Supplement to "Subsampling-based Methods for Inference of the Mean Outcome under Optimal Treatment Regimes"

S1 Detailed estimating procedure in Section 5.2

For any $a_1 \in \{0,1\}$, we calculate

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{\pi_2,a_1} &= \operatorname*{arg\,min}_{\boldsymbol{\xi}\in\mathbb{R}^{3(K+4)}} \sum_{i\in\mathcal{I}} \left(A_i^{(2)} - \sum_{j=1}^{K+4} N_j(X_i^{(2)})\xi_j - \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(X_{i,k}^{(1)})\xi_{j+k(K+4)} \right)^2 \mathbb{I}(A_i^{(1)} = a_1), \\
\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{h_{2,1},a_1} &= \operatorname*{arg\,min}_{\boldsymbol{\xi}\in\mathbb{R}^{3(K+4)}} \sum_{i\in\mathcal{I}} \left(Y_i - \sum_{j=1}^{K+4} N_j(X_i^{(2)})\xi_j - \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(X_{i,k}^{(1)})\xi_{j+k(K+4)} \right)^2 \mathbb{I}(A_i^{(2)} = 1, A_i^{(1)} = a_1), \\
\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{h_{2,0},a_1} &= \operatorname*{arg\,min}_{\boldsymbol{\xi}\in\mathbb{R}^{3(K+4)}} \sum_{i\in\mathcal{I}} \left(Y_i - \sum_{j=1}^{K+4} N_j(X_i^{(2)})\xi_j - \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(X_{i,k}^{(1)})\xi_{j+k(K+4)} \right)^2 \mathbb{I}(A_i^{(2)} = 0, A_i^{(1)} = a_1),
\end{aligned}$$

and compute

$$\begin{aligned}
&\widehat{\pi}_{\mathcal{I},2}((a_1,1),\bar{\boldsymbol{x}}_2) = \min\left(\sum_{j=1}^{K+4} N_j(x_2)\widehat{\xi}_{\mathcal{I},j}^{\pi_2,1} + \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(x_{1,k})\widehat{\xi}_{\mathcal{I},j+k(K+4)}^{\pi_2,1}, 0.05 \right) \mathbb{I}(A_i^{(1)} = a_1), \\
&\widehat{\pi}_{\mathcal{I},2}((a_1,0),\bar{\boldsymbol{x}}_2) = \min\{1 - \widehat{\pi}_{\mathcal{I},2}((a_1,1),\bar{\boldsymbol{x}}_2), 0.05\}, \\
&\widehat{h}_{\mathcal{I},2}((a_1,a_2),\bar{\boldsymbol{x}}_2) = \left(\sum_{j=1}^{K+4} N_j(x_2)\widehat{\xi}_{\mathcal{I},j}^{h_{2,a_2},a_1} + \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(x_{1,k})\widehat{\xi}_{\mathcal{I},j+k(K+4)}^{h_{2,a_2},a_1} \right) \mathbb{I}(A_i^{(1)} = a_1), \\
&\widehat{d}_{\mathcal{I},2}(a_1,\bar{\boldsymbol{x}}_2) = I[\widehat{h}_{\mathcal{I},2}\{(a_1,1),\bar{\boldsymbol{x}}_2\} > \widehat{h}_{\mathcal{I},2}\{(a_1,0),\bar{\boldsymbol{x}}_2\}].
\end{aligned}$$

Then we construct the pseudo outcome

$$\widehat{V}_{i,\mathcal{I}} = \frac{\mathrm{g}\{A_i^{(2)}, \widehat{d}_{\mathcal{I},2}(A_i^{(1)}, \bar{\boldsymbol{X}}_i^{(2)})\}}{\widehat{\pi}_{\mathcal{I},2}(\bar{\boldsymbol{A}}_i^{(2)}, \bar{\boldsymbol{X}}_i^{(2)})}\{Y_i - \widehat{h}_{\mathcal{I},2}(\bar{\boldsymbol{A}}_i^{(2)}, \bar{\boldsymbol{X}}_i^{(2)})\} + \widehat{h}_{\mathcal{I},2}[\{A_i^{(1)}, \widehat{d}_{\mathcal{I},2}(A_i^{(1)}, \bar{\boldsymbol{X}}_i^{(2)})\}, \bar{\boldsymbol{X}}_i^{(2)}], \forall i \in \mathcal{I}_0,$$

and compute

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{\pi_1} &= \operatorname*{arg\,min}_{\boldsymbol{\xi}\in\mathbb{R}^{2(K+4)}} \sum_{i\in\mathcal{I}} \left(A_i^{(1)} - \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(X_{i,k}^{(1)})\xi_{j+(k-1)(K+4)} \right)^2, \\
\widehat{\boldsymbol{\xi}}_{\mathcal{I}}^{h_{a_1}} &= \operatorname*{arg\,min}_{\boldsymbol{\xi}\in\mathbb{R}^{2(K+4)}} \sum_{i\in\mathcal{I}} \left(\widehat{V}_{i,\mathcal{I}} - \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(X_{i,k}^{(1)})\xi_{j+(k-1)(K+4)} \right)^2 \mathbb{I}(A_i^{(1)} = a_1),
\end{aligned}$$

for $a_1 = \{0, 1\}$. Finally, we set

$$\begin{aligned}
\widehat{\pi}_{\mathcal{I}}(1, \boldsymbol{x}_1) &= \min\left(\sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(x_{1,k})\widehat{\xi}^{\pi_1}_{\mathcal{I},j+(k-1)(K+4)}, 0.05\right), \widehat{\pi}_{\mathcal{I}}(0, \boldsymbol{x}_1) = \min\{1 - \widehat{\pi}_{\mathcal{I}}(1, \boldsymbol{x}_1), 0.05\}, \\
\widehat{h}_{\mathcal{I}}(1, \boldsymbol{x}_1) &= \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(x_{1,k})\widehat{\xi}^{h_1}_{\mathcal{I},j+(k-1)(K+4)}, \widehat{h}_{\mathcal{I}}(0, \boldsymbol{x}_1) = \sum_{j=1}^{K+4}\sum_{k=1}^{2} N_j(x_{1,k})\widehat{\xi}^{h_0}_{\mathcal{I},j+(k-1)(K+4)}, \\
\widehat{d}_{\mathcal{I}}(\boldsymbol{x}_1) &= I\{\widehat{h}_{\mathcal{I}}(1, \boldsymbol{x}_1) > \widehat{h}_{\mathcal{I}}(0, \boldsymbol{x}_1)\}.
\end{aligned}$$

Similar to Section 5.1, we select the number of interior knots K via 5-folded cross-validation.

S2 Proofs

S2.1 Proof of Lemma 2.1

Under (A1) and (A2), we have

$$\begin{aligned}
\tau(\boldsymbol{x}) &= \mathrm{E}(Y_0|A_0 = 1, \boldsymbol{X}_0 = x) - \mathrm{E}(Y_0|A_0 = 0, \boldsymbol{X}_0 = x) = \mathrm{E}\{Y_0^*(1)|A_0 = 1, \boldsymbol{X}_0 = x\} \\
&- \mathrm{E}\{Y_0^*(0)|A_0 = 0, \boldsymbol{X}_0 = x\} = \mathrm{E}\{Y_0^*(1)|\boldsymbol{X}_0 = x\} - \mathrm{E}\{Y_0^*(0)|\boldsymbol{X}_0 = x\}.
\end{aligned}$$

Then, for any treatment regime d,

$$\begin{aligned}
V(d) &= \mathrm{E}Y_0^*(0) + \mathrm{E}\{Y_0^*(1) - Y_0^*(0)\}d(\boldsymbol{X}_0) = \mathrm{E}Y_0^*(0) + \mathrm{E}[\mathrm{E}\{Y_0^*(1) - Y_0^*(0)|\boldsymbol{X}_0\}]d(\boldsymbol{X}_0) \\
&= \mathrm{E}Y_0^*(0) + \mathrm{E}\tau(\boldsymbol{X}_0)d(\boldsymbol{X}_0).
\end{aligned}$$

Therefore, it is immediate to see that $d^{opt,0} \in \mathcal{D}^{opt}$. For any treatment regime d, we have

$$V(d^{opt,0}) - V(d) = \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]. \tag{S2.1}$$

We first show any OTR shall satisfy (1). Assume there exists an treatment regime $d \in \mathcal{D}^{opt}$ such that

$$\Pr(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}) > 0. \tag{S2.2}$$

Note that $\tau(\boldsymbol{x})[\mathbb{I}\{\tau(\boldsymbol{x}) > 0\} - d(\boldsymbol{x})] \geq 0, \forall \boldsymbol{x}$. By (S2.1), we have

$$\begin{aligned}
&V(d^{opt,0}) - V(d) = \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)] \\
\geq\ &\mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}) = \mathrm{E}\tau(\boldsymbol{X}_0)\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}).
\end{aligned}$$

Since $\tau(\boldsymbol{x}) > 0$ for any $\boldsymbol{x} \in \mathbb{X}_1$, it follows from (S2.2) that

$$V(d^{opt,0}) - V(d) = \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)] \geq \mathrm{E}\tau(\boldsymbol{X}_0)\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}) > 0.$$

This contradicts the fact that $V(d^{opt,0}) = V(d)$. Therefore,

$$\Pr(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{2,d}) = 0.$$

Similarly, we can show

$$\Pr(\boldsymbol{X}_0 \in \mathbb{X}_2 \cap \mathbb{X}_{1,d}) = 0.$$

This implies that any $d \in \mathcal{D}^{opt}$ must satisfy (1).

Conversely, for any treatment regime d that satisfies (1), it follows from (S2.2) that

$$\begin{aligned} & V(d^{opt,0}) - V(d) = \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_1) \\ + \;& \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_2) \\ = \;& \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_1 \cap \mathbb{X}_{1,d}) \\ + \;& \mathrm{E}\tau(\boldsymbol{X}_0)[\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\} - d(\boldsymbol{X}_0)]\mathbb{I}(\boldsymbol{X}_0 \in \mathbb{X}_2 \cap \mathbb{X}_{2,d}) = 0. \end{aligned}$$

This shows $d \in \mathcal{D}^{opt}$. The proof is hence completed.

S2.2 Proof of Theorem 2.2

For any functions $d(\cdot)$, $\pi^*(\cdot,\cdot)$ and $h^*(\cdot,\cdot)$, define

$$V(d;\pi^*,h^*) = \mathrm{E}\left(\frac{\mathrm{g}\{A_0, d(\boldsymbol{X}_0)\}}{\pi^*(A_0,\boldsymbol{X}_0)}\{Y_0 - h^*(A_0,\boldsymbol{X}_0)\} + h^*(d(\boldsymbol{X}_0),\boldsymbol{X}_0)\right).$$

It is immediate to see that $V_0 = V(d^{opt};\pi,h)$ for any $d^{opt} \in \mathcal{D}^{opt}$. Let $n_{\mathcal{S}} = |\mathcal{S}_{N_0,s_n}|$. Before proving Theorem 2.2, we present the following lemmas whose proofs are given in Section S2.6.

Lemma S2.1 *Under the conditions in Theorem 2.2, we have*

$$\sup_{a=0,1,\boldsymbol{x}\in\mathbb{X}} |h(a,\boldsymbol{x})| \leq C_0,$$

for some constant $C_0 > 0$.

Lemma S2.2 *Under the conditions in Theorem 2.2, there exist some constants $c_1, c_2, c_3 > 0$ and $0 < p_* < 1$ such that*

$$Pr\left(\frac{\binom{n}{s_n} - n_{\mathcal{S}}}{\binom{n}{s_n}} \le c_1 p_*^{c_2 n^{\beta_0}} \right) \ge 1 - 2\exp(-c_3 n). \tag{S2.3}$$

where β_0 is defined in Theorem 2.1. In addition, we have

$$\max_{i \in \{1,\dots,n\}} Pr\left(\left| \frac{n^{(i)}}{B} - \frac{n - s_n}{n} \right| \le \frac{\sqrt{\log n}}{\sqrt{n}} \right) \ge 1 - 4\exp(-c_4 n) - \frac{2}{B}, \tag{S2.4}$$

for some constant $c_4 > 0$, where $n^{(i)} = \sum_{b=1}^{B} \mathbb{I}(i \in \mathcal{I}_b^c)$.

Theorem 2.1 implies that $\widehat{V}_\infty^* - V_0 = \eta_1 - \mathrm{E}\eta_1 + o_p(n^{-1/2})$. Under (A3), (A4) and the condition $\liminf_n \sigma_n > 0$, it follows from central limit theorem that

$$\frac{\sqrt{n}(\widehat{V}_\infty^* - V_0)}{\sigma_{s_n}} \xrightarrow{d} N(0,1). \tag{S2.5}$$

Assume for now, we've shown

$$\widehat{V}_B = \widehat{V}_\infty^* + o_p(n^{-1/2}), \tag{S2.6}$$

and

$$\widehat{\sigma}_B^2 = \sigma_{s_n}^2 + o_p(1). \tag{S2.7}$$

In view of (S2.5), we have

$$\frac{\sqrt{n}(\widehat{V}_B - V_0)}{\widehat{\sigma}_B} \xrightarrow{d} N(0,1),$$

by Slutsky's theorem and the condition that $\liminf_n \sigma_n > 0$. Therefore, it suffices to show (S2.6) and (S2.7).

In the following, we break the proof into three steps. In the first step, we show $\widehat{V}_B = \widehat{V}_B^* + o_p(n^{-1/2})$ where

$$\widehat{V}_B^* \equiv \frac{1}{B} \sum_{b=1}^{B} \widehat{V}_{\mathcal{I}_b^c}(\widehat{d}_{\mathcal{I}_b}; \pi, h) = \frac{1}{2B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) + \widehat{V}_{\mathcal{I}_b^{c(1)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) \right). \tag{S2.8}$$

Next, we show $\widehat{V}_B^* = \widehat{V}_\infty^* + o_p(n^{-1/2})$. In the last step, we show (S2.7) hold.

Step 1: Recall that $\widehat{V}_B$ is defined as

$$\widehat{V}_B = \frac{1}{2B}\sum_{b=1}^{B}\left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(1)}},\widehat{h}_{\mathcal{I}_b^{(1)}}) + \widehat{V}_{\mathcal{I}_b^{c(1)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(2)}},\widehat{h}_{\mathcal{I}_b^{(2)}})\right).$$

In view of (S2.8), it suffices to show

$$\frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\pi,h) = \frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(1)}},\widehat{h}_{\mathcal{I}_b^{(1)}}) + o_p(n^{-1/2}), \tag{S2.9}$$

$$\frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}_b^{c(1)}}(\widehat{d}_{\mathcal{I}_b};\pi,h) = \frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}_b^{c(1)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(2)}},\widehat{h}_{\mathcal{I}_b^{(2)}}) + o_p(n^{-1/2}). \tag{S2.10}$$

In the following, we prove (S2.9). Let $\mathcal{A}_0$ denote the event defined in (S2.3). On the set $\mathcal{A}_0$, we have

$$n_S \geq \frac{1}{2}\binom{n}{s_n}, \tag{S2.11}$$

for sufficiently large n. Notice that $\mathcal{A}_0$ depends only on the dataset $\{O_i\}_{i\in\mathcal{I}_0}$. By Lemma S2.2, we have $\Pr(\mathcal{A}_0)\to 1$.

For any $\mathcal{I}\subseteq\mathcal{I}_0$ with $|\mathcal{I}| = s_n$, let $\mathcal{P}(\mathcal{I})$ denote the set of partitions, i.e,

$$\mathcal{P}(\mathcal{I}) \equiv \left\{(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)}) : \mathcal{I}^{c(1)}\cup\mathcal{I}^{c(2)} = \mathcal{I}^c, \mathcal{I}^{c(1)}\cap\mathcal{I}^{c(2)} = \emptyset, |\mathcal{I}^{c(1)}| = |\mathcal{I}^{c(2)}| = t_n\right\}.$$

Notice that $|\mathcal{P}(\mathcal{I}_1)| = |\mathcal{P}(\mathcal{I}_2)|$ for any subsets $\mathcal{I}_1,\mathcal{I}_2$ such that $|\mathcal{I}_1| = |\mathcal{I}_2|$. Define $\mathcal{P}_0 = \mathcal{P}(\mathcal{I})$ for any $\mathcal{I}\subseteq\mathcal{I}_0$ such that $|\mathcal{I}| = s_n$. For $j = 1,2$, let $\mathcal{I}^{(j)} = \mathcal{I}\cup\mathcal{I}^{c(j)}$, we have

$$\begin{aligned}
& \mathrm{E}\left|\frac{1}{B}\sum_{b=1}^{B}\left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\pi,h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(1)}},\widehat{h}_{\mathcal{I}_b^{(1)}})\right)\right|\mathbb{I}(\mathcal{A}_0) \\
\leq\ & \frac{1}{B}\sum_{b=1}^{B}\mathrm{E}\left|\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\pi,h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(1)}},\widehat{h}_{\mathcal{I}_b^{(1)}})\right|\mathbb{I}(\mathcal{A}_0) \\
=\ & \mathrm{E}\left|\widehat{V}_{\mathcal{I}_1^{c(2)}}(\widehat{d}_{\mathcal{I}_1};\pi,h) - \widehat{V}_{\mathcal{I}_1^{c(2)}}(\widehat{d}_{\mathcal{I}_1};\widehat{\pi}_{\mathcal{I}_1^{(1)}},\widehat{h}_{\mathcal{I}_1^{(1)}})\right|\mathbb{I}(\mathcal{A}_0) && \text{(S2.12)} \\
=\ & \mathrm{E}\frac{1}{n_S\mathcal{P}_0}\sum_{\substack{\mathcal{I}\in\mathcal{S}_{N_0,s_n}\\(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}}\left|\widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\pi,h) - \widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\widehat{\pi}_{\mathcal{I}^{(1)}},\widehat{h}_{\mathcal{I}^{(1)}})\right|\mathbb{I}(\mathcal{A}_0) && \text{(S2.13)} \\
\leq\ & \frac{2}{\binom{n}{s_n}\mathcal{P}_0}\mathrm{E}\sum_{\substack{\mathcal{I}\in\mathcal{S}_{N_0,s_n}\\(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}}\left|\widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\pi,h) - \widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\widehat{\pi}_{\mathcal{I}^{(1)}},\widehat{h}_{\mathcal{I}^{(1)}})\right| && \text{(S2.14)} \\
\leq\ & \frac{2}{\binom{n}{s_n}\mathcal{P}_0}\mathrm{E}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0,|\mathcal{I}|=s_n\\(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}}\left|\widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\pi,h) - \widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}};\widehat{\pi}_{\mathcal{I}^{(1)}},\widehat{h}_{\mathcal{I}^{(1)}})\right|. && \text{(S2.15)}
\end{aligned}$$

where the first equality is due to the fact that $(\mathcal{I}_1, \mathcal{I}_1^{c(1)}, \mathcal{I}_1^{c(2)}), \ldots, (\mathcal{I}_B, \mathcal{I}_B^{c(1)}, \mathcal{I}_B^{c(2)})$ are independent and identically distributed conditional on $\{O_i\}_{i\in\mathcal{I}_0}$, the second equality is due to the fact that

$$\begin{aligned} & \mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_0}} \left| \widehat{V}_{\mathcal{I}_1^{c(2)}}(\widehat{d}_{\mathcal{I}_1}; \pi, h) - \widehat{V}_{\mathcal{I}_1^{c(2)}}(\widehat{d}_{\mathcal{I}_1}; \widehat{\pi}_{\mathcal{I}_1^{(1)}}, \widehat{h}_{\mathcal{I}_1^{(1)}}) \right| \\ = & \frac{1}{n_S \mathcal{P}_0} \sum_{\substack{\mathcal{I}\in\mathcal{S}_{N_0, s_n} \\ (\mathcal{I}^{c(1)}, \mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}} \left| \widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}}; \pi, h) - \widehat{V}_{\mathcal{I}^{c(2)}}(\widehat{d}_{\mathcal{I}}; \widehat{\pi}_{\mathcal{I}^{(1)}}, \widehat{h}_{\mathcal{I}^{(1)}}) \right|, \end{aligned}$$

where $\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_0}}$ denotes the conditional expectation given $\{O_i\}_{i\in\mathcal{I}_0}$, and the second inequality follows by (S2.11).

