Estimation of subsidiary performance metrics under optimal policies

Zhaoqi Li*, Houssam Nassif†, Alex Luedtke*

*University of Washington, †Meta

Supplementary Material

This supplementary material provides proofs for lemmas and theorems in Sections 2 and 3, a detailed description of multiplier bootstrap used in our simulations, as well as additional simulation experiments.

A. Proofs for Section 2

Let $Q_{X,0}$ be the marginal distribution of X under P_0, and let $Q_{Y^*,0}$ and $Q_{Y^\dagger,0}$ be respectively the conditional distribution of Y^* and $Y^\dagger$ given A, X under P_0. Let $\{P_\epsilon : \epsilon \in \mathbb{R}\} \subset \mathcal{M}$ be a parametric submodel that is such that $P_\epsilon = P_0$ when $\epsilon = 0$. This submodel is defined so that the marginal distribution of X and the conditional distributions of $Y^\dagger$ and Y^* given (A, X) satisfy

$$dQ_{X,\epsilon}(x) = (1 + \epsilon S_X(x))dQ_{X,0}(x), \tag{S1.1}$$

where $\mathbb{E}_0\left[S_X(x)\right]=0$ and $\sup_x |S_X(x)| \leq m < \infty$,

$$dQ_{Y^\dagger,\epsilon}(z \mid a,x) = \left(1+\epsilon S_{Y^\dagger}(z \mid a,x)\right) dQ_{Y^\dagger,0}(z \mid a,x), \tag{S1.2}$$

where $\mathbb{E}_0\left[S_{Y^\dagger} \mid A,X\right]=0$ P_0-a.s. and $\sup_{x,a,z} |S_{Y^\dagger}(z \mid a,x)| < \infty$, and

$$dQ_{Y^*,\epsilon}(y \mid a,x) = \left(1+\epsilon S_{Y^*}(y \mid a,x)\right) dQ_{Y^*,0}(y \mid a,x), \tag{S1.3}$$

where $\mathbb{E}_0\left[S_{Y^*} \mid A,X\right]=0$ P_0-a.s. and $\sup_{x,a,y} |S_{Y^*}(y \mid a,x)| < \infty$.

We let $q_{b,\epsilon}(x) = q_b(P_\epsilon)(x)$ and $s_{b,\epsilon}(x) = s_b(P_\epsilon)(x)$.

A.1 Proof of Lemma 1

Proof of Lemma 1. Note that $\pi^*_P(x) = \mathbb{I}\{q_b(P)(x) > 0\}$ for all $x \in \mathcal{X}$. Following the idea of the proof of Theorem 3 in Luedtke and Van Der Laan [2016], we observe that

$$\Psi^*(P) - \mathbb{E}_P\mathbb{E}_P[Y^\dagger \mid A=0,X] = \mathbb{E}_P\left[\pi^*_P(X)s_b(P)(X)\right].$$

By a telescoping argument,

$$\begin{aligned}
\Psi^*(P_\epsilon) - \Psi^*(P_0) &= \mathbb{E}_{P_\epsilon}\mathbb{E}_{P_\epsilon}[Y^\dagger|A=\pi^*_{P_\epsilon}(X),X] - \mathbb{E}_{P_0}\mathbb{E}_{P_0}[Y^\dagger|A=\pi^*(X),X] \\
&= \mathbb{E}_{P_\epsilon}\mathbb{E}_{P_\epsilon}[Y^\dagger|A=\pi^*_{P_\epsilon}(X),X] - \mathbb{E}_{P_\epsilon}\mathbb{E}_{P_\epsilon}[Y^\dagger|A=\pi^*(X),X] \\
&\quad + \mathbb{E}_{P_\epsilon}\mathbb{E}_{P_\epsilon}[Y^\dagger|A=\pi^*(X),X] - \mathbb{E}_{P_0}\mathbb{E}_{P_0}[Y^\dagger|A=\pi^*(X),X] \\
&= \mathbb{E}_{P_\epsilon}[(\mathbb{I}(q_{b,\epsilon}>0) - \mathbb{I}(q_{b,0}>0))\cdot s_{b,\epsilon}] + \Psi_{\pi^*}(P_\epsilon) - \Psi_{\pi^*}(P_0).
\end{aligned} \tag{S1.4}$$

It is known that for a fixed π, Ψ_π is pathwise differentiable with gradient $D(\pi, P_0)$. We shall now show that the first term is $o(\epsilon)$. Letting $B_1 := \{x \in \mathcal{X} : q_{b,0}(x) = 0\}$, we have

$$\begin{aligned}
&\mathbb{E}_{P_\epsilon}\left[\left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right)\right) s_{b,\epsilon}\right] \\
&= \int_{\mathcal{X}\setminus B_1} \left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right)\right) s_{b,\epsilon} dQ_{X,\epsilon} \\
&\quad + \int_{B_1} \left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right)\right) s_{b,\epsilon} dQ_{X,\epsilon}.
\end{aligned}$$

