015	0.949	0.943	0.652
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.001	0.014	0.013	0.936	0.882	0.658	-0.001	0.015	0.013	0.914	0.939	0.639
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.001	0.014	0.013	0.936	0.882	0.658	-0.001	0.015	0.013	0.914	0.939	0.639

Table S4: The results of Simulation (Y3) on 1000 simulation runs under $p = 100, M = 1000, 2000, 5000$.

	Bias	SE	SD	CP	Eff(M)	Eff(S)	Bias	SE	SD	CP	Eff(M)	Eff(S)
	MCAR(K1)						MAR(K2)					
	$M = 1000$											
$\overline{Y}$	0.001	0.088	0.088	0.943	-	-	0.042	0.090	0.089	0.940	-	-
$\overline{Y}_{\text{IPW}}$	0.001	0.088	0.088	0.943	0.000	0.000	0.001	0.127	0.119	0.925	-0.651	-0.421
$\overline{Y}_{\text{cc}}$	-0.000	0.027	0.028	0.950	0.902	0.686	0.000	0.027	0.028	0.951	0.923	0.693
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.004	0.031	0.031	0.950	0.875	0.650	-0.003	0.033	0.035	0.953	0.890	0.635
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.004	0.031	0.031	0.950	0.875	0.650	-0.002	0.034	0.036	0.952	0.885	0.625
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.004	0.030	0.032	0.957	0.878	0.654	-0.003	0.033	0.037	0.963	0.892	0.637
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.004	0.030	0.032	0.959	0.878	0.653	-0.001	0.033	0.038	0.967	0.886	0.627
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.004	0.030	0.032	0.954	0.875	0.650	-0.002	0.034	0.032	0.926	0.885	0.625
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.004	0.030	0.032	0.954	0.875	0.650	-0.002	0.034	0.032	0.926	0.885	0.625
	$M = 2000$											
$\overline{Y}$	0.002	0.064	0.062	0.943	-	-	0.048	0.065	0.063	0.878	-	-
$\overline{Y}_{\text{IPW}}$	0.002	0.064	0.062	0.943	0.000	0.000	0.003	0.091	0.088	0.938	-0.257	-0.394
$\overline{Y}_{\text{cc}}$	-0.000	0.019	0.020	0.949	0.909	0.699	0.000	0.019	0.020	0.944	0.944	0.705
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.002	0.022	0.024	0.958	0.926	0.664
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.024	0.955	0.923	0.656
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.002	0.020	0.022	0.954	0.898	0.682	-0.002	0.022	0.024	0.958	0.927	0.665
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.002	0.020	0.022	0.955	0.898	0.682	-0.001	0.022	0.024	0.957	0.925	0.659
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.022	0.932	0.920	0.648
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.002	0.020	0.022	0.950	0.896	0.680	-0.001	0.023	0.022	0.932	0.920	0.648
	$M = 5000$											
$\overline{Y}$	-0.001	0.040	0.039	0.939	-	-	0.045	0.042	0.040	0.793	-	-
$\overline{Y}_{\text{IPW}}$	-0.001	0.040	0.039	0.939	0.000	0.000	0.001	0.061	0.057	0.931	0.002	-0.466
$\overline{Y}_{\text{cc}}$	-0.000	0.013	0.012	0.944	0.897	0.679	0.000	0.013	0.012	0.946	0.956	0.692
$\hat{\theta}_{\mathcal{L}}^{\text{sp,T}}$	-0.002	0.014	0.013	0.935	0.882	0.659	-0.001	0.014	0.014	0.949	0.944	0.653
$\hat{\theta}_{\text{IPW}}^{\text{sp,T}}$	-0.002	0.014	0.013	0.935	0.882	0.659	-0.001	0.015	0.015	0.946	0.943	0.651
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MAR}}$	-0.001	0.014	0.013	0.936	0.882	0.659	-0.001	0.014	0.015	0.949	0.944	0.654
$\hat{\theta}_{\text{IPW}}^{\text{sp,MAR}}$	-0.001	0.014	0.013	0.936	0.882	0.659	-0.001	0.015	0.015	0.949	0.943	0.652
$\hat{\theta}_{\mathcal{L}}^{\text{sp,MCAR}}$	-0.001	0.014	0.013	0.936	0.882	0.658	-0.001	0.015	0.013	0.914	0.939	0.639
$\hat{\theta}_{\text{IPW}}^{\text{sp,MCAR}}$	-0.001	0.014	0.013	0.936	0.882	0.658	-0.001	0.015	0.013	0.914	0.939	0.639

References

[1] Arun Kumar Kuchibhotla and Abhishek Chakrabortty. Moving beyond sub-gaussianity in high-dimensional statistics: Applications in covariance estimation and linear regression. *Information and Inference: A Journal of the IMA*, 11(4):1389–1456, 2022.

[2] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[3] Yuqian Zhang, Abhishek Chakrabortty, and Jelena Bradic. Double robust semi-supervised inference for the mean: selection bias under MAR labeling with decaying overlap. *Information and Inference: A Journal of the IMA*, 12(3):2066–2159, 2023.