Let $\mathcal{I}_*$ be a random subset uniformly sampled from $\{\mathcal{I} \subseteq \mathcal{I}_0 : |\mathcal{I}| = s_n\}$, independent of $\{O_i\}_{i\in\mathcal{I}_0}$. Given $\mathcal{I}_*$, let $\mathcal{I}_*^{c(1)}$ and $\mathcal{I}_*^{c(2)}$ denote the random partition of $\mathcal{I}_*^c$ generated by the algorithm in Section 2.3. Notice that $(\mathcal{I}_*, \mathcal{I}_*^{c(1)}, \mathcal{I}_*^{c(2)})$ is independent of $\{O_i\}_{i\in\mathcal{I}_0}$. So far, we have shown

$$\begin{aligned} & \mathrm{E} \left| \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) \right) \right| \mathbb{I}(\mathcal{A}_0) \\ \leq & 2\mathrm{E} \left| \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*}; \pi, h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}_{\mathcal{I}_*^{(1)}}, \widehat{h}_{\mathcal{I}_*^{(1)}}) \right|. \end{aligned}$$

It follows from triangle inequality that

$$\begin{aligned} & \mathrm{E} \left| \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) \right) \right| \mathbb{I}(\mathcal{A}_0) \qquad \text{(S2.16)} \\ \leq & \underbrace{\mathrm{E} \left| \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*}; \pi, h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}_{\mathcal{I}_*^{(1)}}, \widehat{h}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*}; \pi, h) + V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}_{\mathcal{I}_*^{(1)}}, \widehat{h}_{\mathcal{I}_*^{(1)}}) \right|}_{\eta_5} \\ + & \underbrace{\mathrm{E} \left| V(\widehat{d}_{\mathcal{I}_*}; \pi, h) - V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}_{\mathcal{I}_*^{(1)}}, \widehat{h}_{\mathcal{I}_*^{(1)}}) \right|}_{\eta_6}. \end{aligned}$$

Below, we prove $\eta_5, \eta_6 = o(n^{-1/2})$. This implies for any $\varepsilon > 0$,

$$\begin{aligned} & \Pr\left(\sqrt{n} \left| \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) \right) \right| > \varepsilon \right) \qquad \text{(S2.17)} \\ \leq & \Pr(\mathcal{A}_0^c) + \Pr\left(\left\{ \sqrt{n} \left| \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) \right) \right| > \varepsilon \right\} \cap \mathcal{A}_0 \right) \\ \leq & \Pr(\mathcal{A}_0^c) + \frac{\sqrt{n}}{\varepsilon} \mathrm{E} \left| \frac{1}{B} \sum_{b=1}^{B} \left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \pi, h) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) \right) \right| \mathbb{I}(\mathcal{A}_0) = o(1). \end{aligned}$$

Hence, (S2.9) is proven.

By Cauchy-Schwarz inequality, we have

$$\eta_5^2 \le \mathrm{E}\left|\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*};\pi,h) + V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\right|^2.$$

Conditional on $\{O_i\}_{i\in\mathcal{I}_*^{(1)}}$, $\mathcal{I}_*$ and $\mathcal{I}_*^{(1)}$,

$$\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*};\pi,h) + V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})$$

corresponds to a sum of i.i.d mean zero random variables. Therefore, we have

$$\begin{aligned}\eta_5^2 &= \mathrm{EVar}\left(\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\middle|\{O_i\}_{i\in\mathcal{I}_*}^{(1)},\mathcal{I}_*,\mathcal{I}_*^{(1)}\right) \quad (S2.18)\\ &= \mathrm{E}\frac{1}{|\mathcal{I}_*^{c(2)}|}\mathrm{Var}\left(\widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\middle|\{O_i\}_{i\in\mathcal{I}_*}^{(1)},\mathcal{I}_*,\mathcal{I}_*^{(1)}\right),\end{aligned}$$

where

$$\widehat{V}_{\{0\}}(d;\pi^*,h^*) = \frac{\mathrm{g}\{A_0,d(\boldsymbol{X}_0)\}}{\pi^*(A_0,\boldsymbol{X}_0)}\{Y - h^*(A_0,\boldsymbol{X}_0)\} + h^*(d(\boldsymbol{X}_0),\boldsymbol{X}_0), \quad (S2.19)$$

for any regime d and functions π^*, h^*.

Notice that $|\mathcal{I}_*^{c(2)}| = t_n = (n-s_n)/2$. Under the condition that $s_n = o(n)$, we have

$$|\mathcal{I}_*^{c(2)}| \asymp n. \quad (S2.20)$$

It thus follows from (S2.18) that

$$\eta_5^2 \asymp \frac{1}{n}\mathrm{EVar}\left(\widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\middle|\{O_i\}_{i\in\mathcal{I}_*}^{(1)},\mathcal{I}_*,\mathcal{I}_*^{(1)}\right).$$

For any random variable $\mathbb{Z}$, we have $\mathrm{Var}(\mathbb{Z}) \le \mathrm{E}\mathbb{Z}^2$. Hence, we have

$$n\eta_5^2 \asymp \mathrm{E}\left|\widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\{0\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\right|^2. \quad (S2.21)$$

By the definition of $\widehat{V}_{\{0\}}$ (see (S2.19)) and Cauchy-Schwarz inequality, the right-hand side (RHS) of (S2.21) can be upper bounded by

$$
\begin{aligned}
& \underbrace{3\,\mathrm{E}\left|\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}Y_0-\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}Y_0\right|^2}_{\eta_5^{(1)}} \qquad \text{(S2.22)} \\
+ \; & \underbrace{3\,\mathrm{E}\left|\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}\widehat{h}_{\mathcal{I}_*^{c(1)}}(A_0,\boldsymbol{X}_0)-\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}h(A_0,\boldsymbol{X}_0)\right|^2}_{\eta_5^{(2)}} \\
+ \; & \underbrace{3\,\mathrm{E}|\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\{\widehat{h}_{\mathcal{I}_*^{(1)}}(1,\boldsymbol{X}_0)-h(1,\boldsymbol{X}_0)\}+\{1-\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}\{\widehat{h}_{\mathcal{I}_*^{c(1)}}(0,\boldsymbol{X}_0)-h(0,\boldsymbol{X}_0)\}|^2}_{\eta_5^{(3)}}.
\end{aligned}
$$

In the following, we show $\eta_5^{(1)},\eta_5^{(2)},\eta_5^{(3)}=o(1)$. This together with (S2.21) implies $\eta_5=o(n^{-1/2})$.

It follows from Condition (A1) and (A4) that

$$
\sup_{\boldsymbol{x}\in\mathbb{X},a=0,1}\mathrm{E}(Y_0^2|A_0=a,\boldsymbol{X}_0=\boldsymbol{x})\le\bar{c}^*, \qquad \text{(S2.23)}
$$

for some constant $\bar{c}^*>0$. Notice that

$$
\begin{aligned}
\eta_5^{(1)} \;&\le\; \mathrm{E}\left|\frac{Y_0}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}-\frac{Y_0}{\pi(A_0,\boldsymbol{X}_0)}\right|^2=\mathrm{E}(\mathrm{E}^{A_0,\boldsymbol{X}_0}Y_0^2)\left|\frac{1}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}-\frac{1}{\pi(A_0,\boldsymbol{X}_0)}\right|^2 \\
&\le\; \bar{c}^*\mathrm{E}\left|\frac{1}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}-\frac{1}{\pi(A_0,\boldsymbol{X}_0)}\right|^2=\bar{c}^*\mathrm{E}\left|\frac{\pi(A_0,\boldsymbol{X}_0)-\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)\pi(A_0,\boldsymbol{X}_0)}\right|^2 \\
&\le\; \frac{\bar{c}^*}{c_0c^*}\mathrm{E}|\pi(A_0,\boldsymbol{X}_0)-\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)|^2=o(1), \qquad \text{(S2.24)}
\end{aligned}
$$

where the second inequality is due to (S2.23), the third inequality is due to (A3) and the condition that $\Pr(\inf_{\mathcal{I}\subseteq\mathcal{I}_0,\boldsymbol{x}\in\mathbb{X},a=0,1}\widehat{\pi}_{\mathcal{I}}(a,\boldsymbol{x})\ge c^*)=1$, and the last equality follows by condition that $\max_{a=0,1}\mathrm{E}|\widehat{\pi}_{\mathcal{I}}(a,\boldsymbol{X}_0)-\pi(a,\boldsymbol{X}_0)|^2=o(|\mathcal{I}|^{-1/2})$. This shows $\eta_5^{(1)}=o(1)$.

By Cauchy-Schwarz inequality, we have

$$\eta_5^{(2)} \le \underbrace{2\,\mathrm{E}\left|\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}h(A_0,\boldsymbol{X}_0)-\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}h(A_0,\boldsymbol{X}_0)\right|^2}_{\eta_5^{(4)}}$$

$$+ \underbrace{2\,\mathrm{E}\left|\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}\widehat{h}_{\mathcal{I}_*^{c(1)}}(A_0,\boldsymbol{X}_0)-\frac{\mathbb{I}\{A_0=\widehat{d}_{\mathcal{I}_*}(\boldsymbol{X}_0)\}}{\widehat{\pi}_{\mathcal{I}_*^{(1)}}(A_0,\boldsymbol{X}_0)}h(A_0,\boldsymbol{X}_0)\right|^2}_{\eta_5^{(5)}}.$$

Similar to (S2.24), we can show $\eta_5^{(4)}=o(1)$. Besides, under the conditions in (7), we have

$$\eta_5^{(5)} \le \frac{1}{c^*}\mathrm{E}|\widehat{h}_{\mathcal{I}_*^{c(1)}}(A_0,\boldsymbol{X}_0)-h(A_0,\boldsymbol{X}_0)|^2=o(1).$$

This shows $\eta_5^{(2)}=o(1)$. Under the condition that $\max_{a=0,1}\mathrm{E}|\widehat{h}_{\mathcal{I}}(a,\boldsymbol{X}_0)-h(a,\boldsymbol{X}_0)|^2 = o(|\mathcal{I}|^{-1/2})$, we have $\eta_5^{(3)}=o(1)$. In view of (S2.21) and (S2.22), we've shown

$$\eta_5 = o(n^{-1/2}). \tag{S2.25}$$

We now show $\eta_6=o(n^{-1/2})$. Note that for any regime d and functions π^*, h^*,

$$V(d;\pi^*,h^*)=\mathrm{E}\left(\frac{\mathbb{I}\{A_0=d(\boldsymbol{X}_0)\}}{\pi^*(A_0,\boldsymbol{X}_0)}\{Y-h^*(A_0,\boldsymbol{X}_0)\}+h^*(d(\boldsymbol{X}_0),\boldsymbol{X}_0)\right)$$
$$= \mathrm{E}h(d(\boldsymbol{X}_0),\boldsymbol{X}_0)+\mathrm{E}\left(\frac{\pi(1,\boldsymbol{X}_0)}{\pi^*(1,\boldsymbol{X}_0)}-1\right)d(\boldsymbol{X}_0)\{h(1,\boldsymbol{X}_0)-h^*(1,\boldsymbol{X}_0)\}$$
$$+ \mathrm{E}\left(\frac{\pi(0,\boldsymbol{X}_0)}{\pi^*(0,\boldsymbol{X}_0)}-1\right)\{1-d(\boldsymbol{X}_0)\}\{h(0,\boldsymbol{X}_0)-h^*(0,\boldsymbol{X}_0)\}.$$

Therefore,

$$|V(d;\pi^*,h^*)-V(d;\pi,h)| \le \sum_{a=0,1}\mathrm{E}\left|\left(\frac{\pi(a,\boldsymbol{X}_0)}{\pi^*(a,\boldsymbol{X}_0)}-1\right)\{h(a,\boldsymbol{X}_0)-h^*(a,\boldsymbol{X}_0)\}\right|$$
$$\le \frac{1}{\inf_{a=0,1,\boldsymbol{x}\in\mathbb{X}}\pi^*(a,\boldsymbol{x})}\sum_{a=0,1}\mathrm{E}|\pi(a,\boldsymbol{X}_0)-\pi^*(a,\boldsymbol{X}_0)||h(a,\boldsymbol{X}_0)-h^*(a,\boldsymbol{X}_0)|$$
$$\le \frac{1}{\inf_{a=0,1,\boldsymbol{x}\in\mathbb{X}}\pi^*(a,\boldsymbol{x})}\sum_{a=0,1}\frac{1}{2}\mathrm{E}\left(|\pi(a,\boldsymbol{X}_0)-\pi^*(a,\boldsymbol{X}_0)|^2+|h(a,\boldsymbol{X}_0)-h^*(a,\boldsymbol{X}_0)|^2\right),$$

where the last inequality follows by Cauchy-Schwarz inequality. Hence, under the conditions in (6) and (7), we have

$$\begin{aligned}
&\eta_6 = \mathrm{E}|V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*};\pi,h)| \\
\leq\ & \frac{1}{2c^*}\sum_{a=0,1}\mathrm{E}\left(|\pi(a,\boldsymbol{X}_0)-\widehat{\pi}_{\mathcal{I}_*^{(1)}}(a,\boldsymbol{X}_0)|^2 + |h(a,\boldsymbol{X}_0)-\widehat{h}_{\mathcal{I}_*^{(1)}}(a,\boldsymbol{X}_0)|^2\right) \\
\leq\ & \frac{1}{c^*}\max_{a=0,1}\mathrm{E}\left(|\pi(a,\boldsymbol{X}_0)-\widehat{\pi}_{\mathcal{I}_*^{(1)}}(a,\boldsymbol{X}_0)|^2 + |h(a,\boldsymbol{X}_0)-\widehat{h}_{\mathcal{I}_*^{(1)}}(a,\boldsymbol{X}_0)|^2\right) = o(|\mathcal{I}_*^{(1)}|^{-1/2}).
\end{aligned}$$

Besides, similar to (S2.20), we can show $|\mathcal{I}_*^{(1)}| \asymp n$. Hence, we obtain $\eta_6 = o(n^{-1/2})$. By Markov's inequality, this together with (S2.25) yields (S2.9). Similarly, we can show (S2.10) holds. Therefore, we have $\widehat{V}_B = \widehat{V}_B^* + o_p(n^{-1/2})$.

Step 2: Recall that $\widehat{V}_B^*$ is defined as

$$\widehat{V}_B^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}_b^c}(\widehat{d}_{\mathcal{I}_b};\pi,h).$$

The expectation and variance of $\widehat{V}_B^*$ conditional on $\{O_i\}_{i\in\mathcal{I}_0}$ are given by

$$\begin{aligned}
\mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0}) &= \frac{1}{n_{\mathcal{S}}}\sum_{\mathcal{I}\in\mathcal{S}_{N_0,sn}}\widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h), \\
\mathrm{Var}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0}) &= \frac{n_{\mathcal{S}}-1}{n_{\mathcal{S}}B}\widehat{s.e}^2\left(\left\{\widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right\}_{\mathcal{I}\in\mathcal{S}_{N_0,sn}}\right).
\end{aligned}$$

For any $\varepsilon > 0$, we have

$$\begin{aligned}
&\Pr(\sqrt{n}|\widehat{V}_B^* - \widehat{V}_\infty^*| > 2\varepsilon) \leq \Pr\left(\sqrt{n}|\mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0}) - \widehat{V}_\infty^*| > \varepsilon\right) \\
+\ & \Pr\left(\sqrt{n}|\widehat{V}_B^* - \mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0})| > \varepsilon\right) \leq \Pr\left(\left\{\sqrt{n}|\mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0}) - \widehat{V}_\infty^*| > \varepsilon\right\}\cap\mathcal{A}_0\right) \\
+\ & \Pr\left(\sqrt{n}|\widehat{V}_B^* - \mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0})| > \varepsilon\right) + \Pr(\mathcal{A}_0^c) \leq \frac{\sqrt{n}}{\varepsilon}\underbrace{\mathrm{E}|\mathrm{E}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0}) - \widehat{V}_\infty^*|\mathbb{I}(\mathcal{A}_0)}_{\zeta_1} \\
+\ & \frac{n}{\varepsilon^2}\underbrace{\mathrm{E}\mathrm{Var}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}_0})}_{\zeta_2} + \Pr(\mathcal{A}_0^c),
\end{aligned}$$

where the first inequality follows by Bonferroni's inequality and the last inequality is due to Markov's inequality.