Under Condition 1, we know that $\Pr_0(q_{b,0}(X) \neq 0) = 1$, so the second term is zero. Then we aim to show that the first term is $o(|\epsilon|)$. Note that

$$\begin{aligned}
&\left|\int_{\mathcal{X}\setminus B_1} \left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right)\right) s_{b,\epsilon} dQ_{X,\epsilon}\right| \\
&\leq \int_{\mathcal{X}\setminus B_1} \left|\left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right)\right) s_{b,\epsilon}\right| dQ_{X,\epsilon} \\
&\leq \int_{\mathcal{X}\setminus B_1} I\left(|q_{b,0}| < |q_{b,\epsilon} - q_{b,0}|\right) |s_{b,\epsilon}| \, dQ_{X,\epsilon}
\end{aligned}$$

by looking at the sign of $q_{b,\epsilon}$ and $q_{b,0}$. Also,

$$\begin{aligned}
q_{b,\epsilon}(x) &= \int y \left(dQ_{Y^*,\epsilon}(y \mid A = 1, X = x) - dQ_{Y^*,\epsilon}(y \mid A = 0, X = x)\right) \\
&= q_{b,0}(x) + \epsilon(\mathbb{E}_0\left[Y^* S_{Y^*}(Y^* \mid 1, X) \mid A = 1, X = x\right] \\
&\qquad - \mathbb{E}_0\left[Y^* S_{Y^*}(Y^* \mid 0, X) \mid A = 0, X = x\right]) \\
&= q_{b,0}(x) + \epsilon \bar{h}(x)
\end{aligned}$$

where

$$\bar{h}(x) = \mathbb{E}_0[Y^* S_{Y^*}(Y^*|1, X)|A = 1, X = x] - \mathbb{E}_0[Y^* S_{Y^*}(Y^*|0, X)|A = 0, X = x].$$

Similarly, $s_{b,\epsilon}(x) = s_{b,0}(x) + \epsilon \cdot \tilde{h}(x)$ where

$$\tilde{h}(x) = \mathbb{E}_0[Y^\dagger S_{Y^\dagger}(Y^\dagger|1, X)|A = 1, X = x] - \mathbb{E}_0[Y^\dagger S_{Y^\dagger}(Y^\dagger|0, X)|A = 0, X = x].$$

Note that $\tilde{h}$ and $\bar{h}$ are uniformly bounded since Y^*, $Y^\dagger$, S_{Y^*}, and $S_{Y^\dagger}$ are bounded. Let $H = \max\{\sup_x |\bar{h}(x)|,$
$\sup_x |\tilde{h}(x)|\}$. Therefore,

$$\begin{aligned}
&\int_{\mathcal{X}\setminus B_1} I\left(|q_{b,0}| < |q_{b,\epsilon} - q_{b,0}|\right) |s_{b,\epsilon}| \, dQ_{X,\epsilon} \\
&\leq \int_{\mathcal{X}\setminus B_1} I\left(|q_{b,0}| < H|\epsilon|\right) \left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,\epsilon} \\
&\leq (1 + m|\epsilon|) \int_{\mathcal{X}\setminus B_1} I\left(|q_{b,0}| < H|\epsilon|\right) \left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0} \\
&= (1 + m|\epsilon|) \int_{\mathcal{X}\setminus B_1} I\left(0 < |q_{b,0}| < H|\epsilon|\right) \left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0}.
\end{aligned}$$

Denote $\tilde{\mathcal{X}} = \mathcal{X} \setminus B_1$. Under the first condition, define the set

$$B_{2,t} = \{x \in \tilde{\mathcal{X}} : |s_{b,0}(x)| < Ct^{-1}|q_{b,0}(x)|\}.$$

Then

$$\begin{aligned}
&\int_{\mathcal{X}\setminus B_1} I\left(|q_{b,0}| < H|\epsilon|\right) \left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0} \\
&= \int_{\tilde{\mathcal{X}}} I\left(0 < |q_{b,0}| < H|\epsilon|\right) \left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0}
\end{aligned}$$

$$= \int_{B_{2,t}} I\left(0 < |q_{b,0}| < H|\epsilon|\right)\left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0}$$
$$+ \int_{\tilde{\mathcal{X}}\setminus B_{2,t}} I\left(0 < |q_{b,0}| < H|\epsilon|\right)\left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0}.$$

On one hand, note that for $x \in B_{2,t}$ and under the fact that $|q_{b,0}(x)| \leq H|\epsilon|$ we have $|s_b(x)| \leq CHt^{-1}|\epsilon|$. define C_2 such that $P_0(0 < |q_{b,0}(X)| < t) \leq C_2 t^{\gamma}$ for any $t > 0$, the first term

$$\int_{B_{2,t}} I\left(0 < |q_{b,0}| < H|\epsilon|\right)\left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0} \tag{S1.5}$$
$$\leq \int_{B_{2,t}} I\left(0 < |q_{b,0}| < H|\epsilon|\right)\left(CHt^{-1}|\epsilon| + H|\epsilon|\right) dQ_{X,0}$$
$$\leq \left(CHt^{-1}|\epsilon| + H|\epsilon|\right) P_0\left(0 < |q_{b,0}(X)| < H|\epsilon|\right)$$
$$\leq \left(Ct^{-1}|\epsilon| + H|\epsilon|\right) C_2 (H|\epsilon|)^{\gamma} \tag{S1.6}$$