By Lemma S2.2, to prove $\widehat{V}_B^* - \widehat{V}_\infty^* = o_p(n^{-1/2})$, it suffices to show $\zeta_1 = o(n^{-1/2})$ and $\zeta_2 = o(n^{-1})$. We first show $\zeta_1 = o(n^{-1/2})$. It follows from triangle inequality that

$$\begin{aligned}
\zeta_1 \quad \leq \quad & \underbrace{\mathrm{E}\left|\frac{1}{n_{\mathcal{S}}}\sum_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h) - \frac{1}{n_{\mathcal{S}}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right|\mathbb{I}(\mathcal{A}_0)}_{\zeta_1^{(1)}} \\
+ \quad & \underbrace{\mathrm{E}\left|\frac{1}{n_{\mathcal{S}}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h) - \frac{1}{\binom{n}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right|\mathbb{I}(\mathcal{A}_0)}_{\zeta_1^{(2)}}
\end{aligned}$$

Recall that for any regime d and any $\mathcal{I}\subseteq\mathcal{I}_0$, $\widehat{V}_{\mathcal{I}^c}(d;\pi,h)$ is defined as

$$\frac{1}{|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}\left(\frac{\mathbb{I}\{A_i = d(\boldsymbol{X}_i)\}}{\pi(A_i,\boldsymbol{X}_i)}\{Y_i - h(A_i,\boldsymbol{X}_i)\} + h(d(\boldsymbol{X}_i),\boldsymbol{X}_i)\right).$$

By Condition (A1), (A3) and Lemma S2.2, we have

$$\begin{aligned}
& |\widehat{V}_{\mathcal{I}^c}(d;\pi,h)| \leq \frac{1}{|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}\left(\frac{1}{c^*}(|Y_i| + |h(A_i,\boldsymbol{X}_i)|) + |h(d(\boldsymbol{X}_i),\boldsymbol{X}_i)|\right) \qquad \text{(S2.26)}\\
\leq \quad & \frac{1}{|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}\left(\frac{1}{c^*}|Y_i| + \frac{C_0}{c^*} + C_0\right) = \frac{C_0}{c^*} + C_0 + \frac{1}{c^*|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}|A_iY_i + (1-A_i)Y_i| \\
= \quad & \frac{C_0}{c^*} + C_0 + \frac{1}{c^*|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}|A_iY_i^*(1) + (1-A_i)Y_i^*(0)| \\
\leq \quad & \frac{C_0}{c^*} + C_0 + \frac{1}{c^*|\mathcal{I}^c|}\sum_{i\in\mathcal{I}^c}(|Y_i^*(1)| + |Y_i^*(0)|) \leq \frac{C_0}{c^*} + C_0 + \frac{1}{c^*|\mathcal{I}^c|}\sum_{i=1}^{n}(|Y_i^*(1)| + |Y_i^*(0)|).
\end{aligned}$$

Combining this together with (S2.11) and the definition of $\mathcal{A}_0$, we have

$$\begin{aligned}
& \zeta_1^{(1)} = \mathrm{E}\left|\frac{1}{n_{\mathcal{S}}}\sum_{\substack{\mathcal{I}\in\mathcal{S}^c_{N_0,s_n}\\|\mathcal{I}|=s_n}} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right|\mathbb{I}(\mathcal{A}_0) \leq \frac{2\left(\binom{n}{s_n} - |\mathcal{S}_{N_0,s_n}|\right)}{\binom{n}{s_n}}\mathrm{E}\max_{\substack{\mathcal{I}\subseteq\mathcal{I}_0^c\\|\mathcal{I}|=s_n}}\left|\widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right|\mathbb{I}(\mathcal{A}_0) \\
\leq \quad & 2c_1p_*^{c_2n^{\beta_0}}\mathrm{E}\max_{\substack{\mathcal{I}\subseteq\mathcal{I}_0^c\\|\mathcal{I}|=s_n}}\left|\widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}};\pi,h)\right| \leq 2c_1p_*^{c_2n^{\beta_0}}\mathrm{E}\left(\frac{C_0}{c^*} + C_0 + \frac{1}{c^*(n-s_n)}\sum_{i=1}^{n}(|Y_i^*(1)| + |Y_i^*(0)|)\right) \\
= \quad & 2c_1p_*^{c_2n^{\beta_0}}\left(\frac{C_0}{c^*} + C_0 + \frac{n}{c^*(n-s_n)}\{\mathrm{E}|Y_0^*(0)| + \mathrm{E}|Y_0^*(1)|\}\right).
\end{aligned}$$

By Condition (A4) and Cauchy-Schwarz inequality, we have

$$\max_{a=0,1} \mathrm{E}|Y_0^*(a)| \leq \sqrt{\max_{a=0,1} \mathrm{E}\{Y_0^*(a)\}^2} = \sqrt{\sup_{\boldsymbol{x}\in\mathbb{X},a=0,1} \mathrm{E}[\{Y_0^*(a)\}^2|\boldsymbol{X}_0 = \boldsymbol{x}]} = O(1).$$

Besides, since $s_n = o(n)$, we have

$$n/(n - s_n) \leq 2, \tag{S2.27}$$

for sufficiently large n. Therefore, we have

$$\zeta_1^{(1)} \leq \bar{c}^* p_*^{c_2 n^{\beta_0}},$$

for some constant $\bar{c}^* > 0$. Since $0 < p_* < 1$, $0 < \beta_0 < 1$ and $c_2 > 0$, we have $\zeta_1^{(1)} = o(n^{-1/2})$. Similarly, we can show $\zeta_1^{(2)} = o(n^{-1/2})$. Therefore, we have $\zeta_1 = o(n^{-1/2})$.

Now we show $\zeta_2 = o(n^{-1})$. Recall that

$$\mathrm{Var}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}}) = \frac{n_{\mathcal{S}} - 1}{n_{\mathcal{S}} B} \widehat{s.e}^2 \left(\left\{ \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}}; \pi, h) \right\}_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}} \right),$$

we have

$$\mathrm{Var}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}}) \leq \frac{1}{n_{\mathcal{S}} B} \sum_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}} \widehat{V}_{\mathcal{I}^c}^2(\widehat{d}_{\mathcal{I}}; \pi, h). \tag{S2.28}$$

Besides, it follows from (S2.26) and Cauchy-Schwarz inequality that

$$\begin{aligned}
|\widehat{V}_{\mathcal{I}^c}(d; \pi, h)|^2 &\leq \left(\frac{C_0}{c^*} + C_0 + \frac{1}{c^*|\mathcal{I}^c|} \sum_{i=1}^n (|Y_i^*(1)| + |Y_i^*(0)|) \right)^2 \\
&\leq \frac{3C_0^2}{(c^*)^2} + 3C_0^2 + \frac{3}{(c^*)^2|\mathcal{I}^c|^2} \left(\sum_{i=1}^n (|Y_i^*(1)| + |Y_i^*(0)|) \right)^2 \\
&\leq \frac{3C_0^2}{(c^*)^2} + 3C_0^2 + \frac{6n}{(c^*)^2|\mathcal{I}^c|^2} \sum_{i=1}^n \{|Y_i^*(1)|^2 + |Y_i^*(0)|^2\}.
\end{aligned}$$

Combining this together with (S2.28) yields

$$\begin{aligned}
\zeta_2 = \mathrm{E}\mathrm{Var}(\widehat{V}_B^*|\{O_i\}_{i\in\mathcal{I}}) &\leq \mathrm{E}\frac{1}{n_{\mathcal{S}} B} \sum_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}} \left(\frac{3C_0^2}{(c^*)^2} + 3C_0^2 + \frac{6n}{(c^*)^2|\mathcal{I}^c|^2} \sum_{i=1}^n \{|Y_i^*(1)|^2 + |Y_i^*(0)|^2\} \right) \\
&\leq \frac{1}{B} \mathrm{E} \max_{\substack{\mathcal{I}\subseteq\mathcal{I}_0 \\ |\mathcal{I}|=s_n}} \left(\frac{3C_0^2}{(c^*)^2} + 3C_0^2 + \frac{6n}{(c^*)^2|\mathcal{I}^c|^2} \sum_{i=1}^n \{|Y_i^*(1)|^2 + |Y_i^*(0)|^2\} \right) \\
&\leq \frac{1}{B} \left(\frac{3C_0^2}{(c^*)^2} + 3C_0^2 + \frac{6n^2}{(c^*)^2|n - s_n|^2} \mathrm{E}\{|Y_0^*(1)|^2 + |Y_0^*(0)|^2\} \right).
\end{aligned}$$

Notice that we require $B \gg n$. It follows from (S2.27) and Condition (A4) that

$$\zeta_2 = O(B^{-1}) = o(n^{-1}).$$

Therefore, we've shown $\widehat{V}_B^* - \widehat{V}_\infty^* = o_p(n^{-1/2})$.

Step 3: Recall that $\widehat{\sigma}_B^2$ is defined as

$$\widehat{\sigma}_B^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 - \frac{n}{n-1}(\overline{V})^2,$$

where

$$\widehat{V}^{(i)} = \frac{1}{n^{(i)}}\sum_{b=1}^{B}\left(\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(1)}},\widehat{h}_{\mathcal{I}_b^{(1)}})\mathbb{I}(i \notin \mathcal{I}_b^{(1)}) + \widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(2)}},\widehat{h}_{\mathcal{I}_b^{(2)}})\mathbb{I}(i \notin \mathcal{I}_b^{(2)})\right),$$

and $\overline{V} = \sum_{i=1}^n \widehat{V}^{(i)}/n$. Let $\mathcal{A}_i$ denote the event defined in (S2.4). Since $s_n = o(n)$, when $\mathcal{A}_i$ holds, we have for sufficiently large n,

$$n^{(i)} \geq \frac{B}{2}. \tag{S2.29}$$

For any $i \in \mathcal{I}_0$, define

$$\widehat{V}^{*(i)} = \frac{1}{n^{(i)}}\sum_{b=1}^{B}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h)\mathbb{I}(i \notin \mathcal{I}_b).$$

Below, we first show

$$\max_{i\in\mathcal{I}_0}\mathrm{E}|\widehat{V}^{(i)} - \widehat{V}^{*(i)}|^2\mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) = o(1). \tag{S2.30}$$

By definition, we have

$$\left|\widehat{V}^{(i)} - \widehat{V}^{*(i)}\right| \leq \frac{1}{n^{(i)}}\sum_{j=1}^{2}\sum_{b=1}^{B}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h) - \widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(j)}},\widehat{h}_{\mathcal{I}_b^{(j)}})\right|\mathbb{I}(i \notin \mathcal{I}_b^{(j)}).$$

It follows from Cauchy-Schwarz inequality that

$$\left|\widehat{V}^{(i)} - \widehat{V}^{*(i)}\right|^2 \leq \frac{2B}{(n^{(i)})^2}\sum_{j=1}^{2}\sum_{b=1}^{B}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h) - \widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(j)}},\widehat{h}_{\mathcal{I}_b^{(j)}})\right|^2\mathbb{I}(i \notin \mathcal{I}_b^{(j)}).$$

Under the event defined in $\mathcal{A}_i$, it follows from (S2.29) that

$$\left|\widehat{V}^{(i)} - \widehat{V}^{*(i)}\right|^2 \leq \frac{8}{B}\sum_{j=1}^{2}\sum_{b=1}^{B}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h) - \widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(j)}},\widehat{h}_{\mathcal{I}_b^{(j)}})\right|^2\mathbb{I}(i \notin \mathcal{I}_b^{(j)}).$$

Therefore, we have

$$\mathrm{E}\left|\widehat{V}^{(i)}-\widehat{V}^{*(i)}\right|^2\mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_0)\le\frac{8}{B}\sum_{j=1}^{2}\sum_{b=1}^{B}\mathrm{E}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\widehat{\pi}_{\mathcal{I}_b^{(j)}},\widehat{h}_{\mathcal{I}_b^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}_b^{(j)})\mathbb{I}(\mathcal{A}_0).$$

For any $\mathcal{I}\subseteq\mathcal{I}_0$ with $|\mathcal{I}|=s_n$, notice that either $i\notin\mathcal{I}^{(1)}$ or $i\notin\mathcal{I}^{(2)}$ implies $i\notin\mathcal{I}$. Similar to (S2.12)-(S2.15), we can show

$$\begin{aligned}
\mathrm{E}\left|\widehat{V}^{(i)}-\widehat{V}^{*(i)}\right|^2\mathbb{I}(\mathcal{A}_i\cap\mathcal{A}_0)&\le 8\sum_{j=1}^{2}\mathrm{E}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_1};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_1};\widehat{\pi}_{\mathcal{I}_1^{(j)}},\widehat{h}_{\mathcal{I}_1^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}_1^{(j)})\mathbb{I}(\mathcal{A}_0)\\
&\le\frac{8}{\mathcal{P}_0|\mathcal{S}_{N_0,s_n}|}\sum_{j=1}^{2}\sum_{\substack{\mathcal{I}\subseteq\mathcal{S}_{N_0,s_n},i\notin\mathcal{I}\\(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}}\mathrm{E}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\widehat{\pi}_{\mathcal{I}^{(j)}},\widehat{h}_{\mathcal{I}^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}^{(j)})\mathbb{I}(\mathcal{A}_0)\\
&\le\frac{16}{\mathcal{P}_0\binom{n}{s_n}}\sum_{j=1}^{2}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)},|\mathcal{I}|=s_n\\(\mathcal{I}^{c(1)},\mathcal{I}^{c(2)})\in\mathcal{P}(\mathcal{I})}}\mathrm{E}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\widehat{\pi}_{\mathcal{I}^{(j)}},\widehat{h}_{\mathcal{I}^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}^{(j)})\\
&\le 16\sum_{j=1}^{2}\mathrm{E}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(j)}},\widehat{h}_{\mathcal{I}_*^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}_*^{(j)}),
\end{aligned}$$

where $\mathcal{I}_*$ denotes a random subset uniformly sampled from $\{\mathcal{I}:\mathcal{I}\subseteq\mathcal{I}_{(-i)},|\mathcal{I}|=s_n\}$, independent of the data $\{O_i\}_{i\in\mathcal{I}_0}$, and $(\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)})$ denote the random partition of $\mathcal{I}_*^c$ generated our algorithm.

Given $(\mathcal{I}_*,\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)})$, the indicator $\mathbb{I}(i\notin\mathcal{I}_*^{(j)})$ fixed. If $i\in\mathcal{I}_*^{(j)}$, then we have

$$\mathrm{E}^{\mathcal{I}_*,\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)}}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(j)}},\widehat{h}_{\mathcal{I}_*^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}_*^{(j)})=0,$$

where $\mathrm{E}^{\mathcal{I}_*,\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)}}$ denotes the conditional expectation given $\mathcal{I}_*$, $\mathcal{I}_*^{c(1)}$ and $\mathcal{I}_*^{c(2)}$. Otherwise, using similar arguments in bounding $|\eta_5|$ in Step 1, we can show

$$\sup_{\substack{i\in\mathcal{I}_0,j=1,2\\\mathcal{I}_*\subseteq\mathcal{I}_{(-i)},|\mathcal{I}^*|=s_n\\(\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)})\in\mathcal{P}(\mathcal{I}_*)}}\mathrm{E}^{\mathcal{I}_*,\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)}}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(j)}},\widehat{h}_{\mathcal{I}_*^{(j)}})\right|^2=o(1).$$

Therefore, we have

$$\sup_{\substack{i\in\mathcal{I}_0,j=1,2\\\mathcal{I}_*\subseteq\mathcal{I}_{(-i)},|\mathcal{I}^*|=s_n\\(\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)})\in\mathcal{P}(\mathcal{I}_*)}}\mathrm{E}^{\mathcal{I}_*,\mathcal{I}_*^{c(1)},\mathcal{I}_*^{c(2)}}\left|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\pi,h)-\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(j)}},\widehat{h}_{\mathcal{I}_*^{(j)}})\right|^2\mathbb{I}(i\notin\mathcal{I}_*^{(j)})=o(1).$$

This implies (S2.30) holds.

Similar to (S2.26), we have for any $i \in \mathcal{I}_0$ and any $\mathcal{I} \subseteq \mathcal{I}_0$,

$$|\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}}; \pi, h)| \leq \frac{C_0}{c^*} + C_0 + \frac{1}{c^*}\{|Y_i^*(0)| + |Y_i^*(1)|\}.$$

Therefore, we have for any $i \in \mathcal{I}$,

$$\begin{aligned} |\widehat{V}^{*(i)}| &\leq \left|\frac{C_0}{c^*} + C_0 + \frac{1}{c^*}\{|Y_i^*(0)| + |Y_i^*(1)|\}\right| \left(\frac{1}{n^{(i)}} \sum_{b=1}^{B} \mathbb{I}(i \notin \mathcal{I}_b)\right) \\ &= \left|\frac{C_0}{c^*} + C_0 + \frac{1}{c^*}\{|Y_i^*(0)| + |Y_i^*(1)|\}\right|. \end{aligned}$$

By Condition (A4) and Cauchy-Schwarz inequality, we have

$$\max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{*(i)}|^2 \leq \frac{4C_0^2}{(c^*)^2} + 4C_0^2 + \frac{\mathrm{E}(4|Y_i^*(0)|^2 + 4|Y_i^*(1)|^2)}{(\bar{c}^*)^2} = O(1). \tag{S2.31}$$

This together with (S2.30) yields

$$\begin{aligned} &\max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} + \widehat{V}^{*(i)}|^2 \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) = \max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} - \widehat{V}^{*(i)} + 2\widehat{V}^{*(i)}|^2 \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) \\ \leq\ & 2\max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} - \widehat{V}^{*(i)}|^2 \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) + 8\max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{*(i)}|^2 = O(1). \end{aligned} \tag{S2.32}$$

In view of (S2.30) and (S2.32), it follows from Cauchy-Schwarz inequality that

$$\begin{aligned} &\max_{i \in \mathbb{I}_0} \mathrm{E}|(\widehat{V}^{(i)})^2 - (\widehat{V}^{*(i)})^2| \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) = \max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} - \widehat{V}^{*(i)}||\widehat{V}^{(i)} + \widehat{V}^{*(i)}| \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) \\ &\leq \sqrt{\max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} + \widehat{V}^{*(i)}|^2 \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) \max_{i \in \mathbb{I}_0} \mathrm{E}|\widehat{V}^{(i)} - \widehat{V}^{*(i)}|^2 \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0)} = o(1). \end{aligned}$$

Hence,

$$\begin{aligned} &\mathrm{E}\left|\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{*(i)}\}^2\right| \mathbb{I}(\mathcal{A}_0)\mathbb{I}(\cap_{i=1}^{n}\mathcal{A}^{(i)}) \\ \leq\ & \frac{n}{n-1}\max_{i \in \mathbb{I}_0} \mathrm{E}|(\widehat{V}^{(i)})^2 - (\widehat{V}^{*(i)})^2| \mathbb{I}(\mathcal{A}^{(i)} \cap \mathcal{A}_0) = o(1). \end{aligned} \tag{S2.33}$$

Notice that $B \gg n$. By Lemma S2.2 and Bonferroni's inequality, we have

$$\begin{aligned} \Pr\left\{\mathcal{A}_0^c \cup \left(\cup_{i=1}^{n} \mathcal{A}^{(i)^c}\right)\right\} &\leq \Pr(\mathcal{A}_0^c) + \sum_{i=1}^{n} \Pr(\mathcal{A}^{(i)^c}) \\ &\leq \frac{n}{B} + 4n\exp(-c_4 n) + 2\exp(-c_3 n) \to 0. \end{aligned} \tag{S2.34}$$

This together with (S2.33) implies that

$$\Pr\left(\left|\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{*(i)}\}^2\right| > \varepsilon\right) \le \Pr\left\{\mathcal{A}_0^c \cup \left(\cup_{i=1}^n \mathcal{A}^{(i)^c}\right)\right\}$$
$$+ \quad \frac{1}{\varepsilon}\mathrm{E}\left|\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{*(i)}\}^2\right| \mathbb{I}(\mathcal{A}_0)\mathbb{I}(\cap_{i=1}^n \mathcal{A}^{(i)}) \to 0,$$

for any $\varepsilon > 0$. Therefore, we've shown

$$\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{*(i)}\}^2 + o_p(1). \tag{S2.35}$$

Conditional on the event defined in $\mathcal{A}^{(i)}$, we have

$$\underbrace{\frac{n-s_n}{n} - \frac{\sqrt{\log n}}{\sqrt{n}}}_{p_L} \le \frac{n^{(i)}}{B} \le \underbrace{\frac{n-s_n}{n} + \frac{\sqrt{\log n}}{\sqrt{n}}}_{p_U}.$$

Let

$$\widehat{V}_L^{(i)} = \frac{1}{Bp_U}\sum_{b=1}^{B}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h)\mathbb{I}(i \notin \mathcal{I}_b) \text{ and } \widehat{V}_U^{(i)} = \frac{1}{Bp_L}\sum_{b=1}^{B}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}_b};\pi,h)\mathbb{I}(i \notin \mathcal{I}_b),$$

we have

$$\{\widehat{V}_L^{(i)}\}^2 \le \{\widehat{V}^{*(i)}\}^2 \le \{\widehat{V}_U^{(i)}\}^2.$$