for $t < 1$. For the second term, let $C_3 := \sup_x |s_{b,0}(x)|$, we have

$$\int_{\tilde{\mathcal{X}}\setminus B_{2,t}} I\left(0 < |q_{b,0}| < H|\epsilon|\right)\left(|s_{b,0}| + H|\epsilon|\right) dQ_{X,0}$$
$$\leq (C_3 + H|\epsilon|) P_0(|s_{b,0}(X)| > Ct^{-1}|q_{b,0}(X)|)$$
$$\leq (C_3 + H|\epsilon|) t^{\zeta}$$

where the last inequality follows from Condition 1. Therefore, the sum is bounded by

$$\left(Ct^{-1}|\epsilon| + H|\epsilon|\right) C_2 (H|\epsilon|)^{\gamma} + (C_3 + H|\epsilon|) t^{\zeta}.$$

Taking $t = |\epsilon|^{\frac{1+\gamma}{\zeta+1}}$ gives that this is $O(|\epsilon|^{1+\gamma-\frac{1+\gamma}{\zeta+1}})$, which is $o(|\epsilon|)$ given that $\gamma > \frac{1}{\zeta}$. Combining all of the results above gives

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbb{E}_{P_\epsilon} \left[\left(I\left(q_{b,\epsilon} > 0\right) - I\left(q_{b,0} > 0\right) \right) s_{b,\epsilon} \right] = 0.$$

Therefore, Ψ^* is pathwise differentiable, and, per (S1.4), has the same canonical gradient as the parameter Ψ_{π^*}, namely $D(\pi^*, P_0)$. □

A.2 Proof of Theorem 1

Proof of Theorem 1. We would first like to show that $\psi_{OS,n}$ is an asymptotically linear estimator of ψ_0. For simplicity of notation, we let $\pi_n^* := \pi_{\widehat{P}_n}^*$ and drop the dependence of π in the definition of Ψ_π in this proof. Note that $\psi_{OS,n} - \psi_0 = (P_n - P_0)D(P_0) + (P_n - P_0)[D(\widehat{P}_n) - D(P_0)] + R(\widehat{P}_n, P_0)$. Note that the first term $(P_n - P_0)D(P_0)$ is the linear term and $(P_n - P_0)[D(\widehat{P}_n) - D(P_0)] = o_{P_0}(n^{-1/2})$ under the Donsker condition and the fact that $\|D(\widehat{P}_n) - D(P_0)\|_2 \xrightarrow{p} 0$ (Lemma 19.24 of Van der Vaart [2000]). To show that $\psi_{OS,n}$ is asymptotically linear, we only need to argue that the remainder term $R(\widehat{P}_n, P_0)$ is $o_{P_0}(n^{-1/2})$. Note that

$$\begin{aligned} P_0 D(\widehat{P}_n) &= \mathbb{E}_0 \left[\frac{\mathbb{I}\{A = \pi_n^*(X)\}}{p_n(A|X)} (Y^\dagger - s(A, X)) + s(\pi_n^*(X), X) - \Psi(\widehat{P}_n) \right] \\ &= \mathbb{E}_0 \left[\frac{\mathbb{I}\{A = \pi_n^*(X)\}}{p_n(A|X)} (s_0(A, X) - s(A, X)) + s(\pi_n^*(X), X) - \Psi(\widehat{P}_n) \right], \end{aligned}$$

by the law of total expectation. Therefore,

$$
\begin{aligned}
R(\widehat{P}_n, P_0) &= \Psi(\widehat{P}_n) - \Psi(P_0) + P_0 D(\widehat{P}_n) \\
&= \int \left\{ \frac{\mathbb{I}\{a = \pi_n^*(x)\}}{p_n(a|x)} (s_0(a,x) - s_n(a,x)) + s_n(\pi_n^*(x), x) - s_0(\pi^*(x), x) \right\} dP_0(a,x) \\
&= \int \left(\frac{\mathbb{I}\{a = \pi_n^*(x)\}}{p_n(a|x)} - 1 \right) [s_0(\pi_n^*(x), x) - s_n(\pi_n^*(x), x)] dP_0(a,x) + \Psi_{\pi_n^*}(P_0) - \Psi_{\pi^*}(P_0) \\
&= \iint \left(\frac{\mathbb{I}\{a = \pi_n^*(x)\}}{p_n(a|x)} - 1 \right) [s_0(\pi_n^*(x), x) - s_n(\pi_n^*(x), x)] p_0(a|x) da\, dP_0(x) \\
&\quad + \Psi_{\pi_n^*}(P_0) - \Psi_{\pi^*}(P_0) \\
&= \int \left(\frac{p_0(\pi_n^*(x)|x)}{p_n(\pi_n^*(x)|x)} - 1 \right) [s_0(\pi_n^*(x), x) - s_n(\pi_n^*(x), x)] dP_0(x) \\
&\quad + \Psi_{\pi_n^*}(P_0) - \Psi_{\pi^*}(P_0) \\
&=: R_{1n} + R_{2n}.
\end{aligned}
$$

The first term R_{1n} is $o_{P_0}(n^{-1/2})$ under under Condition 4 — see Proposition 1. As for the second term R_{2n}, Proposition 2 shows that it is $o_{P_0}(n^{-1/2})$ under the margin condition. □

Proposition 1. *Under Condition 4, $R_{1n} = o_{P_0}(n^{-1/2})$.*

Proof. By Jensen's inequality, the fact that $\pi_n^*(x) \in \{0, 1\}$ for all x, the fact that $(b + c) \leq 2 \max\{b, c\}$ for $b, c \in \mathbb{R}$, and Cauchy-Schwarz, we have that

$$
\begin{aligned}
|R_{1n}| &= \left| \int \left(\frac{p_0(\pi_n^*(x)|x)}{p_n(\pi_n^*(x)|x)} - 1 \right) [s_0(\pi_n^*(x), x) - s_n(\pi_n^*(x), x)] dP_0(x) \right| \\
&\leq \int \left| \left(\frac{p_0(\pi_n^*(x)|x)}{p_n(\pi_n^*(x)|x)} - 1 \right) [s_0(\pi_n^*(x), x) - s_n(\pi_n^*(x), x)] \right| dP_0(x) \\
&\leq \int \sum_{a=0}^{1} \left| \left(\frac{p_0(a|x)}{p_n(a|x)} - 1 \right) [s_0(a, x) - s_n(a, x)] \right| dP_0(x)
\end{aligned}
$$

$$= \sum_{a=0}^{1} \int \left| \left(\frac{p_0(a|x)}{p_n(a|x)} - 1 \right) [s_0(a,x) - s_n(a,x)] \right| dP_0(x)$$
$$\leq 2 \max_{a \in \{0,1\}} \int \left| \left(\frac{p_0(a|x)}{p_n(a|x)} - 1 \right) [s_0(a,x) - s_n(a,x)] \right| dP_0(x)$$
$$\leq 2 \max_{a \in \{0,1\}} \left\{ \left\| \frac{p_0(a \mid X)}{p_n(a \mid X)} - 1 \right\|_{2,P_0} \|s_n(a,X) - s_0(a,X)\|_{2,P_0} \right\}.$$

□

The following proposition shows that the second term R_{2n} is $o_{P_0}(n^{-1/2})$ under our margin condition.

Proposition 2. *Assume Conditions 1, 2, and 3 hold. Then, for any $\epsilon > 0$, $|R_{2n}| = o_{P_0}(n^{-1/2})$.*

Proof. We adopt the idea in proof of Theorem 8 of Luedtke and Van Der Laan [2016]. Let $B'_{3,u} = \{x \in \mathcal{X} : |s_{b,0}(x)| < C_1 u |q_{b,0}(x)|\}$ and $A_u = \{x \in \mathcal{X} : C_1 u |q_{b,0}(x)| \leq |s_{b,0}(x)| < C_1(u+1)\,|q_{b,0}(x)|\}$. Then for any $t > 0$,

$$|\Psi_{\pi_n^*}(P_0) - \Psi_{\pi^*}(P_0)|$$
$$= \mathbb{E}_{P_0}\left[s_{b,0}(X)(\pi_n^*(X) - \pi^*(X))\right]$$
$$\leq \mathbb{E}_0\left[|s_{b,0}(X)|\, I\left(\pi^*(X) \neq \pi_n^*(X)\right)\right]$$
$$= \sum_{u=0}^{\infty} \mathbb{E}_0[|s_{b,0}(X)| I(\pi^*(X) \neq \pi_n^*(X)) I(A_u)]$$
$$\leq \sum_{u=0}^{\infty} \mathbb{E}_0[|s_{b,0}(X)| I(|q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|) I(A_u)].$$

where the last inequality follows from the fact that for any $x \in \mathcal{X}$, $\pi^*(x) \neq \pi_n^*(x)$ implies that $|q_{b,n}(x) - q_{b,0}(x)| \geq |q_{b,0}(x)|$. From Condition 1 we know that $q_{b,0}(X) \neq 0$ with P_0-probability 1, so

$$\begin{aligned}
&\sum_{u=0}^{\infty} \mathbb{E}_0[|s_{b,0}(X)|I(|q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|)I(A_u)] \\
&= \sum_{u=0}^{\infty} \mathbb{E}_0[|s_{b,0}(X)|I(0 < |q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|)I(A_u)].
\end{aligned}$$