Define

$$\widehat{V}_\infty^{*(i)} = \frac{1}{\binom{n-1}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\mathbb{I}(i\notin\mathcal{I}) = \frac{1}{\binom{n-1}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h).$$

Below, we show

$$\max_{i\in\mathcal{I}_0}\mathrm{E}|\{\widehat{V}_\infty^{*(i)}\}^2 - \{\widehat{V}^{*(i)}\}^2|\mathbb{I}(\mathcal{A}_0\cap\mathcal{A}^{(i)}) = o(1). \tag{S2.36}$$

To prove this, it suffices to show

$$\max_{i\in\mathcal{I}_0}\mathrm{E}|\{\widehat{V}_\infty^{*(i)}\}^2 - \{\widehat{V}_L^{*(i)}\}^2|\mathbb{I}(\mathcal{A}_0\cap\mathcal{A}^{(i)}) = o(1), \tag{S2.37}$$

and

$$\max_{i\in\mathcal{I}_0}\mathrm{E}|\{\widehat{V}_{\infty}^{*(i)}\}^2-\{\widehat{V}_{U}^{*(i)}\}^2|\mathbb{I}(\mathcal{A}_0\cap\mathcal{A}^{(i)})=o(1), \tag{S2.38}$$

We first show (S2.37). Notice that the mean and variance of $\widehat{V}_L^{*(i)}$ conditional on $\{O_j\}_{j\in\mathcal{I}_0}$ are given by

$$\begin{aligned}
\mathrm{E}\left(\widehat{V}_L^{*(i)}|\{O_j\}_{j\in\mathcal{I}_0}\right) &= \frac{1}{p_U|\mathcal{S}_{N_0,s_n}|}\sum_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\mathbb{I}(i\notin\mathcal{I}),\\
\mathrm{Var}\left(\widehat{V}_L^{*(i)}|\{O_j\}_{j\in\mathcal{I}_0}\right) &= \frac{|\mathcal{S}_{N_0,s_n}|-1}{Bp_U^2|\mathcal{S}_{N_0,s_n}|}\widehat{s.e}^2\left(\left\{\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\mathbb{I}(i\notin\mathcal{I})\right\}_{\mathcal{I}\in\mathcal{S}_{N_0,s_n}}\right).
\end{aligned}$$

Using similar arguments in bounding ζ_1 and ζ_2 in Step 2, we can show

$$\begin{aligned}
&\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_L^{*(i)}-\frac{1}{p_U\binom{n}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\right|^2\mathbb{I}(\mathcal{A}_0)\\
=\;&\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_L^{*(i)}-\frac{1}{p_U\binom{n}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\mathbb{I}(i\notin\mathcal{I})\right|^2\mathbb{I}(\mathcal{A}_0)\\
\le\;&\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\mathrm{E}\left(\widehat{V}_L^{*(i)}|\{O_j\}_{j\in\mathcal{I}_0}\right)-\frac{1}{p_U\binom{n}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}}\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}};\pi,h)\mathbb{I}(i\notin\mathcal{I})\right|^2\mathbb{I}(\mathcal{A}_0)\\
+\;&\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\mathrm{Var}\left(\widehat{V}_L^{*(i)}|\{O_j\}_{j\in\mathcal{I}_0}\right)\right|^2\mathbb{I}(\mathcal{A}_0)=o(1).
\end{aligned}$$

By the definition of $\widehat{V}_{\infty}^{*(i)}$, this implies

$$\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_L^{*(i)}-\frac{(n-s_n)}{np_U}\widehat{V}_{\infty}^{*(i)}\right|^2\mathbb{I}(\mathcal{A}_0)=o(1). \tag{S2.39}$$

Besides, similar to (S2.31), we can show

$$\max_{i\in\mathcal{I}_0}\mathrm{E}|\widehat{V}_{\infty}^{*(i)}|=O(1). \tag{S2.40}$$

Therefore, by the condition $s_n=o(n)$, we have

$$\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_{\infty}^{*(i)}-\frac{n-s_n}{np_U}\widehat{V}_{\infty}^{*(i)}\right|^2\le\frac{(\log n)/n}{p_U^2}\mathrm{E}|\widehat{V}_{\infty}^{*(i)}|=o(1).$$

This together with (S2.39) yields

$$\max_{i\in\mathcal{I}_0} \mathrm{E}\left|\widehat{V}_L^{*(i)} - \widehat{V}_\infty^{*(i)}\right|^2 \mathbb{I}(\mathcal{A}_0) = o(1). \tag{S2.41}$$

In addition, combining (S2.31) with (S2.40) yields

$$\max_{i\in\mathcal{I}_0} \mathrm{E}\left|\widehat{V}_L^{*(i)} + \widehat{V}_\infty^{*(i)}\right|^2 = O(1).$$

This together with (S2.41) gives

$$\begin{aligned} &\max_{i\in\mathcal{I}_0} \mathrm{E}\left|\{\widehat{V}_L^{*(i)}\}^2 - \{\widehat{V}_\infty^{*(i)}\}^2\right| \mathbb{I}(\mathcal{A}_0) = \max_{i\in\mathcal{I}_0} \mathrm{E}\left|\widehat{V}_L^{*(i)} + \widehat{V}_\infty^{*(i)}\right| \left|\widehat{V}_L^{*(i)} - \widehat{V}_\infty^{*(i)}\right| \mathbb{I}(\mathcal{A}_0) \\ \leq\ & \sqrt{\mathrm{E}\left|\widehat{V}_L^{*(i)} - \widehat{V}_\infty^{*(i)}\right|^2 \mathbb{I}(\mathcal{A}_0)} \sqrt{\mathrm{E}\left|\widehat{V}_L^{*(i)} + \widehat{V}_\infty^{*(i)}\right|^2} = o(1). \end{aligned}$$

Hence, we've shown (S2.37) holds. Similarly, we can show (S2.38) holds. This proves (S2.36). Therefore, we have

$$\mathrm{E}\left|\frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}_\infty^{*(i)}\}^2\right| \mathbb{I}\left(\mathcal{A}_0 \cap (\cap_{j=1}^n \mathcal{A}^{(j)})\right) = o(1).$$

It thus follows from (S2.34) and Markov's inequality that

$$\begin{aligned} \Pr\left(\left|\frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}_\infty^{*(i)}\}^2\right| > \varepsilon\right) &\leq \Pr\left(\mathcal{A}_0^c \cup (\cup_{j=1}^n \mathcal{A}^{(i)^c})\right) \\ +\mathrm{E}\left|\frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}_\infty^{*(i)}\}^2\right| &\mathbb{I}\left(\mathcal{A}_0 \cap (\cap_{j=1}^n \mathcal{A}^{(j)})\right) \to 0, \end{aligned}$$

for any $\varepsilon > 0$. This implies

$$\frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^n \{\widehat{V}_\infty^{*(i)}\}^2 = o_p(1). \tag{S2.42}$$

Notice that we have $\widehat{V}_\infty^{*(i)} = \widehat{V}_{\{i\}}(\widehat{d}_{s_n}^{(-i)}; \pi, h)$. Based on the ANOVA decomposition (see (26)), we have

$$\widehat{d}_{s_n}^{(-i)}(\cdot) = p_{s_n}(\cdot) + \sum_{k=1}^{s_n} \frac{\binom{n-1-k}{s_n-k}}{\binom{n-1}{s_n}} \sum_{\{j_1,\ldots,j_k\}\subseteq\mathcal{I}_{(-i)}} d_{s_n,k}(O_{j_1},\ldots,O_{j_k};\cdot).$$

Using similar arguments in bounding $\eta_3^{(2)}$ and $\eta_3^{(3)}$ in the proof of Theorem 2.1, we can show

$$\max_{i\in\mathcal{I}_0} \mathrm{E}|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2 = o(1). \tag{S2.43}$$

By Condition (A1), (A3), (A4), Lemma S2.2 and Cauchy-Schwarz inequality, we have

$$
\begin{aligned}
& \mathrm{E}|\widehat{V}_{\{i\}}(\widehat{d}_{s_n}^{(-i)};\pi,h) - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)|^2 \\
\le\ & 4\mathrm{E}\frac{A_i|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2}{\pi^2(1,\boldsymbol{X}_i)}|Y_i^*(1) - h(1,\boldsymbol{X}_i)|^2 \\
+\ & 4\mathrm{E}\frac{(1-A_i)|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2}{\pi^2(0,\boldsymbol{X}_i)}|Y_i^*(0) - h(0,\boldsymbol{X}_i)|^2 \\
+\ & 4\mathrm{E}h^2(0,\boldsymbol{X}_i)|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2 + 4\mathrm{E}h^2(1,\boldsymbol{X}_i)|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2 \\
\le\ & \max_{a=0,1}\frac{8}{c^*}\mathrm{E}\{\mathrm{E}^{\boldsymbol{X}_i}|Y_i^*(a) - h(a,\boldsymbol{X}_i)|^2\}|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2 \\
+\ & 8\mathrm{E}|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2 \le O(1)\mathrm{E}|\widehat{d}_{s_n}^{(-i)}(\boldsymbol{X}_i) - p_{s_n}(\boldsymbol{X}_i)|^2,
\end{aligned}
$$

where $O(1)$ denotes a universal constant independent of i. This together with (S2.43) yields

$$
\max_{i\in\mathcal{I}_0}\mathrm{E}|\widehat{V}_{\{i\}}(\widehat{d}_{s_n}^{(-i)};\pi,h) - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)|^2 = o(1). \tag{S2.44}
$$

By Markov's inequality, we obtain

$$
\frac{1}{n-1}\sum_{i=1}^n\{\widehat{V}_\infty^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^n\widehat{V}_{\{i\}}^2(p_{s_n};\pi,h) = o_p(1).
$$

Combining this with (S2.35) and (S2.42), we've shown

$$
\frac{1}{n-1}\sum_{i=1}^n\{\widehat{V}^{(i)}\}^2 = \frac{1}{n-1}\sum_{i=1}^n\widehat{V}_{\{i\}}^2(p_{s_n};\pi,h) + o_p(1). \tag{S2.45}
$$

In addition, it follows from Cauchy-Schwarz inequality that

$$
\begin{aligned}
& \mathrm{E}\left|\left(\frac{1}{n}\sum_{i=1}^n\widehat{V}^{(i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n\widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)^2\right|\mathcal{I}\left(\mathcal{A}_0\cap(\cap_{i=1}^n\mathcal{A}^{(i)})\right) \\
=\ & \mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n\left(\widehat{V}^{(i)} - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)\right|\left|\frac{1}{n}\sum_{i=1}^n\left(\widehat{V}^{(i)} + \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)\right|\mathcal{I}\left(\mathcal{A}_0\cap(\cap_{i=1}^n\mathcal{A}^{(i)})\right) \\
\le\ & \sqrt{\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n\left(\widehat{V}^{(i)} - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)\right|^2\mathcal{I}\left(\mathcal{A}_0\cap(\cap_{i=1}^n\mathcal{A}^{(i)})\right)}\sqrt{\mathrm{E}\left|\frac{1}{n}\sum_{i=1}^n\left(\widehat{V}^{(i)} + \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)\right|^2} \\
\le\ & \sqrt{\frac{1}{n}\mathrm{E}\sum_{i=1}^n\left|\widehat{V}^{(i)} - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2\mathbb{I}(\mathcal{A}_0\cap\mathcal{A}^{(i)})}\sqrt{\frac{1}{n}\mathrm{E}\sum_{i=1}^n\left|\widehat{V}^{(i)} + \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2} \\
\le\ & \sqrt{\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}^{(i)} - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2\mathbb{I}(\mathcal{A}_0\cap\mathcal{A}^{(i)})}\sqrt{\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}^{(i)} + \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2}.
\end{aligned}
$$

By (S2.33), (S2.36) and (S2.44), we have

$$\max_{i\in\mathcal{I}_0} \mathrm{E}\left|\widehat{V}^{(i)} - \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2 \mathbb{I}(\mathcal{A}_0 \cap \mathcal{A}^{(i)}) = o(1).$$

Besides, similar to (S2.32), we can show

$$\max_{i\in\mathcal{I}_0} \mathrm{E}\left|\widehat{V}^{(i)} + \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right|^2 = O(1).$$

This implies

$$\mathrm{E}\left|\left(\frac{1}{n}\sum_{i=1}^n \widehat{V}^{(i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)^2\right| \mathcal{I}\left(\mathcal{A}_0 \cap (\cap_{i=1}^n \mathcal{A}^{(i)})\right) = o(1).$$

By Markov's inequality, we have

$$\begin{aligned}\Pr\left(\left|\left(\frac{1}{n}\sum_{i=1}^n \widehat{V}^{(i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)^2\right| > \varepsilon\right) &\le \Pr\left(\mathcal{A}_0^c \cup (\cup_{j=1}^n \mathcal{A}^{(i)c})\right)\\ &+\mathrm{E}\left|\left(\frac{1}{n}\sum_{i=1}^n \widehat{V}^{(i)}\right)^2 - \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)^2\right| \mathbb{I}\left(\mathcal{A}_0 \cap (\cap_{j=1}^n \mathcal{A}^{(j)})\right) \to 0,\end{aligned}$$

for any $\varepsilon > 0$. Therefore,

$$\left(\frac{1}{n}\sum_{i=1}^n \widehat{V}^{(i)}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^n \widehat{V}_{\{i\}}(p_{s_n};\pi,h)\right)^2 + o_p(1) = \eta_1^2 + o_p(1).$$

Combining this together with (S2.45), we have

$$\widehat{\sigma}_B^2 = \frac{1}{n-1}\sum_{i=1}^n \widehat{V}_{\{i\}}^2(p_{s_n};\pi,h) - \frac{n}{n-1}\eta_1^2 + o_p(1).$$

Under the given conditions, it follows from law of larger numbers that

$$\frac{1}{n}\sum_{i=1}^n \widehat{V}_{\{i\}}^2(p_{s_n};\pi,h) = \mathrm{E}\left(\frac{\mathrm{g}\{A_0, p_{s_n}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}\{Y_0 - h(A_0,\boldsymbol{X}_0)\} + h\{p_{s_n}(\boldsymbol{X}_0),\boldsymbol{X}_0\}\right)^2 + o_p(1),$$

and

$$\eta_1 = \mathrm{E}\left(\frac{\mathrm{g}\{A_0, p_{s_n}(\boldsymbol{X}_0)\}}{\pi(A_0,\boldsymbol{X}_0)}\{Y_0 - h(A_0,\boldsymbol{X}_0)\} + h\{p_{s_n}(\boldsymbol{X}_0),\boldsymbol{X}_0\}\right) + o_p(1).$$

Therefore, we have $\widehat{\sigma}_B^2 = \sigma_{s_n}^2 + o_p(1)$. The proof is hence completed.

S2.3 Proof of Theorem 3.1

Let $\nu_0 = \Pr\{\tau(\boldsymbol{X}_0) = 0\}$ and $d_0 = d(\boldsymbol{X}_0)$ for any function d. Since e_0 is independent of A_0 and $\boldsymbol{X}_0$, we have

$$\widetilde{\sigma}_0^2(\widehat{d}_{\mathcal{I}_{(j)}}; \pi, h) = \mathrm{Var}\left(\frac{\mathrm{g}(A_0, \widehat{d}_{\mathcal{I}_{(j)},0})}{\pi(A_0, \boldsymbol{X}_0)} e_0 + h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) \middle| \{O_i\}_{i\in\mathcal{I}_{(j)}} \right) \tag{S2.46}$$

$$= \mathrm{Var}\left(\frac{\mathrm{g}(A_0, \widehat{d}_{\mathcal{I}_{(j)},0})}{\pi(A_0, \boldsymbol{X}_0)} e_0 \middle| \{O_i\}_{i\in\mathcal{I}_{(j)}} \right) + \mathrm{Var}\{h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) | \{O_i\}_{i\in\mathcal{I}_{(j)}}\}.$$

By definition, we have

$$h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) = \widehat{d}_{\mathcal{I}_{(j)},0} h(1, \boldsymbol{X}_0) + (1 - \widehat{d}_{\mathcal{I}_{(j)},0}) h(0, \boldsymbol{X}_0) = h(0, \boldsymbol{X}_0) + \tau(\boldsymbol{X}_0)\mathbb{I}\{\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) > 0\}.$$

For any $d^{opt} \in \mathcal{D}^{opt}$ and $\varepsilon > 0$, it follows from Markov's inequality that

$$\begin{aligned} & \Pr\left\{ \mathrm{E}\left(|h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)|^2 | \{O_i\}_{i\in\mathcal{I}_{(j)}} \right) > \varepsilon \right\} \\ \leq\ & \frac{1}{\varepsilon^2} \mathrm{E}|h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)|^2 = \frac{1}{\varepsilon^2} \mathrm{E}\tau^2(\boldsymbol{X}_0) |\mathbb{I}\{\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) > 0\} - \mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\}|. \end{aligned}$$

Here, $h(d_0^{opt}, \boldsymbol{X}_0) = h(0, \boldsymbol{X}_0) + \max\{\tau(\boldsymbol{X}_0), 0\}$ is independent of d^{opt}.