For any $x \in A_u$, $|s_{b,0}(x)| \leq C_1(u+1)|q_{b,0}(x)|$, so for each u,

$$\begin{aligned}
&\mathbb{E}_0\left[|s_{b,0}(X)|\, I\left(0 < |q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|\right) I(A_u)\right] \\
&\leq C_1 \mathbb{E}_0\left[(u+1)\, |q_{b,0}(X)|\, I\left(0 < |q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|\right) I(A_u)\right] \\
&\leq C_1 \mathbb{E}_0\left[(u+1)\, |q_{b,n}(X) - q_{b,0}(X)|\, I\left(0 < |q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|\right) I(A_u)\right] \\
&\leq C_1 \mathbb{E}_0\left[(u+1) \max_{x \in \mathcal{X}} \|q_{b,n}(x) - q_{b,0}(x)\|\, I\left(0 < |q_{b,0}(X)| \leq \max_{x \in \mathcal{X}} \|q_{b,n}(x) - q_{b,0}(x)\|\right) I(A_u)\right] \\
&= C_1(u+1) \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0} \mathbb{E}_0\left[I\left(0 < |q_{b,0}(X)| \leq \max_{x \in \mathcal{X}} \|q_{b,n}(x) - q_{b,0}(x)\|\right) I(A_u)\right] \\
&= C_1(u+1) \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0} P_0(0 < |q_{b,0}(X)| \leq \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0}, A_u).
\end{aligned}$$

For an event $\mathcal{E} \subseteq \mathcal{X}$, let $\mathbb{P}^\infty(\mathcal{E}) := P_0(0 < |q_{b,0}(X)| \leq \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0}, \mathcal{E})$. Then, for any $k \in \mathbb{N}$,

$$\begin{aligned}
&\sum_{u=0}^{k} \mathbb{E}_0[|s_{b,0}(X)|I(0 < |q_{b,0}(X)| \leq |q_{b,n}(X) - q_{b,0}(X)|)I(A_u)] \\
&\leq \sum_{u=0}^{k} C_1(u+1) \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0} \mathbb{P}^\infty(A_u) \\
&= \sum_{u=0}^{k} C_1(u+1) \left\|q_{b,n} - q_{b,0}\right\|_{\infty,P_0} [\mathbb{P}^\infty(B'_{3,u+1}) - \mathbb{P}^\infty(B'_{3,u})]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{u=0}^{k} C_1(u+1)\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,u+1}) - \sum_{u=0}^{k} C_1(u+1)\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,u}) \\
&= \sum_{u=1}^{k+1} C_1 u\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,u}) - \sum_{u=0}^{k} C_1(u+1)\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,u}) \\
&= C_1(k+1)\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,k+1}) - \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\mathbb{P}^{\infty}(B'_{3,u}) \\
&= \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\left[\mathbb{P}^{\infty}(B'_{3,k+1}) - \mathbb{P}^{\infty}(B'_{3,u})\right] \\
&\leq \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\left[\mathbb{P}^{\infty}(\mathcal{X}) - \mathbb{P}^{\infty}(B'_{3,u})\right] \\
&= \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\left[\mathbb{P}^{\infty}(B'^{c}_{3,u})\right] \\
&= \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}\left[\mathbb{P}_0(0<|q_{b,0}(X)|\leq \left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}, B'^{c}_{3,u})\right] \\
&\leq \sum_{u=0}^{k} C_1\left\|q_{b,n}-q_{b,0}\right\|_{\infty,P_0}^{1+\gamma/2} u^{-\zeta/2}.
\end{aligned}
$$

where the last step follows from Holder's inequality. Since $\zeta > 2$, let $k \to \infty$ and the infinite sum converges. Therefore,

$$
\begin{aligned}
&\left|\Psi_{\pi_n^*}(P_0) - \Psi_{\pi^*}(P_0)\right| \\
&= \sum_{u=1}^{\infty} \mathbb{E}_0[|s_{b,0}(X)|I(\pi^*(X)\neq \pi_n^*(X))|A_u]\mathbb{P}(A_u) \\
&= \lim_{k\to\infty}\sum_{u=1}^{k} \mathbb{E}_0[|s_{b,0}(X)|I(\pi^*(X)\neq \pi_n^*(X))|A_u]\mathbb{P}(A_u) \lesssim \left\|q_{b,n}-q_{b,0}\right\|_{p,P_0}^{1+\gamma/2}.
\end{aligned}
$$

Note that under Condition 3, we have $\|q_{b,n}-q_{b,0}\|_{\infty,P_0}^{1+\gamma/2} = o_{P_0}(n^{-1/2})$ for any $\gamma > 0$, so $|R_{2n}| = o_{P_0}(n^{-1/2})$. □

B. Proofs for Section 3

For notational simplicity, throughout this section and later we denote $\psi_\pi := \Psi_\pi(P_0)$ for some policy $\pi \in \Pi$.

B.1 Proof of Lemma 2

Proof of Lemma 2. We have that

$$\left\{\Pi^* \subseteq \widehat{\Pi}_\beta\right\} = \left\{\omega_{\pi'} < \sup_{\pi\in\Pi} \omega_\pi, \forall \pi' \in \widehat{\Pi}_\beta^C\right\}.$$