Since $|\mathbb{I}\{\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) > 0\} - \mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\}| \leq \mathbb{I}\{|\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)| \geq \tau(\boldsymbol{X}_0)\}$, by Condition (A6) and Markov's inequality, we have

$$\begin{aligned} \mathrm{E}\tau^2(\boldsymbol{X}_0)|\mathbb{I}\{\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) > 0\} - \mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\}| &\leq \mathrm{E}\mathbb{I}\{|\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)| \geq \tau(\boldsymbol{X}_0)\} \\ &\leq \mathrm{E}|\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0) - \tau(\boldsymbol{X}_0)|^2 \to 0, \end{aligned}$$

as $j \to \infty$. This implies that

$$\mathrm{E}\left(|h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)|^2 | \{O_i\}_{i\in\mathcal{I}_{(j)}} \right) \xrightarrow{P} 0, \qquad \text{as } j \to \infty.$$

Let $\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}$ denote the conditional expectation given $\{O_i\}_{i\in\mathcal{I}_{(j)}}$, it follows from Jensen's inequality and Cacuhy-Schwarz inequality, we have as $j \to \infty$,

$$\begin{aligned} & \mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}} |h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0) - \mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)\}|^2 \\ \leq\ & 2\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}} |h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)|^2 + 2|\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)\}|^2 \\ \leq\ & 4\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}} |h(\widehat{d}_{\mathcal{I}_{(j)},0}, \boldsymbol{X}_0) - h(d_0^{opt}, \boldsymbol{X}_0)|^2 = o_p(1). \end{aligned} \tag{S2.47}$$

Moreover, it follows from Lemma S2.1 that

$$|h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)+h(d_0^{opt},\boldsymbol{X}_0)-\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)-h(d_0^{opt},\boldsymbol{X}_0)\}|\leq 8C_0.$$

This together with (S2.47) yields

$$\begin{aligned}
&\left|\mathrm{Var}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)|\{O_i\}_{i\in\mathcal{I}_{(j)}}\}-\mathrm{Var}\{h(d_0^{opt},\boldsymbol{X}_0)\}\right|\\
=\;& \mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\left|h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)-h(d_0^{opt},\boldsymbol{X}_0)-[\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)-h(d_0^{opt},\boldsymbol{X}_0)\}]\right|\\
\times\;& \left|h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)+h(d_0^{opt},\boldsymbol{X}_0)-[\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)+h(d_0^{opt},\boldsymbol{X}_0)\}]\right|\\
\leq\;& 8C_0\sqrt{\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\left|h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)-h(d_0^{opt},\boldsymbol{X}_0)-[\mathrm{E}^{\{O_i\}_{i\in\mathcal{I}_{(j)}}}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)-h(d_0^{opt},\boldsymbol{X}_0)\}]\right|^2}\\
=\;& o_p(1),
\end{aligned}$$

as $j\to\infty$. By Lemma (S2.1), $|\mathrm{Var}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)|\{O_i\}_{i\in\mathcal{I}_{(j)}}\}-\mathrm{Var}\{h(d_0^{opt},\boldsymbol{X}_0)\}|$ is uniformly bounded. Thus, we have as $j\to\infty$,

$$\mathrm{E}\left|\mathrm{Var}\{h(\widehat{d}_{\mathcal{I}_{(j)},0},\boldsymbol{X}_0)|\{O_i\}_{i\in\mathcal{I}_{(j)}}\}-\mathrm{Var}\{h(d_0^{opt},\boldsymbol{X}_0)\}\right|=o(1). \tag{S2.48}$$

With some calculations, we have

$$\begin{aligned}
&\mathrm{Var}\left(\frac{\mathrm{g}(A_0,\widehat{d}_{\mathcal{I}_{(j)},0})}{\pi(A_0,\boldsymbol{X}_0)}e_0\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)=\sigma_0^2\mathrm{E}\left(\frac{\mathrm{g}^2\{A_0,\widehat{d}_{\mathcal{I}_{(j)},0}\}}{\pi^2(A_0,\boldsymbol{X}_0)}\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right) \qquad (\text{S2.49})\\
=\;&\sigma_0^2\mathrm{E}\left(\frac{\widehat{d}^2_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{(1-\widehat{d}_{\mathcal{I}_{(j)},0})^2}{\pi(0,\boldsymbol{X}_0)}\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)=\sigma_0^2\mathrm{E}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(0,\boldsymbol{X}_0)}\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right),
\end{aligned}$$

where the last equality is due to that $\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\in\{0,1\}$.

In the following, we show

$$\lim_j \mathrm{E}\left|\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}\right|=0. \tag{S2.50}$$

Notice that

$$
\begin{aligned}
& \mathrm{E}\left|\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}\right| \qquad \text{(S2.51)}\\
\leq\ & \underbrace{\mathrm{E}\left|\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}\right|\mathbb{I}\{0<\tau(\boldsymbol{X}_0)\leq j^{-1/4}\}}_{\zeta_3}\\
+\ & \underbrace{\mathrm{E}\left|\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}\right|\mathbb{I}\{\tau(\boldsymbol{X}_0)>j^{-1/4}\}}_{\zeta_4}.
\end{aligned}
$$

It follows from Condition (A3) and (A5) that

$$
\zeta_3 \leq \frac{1}{c_0}\mathrm{E}\mathbb{I}\{0<\tau(\boldsymbol{X}_0)\leq j^{-1/4}\} \leq \frac{\bar{c}}{c_0}j^{-1/(4\alpha)} \to 0, \qquad \text{as } \ j\to\infty.
$$

In addition, similar to (32), we have

$$
\begin{aligned}
\zeta_4 \leq \frac{1}{c_0}\mathrm{E}|\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)-\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}|\mathbb{I}\{\tau(\boldsymbol{X}_0)>j^{-1/4}\} &\leq \frac{1}{c_0}\mathrm{E}\frac{|\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)-\tau(\boldsymbol{X}_0)|^2}{|\tau(\boldsymbol{X}_0)|^2}\mathbb{I}\{\tau(\boldsymbol{X}_0)>j^{-1/4}\}\\
&\leq \frac{j^{1/2}}{c_0}\mathrm{E}|\widehat{\tau}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)-\tau(\boldsymbol{X}_0)|^2 = O(j^{-\kappa_0+1/2})\to 0, \qquad \text{as } \ j\to\infty,
\end{aligned}
$$

the last inequality is due to the relation that $\kappa_0>(\alpha+2)/(2\alpha+2)>1/2$. By (S2.51), we've shown (S2.50) holds. By Jensen's inequality, this further implies

$$
\lim_j \mathrm{E}\left|\mathrm{E}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}\,\middle|\,\{O_i\}_{i\in\mathcal{I}_j}\right)\right|=0. \qquad \text{(S2.52)}
$$

Similarly, we can show

$$
\begin{aligned}
&\lim_j \mathrm{E}\left|\mathrm{E}\left(\frac{\{1-\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\}\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)}-\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)}\,\middle|\,\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)\right|=0,\\
&\lim_j \mathrm{E}\left|\mathrm{E}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(1,\boldsymbol{X}_0)}\,\middle|\,\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)\right|=0,\\
&\lim_j \mathrm{E}\left|\mathrm{E}\left(\frac{\{1-\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{X}_0)\}\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(0,\boldsymbol{X}_0)}\,\middle|\,\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)\right|=0.
\end{aligned}
$$

Combining this together with (S2.52) yields

$$\mathrm{E}\left|\mathrm{E}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(0,\boldsymbol{X}_0)}\right|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)-\mathrm{E}\left(\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}+\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)}\right)$$
$$-\mathrm{E}\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(0,\boldsymbol{X}_0)}\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)\Bigg|=o(1),\qquad \text{as}\quad j\to\infty.$$

Since $l_n\to\infty$, we have

$$\max_{j\geq l_n}\mathrm{E}\left|\mathrm{E}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(0,\boldsymbol{X}_0)}\right|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)-\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}$$
$$-\quad\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)}-\mathrm{E}\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(1,\boldsymbol{X}_0)}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)},0}}{\pi(0,\boldsymbol{X}_0)}\middle|\{O_i\}_{i\in\mathcal{I}_{(j)}}\right)\Bigg|=o(1).$$

This together with (S2.47), (S2.48) and (S2.49) yields

$$\sup_{j\geq l_n}\mathrm{E}\left|\widetilde{\sigma}_0^2(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)-\nu_1-\sigma_0^2\underbrace{\int_{\boldsymbol{x}\in\mathbb{X}_0}\left(\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{x})}{\pi(1,\boldsymbol{x})}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{x})}{\pi(0,\boldsymbol{x})}\right)dF_X(\boldsymbol{x})}_{\kappa_j}\right|=o(1), \quad \text{(S2.53)}$$

where $\mathbb{X}_0=\{\boldsymbol{x}\in\mathbb{X}:\tau(\boldsymbol{x})=0\}$, $F_X(\cdot)$ denotes the cumulative distribution function of $\boldsymbol{X}_0$, and

$$\nu_1=\mathrm{Var}\{h(d_0^{opt},\boldsymbol{X}_0)\}+\sigma_0^2\mathrm{E}\left(\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)}+\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)}\right).$$

It follows from (S2.47) and (S2.49),

$$\inf_{j\geq l_n}\widetilde{\sigma}_0^2(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\geq\sigma_0^2>0. \quad \text{(S2.54)}$$

In addition, by Condition (A3) and Lemma S2.1, we obtain

$$\sup_{j\geq l_n}\widetilde{\sigma}_0^2(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\leq C_0^2+\frac{\sigma_0^2}{c_0}. \quad \text{(S2.55)}$$

Notice that

$$\mathrm{E}\frac{(n-l_n)^2}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}=\mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\widetilde{\sigma}_0^2(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}.$$

By (S2.53), (S2.54) and (S2.55), we have

$$\begin{aligned}
& \mathrm{E}\left|\frac{(n-l_n)^2}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}-\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)(\nu_1+\sigma_0^2\kappa_j)}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}\right| \\
\leq\ & \mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)|\widetilde{\sigma}_0^{2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)-\nu_1-\sigma_0^2\kappa_j|}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2} \\
\leq\ & \frac{(C_0^2+\sigma_0^2/c_0)^2}{\sigma_0^4}\sup_{j\geq l_n}\mathrm{E}\left|\widetilde{\sigma}_0^{2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)-\nu_1-\sigma_0^2\kappa_j\right|=o(1).
\end{aligned} \tag{S2.56}$$

In the following, we provide a lower bound for

$$\begin{aligned}
& \mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)(\nu_1+\sigma_0^2\kappa_j)}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}=\mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\nu_1}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2} \\
+\ & \int_{\boldsymbol{x}\in\mathbb{X}_0}\mathrm{E}\frac{\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}\underbrace{\left(\frac{\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{x})}{\pi(1,\boldsymbol{x})}+\frac{1-\widehat{d}_{\mathcal{I}_{(j)}}(\boldsymbol{x})}{\pi(0,\boldsymbol{x})}\right)}_{\kappa_j(\boldsymbol{x})}dF_X(\boldsymbol{x}).
\end{aligned} \tag{S2.57}$$

It follows from Cauchy-Schwarz inequality that

$$\mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\nu_1}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}\geq\mathrm{E}\frac{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\nu_1}{(n-l_n)\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)}=\nu_1. \tag{S2.58}$$

Similarly, we have

$$\mathrm{E}\frac{\displaystyle\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\kappa_j(\boldsymbol{x})}{\left(\displaystyle\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2}\geq\mathrm{E}\frac{\displaystyle\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\kappa_j(\boldsymbol{x})}{\displaystyle\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-2}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\kappa_j(\boldsymbol{x})\sum_{j=l_n}^{n-1}\kappa_j^{-1}(\boldsymbol{x})}=\mathrm{E}\frac{1}{\displaystyle\sum_{j=l_n}^{n-1}\kappa_j^{-1}(\boldsymbol{x})} \tag{S2.59}$$

For any $y_{l_n},y_{l_n+1},\ldots,y_{n-1},z_{l_n},z_{l_n+1},\ldots,z_{n-1}\in\mathbb{R}^+$, it follows from Taylor's theorem that

$$\begin{aligned}
& \frac{1}{\sum_{j=l_n}^{n-1}y_j^{-1}}=\frac{1}{\sum_{j=l_n}^{n-1}z_j^{-1}}+\sum_{i=l_n}^{n-1}\frac{z_i^{-2}(y_i-z_i)}{(\sum_{j=l_n}^{n-1}z_j^{-1})^2}+\frac{\{\sum_{i=l_n}^{n-1}z_i^{*-2}(y_i-z_i)\}^2}{(\sum_{j=l_n}^{n-1}z_j^{*-1})^3} \\
-\ & \sum_{i=l_n}^{n-1}\frac{z_i^{*-3}(y_i-z_i)^2}{(\sum_{j=l_n}^{n-1}z_j^{*-1})^2}\geq\frac{1}{\sum_{j=l_n}^{n-1}z_j^{-1}}+\sum_{i=l_n}^{n-1}\frac{z_i^{-2}(y_i-z_i)}{(\sum_{j=l_n}^{n-1}z_j^{-1})^2}-\sum_{i=l_n}^{n-1}\frac{z_i^{*-3}(y_i-z_i)^2}{(\sum_{j=l_n}^{n-1}z_j^{*-1})^2},
\end{aligned}$$

where z_j^* lies between y_j and z_j, for $j = l_n \dots, n-1$. Take $y_j = \kappa_j(\boldsymbol{x})$, $z_j = \mathrm{E}\kappa_j(\boldsymbol{x})$, we obtain

$$\mathrm{E}\frac{1}{\sum_{j=l_n}^{n-1}\kappa_j^{-1}(\boldsymbol{x})} \geq \frac{1}{\sum_{j=l_n}^{n-1}\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1}} - \sum_{i=l_n}^{n-1}\frac{z_i^{*-3}\mathrm{Var}\{\kappa_i(\boldsymbol{x})\}}{(\sum_{j=l_n}^{n-1} z_j^{*-1})^2}, \tag{S2.60}$$

where z_j^* lies between $1/\pi(0,\boldsymbol{x})$ and $1/\pi(1,\boldsymbol{x})$.

Recall that $\mathrm{E}\kappa_j(\boldsymbol{x}) = p_j(\boldsymbol{x})/\pi(0,\boldsymbol{x}) + \{1-p_j(\boldsymbol{x})\}/\pi(1,\boldsymbol{x})$. Condition (A7), we have for any $\boldsymbol{x} \in \mathbb{X}_0$, $\mathrm{E}\kappa_j(\boldsymbol{x}) \to \sum_{a=0,1} 1/\{2\pi(a,\boldsymbol{x})\}$ as $j \to \infty$. By Condition (A3), $\pi(a,\boldsymbol{x})$ are lower bounded by $c_0 > 0$. Therefore $\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1} \to [\sum_{a=0,1} 1/\{2\pi(a,\boldsymbol{x})\}]^{-1}$, $\forall \boldsymbol{x} \in \mathbb{X}_0$, as $j \to \infty$. Since $l_n \to \infty$, we have $\sum_{j=l_n}^{n-1}\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1}/(n-l_n) \to [\sum_{a=0,1} 1/\{2\pi(a,\boldsymbol{x})\}]^{-1}$, $\forall \boldsymbol{x} \in \mathbb{X}_0$, and hence $(n-l_n)/[\sum_{j=l_n}^{n-1}\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1}] \to \sum_{a=0,1} 1/\{2\pi(a,\boldsymbol{x})\}$, $\forall \boldsymbol{x} \in \mathbb{X}_0$. Since $(n-l_n)/[\sum_{j=l_n}^{n-1}\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1}]$ is uniformly bounded for any $\boldsymbol{x}$, it follows from dominated convergence theorem that

$$\int_{\boldsymbol{x}\in\mathbb{X}_0} \frac{(n-l_n)dF_X(\boldsymbol{x})}{\sum_{j=l_n}^{n-1}\{\mathrm{E}\kappa_j(\boldsymbol{x})\}^{-1}} \to \sum_{a=0,1}\int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{dF_X(\boldsymbol{x})}{2\pi(a,\boldsymbol{x})} = \int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{dF_X(\boldsymbol{x})}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})}. \tag{S2.61}$$

Similarly, we can show

$$\left|\int_{\boldsymbol{x}\in\mathbb{X}_0}\sum_{i=l_n}^{n-1}\frac{(n-l_n)z_i^{*-3}\mathrm{Var}\{\kappa_i(\boldsymbol{x})\}}{(\sum_{j=l_n}^{n-1} z_j^{*-1})^2}dF_X(\boldsymbol{x}) - \int_{\boldsymbol{x}\in\mathbb{X}_0}\sum_{i=l_n}^{n-1}\frac{(n-l_n)z_i^{*-3}}{4(\sum_{j=l_n}^{n-1} z_j^{*-1})^2}dF_X(\boldsymbol{x})\right| = o(1).$$

This together with (S2.60) and (S2.61) yields

$$\int_{\boldsymbol{x}\in\mathbb{X}_0}\mathrm{E}\frac{(n-l_n)dF_X(\boldsymbol{x})}{\sum_{j=l_n}^{n-1}\kappa_j^{-1}(\boldsymbol{x})} \geq \int_{\boldsymbol{x}\in\mathbb{X}_0}\left(\frac{1}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} - \frac{(n-l_n)\sum_{i=l_n}^{n-1} z_i^{*-3}}{4(\sum_{j=l_n}^{n-1} z_j^{*-1})^2}\right)dF_X(\boldsymbol{x}) + o(1) \tag{S2.62}$$

Since z_j^* lies between $1/\pi(0,\boldsymbol{x})$ and $1/\pi(1,\boldsymbol{x})$, we have

$$\sum_{i=l_n}^{n-1}\frac{(n-l_n)z_i^{*-3}}{4(\sum_{j=l_n}^{n-1} z_j^{*-1})^2} \leq \sum_{i=l_n}^{n-1}\frac{(n-l_n)z_i^{*-2}}{4(\sum_{j=l_n}^{n-1} z_j^{*-1})^2}.$$

In addition, it follows from Pólya-Szegö's inequality (Pólya, 1964) and Cauchy-Schwarz inequality that

$$\begin{aligned}
& \frac{(n-l_n)\sum_{i=l_n}^{n-1} z_i^{*-2}}{(\sum_{j=l_n}^{n-1} z_j^{*-1})^2} \leq \frac{1}{4}\left(\sqrt{\frac{\pi(1,\boldsymbol{x})}{\pi(0,\boldsymbol{x})}} + \sqrt{\frac{\pi(0,\boldsymbol{x})}{\pi(1,\boldsymbol{x})}}\right)^2 \leq \frac{1}{2}\left(\frac{\pi(1,\boldsymbol{x})}{\pi(0,\boldsymbol{x})} + \frac{\pi(0,\boldsymbol{x})}{\pi(1,\boldsymbol{x})}\right) \\
= \; & \frac{\pi^2(1,\boldsymbol{x}) + \pi^2(0,\boldsymbol{x})}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} \leq \frac{\{\pi(1,\boldsymbol{x}) + \pi(0,\boldsymbol{x})\}^2}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} = \frac{1}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})}.
\end{aligned}$$

Combining this together with (S2.62) that

$$\begin{aligned}\int_{\boldsymbol{x}\in\mathbb{X}_0} \mathrm{E}\frac{(n-l_n)dF_X(\boldsymbol{x})}{\sum_{j=l_n}^{n-1}\kappa_j^{-1}(\boldsymbol{x})} &\geq \int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{dF_X(\boldsymbol{x})}{2\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} - \int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{dF_X(\boldsymbol{x})}{8\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} + o(1)\\ &= \int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{3dF_X(\boldsymbol{x})}{8\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} + o(1) = \mathrm{E}\frac{3}{8\pi(0,\boldsymbol{X}_0)\pi(1,\boldsymbol{X}_0)} + o(1).\end{aligned}$$

In view of (S2.56)-(S2.59), we've shown

$$\mathrm{E}\frac{(n-l_n)^2}{\left(\sum_{j=l_n}^{n-1}\widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}};\pi,h)\right)^2} \geq \nu_1 + \int_{\boldsymbol{x}\in\mathbb{X}_0}\frac{3dF_X(\boldsymbol{x})}{8\pi(0,\boldsymbol{x})\pi(1,\boldsymbol{x})} + o(1). \tag{S2.63}$$

Now, let's consider

$$\sigma_{s_n}^2 = \widetilde{\sigma}_0^2(p_{s_n};\pi,h) = \mathrm{Var}\left(\frac{\mathrm{g}(A_0,p_{s_n,0})}{\pi(A_0,\boldsymbol{X}_0)}e_0\right) + \mathrm{Var}\{h(p_{s_n},\boldsymbol{X}_0)\}. \tag{S2.64}$$

Similar to (S2.48), we can show

$$\mathrm{Var}\{h(p_{s_n},\boldsymbol{X}_0)\} = \mathrm{Var}\{h(d_0^{opt},\boldsymbol{X}_0)\} + o(1). \tag{S2.65}$$