Therefore,

$$\begin{aligned}
&\left\{\Pi^* \subseteq \widehat{\Pi}_\beta\right\}^C \\
&= \left\{\exists \pi' \in \widehat{\Pi}_\beta^C : \omega_{\pi'} = \sup_{\pi\in\Pi} \omega_\pi\right\} \\
&\subseteq \left\{\exists \pi' \in \widehat{\Pi}_\beta^C : \left[\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\beta}{n^{1/2}} + L_n\right] > \sup_{\pi\in\Pi} \omega_\pi,\right\} \\
&= \left\{\exists \pi' \in \widehat{\Pi}_\beta^C : \left[\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\beta}{n^{1/2}}\right] > \sup_{\pi\in\Pi} \omega_\pi - L_n\right\}, \qquad \text{(S2.7)}
\end{aligned}$$

where the inclusion follows from the definition of $\widehat{\Pi}_\beta$. Let $\mathcal{A}$ denote the event $\{L_n \le \sup_{\pi\in\Pi} \omega_\pi\} \cap \left[\cap_{\pi\in\Pi} \left\{\omega_\pi \le \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\}\right]$. Hence, (S2.7) shows that

$$\begin{aligned}
&\left\{\Pi^* \not\subseteq \widehat{\Pi}_\beta\right\}^C \\
&\subseteq \left[\left\{\exists \pi' \in \widehat{\Pi}_\beta^C : \left[\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\beta}{n^{1/2}}\right] > \sup_{\pi\in\Pi} \omega_\pi - L_n\right\} \cap \mathcal{A}\right] \cup \mathcal{A}^C
\end{aligned}$$

$$\subseteq \left[\left\{\exists \pi' \in \widehat{\Pi}_\beta^C : \omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\beta}{n^{1/2}} > 0\right\} \cap \mathcal{A}\right] \cup \mathcal{A}^C$$

$$= \mathcal{A}^C,$$

where the final equality used that the leading event in the union above is equal to the null set since under $\mathcal{A}$, we have $\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\beta}{n^{1/2}} \leq 0$ for each $\pi \in \Pi$. Also, note that by Lemma 4, $\Pr\left(\cap_{\pi \in \Pi}\left\{\omega_\pi \leq \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\}\right) \to 1 - \beta/2$, and by definition of L_n, $\limsup_n \Pr\left(\{L_n < \sup_{\pi \in \Pi} \omega_\pi\}\right) \geq 1 - \beta/2$. Hence, by a union bound,

$$\limsup_n P\left\{\Pi^* \not\subseteq \widehat{\Pi}_\beta\right\} \leq \beta.$$

□

Lemma 4 in the following shows a uniform confidence band for $\{\omega_\pi : \pi \in \Pi\}$ which helps prove the validity of the candidate policy set $\widehat{\Pi}_\beta$.

Lemma 4. *If $\inf_{\pi \in \Pi} \sigma_\pi(P_0) > 0$, and $\widehat{\sigma}_\pi$ is a consistent estimator of $\sigma_\pi(P_0)$ for each $\pi \in \Pi$, an asymptotically valid uniform β-level confidence band is given by $\left\{\widehat{\omega}_\pi \pm \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} : \pi \in \Pi\right\}$.*

Proof of Lemma 4. To see that this is the case, note that t_β is the $1 - \beta/2$ quantile of $\sup_{f \in \mathcal{F}} \mathbb{G} f$, and also

$$P\left(\cap_{\pi \in \Pi}\left\{\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \leq \omega_\pi \leq \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\}\right)$$

$$= P\left(\cap_{\pi\in\Pi}\left\{-t_\beta \le n^{1/2}\frac{\widehat{\omega}_\pi - \omega_\pi}{\widehat{\sigma}_\pi} \le t_\beta\right\}\right)$$
$$\to P\left(\cap_{\pi\in\Pi}\left\{-t_\beta \le \mathbb{G}f \le t_\beta\right\}\right)$$
$$= P\left(\cap_{\pi\in\Pi}\left[\left\{-t_\beta \le \inf_{f\in\mathcal{F}} \mathbb{G}f\right\} \cap \left\{\sup_{f\in\mathcal{F}} \mathbb{G}f \le t_\beta\right\}\right]\right)$$
$$= 1-\beta,$$

where the convergence follows from the fact that $n^{1/2}\frac{\widehat{\omega}_\pi-\omega_\pi}{\widehat{\sigma}_\pi} \rightsquigarrow \mathbb{G}f$ by Lemma 5 and Slutsky's Theorem. □

Lemma 5 ($\mathcal{F}$ is P_0-Donsker)**.** *Assume that Conditions 8 and 9 hold and also that*

(i) Π *satisfies the uniform entropy bound, that is,*

$$\int_0^\infty \sup_{Q_X} \sqrt{\log N\left(\varepsilon, \Pi, L^2(Q_X)\right)} d\varepsilon < \infty,$$

where the supremum is over all finitely supported measures on $\mathcal{X}$*;*

(ii) there exists $L > 0$ *such that, for all finitely supported distributions* Q *of* (X, A, Y) *with support on* $\mathcal{X} \times \{0,1\} \times \mathcal{Y}$*, the gradient map* $\pi \mapsto D_\pi$ *is L-Lipschitz, in the sense that, for any* $\pi, \pi' \in \Pi$*,* $\|D_\pi - D_{\pi'}\|_{L^2(Q)} \le L\|\pi - \pi'\|_{L^2(Q_X)}$*, where* Q_X *is the marginal distribution of* X *under* Q*;*