With some calculations, we have

$$\begin{aligned}\mathrm{Var}\left(\frac{\mathrm{g}(A_0,p_{s_n,0})}{\pi(A_0,\boldsymbol{X}_0)}e_0\right) &= \sigma_0^2\mathrm{E}\frac{\mathrm{g}^2(A_0,p_{s_n,0})}{\pi^2(A_0,\boldsymbol{X}_0)} = \sigma_0^2\mathrm{E}\left(\frac{A_0p_{s_n}^2(\boldsymbol{X}_0)}{\pi^2(1,\boldsymbol{X}_0)} + \frac{(1-A_0)\{1-p_{s_n}(\boldsymbol{X}_0)\}^2}{\pi^2(0,\boldsymbol{X}_0)}\right)\\ &= \sigma_0^2\mathrm{E}\left(\frac{p_{s_n}^2(\boldsymbol{X}_0)}{\pi(1,\boldsymbol{X}_0)} + \frac{\{1-p_{s_n}(\boldsymbol{X}_0)\}^2}{\pi(0,\boldsymbol{X}_0)}\right).\end{aligned}$$

Using similar arguments in (S2.53), we can show

$$\begin{aligned}&\mathrm{E}\frac{p_{s_n}^2(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)} = \mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)} + o(1),\\ &\mathrm{E}\frac{\{1-p_{s_n}(\boldsymbol{X}_0)\}^2\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)} = \mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(0,\boldsymbol{X}_0)} + o(1),\\ &\mathrm{E}\frac{p_{s_n}^2(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)<0\}}{\pi(1,\boldsymbol{X}_0)} = o(1) \quad \text{and} \quad \mathrm{E}\frac{\{1-p_{s_n}(\boldsymbol{X}_0)\}^2\mathbb{I}\{\tau(\boldsymbol{X}_0)>0\}}{\pi(1,\boldsymbol{X}_0)} = o(1).\end{aligned}$$

In addition, by Condition (A7), we have

$$\begin{aligned}\mathrm{E}\frac{p_{s_n}^2(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}}{\pi(1,\boldsymbol{X}_0)} &= \frac{1}{4}\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}}{\pi(1,\boldsymbol{X}_0)} + o(1),\\ \mathrm{E}\frac{\{1-p_{s_n}(\boldsymbol{X}_0)\}^2\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}}{\pi(0,\boldsymbol{X}_0)} &= \frac{1}{4}\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}}{\pi(0,\boldsymbol{X}_0)} + o(1).\end{aligned}$$

Therefore,

$$\mathrm{Var}\left(\frac{\mathrm{g}(A_0, p_{s_n,0})}{\pi(A_0, \boldsymbol{X}_0)}e_0\right) = \sigma_0^2\mathrm{E}\left(\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0) > 0\}}{\pi(1, \boldsymbol{X}_0)} + \frac{\mathbb{I}\{\tau(\boldsymbol{X}_0) < 0\}}{\pi(0, \boldsymbol{X}_0)}\right) + \sigma_0^2\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0) = 0\}}{4\pi(0, \boldsymbol{X}_0)\pi(1, \boldsymbol{X}_0)} + o(1).$$

Combining this together with (S2.64) and (S2.65) yields

$$\widetilde{\sigma}_0^2(p_{s_n}; \pi, h) \to \nu_1 + \sigma_0^2\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0) = 0\}}{4\pi(0, \boldsymbol{X}_0)\pi(1, \boldsymbol{X}_0)}. \tag{S2.66}$$

By (S2.63) and Condition (A3), we have

$$\mathrm{E}\frac{(n - l_n)^2}{\left(\sum_{j=l_n}^{n-1} \widetilde{\sigma}_0^{-1}(\widehat{d}_{\mathcal{I}_{(j)}}; \pi, h)\right)^2} - \widetilde{\sigma}_0^2(p_{s_n}; \pi, h) \geq \sigma_0^2\mathrm{E}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0) = 0\}}{8\pi(0, \boldsymbol{X}_0)\pi(1, \boldsymbol{X}_0)} + o(1) = \frac{\sigma_0^2\nu_0}{8c_0^2} + o(1).$$

In view of (9) and (14), we've shown

$$n\mathrm{E}L^2(\widehat{V}^{on}, \alpha) - n\mathrm{E}L^2(\widehat{V}_B, \alpha) \geq \frac{z_{\alpha/2}^2\sigma_0^2\nu_0}{2c_0^2} + o(1),$$

This completes the proof.

S2.4 Proof of Theorem 3.2

Recall that $\nu_0 = \Pr\{\tau(\boldsymbol{X}_0) = 0\}$. By (9), (11) and (16), it suffices to show

$$\inf_{d^{opt} \in \mathcal{D}^{opt}} \widetilde{\sigma}_0^2(d^{opt}; \pi, h) \geq \widetilde{\sigma}_0^2(p_{s_n}; \pi, h) + c^{**}\sigma_0^2\nu_0 + o(1). \tag{S2.67}$$

Similar to (S2.46) and (S2.49), we have

$$\begin{aligned}\widetilde{\sigma}_0^2(d^{opt}; \pi, h) &= \mathrm{Var}\left(\frac{\mathrm{g}\{A_0, d^{opt}(\boldsymbol{X}_0)\}}{\pi(A_0, \boldsymbol{X}_0)}e_0\right) + \mathrm{Var}\{h(d^{opt}(\boldsymbol{X}_0), \boldsymbol{X}_0)\} \\ &= \mathrm{Var}\{h(d^{opt}(\boldsymbol{X}_0), \boldsymbol{X}_0)\} + \sigma_0^2\mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1, \boldsymbol{X}_0)} + \frac{1 - d^{opt}(\boldsymbol{X}_0)}{\pi(0, \boldsymbol{X}_0)}\right).\end{aligned} \tag{S2.68}$$

By Lemma 2.1, we have

$$\begin{aligned}\mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1, \boldsymbol{X}_0)} + \frac{1 - d^{opt}(\boldsymbol{X}_0)}{\pi(0, \boldsymbol{X}_0)}\right) &= \mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1, \boldsymbol{X}_0)} + \frac{1 - d^{opt}(\boldsymbol{X}_0)}{\pi(0, \boldsymbol{X}_0)}\right)\mathbb{I}\{\tau(\boldsymbol{X}_0) = 0\} \\ &+\mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)\mathbb{I}\{\tau(\boldsymbol{X}_0 > 0)\}}{\pi(1, \boldsymbol{X}_0)} + \frac{\{1 - d^{opt}(\boldsymbol{X}_0)\}\mathbb{I}\{\tau(\boldsymbol{X}_0) < 0\}}{\pi(0, \boldsymbol{X}_0)}\right).\end{aligned}$$

This together with (S2.68) gives

$$\widetilde{\sigma}_0^2(d^{opt}; \pi, h) = \nu_1 + \mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1, \boldsymbol{X}_0)} + \frac{1 - d^{opt}(\boldsymbol{X}_0)}{\pi(0, \boldsymbol{X}_0)}\right)\mathbb{I}\{\tau(\boldsymbol{X}_0) = 0\}$$

In view of (S2.66), to prove (S2.67), it suffices to show

$$\inf_{d^{opt}\in\mathcal{D}^{opt}} \mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1,\boldsymbol{X}_0)}+\frac{1-d^{opt}(\boldsymbol{X}_0)}{\pi(0,\boldsymbol{X}_0)}\right)\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\} \geq \frac{1}{4}\mathrm{E}\sum_{a=0,1}\frac{\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}}{\pi(a,\boldsymbol{X}_0)}+c^{**}\nu_0.$$

Notice that $d^{opt}(\boldsymbol{X}_0)\in\{0,1\}$. For any $\boldsymbol{x}\in\mathbb{X}_0$, we have

$$\frac{d^{opt}(\boldsymbol{x})}{\pi(1,\boldsymbol{x})}+\frac{1-d^{opt}(\boldsymbol{x})}{\pi(0,\boldsymbol{x})}-\frac{1}{4}\left(\frac{1}{\pi(0,\boldsymbol{x})}+\frac{1}{\pi(1,\boldsymbol{x})}\right)\geq \min_{a=0,1}\frac{1}{\pi(a,\boldsymbol{x})}-\frac{1}{4}\left(\frac{1}{\pi(0,\boldsymbol{x})}+\frac{1}{\pi(1,\boldsymbol{x})}\right)\geq c^{**}.$$

Thus, for any $d^{opt}\in\mathcal{D}^{opt}$, we have

$$\mathrm{E}\left(\frac{d^{opt}(\boldsymbol{X}_0)}{\pi(1,\boldsymbol{X}_0)}+\frac{1-d^{opt}(\boldsymbol{X}_0)}{\pi(0,\boldsymbol{X}_0)}-\sum_{a=0,1}\frac{1}{4\pi(a,\boldsymbol{X}_0)}\right)\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}\geq c^{**}\mathrm{E}\mathbb{I}\{\tau(\boldsymbol{X}_0)=0\}=c^{**}\nu_0.$$

The proof is hence completed.

S2.5 Proof of Theorem 4.1

For notational convenience, we use a shorthand and write $\widehat{V}_i^{(k)}(d;\pi,h)$ as $\widehat{V}_i^{(k)}(d)$ for any $k=2,\dots,K, i=0,1,\dots,n$ and any dynamic treatment regime d. In addition, for any $d=\{d_k\}_{k=1}^K$, $\pi^*=\{\pi_k^*\}_{k=1}^K$, $h^*=\{h_k^*\}_{k=1}^K$ and $i=0,1,\dots,n$, let $d_{k,i}=d_k(\bar{\boldsymbol{A}}_i^{(k-1)},\bar{\boldsymbol{X}}_i^{(k)})$, $\pi_{k,i}^*=\pi_k^*(\bar{\boldsymbol{A}}_i^{(k)},\bar{\boldsymbol{X}}_i^{(k)})$, $h_{k,i}^*=h_k^*(\bar{\boldsymbol{A}}_i^{(k)},\bar{\boldsymbol{X}}_i^{(k)})$, $\forall k=2,\dots,K$ and $d_{1,i}=d_1(\boldsymbol{X}_i^{(1)})$, $\pi_{1,i}^*=\pi_1^*(\boldsymbol{X}_i^{(1)})$, $h_{1,i}^*=h_1^*(\boldsymbol{X}_i^{(i)})$.

Similar to Lemma S2.1, under (C1), (C2) and (C4), we have

$$\sup_{\bar{\boldsymbol{x}}_K\in\bar{\mathbb{X}}^{(K)},\bar{\boldsymbol{a}}_K\in\{0,1\}^K}|h_K(\bar{\boldsymbol{a}}_K,\bar{\boldsymbol{x}}_K)|=O(1). \tag{S2.69}$$

Recall that for any dynamic treatment regime d,

$$\widehat{V}_i^{(K)}(d)=\frac{\mathrm{g}(A_i^{(K)},d_{K,i})}{\pi_{K,i}}(Y_i-h_{K,i})+h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}),\bar{\boldsymbol{X}}_i^{(K)}\}.$$

By (S2.69) and Condition (C1)-(C4), we have

$$
\begin{aligned}
&\sup_{\substack{\bar{\boldsymbol{x}}_{K-1}\in\bar{\mathbb{X}}^{(K-1)}\\ \bar{\boldsymbol{a}}_{K-1}\in\{0,1\}^{K-1}, i\in\mathcal{I}_0}} \mathrm{E}\left(\{\widehat{V}_i^{(K)}(d)\}^2|\bar{\boldsymbol{X}}_i^{(K-1)}=\bar{\boldsymbol{x}}_{K-1}, \bar{\boldsymbol{A}}_i^{(K-1)}=\bar{\boldsymbol{a}}_{K-1}\right)\\
\le\ & 2\sup_{\substack{\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\\ i\in\mathcal{I}_0}} \mathrm{E}\left\{\left(\frac{\mathrm{g}(A_i^{(K)},d_{K,i})}{\pi_{K,i}}(Y_i-h_{K,i})\right)^2\middle|\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\right\}+O(1)\\
\le\ & \frac{2}{c_0^2}\sup_{\substack{\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\\ i\in\mathcal{I}_0}} \mathrm{E}\{(Y_i-h_{K,i})^2|\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\}+O(1)\\
\le\ & \frac{4}{c_0^2}\sup_{\substack{\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\\ i\in\mathcal{I}_0}} \mathrm{E}\{Y_i^2|\bar{\boldsymbol{X}}_i^{(K-1)},\bar{\boldsymbol{A}}_i^{(K-1)}\}+O(1)\le \frac{4}{c_0^2}\sup_{\substack{\bar{\boldsymbol{X}}_i^{(K)},\bar{\boldsymbol{A}}_i^{(K)}\\ i\in\mathcal{I}_0}} \mathrm{E}\{Y_i^2|\bar{\boldsymbol{X}}_i^{(K)},\bar{\boldsymbol{A}}_i^{(K)}\}+O(1)\\
\le\ & \frac{4}{c_0^2}\sup_{\substack{\bar{\boldsymbol{X}}_i^{(K)*}(\bar{\boldsymbol{A}}_i^{(K-1)}),\bar{\boldsymbol{A}}_i^{(K)}\\ i\in\mathcal{I}_0}} \mathrm{E}[\{Y_i^*(\bar{\boldsymbol{A}}_i^{(K)})\}^2|\bar{\boldsymbol{X}}_i^{(K)*}(\bar{\boldsymbol{A}}_i^{(K-1)}),\bar{\boldsymbol{A}}_i^{(K)}]+O(1)=O(1),
\end{aligned}
$$

where the first and the third inequalities are due to Cauchy-Schwarz inequality.

Notice that

$$
\mathrm{E}\left(\widehat{V}_i^{(K)}(d^{opt})|\bar{\boldsymbol{X}}_i^{(K-1)}=\bar{\boldsymbol{x}}_{K-1},\bar{\boldsymbol{A}}_i^{(K-1)}=\bar{\boldsymbol{a}}_{K-1}\right)=h(\bar{\boldsymbol{a}}_{K-1},\bar{\boldsymbol{x}}_{K-1}),
$$

for any $d^{opt}\in\mathcal{D}^{opt}$. By Jensen's inequality, we obtain

$$
\sup_{\bar{\boldsymbol{x}}_{K-1}\in\bar{\mathbb{X}}^{(K-1)},\bar{\boldsymbol{a}}_{K-1}\in\{0,1\}^{K-1}} |h_{K-1}(\bar{\boldsymbol{a}}_{K-1},\bar{\boldsymbol{x}}_{K-1})|=O(1).
$$

Similarly, we can show there exists some constant $C_0>0$ such that

$$
\max_{k=1,\dots,K}\sup_{\substack{\bar{\boldsymbol{x}}_k\in\bar{\mathbb{X}}^{(k)}\\ \bar{\boldsymbol{a}}_k\in\{0,1\}^k}} |h_k(\bar{\boldsymbol{a}}_k,\bar{\boldsymbol{x}}_k)|\le C_0, \tag{S2.70}
$$

and

$$
\sup_{\substack{d=\{d_k\}_{k=1}^K\\ i\in\mathcal{I}_0\\ k=2,\dots,K}}\sup_{\substack{\bar{\boldsymbol{x}}_{k-1}\in\bar{\mathbb{X}}^{(k-1)}\\ \bar{\boldsymbol{a}}_{k-1}\in\{0,1\}^{k-1}}} \mathrm{E}\left(\{\widehat{V}_i^{(k)}(d)\}^2|\bar{\boldsymbol{X}}_i^{(k-1)}=\bar{\boldsymbol{x}}_{k-1},\bar{\boldsymbol{A}}_i^{(k-1)}=\bar{\boldsymbol{a}}_{k-1}\right)\le C_0. \tag{S2.71}
$$

To prove Theorem 4.1, we break the proof into four steps. In the first step, we show $\widehat{V}_B=\widehat{V}_B^*+o_p(n^{-1/2})$ where

$$
\widehat{V}_B^*=\frac{1}{B}\sum_{b=1}^{B}\widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}}).
$$

In the second step, we show $\widehat{V}_B^* = \widehat{V}_\infty^* + o_p(n^{-1/2})$ where

$$\widehat{V}_\infty^* = \frac{1}{\binom{n}{s_n}} \sum_{\mathcal{I}\subseteq\mathcal{I}_0, |\mathcal{I}|=s_n} \widehat{V}_{\mathcal{I}^c}(\widehat{d}_{\mathcal{I}}).$$

In the third step, we show $\sqrt{n}(\widehat{V}_\infty^* - V_0)/\sigma_{s_n} \xrightarrow{d} N(0,1)$. In the last step, we show $\widehat{\sigma}_B^2 = \sigma_{s_n}^2 + o_p(1)$. The proof is hence completed.

Step 1: By the definitions of $\widehat{V}_B$ and $\widehat{V}_B^*$, we need to show

$$\frac{1}{B}\sum_{b=1}^{B}\left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b})\right) = o_p(n^{-1/2}),$$
$$\frac{1}{B}\sum_{b=1}^{B}\left(\widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b}; \widehat{\pi}_{\mathcal{I}_b^{(1)}}, \widehat{h}_{\mathcal{I}_b^{(1)}}) - \widehat{V}_{\mathcal{I}_b^{c(2)}}(\widehat{d}_{\mathcal{I}_b})\right) = o_p(n^{-1/2}).$$

For any $d = \{d_k\}_{k=1}^K$, $\pi^* = \{\pi_k^*\}_{k=1}^K$ and $h^* = \{h_k^*\}_{k=1}^K$, define $V(d;\pi^*,h^*) = \mathrm{E}\widehat{V}_{\{0\}}(d;\pi^*,h^*)$ where

$$\widehat{V}_{\{0\}}(d;\pi^*,h^*) = \frac{\mathrm{g}\{A_0^{(1)}, d_1(\bar{\boldsymbol{X}}_0^{(1)})\}}{\pi_1^*(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(1)})}\{\widehat{V}_0^{(2)}(d;\pi^*,h^*) - h_1^*(A_0^{(1)}, \bar{\boldsymbol{X}}_0^{(1)})\} + h_1^*\{d_1(\bar{\boldsymbol{X}}_0^{(1)}), \bar{\boldsymbol{X}}_0^{(1)}\}.$$

Using similar arguments in (S2.12)-(S2.17), it suffices to show $\eta_7, \eta_8 = o(n^{-1/2})$ where

$$\begin{aligned}\eta_7 &= \mathrm{E}\left|\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\pi,h) - \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*};\pi,h) + V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\right|,\\ \eta_8 &= \mathrm{E}\left|V(\widehat{d}_{\mathcal{I}_*};\pi,h) - V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}},\widehat{h}_{\mathcal{I}_*^{(1)}})\right|,\end{aligned}$$

where $\mathcal{I}_*$ denotes a random subset uniformly sampled from the set $\{\mathcal{I}\subseteq\mathcal{I}_0 : |\mathcal{I}| = s_n\}$, $\mathcal{I}_*^{c(1)}$ and $\mathcal{I}_*^{c(2)}$ correspond to a random partition of $\mathcal{I}_*^c$ with $|\mathcal{I}_*^{c(1)}| = |\mathcal{I}_*^{c(2)}| = t_n = (n-s_n)/2$, and $\mathcal{I}_*^{(j)} = \mathcal{I}_*^{c(j)} \cup \mathcal{I}_*$ for $j = 1, 2$.