(iii) $\sup_{\pi\in\Pi} \operatorname{ess\,sup}_{x\in\mathcal{X}, a\in\{0,1\}, y\in\mathcal{Y}} |D_\pi(P_0)(x,a,y)| < \infty.$

Then, the set $\mathcal{F} := \{D_\pi(P_0)/\sigma_\pi(P_0) : \pi \in \Pi\}$ *is* P_0*-Donsker.*

Proof of Lemma 5. We would like to use Theorem 2.5.2 of Van Der Vaart and Wellner [2013]. First, by (iii) and Condition 9,

$$C := \frac{\sup_{\pi\in\Pi} \operatorname{ess\,sup}_{x\in\mathcal{X},a\in\{0,1\},y\in\mathcal{Y}} |D_\pi(P_0)(x,a,y)|}{\inf_{\pi\in\Pi}\sigma_\pi(P_0)} < \infty.$$

Hence, an envelope function for $\mathcal{F}$ is the constant function $F(x,a,y) = C$. By (ii) and properties of covering numbers, for any Q as stated in (ii) and implied marginal distribution Q_X, we have that $N(C\varepsilon, \mathcal{F}, L^2(Q)) \leq N(C\varepsilon/L, \Pi, L^2(Q_X))$. Combining this with (i) shows that $\mathcal{F}$ satisfies the uniform entropy bound in the sense that $\int_0^\infty \sup_Q \sqrt{\log N(\varepsilon, \mathcal{F}, L^2(Q))}d\varepsilon < \infty$, where the supremum is over all finitely supported measures on $\mathcal{X} \times \{0,1\} \times \mathcal{Y}$. Hence, $\mathcal{F}$ is P_0-Donsker by Theorem 2.5.2 of Van Der Vaart and Wellner [2013]. □

B.2 Proof of Theorem 2

This subsection shows the proof of Theorem 2, which gives the asymptotic coverage of the confidence interval for the union bounding method.

Proof of Theorem 2. We have that

$$\left\{[\inf_{\pi\in\Pi^*}\psi_\pi, \sup_{\pi\in\Pi^*}\psi_\pi] \not\subseteq \mathrm{CI}_n\right\}$$

$$= \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right] \right\} \cup \left\{ \sup_{\pi\in\Pi^*} \psi_\pi > \sup_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right] \right\}$$
$$\subseteq \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\}$$
$$\cup \left\{ \sup_{\pi\in\Pi^*} \psi_\pi > \sup_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\} \cup \left\{ \Pi^* \not\subseteq \widehat{\Pi}_\beta \right\}.$$

Hence, by a union bound and the fact that $\limsup_n (a_n + b_n + c_n) \leq \limsup_n a_n + \limsup_n b_n + \limsup_n c_n$, we see that

$$\limsup_n P \left\{ \mathrm{CI}_n \not\subseteq [\inf_{\pi\in\Pi^*} \psi_\pi, \sup_{\pi\in\Pi^*} \psi_\pi] \right\}$$
$$\leq \limsup_n P \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\}$$
$$+ \limsup_n P \left\{ \sup_{\pi\in\Pi^*} \psi_\pi > \sup_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\}$$
$$+ \limsup_n P \left\{ \Pi^* \not\subseteq \widehat{\Pi}_\beta \right\}.$$

The third term is upper bounded by β by Lemma 2. In what follows we will show that the first term on the right-hand side is no more than $(\alpha - \beta)/2$. Similar arguments can be used to show that the second term is also no more than $(\alpha - \beta)/2$. By a union bound argument, the sum of three terms is upper bounded by α, which completes the proof.

We begin by noting that, for any $n \in \mathbb{N}$,

$$\left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\}$$

$$\subseteq \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\Pi^*} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\}$$
$$\subseteq \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\Pi^*} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right] \right\}.$$

By Lemma 7 and $\pi \mapsto \psi_\pi$ is continuous, there exists a π^ℓ such that $\psi_{\pi^\ell} = \inf_{\pi\in\Pi^*} \psi_\pi$. Combining this with the above, we see that

$$\left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\} \subseteq \left\{ \psi_{\pi^\ell} < \inf_{\pi\in\Pi^*} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right] \right\}$$
$$\subseteq \left\{ \psi_{\pi^\ell} < \widehat{\psi}_{\pi^\ell} - \frac{\widehat{\kappa}_{\pi^\ell} z_{\alpha,\beta}}{n^{1/2}} \right\}.$$

Then

$$P\left(\psi_{\pi^\ell} < \widehat{\psi}_{\pi^\ell} - \frac{\widehat{\kappa}_{\pi^\ell} z_{\alpha,\beta}}{n^{1/2}} \right) = P\left(n^{1/2} \frac{\widehat{\psi}_{\pi^\ell} - \psi_{\pi^\ell}}{\widehat{\kappa}_{\pi^\ell}} > z_{\alpha,\beta} \right).$$