Define the functions $\widehat{\pi}_{\mathcal{I}_*^{(1)}}^{(l)} = \{\widehat{\pi}_{\mathcal{I}_*^{(1)},k}^{(l)}\}_{k=1}^K$, $\widehat{h}_{\mathcal{I}_*^{(1)}}^{(l)} = \{\widehat{h}_{\mathcal{I}_*^{(1)},k}^{(l)}\}_{k=1}^K$ as follows:

$$\widehat{\pi}_{\mathcal{I}_*^{(1)},k}^{(l)} = \pi_k\mathbb{I}(l<k) + \widehat{\pi}_{\mathcal{I}_*^{(1)},k}\mathbb{I}(l\ge k) \text{ and } \widehat{h}_{\mathcal{I}_*^{(1)},k}^{(l)} = h_k\mathbb{I}(l<k) + \widehat{h}_{\mathcal{I}_*^{(1)},k}\mathbb{I}(l\ge k),$$

for any $k = 1,\dots,K$, $l = 0,\dots,K$. Notice that for $l = 0,1,2,\dots,K-1$, $\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}_{\mathcal{I}_*^{(1)}}^{(l+1)},\widehat{h}_{\mathcal{I}_*^{(1)}}^{(l+1)}) -$

$\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})$ equals

$$
\begin{aligned}
&\frac{1}{t_n}\sum_{i\in\mathcal{I}_*^{c(2)}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_i^{(j)},\widehat{d}_{\mathcal{I}_*,j,i})}{\pi_{j,i}}\Bigg(\frac{\mathrm{g}(A_i^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,i})}{\widehat{\pi}_{\mathcal{I}_*,l+1,i}}\{\widehat{V}_i^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{h}_{\mathcal{I}_*^{(1)},l+1,i}\}\\
+&\ \widehat{h}_{\mathcal{I}_*^{(1)},l+1}\{(\bar{\boldsymbol{A}}_i^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,i}),\bar{\boldsymbol{X}}_i^{(l+1)}\}-\frac{\mathrm{g}(A_i^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,i})}{\pi_{l+1,i}}\{\widehat{V}_i^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-h_{l+1,i}\}\\
-&\ h_{l+1}\{(\bar{\boldsymbol{A}}_i^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,i}),\bar{\boldsymbol{X}}_i^{(l+1)}\}\Bigg),
\end{aligned}
$$

where $\widehat{V}_i^{(K+1)}(\widehat{d}_{\mathcal{I}_*};\pi,h)=Y_i$ and $\bar{\boldsymbol{A}}_i^{(0)}=\emptyset$, for $i=0,1,\ldots,n$.

By (S2.70) and (S2.71), using similar arguments in bounding η_5 in the proof of Theorem 2.2, we can show

$$
\begin{aligned}
\max_{l=0,\ldots,K-1}\mathrm{E}\Big|&\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})-\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\\
&-V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})+V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\Big|=o(n^{-1/2}).
\end{aligned}
$$

By triangle inequality, we obtain

$$
\begin{aligned}
\eta_7\leq&\sum_{l=0,\ldots,K-1}\mathrm{E}\Big|\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})-\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\\
-&\ V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})+V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\Big|\leq K\max_{l=0,\ldots,K-1}\mathrm{E}\Big|\widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})\\
-&\ \widehat{V}_{\mathcal{I}_*^{c(2)}}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})-V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})+V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\Big|=o(n^{-1/2}).
\end{aligned}
$$

Now we show $\eta_8=o(n^{-1/2})$. Notice that

$$
\mathrm{E}\{\widehat{V}^{(K)}_{\{0\}}(\widehat{d}_{\mathcal{I}})|\widehat{d}_{\mathcal{I}},\bar{\boldsymbol{A}}_0^{(K)},\bar{\boldsymbol{X}}_0^{(K)}\}=h_K\{(\bar{\boldsymbol{A}}_0^{(K-1)},\widehat{d}_{\mathcal{I},K,0}),\bar{\boldsymbol{X}}_0^{(K)}\},
$$

for any $\mathcal{I}\subseteq\mathcal{I}_0$ with $|\mathcal{I}|=s_n$. Therefore, we have for any $d^{opt}\in\mathcal{D}^{opt}$,

$$
\begin{aligned}
&\mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(K)}_{\{0\}}(\widehat{d}_{\mathcal{I}})-\widehat{V}^{(K)}_{\{0\}}(d^{opt})\,\middle|\,\widehat{d}_{\mathcal{I}},\bar{\boldsymbol{A}}_0^{(K)},\bar{\boldsymbol{X}}_0^{(K)}\right)\right|\\
=&\ \mathrm{E}\left|\tau_K(\bar{\boldsymbol{A}}_0^{(K-1)},\bar{\boldsymbol{X}}_0^{(K)})\right|\left|\widehat{d}_{\mathcal{I},K,0}-\mathbb{I}\{\tau_K(\bar{\boldsymbol{A}}_0^{(K-1)},\bar{\boldsymbol{X}}_0^{(K)})>0\}\right|.
\end{aligned}
$$

Under Condition (C5) and (C6), using similar arguments in bounding $\eta_4^{(1)}$ in the proof of Theorem 2.1, we have

$$
\mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(K)}_{\{0\}}(\widehat{d}_{\mathcal{I}})-\widehat{V}^{(K)}_{\{0\}}(d^{opt})\,\middle|\,\widehat{d}_{\mathcal{I}},\bar{\boldsymbol{A}}_0^{(K)},\bar{\boldsymbol{X}}_0^{(K)}\right)\right|=o(n^{-1/2}). \tag{S2.72}
$$

Assume for now, we've shown

$$\mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(k+1)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k+1)}_{\{0\}}(d^{opt})\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k+1)}_0, \bar{\boldsymbol{X}}^{(k+1)}_0\right)\right| = o(n^{-1/2}). \tag{S2.73}$$

By the definition of $\widehat{V}^{(k)}_{\{0\}}(d)$, we have

$$\widehat{V}^{(k)}_{\{0\}}(d) = \frac{\mathrm{g}(A^{(k)}_0, d_{k,0})}{\pi_{k,0}}\{\widehat{V}^{(k+1)}_{\{0\}}(d) - h_{k,0}\} + h_k\{(\bar{\boldsymbol{A}}^{(k-1)}_0, d_{k,0}), \bar{\boldsymbol{X}}^{(k)}_0\}.$$

Therefore,

$$\begin{aligned}
& \mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(k)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k)}_{\{0\}}(d^{opt})\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k)}_0, \bar{\boldsymbol{X}}^{(k)}_0\right)\right| \\
\leq\ & \underbrace{\mathrm{E}\left|\mathrm{E}\left\{\frac{\mathrm{g}(A^{(k)}_0, \widehat{d}_{\mathcal{I},k,0})}{\pi_{k,0}}\left(\widehat{V}^{(k+1)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k+1)}_{\{0\}}(d^{opt})\right)\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k)}_0, \bar{\boldsymbol{X}}^{(k)}_0\right\}\right|}_{\eta_9} \\
+\ & \underbrace{\mathrm{E}\left|h_k\{(\bar{\boldsymbol{A}}^{(k-1)}_0, \widehat{d}_{\mathcal{I},k,0}), \bar{\boldsymbol{X}}^{(k)}_0, \} - h_k\{(\bar{\boldsymbol{A}}^{(k-1)}_0, d^{opt}_{k,0}), \bar{\boldsymbol{X}}^{(k)}_0\}\right|}_{\eta_{10}}.
\end{aligned}$$

By Condition (C3) and (S2.73), we have

$$\begin{aligned}
\eta_9 \ &\leq\ \mathrm{E}\left|\mathrm{E}\left\{\frac{\mathrm{g}(A^{(k)}_0, \widehat{d}_{\mathcal{I},k,0})}{\pi_{k,0}}\left(\widehat{V}^{(k+1)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k+1)}_{\{0\}}(d^{opt})\right)\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k+1)}_0, \bar{\boldsymbol{X}}^{(k+1)}_0\right\}\right| \\
&=\ \mathrm{E}\frac{\mathrm{g}(A^{(k)}_0, \widehat{d}_{\mathcal{I},k,0})}{\pi_{k,0}}\left|\mathrm{E}\left\{\left(\widehat{V}^{(k+1)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k+1)}_{\{0\}}(d^{opt})\right)\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k+1)}_0, \bar{\boldsymbol{X}}^{(k+1)}_0\right\}\right| \\
&\leq\ \frac{1}{c_0}\mathrm{E}\left|\mathrm{E}\left\{\left(\widehat{V}^{(k+1)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k+1)}_{\{0\}}(d^{opt})\right)\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k+1)}_0, \bar{\boldsymbol{X}}^{(k+1)}_0\right\}\right| = o(n^{-1/2}).
\end{aligned}$$

Under Condition (C5) and (C6), using similar arguments in bounding $\eta_4^{(1)}$ in the proof of Theorem 2.1, we can show

$$\eta_{10} = o(n^{-1/2}). \tag{S2.74}$$

Thus, we've shown

$$\mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(k)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k)}_{\{0\}}(d^{opt})\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k)}_0, \bar{\boldsymbol{X}}^{(k)}_0\right)\right| = o(n^{-1/2}).$$

Since K is a fixed constant, we have

$$\max_{k=2,\dots,K}\mathrm{E}\left|\mathrm{E}\left(\widehat{V}^{(k)}_{\{0\}}(\widehat{d}_{\mathcal{I}}) - \widehat{V}^{(k)}_{\{0\}}(d^{opt})\right| \widehat{d}_{\mathcal{I}}, \bar{\boldsymbol{A}}^{(k)}_0, \bar{\boldsymbol{X}}^{(k)}_0\right)\right| = o(n^{-1/2}). \tag{S2.75}$$

Let $\mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}$ denote the conditional expectation given $\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}$, we have

$$\begin{aligned}
& V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})-V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\\
= \; & \mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\left(\widehat{V}_0^{(l+1)}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})-\widehat{V}_0^{(l+1)}(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\right)\\
= \; & \mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\left(\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\widehat{\pi}_{\mathcal{I}_*^{(1)},l+1,0}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{h}_{\mathcal{I}_*^{(1)},l+1,0}\}\right.\\
+ \; & \widehat{h}_{\mathcal{I}_*^{(1)},l+1}\{(\bar{\boldsymbol{A}}_0^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,0}),\bar{\boldsymbol{X}}_0^{(l+1)}\}-\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\pi_{l+1,0}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-h_{l+1,0}\}\\
- \; & \left. h_{l+1}\{(\bar{\boldsymbol{A}}_0^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,0}),\bar{\boldsymbol{X}}_0^{(l+1)}\}\right).
\end{aligned}$$

By Condition (C3), (23) and (S2.75), we have

$$\begin{aligned}
& \mathrm{E}\left|\mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\widehat{\pi}_{\mathcal{I}_*^{(1)},l+1,0}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{V}_0^{(l+2)}(d^{opt})\}\right|\\
\le \; & \mathrm{E}\left|\mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}},\bar{\boldsymbol{A}}_0^{(l+2)},\bar{\boldsymbol{X}}_0^{(l+2)}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\widehat{\pi}_{\mathcal{I}_*^{(1)},l+1,0}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{V}_0^{(l+2)}(d^{opt})\}\right|\\
\le \; & \frac{1}{c^*c_0^l}\mathrm{E}\left|\mathrm{E}^{\mathcal{I}_*,\widehat{d}_{\mathcal{I}_*},\bar{\boldsymbol{A}}_0^{(l+2)},\bar{\boldsymbol{X}}_0^{(l+2)}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{V}_0^{(l+2)}(d^{opt})\}\right|=o(n^{-1/2}). \qquad \text{(S2.76)}
\end{aligned}$$

Similarly, we can show

$$\mathrm{E}\left|\mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}\prod_{j=1}^{l+1}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\{\widehat{V}_0^{(l+2)}(\widehat{d}_{\mathcal{I}_*})-\widehat{V}_0^{(l+2)}(d^{opt})\}\right|=o(n^{-1/2}). \quad \text{(S2.77)}$$

Besides, using similar arguments in bounding η_6 in the proof of Theorem 2.2, we have

$$\begin{aligned}
& \mathrm{E}\left|\mathrm{E}^{\mathcal{I}_*^{(1)},\mathcal{I}_*,\{O_i\}_{i\in\mathcal{I}_*^{(1)}}}\prod_{j=1}^{l}\frac{\mathrm{g}(A_0^{(j)},\widehat{d}_{\mathcal{I}_*,j,0})}{\pi_{j,0}}\left(\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\widehat{\pi}_{\mathcal{I}_*^{(1)},l+1,0}}\{\widehat{V}_0^{(l+2)}(d^{opt})-\widehat{h}_{\mathcal{I}_*^{(1)},l+1,0}\}\right.\right.\\
+ \; & \widehat{h}_{\mathcal{I}_*^{(1)},l+1}\{(\bar{\boldsymbol{A}}_0^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,0}),\bar{\boldsymbol{X}}_0^{(l+1)}\}-\frac{\mathrm{g}(A_0^{(l+1)},\widehat{d}_{\mathcal{I}_*,l+1,0})}{\pi_{l+1,0}}\{\widehat{V}_0^{(l+2)}(d^{opt})-h_{l+1,0}\}\\
- \; & \left.\left. h_{l+1}\{(\bar{\boldsymbol{A}}_0^{(l)},\widehat{d}_{\mathcal{I}_*,l+1,0}),\bar{\boldsymbol{X}}_0^{(l+1)}\}\right)\right|=o(n^{-1/2}).
\end{aligned}$$

Combining this together with (S2.76) and (S2.77) yields

$$\mathrm{E}\left|V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}})-V(\widehat{d}_{\mathcal{I}_*};\widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}},\widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}})\right|=o(n^{-1/2}),$$

for all $l = 0, \ldots, K-1$. Since K is a fixed integer, we have

$$\max_{l=0,\ldots,K-1} \mathrm{E}\left| V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}}) \right| = o(n^{-1/2}).$$

By triangle inequality, we obtain

$$\begin{aligned}
\eta_8 &\leq \sum_{l=0,\ldots,K-1} \mathrm{E}\left| V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}}) \right| \\
&\leq K \max_{l=0,\ldots,K-1} \mathrm{E}\left| V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l+1)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l+1)}_{\mathcal{I}_*^{(1)}}) - V(\widehat{d}_{\mathcal{I}_*}; \widehat{\pi}^{(l)}_{\mathcal{I}_*^{(1)}}, \widehat{h}^{(l)}_{\mathcal{I}_*^{(1)}}) \right| = o(n^{-1/2}).
\end{aligned}$$

This implies $\widehat{V}_B = \widehat{V}_B^* + o_p(n^{-1/2})$.

Step 2: The assertion $\widehat{V}_B^* = \widehat{V}_\infty^* + o_p(n^{-1/2})$ can be proven using similar arguments in the second step of the proof of Theorem 2.2. We omit the details for brevity.

Step 3: For any $i \in \mathcal{I}_0$, $\mathcal{I} \subseteq \mathcal{I}_{(-i)}$ with $|\mathcal{I}| = s_n$, define

$$Q_i = \mathrm{E}\left(\widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}}) | O_i \right).$$

Notice that $Q_1, \ldots, Q_n$ are i.i.d random variables with $\mathrm{Var}(Q_i) = \sigma^2_{s_n}$. We first show

$$\widehat{V}_\infty^* = \frac{1}{n}\sum_{i=1}^n Q_i + o_p(n^{-1/2}). \tag{S2.78}$$

Recall that

$$\widehat{V}_i^{(K)}(\widehat{d}_{\mathcal{I}}) = \frac{g(A_i^{(K)}, \widehat{d}_{\mathcal{I},K,i})}{\pi_{K,i}}(Y_i - h_{K,i}) + h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)}, \widehat{d}_{\mathcal{I},K,i}), \bar{\boldsymbol{X}}_i^{(K)}\}, \tag{S2.79}$$

for any $i \in \mathcal{I}_0$, $\mathcal{I} \subseteq \mathcal{I}_{(-i)}$ with $|\mathcal{I}| = s_n$. Let

$$T_i^{(K)}(\mathcal{I}) = (Y_i - h_{K,i}) \prod_{k=1}^K \frac{g(A_i^{(k)}, \widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}} \quad \text{and} \quad T_i^{(K)} = \mathrm{E}\{T_i^{(K)}(\mathcal{I})|O_i\}.$$

Using similar arguments in bounding $|\eta_3|$ in the proof of Theorem 2.1, we can show that

$$\frac{1}{n\binom{n-1}{s_n}} \sum_{i\in\mathcal{I}_0} \sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)} \\ |\mathcal{I}|=s_n}} \left(T_i^{(K)}(\mathcal{I}) - T_i^{(K)} \right) = o_p(n^{-1/2}). \tag{S2.80}$$

Besides, by Condition (C3) and (S2.72), we have

$$\mathrm{E}\left|\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}[h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},\widehat{d}_{\mathcal{I},K,i}),\bar{\boldsymbol{X}}_i^{(K)}\}-h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}^{opt}),\bar{\boldsymbol{X}}_i^{(K)}\}]\right|$$
$$\leq\frac{1}{c_0^{K-1}}\max_{\substack{\mathcal{I}\subseteq\mathcal{I}_0\\|\mathcal{I}|=s_n}}\mathrm{E}\left|h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},\widehat{d}_{\mathcal{I},K,i}),\bar{\boldsymbol{X}}_i^{(K)}\}-h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}^{opt}),\bar{\boldsymbol{X}}_i^{(K)}\}\right|=o(n^{-1/2}),$$

for any $d^{opt}\in\mathcal{D}^{opt}$. By Markov's inequality, we obtain

$$\begin{aligned}&\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},\widehat{d}_{\mathcal{I},K,i}),\bar{\boldsymbol{X}}_i^{(K)}\}\\ =\ &\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}^{opt}),\bar{\boldsymbol{X}}_i^{(K)}\}+o_p(n^{-1/2}).\end{aligned}$$

Combining this together with (S2.79) and (S2.80) yields

$$\begin{aligned}&\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}\widehat{V}_i^K(\widehat{d}_{\mathcal{I}}) \qquad\qquad (S2.81)\\ =\ &\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\left(T_i^{(K)}+\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}^{opt}),\bar{\boldsymbol{X}}_i^{(K)}\}\right)+o_p(n^{-1/2})\\ =\ &\frac{1}{n}\sum_{i=1}^{n}T_i^{(K)}+\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{K-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}h_K\{(\bar{\boldsymbol{A}}_i^{(K-1)},d_{K,i}^{opt}),\bar{\boldsymbol{X}}_i^{(K)}\}+o_p(n^{-1/2}).\end{aligned}$$

Let $\varepsilon_i^{(j)}=h_j\{(\bar{\boldsymbol{A}}_i^{(j-1)},d_{j,i}^{opt}),\bar{\boldsymbol{X}}_i^{(j)}\}-h_{j-1,i}$. Similarly, we can show for all $j=2,\dots,K-1$,

$$\begin{aligned}&\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{j-2}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}\left(\frac{\mathrm{g}(A_i^{(j-1)},\widehat{d}_{\mathcal{I},j-1,i})}{\pi_{j-1,i}}\varepsilon_i^{(j)}+h_{j-1}\{(\bar{\boldsymbol{A}}_i^{(j-2)},\widehat{d}_{\mathcal{I},j-1,i}),\bar{\boldsymbol{X}}_i^{(j-1)}\}\right)\\ &=\frac{1}{n}\sum_{i=1}^{n}T_i^{(j)}+\frac{1}{n\binom{n-1}{s_n}}\sum_{i\in\mathcal{I}_0}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\|\mathcal{I}|=s_n}}\prod_{k=1}^{j-2}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}h_{j-1}\{(\bar{\boldsymbol{A}}_i^{(j-2)},d_{j-1,i}^{opt}),\bar{\boldsymbol{X}}_i^{(j-1)}\}+o_p(n^{-1/2}),\end{aligned}$$

where

$$T_i^{(j)}=\mathrm{E}\left(\prod_{k=1}^{j-1}\frac{\mathrm{g}(A_i^{(k)},\widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}}\varepsilon_i^{(j)}\,\middle|\,O_i\right),$$

for any $\mathcal{I} \subseteq \mathcal{I}_{(-i)}$ with $|\mathcal{I}| = s_n$. This together with (S2.81) yields

$$
\begin{aligned}
\widehat{V}_\infty^* &= \frac{1}{n\binom{n-1}{s_n}} \sum_{i\in\mathcal{I}_0} \sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\ |\mathcal{I}|=s_n}} \left\{ \prod_{k=1}^{K-1} \frac{\mathrm{g}(A_i^{(k)}, \widehat{d}_{\mathcal{I},k_i})}{\pi_{k,i}} \widehat{V}_i^{(K)}(\widehat{d}_{\mathcal{I}}) \right. \qquad \text{(S2.82)}\\
&- \left. \sum_{j=1}^{K-1} \prod_{k=1}^{j-1} \frac{\mathrm{g}(A_i^{(k)}, \widehat{d}_{\mathcal{I},k_i})}{\pi_{k,i}} \left(\frac{\mathrm{g}(A_i^{(j)}, \widehat{d}_{\mathcal{I},j,i})}{\pi_{j,i}} h_{j,i} - h_{j-1}\{(\bar{\boldsymbol{A}}_i^{(j-2)}, \widehat{d}_{\mathcal{I},j-1,i}), \bar{\boldsymbol{X}}_i^{(j-1)}\} \right) \right\}\\
&= \frac{1}{n} \sum_{k=2}^{K} T_i^{(k)} + \frac{1}{n} \sum_{i=1}^{n} h_1(d_{1,i}^{opt}, \boldsymbol{X}_i^{(1)}) + o_p(n^{-1/2}).
\end{aligned}
$$

Define

$$
\bar{T}_i^{(j)} = \mathrm{E}\left(\prod_{k=1}^{j-1} \frac{\mathrm{g}(A_i^{(k)}, \widehat{d}_{\mathcal{I},k,i})}{\pi_{k,i}} \{h_j\{(\bar{\boldsymbol{A}}_i^{(j-1)}, \widehat{d}_{\mathcal{I},j,i}), \bar{\boldsymbol{X}}_i^{(j)}\} - h_{j-1,i}\} \middle| O_i \right).
$$

By Condition (C3) and (S2.74), we have

$$
\mathrm{E}\left| T_i^{(j)} - \bar{T}_i^{(j)} \right| \le \frac{1}{c_0^{j-1}} \mathrm{E}\left| h_j\{(\bar{\boldsymbol{A}}_i^{(j-1)}, \widehat{d}_{\mathcal{I},j,i}), \bar{\boldsymbol{X}}_i^{(j)}\} - h_j\{(\bar{\boldsymbol{A}}_i^{(j-1)}, d_{j,i}^{opt}), \bar{\boldsymbol{X}}_i^{(j)}\} \right| = o(n^{-\frac{1}{2}}). \tag{S2.83}
$$

In addition, let

$$
\bar{T}_i^{(1)} = \mathrm{E}\left(h_1(\widehat{d}_{\mathcal{I},1,i}, \boldsymbol{X}_i^{(1)}) \middle| O_i \right),
$$

for any $\mathcal{I} \subseteq \mathcal{I}_{(-i)}$ with $|\mathcal{I}| = s_n$. Similar to the proof of Theorem (2.1), we can show

$$
\max_{i\in\mathcal{I}_0} \mathrm{E}\left| h_1(d_{1,i}^{opt}, \bar{\boldsymbol{X}}_i) - \bar{T}_i^{(1)} \right| = o(n^{-1/2}). \tag{S2.84}
$$

Combining this together with (S2.83) and (S2.82), we have

$$
\widehat{V}_\infty^* = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} \bar{T}_i^{(j)} + o_p(n^{-1/2}).
$$

Notice that $Q_i = \sum_{j=1}^{K} \bar{T}_i^{(j)}, \forall i \in \mathcal{I}_0$. Thus, we've shown (S2.78).

Moreover, it follows from (S2.83) and (S2.84) that

$$
\max_{i\in\mathcal{I}_0} \mathrm{E}\left| Q_i - \sum_{j=2}^{K} T_i^{(j)} - h_1(d_{1,i}^{opt}, \boldsymbol{X}_i^{(1)}) \right| = o(n^{-1/2}).
$$

Therefore, we have

$$\mathrm{E}Q_i = \mathrm{E}\left(\sum_{j=2}^{K} T_i^{(j)} + h_1(d_{1,i}^{opt}, \boldsymbol{X}_i^{(1)})\right) + o(n^{-1/2}) = \mathrm{E}h_1(d_{1,i}^{opt}, \boldsymbol{X}_i^{(1)}) + o(n^{-1/2}). \quad \text{(S2.85)}$$

Notice that $\mathrm{E}h_1(d_{1,i}^{opt}, \boldsymbol{X}_i^{(1)}) = V_0$. Under the condition that $\liminf_n \sigma_n > 0$, it follows from (S2.78) and (S2.85) that

$$\frac{\sqrt{n}}{\sigma_{s_n}}(\widehat{V}_\infty^* - V_0) \xrightarrow{d} N(0,1).$$

Step 4: Using similar arguments in Step 3 of the proof of Theorem 2.2, we can show

$$\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}^{(i)}\}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}_\infty^{*(i)}\}^2 + o_p(1),$$

and

$$\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}^{(i)}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_\infty^{*(i)}\right)^2 + o_p(1),$$

where

$$\widehat{V}_\infty^{*(i)} = \frac{1}{\binom{n-1}{s_n}}\sum_{\substack{\mathcal{I}\subseteq\mathcal{I}_{(-i)}\\ |\mathcal{I}|=s_n}} \widehat{V}_{\{i\}}(\widehat{d}_{\mathcal{I}}).$$

This implies that

$$\widehat{\sigma}_B^2 = \frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}_\infty^{*(i)}\}^2 - \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_\infty^{*(i)}\right)^2 + o_p(1). \quad \text{(S2.86)}$$

Since $s_n = o(n)$, by the ANOVA decomposition (Efron and Stein, 1981), we have

$$\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_\infty^{*(i)} - Q_i\right|^2 = o(1).$$

In addition, by (S2.70) and (S2.71), we can show

$$\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_\infty^{*(i)} + Q_i\right|^2 = O(1).$$

By Cauchy-Schwarz inequality, we obtain

$$\mathrm{E}\left|\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}_\infty^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^{n}Q_i^2\right| \le \max_{i\in\mathcal{I}_0}\mathrm{E}\left|\{\widehat{V}_\infty^{*(i)}\}^2 - Q_i^2\right|$$
$$\le \sqrt{\max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_\infty^{*(i)} - Q_i\right|^2 \max_{i\in\mathcal{I}_0}\mathrm{E}\left|\widehat{V}_\infty^{*(i)} + Q_i\right|^2} = o(1).$$

It follows from Markov's inequality that

$$\frac{1}{n-1}\sum_{i=1}^{n}\{\widehat{V}_{\infty}^{*(i)}\}^2 - \frac{1}{n-1}\sum_{i=1}^{n} Q_i^2 = o_p(1).$$

Similarly, we can show

$$\left(\frac{1}{n}\sum_{i=1}^{n}\widehat{V}_{\infty}^{*(i)}\right)^2 = \left(\frac{1}{n}\sum_{i=1}^{n} Q_i\right)^2 + o_p(1).$$

In view of (S2.86), we've shown

$$\widehat{\sigma}_B^2 = \frac{1}{n-1}\sum_{i=1}^{n} Q_i^2 - \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} Q_i\right)^2 + o_p(1).$$

In addition, it follows from the law of large numbers that

$$\frac{1}{n-1}\sum_{i=1}^{n} Q_i^2 - \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{n} Q_i\right)^2 = \sigma_{s_n}^2 + o_p(1).$$

Thus, we have $\widehat{\sigma}_B^2 = \sigma_{s_n}^2 + o_p(1)$. The proof is hence completed.

S2.6 Proofs of Lemmas in the supplementary appendix

S2.6.1 Proof of Lemma S2.1

By Condition (A1) and (A2), we have for any $\boldsymbol{x} \in \mathbb{X}$ and $a = 0, 1$,

$$h(a, \boldsymbol{x}) = \mathrm{E}(Y_0|A_0 = a, \boldsymbol{X}_0 = \boldsymbol{x}) = \mathrm{E}\{Y_0^*(a)|A_0 = a, \boldsymbol{X}_0 = \boldsymbol{x}\} = \mathrm{E}\{Y_0^*(a)|\boldsymbol{X}_0 = \boldsymbol{x}\}.$$

Condition (A4) states that

$$\sup_{\boldsymbol{x}\in\mathbb{X}, a=0,1} \mathrm{E}[\{Y_0^*(a)\}^2|\boldsymbol{X}_0 = \boldsymbol{x}] \le \bar{c}^*,$$

for some constant $\bar{c}^* > 0$. Therefore,

$$\sup_{\boldsymbol{x}\in\mathbb{X}, a=0,1} |h(a, \boldsymbol{x})| \le \sqrt{\sup_{\boldsymbol{x}\in\mathbb{X}, a=0,1} \mathrm{E}[\{Y_0^*(a)\}^2|\boldsymbol{X}_0 = \boldsymbol{x}]} \le \sqrt{\bar{c}^*}.$$

Lemma S2.1 thus holds by setting the constant $C_0 = \sqrt{\bar{c}^*}$.

S2.6.2 Proof of Lemma S2.2

Let $p_0 = \Pr(A_0 = 1)$. By Condition (A3), we have

$$0 < c_0 \leq p_0 \leq 1 - c_0 < 1. \tag{S2.87}$$

Consider the event

$$\mathcal{A}_* = \{c_0 n/2 < n_A < (1 - c_0/2)n\}.$$

It follows from Hoeffding's inequality (Hoeffding, 1963) that

$$\Pr(\mathcal{A}_*^c) \leq \Pr\left(|n_A - np_0| \leq c_0 n/3\right) \leq 2\exp\left(-\frac{18n}{c_0^2}\right) \to 0. \tag{S2.88}$$

Note that the random variable $n_{\mathcal{S}}$ is completely determined by n_A. For $s_n < n_A < n - s_n$, we have

$$\frac{\binom{n}{s_n} - n_{\mathcal{S}}}{\binom{n}{s_n}} = \frac{\sum_{k=0}^{N_0-1}\binom{n_A}{s_n-k}\binom{n-n_A}{k} + \sum_{k=0}^{N_0-1}\binom{n_A}{k}\binom{n-n_A}{s_n-k}}{\binom{n}{s_n}}. \tag{S2.89}$$

Let $m^{(s)} = m(m-1)\cdots(m-s+1)$ for any integers $m \geq s > 0$, we have for any $0 \leq k \leq N_0 - 1 \leq s_n$,

$$\frac{\binom{n_A}{s_n-k}\binom{n-n_A}{k}}{\binom{n}{s_n}} = \binom{s_n}{k}\frac{n_A^{(s_n-k)}(n-n_A)^{(k)}}{n^{(s_n)}} \leq \frac{n_A^{(s_n-k)}(n-n_A)^{(k)}}{n^{(s_n)}} \leq \frac{n_A^{s_n-k}(n-n_A)^k}{(n-s_n+1)^{s_n}}.$$

Since $s_n = o(n)$, for sufficiently large n, we have $n - s_n + 1 \geq (1 - c_0/3)n$. Thus, under the event defined in $\mathcal{A}_*$, we have

$$\frac{\binom{n_A}{s_n-k}\binom{n-n_A}{k}}{\binom{n}{s_n}} \leq \left(\frac{1-c_0/2}{1-c_0/3}\right)^{s_n}, \qquad \forall 0 \leq k \leq N_0 - 1 \leq s_n.$$

Similarly, we can show

$$\frac{\binom{n_A}{k}\binom{n-n_A}{s_n-k}}{\binom{n}{s_n}} \leq \left(\frac{1-c_0/2}{1-c_0/3}\right)^{s_n}, \qquad \forall 0 \leq k \leq N_0 - 1 \leq s_n,$$

under the event defined in $\mathcal{A}_*$.

By (S2.89), we obtain

$$\frac{\binom{n}{s_n} - n_{\mathcal{S}}}{\binom{n}{s_n}} \leq 2N_0 p_*^{s_n},$$

under the event defined in $\mathcal{A}_*$, where $p_* = (1 - c_0/2)/(1 - c_0/3)$. Notice that N_0 is a fixed constant. Under the given conditions, we have $s_n \asymp n^{\beta_0}$. Set $c_3 = 18c_0^{-2}$, it follows from (S2.88) that

$$\Pr\left(\frac{\binom{n}{s_n} - n_{\mathcal{S}}}{\binom{n}{s_n}} \le c_1 p_*^{c_2 n^{\beta_0}} \right) \ge \Pr(\mathcal{A}_*) \ge 1 - 2\exp(-c_3 n) \to 1,$$

for some constants $c_1, c_2 > 0$. This completes the proof of (S2.3).

For any $i \in \{1, \ldots, n\}$, define $\mathcal{S}_{N_0,s_n}^{(i)} = \{\mathcal{I} \in \mathcal{S}_{N_0,s_n} : i \notin \mathcal{I}\}$ and $n_{\mathcal{S}}^{(i)} = |\mathcal{S}_{N_0,s_n}^{(i)}|$. Similar to (S2.3), there exist some constants $c_1^*, c_2^*, c_3^* > 0$ and $0 < p_{**} < 1$ such that

$$\Pr\left(\frac{\binom{n-1}{s_n} - n_{\mathcal{S}}^{(i)}}{\binom{n-1}{s_n}} \le c_1^* p_{**}^{c_2^* n^{\beta_0}} \right) \ge 1 - 2\exp(-c_3^* n). \tag{S2.90}$$

Let $\mathcal{A}^{(i)}$ be the event defined in (S2.90). Set $c_4 = \min(c_3^*, c_3)$, it follows from Bonferroni's inequality that

$$\Pr(\mathcal{A}^{(i)} \cap \mathcal{A}_*) \ge 1 - \Pr(\mathcal{A}^{(i)}) - \Pr(\mathcal{A}_*) \ge 1 - 4\exp(-c_4 n).$$

Under the events defined in $\mathcal{A}^{(i)}$ and $\mathcal{A}_*$, we have

$$\begin{aligned}
&\left| \frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}} - \frac{\binom{n-1}{s_n}}{\binom{n}{s_n}} \right| = \frac{\binom{n-1}{s_n}}{\binom{n}{s_n}} \left| \frac{n_{\mathcal{S}}^{(i)}/\binom{n-1}{s_n} - n_{\mathcal{S}}/\binom{n}{s_n}}{n_{\mathcal{S}}/\binom{n}{s_n}} \right| = \frac{n - s_n}{n} \left| \frac{n_{\mathcal{S}}^{(i)}/\binom{n-1}{s_n} - n_{\mathcal{S}}/\binom{n}{s_n}}{n_{\mathcal{S}}/\binom{n}{s_n}} \right| \\
\le\ & \frac{2(n - s_n)}{n} \left| \frac{n_{\mathcal{S}}^{(i)}}{\binom{n-1}{s_n}} - \frac{n_{\mathcal{S}}}{\binom{n}{s_n}} \right| \le \frac{2(n - s_n)}{n} \left(\left| \frac{n_{\mathcal{S}}^{(i)}}{\binom{n-1}{s_n}} - 1 \right| + \left| \frac{n_{\mathcal{S}}}{\binom{n}{s_n}} - 1 \right| \right) \\
\le\ & 2(c_1 p_*^{c_2 n^{\beta_0}} + c_1^* p_{**}^{c_2^* n^{\beta_0}}) \ll \frac{\sqrt{\log n}}{\sqrt{n}},
\end{aligned}$$

where the first inequality is due to (S2.11), the third inequality follows by the definitions of $\mathcal{A}^{(i)}$ and $\mathcal{A}_*$.

Conditional on $\{O_i\}_{i \in \mathcal{I}_0}$, the random variables $\mathbb{I}(i \notin \mathcal{I}_1), \ldots, \mathbb{I}(i \notin \mathcal{I}_B)$ are independent Bernoulli random variables with mean $\Pr(i \notin \mathcal{I}_b | \{O_i\}_{i \in \mathcal{I}_0}) = n_{\mathcal{S}}^{(i)}/n_{\mathcal{S}}$. Hence, it follows from Hoeffding's inequality that

$$\Pr\left(\left| \frac{n^{(i)}}{B} - \frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}} \right| \le \frac{\sqrt{\log B}}{\sqrt{2B}} \middle| \{O_i\}_{i \in \mathcal{I}_0} \right) \ge 1 - 2\exp\left(-2\log B/2\right) = 1 - \frac{2}{B}.$$

Therefore, we have

$$\Pr\left(\left|\frac{n^{(i)}}{B} - \frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}}\right| \leq \frac{\sqrt{\log B}}{\sqrt{2B}}\right) \geq 1 - \frac{2}{B}. \tag{S2.91}$$

Since $B \gg n$, we have $\sqrt{\log B}/\sqrt{2B} \leq \sqrt{\log n}/\sqrt{2n}$. Under the events defined (S2.91), $\mathcal{A}^{(i)}$ and $\mathcal{A}_*$, we have

$$\left|\frac{n^{(i)}}{B} - \frac{n - s_n}{n}\right| = \left|\frac{n^{(i)}}{B} - \frac{\binom{n-1}{s_n}}{\binom{n}{s_n}}\right| \leq \left|\frac{n^{(i)}}{B} - \frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}}\right| + \left|\frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}} - \frac{\binom{n-1}{s_n}}{\binom{n}{s_n}}\right| \leq \frac{\sqrt{\log n}}{\sqrt{n}}.$$

The proof is hence completed by noting that

$$\begin{aligned}
\Pr\left(\left|\frac{n^{(i)}}{B} - \frac{n - s_n}{n}\right| > \frac{\sqrt{\log n}}{\sqrt{n}}\right) &\leq \Pr\left(\left|\frac{n^{(i)}}{B} - \frac{n_{\mathcal{S}}^{(i)}}{n_{\mathcal{S}}}\right| \leq \frac{\sqrt{\log B}}{\sqrt{2B}}\right) + \Pr\left(\mathcal{A}_*^c \cup (\mathcal{A}^{(i)})^c\right) \\
&\leq \frac{2}{B} + 4\exp(-c_4 n).
\end{aligned}$$

References

Efron, B. and C. Stein (1981). The jackknife estimate of variance. *Ann. Statist. 9*(3), 586–596.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc. 58*, 13–30.

Pólya, G. (1964). G. szeg o. aufgaben und lehrs atze aus der analysis, 2 vol.