By Condition 9, $\widehat{\kappa}_{\pi^\ell}$ is a consistent estimator for $\kappa_{\pi^\ell}(P_0)$. Then with Slutsky's Theorem, $n^{1/2} \frac{\widehat{\psi}_{\pi^\ell} - \psi_{\pi^\ell}}{\widehat{\kappa}_{\pi^\ell}} \rightsquigarrow \mathbb{G} f_{\pi^\ell}$, so by definition of $z_{\alpha,\beta}$,

$$P\left(n^{1/2} \frac{\widehat{\psi}_{\pi^\ell} - \psi_{\pi^\ell}}{\widehat{\kappa}_{\pi^\ell}} > z_{\alpha,\beta} \right) \leq (\alpha - \beta)/2,$$

and so

$$\limsup_{n\to\infty} \mathbb{P} \left\{ \inf_{\pi\in\Pi^*} \psi_\pi < \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\} \leq (\alpha - \beta)/2.$$

By a symmetric argument, we also have

$$\limsup_{n\to\infty} P \left\{ \sup_{\pi\in\Pi^*} \psi_\pi > \sup_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} \right], \Pi^* \subseteq \widehat{\Pi}_\beta \right\} \leq (\alpha - \beta)/2.$$

Therefore, an asymptotic $1-\alpha$ confidence interval for $[\psi_0^l, \psi_0^u]$ is

$$\left[\inf_{\pi \in \widehat{\Pi}_\beta}\left\{\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right\}, \sup_{\pi \in \widehat{\Pi}_\beta}\left\{\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right\}\right].$$

□

In the following lemma, for some subset $\mathcal{G}$ of a space $L^2(Q)$, define the covering number $N(\epsilon, \mathcal{G}, L^2(Q))$ to be the minimal cardinality of an ϵ-cover of $\mathcal{G}$ with respect to the $L^2(Q)$ metric Van Der Vaart and Wellner [2013]. Before stating the lemma, we recall that $\mathcal{F} := \{D_\pi(P_0)/\sigma_\pi(P_0) : \pi \in \Pi\}$.

Lemma 6. Π^* *is a closed subset of* $L^2(P_0)$.

Proof. Let $(\pi_k)_{k=1}^\infty$ be a Π^*-valued sequence that converges to some π^* in $L^2(P)$. Since $\pi \mapsto \omega_\pi$ is a continuous map from $\{0,1\}^{\mathcal{X}}$ to $\mathbb{R}$ when the domain is equipped with the $L^2(P)$-topology, $\omega_{\pi_k} \to \omega_{\pi^*}$. As $\pi_k \in \Pi^*$ for all k, $\omega_{\pi_k} = \sup_{\pi \in \Pi} \omega_\pi$ for all k. Hence, $\omega_{\pi^*} = \sup_{\pi \in \Pi} \omega_\pi$. As Π is closed, this shows that $\pi^* \in \Pi^*$. Hence, Π^* is a closed subset of $L^2(P)$. □

Lemma 7. *If* Π^* *is closed in* $L^2(P_0)$ *and* Π^* *is* P_0*-Donsker,* Π^* *is compact.*

Proof of Lemma 7. Since Π^* is P_0-Donsker following from Π being P_0-Donsker, then Π^* is totally bounded in $L^2(P_0)$. Also, since $L^2(P_0)$ is complete, Π^* being closed implies that Π^* is complete. And totally bounded and complete subsets of a metric space are compact, so Π^* is compact. □

B.3 Proof of Lemma 3

Proof of Lemma 3. To show this lemma, we first define two events $\{\Pi^* \subseteq \widehat{\Pi}_\beta\}$ and $\{\omega_{\pi^*} - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi \le \frac{4t_\beta}{n^{1/2}} \sup_{\pi\in\Pi} \widehat{\sigma}_\pi\}$. These events ensure that all Ω-optimal policies are contained in $\widehat{\Pi}_\beta$, and $\widehat{\Pi}_\beta$ only contains nearly optimal policies. Lemma 2 and 8 ensure that both events happen with probability at least $1-\beta$ asymptotically. Lemma 9 ensures that our confidence interval shrinks at an $n^{-1/2}$ rate under these events. □

Lemma 8. *For any $\beta > 0$, $\liminf_{n\to\infty} \mathbb{P}\left(\omega_{\pi^*} - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi \le \frac{4t_\beta}{n^{1/2}} \sup_{\pi\in\Pi} \widehat{\sigma}_\pi\right) \ge 1-\beta$.*

Proof of Lemma 8. Note that by the definition of $\widehat{\Pi}_\beta$, we have

$$\inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right] \ge \sup_{\pi\in\Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right],$$

so

$$\omega_{\pi^*} - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi \le \omega_{\pi^*} - \sup_{\pi\in\Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right] + \inf_{\pi\in\widehat{\Pi}_\beta} \left[\widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right] - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi.$$

Hence,

$$\begin{aligned}
&\left\{\omega_{\pi^*} - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi > \frac{4t_\beta}{n^{1/2}} \sup_{\pi\in\Pi} \widehat{\sigma}_\pi\right\} \\
&\subseteq \left\{\omega_{\pi^*} - \sup_{\pi\in\Pi} \left\{\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\} + \inf_{\pi\in\widehat{\Pi}_\beta} \left\{\widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\} - \inf_{\pi\in\widehat{\Pi}_\beta} \omega_\pi > \frac{4t_\beta}{n^{1/2}} \sup_{\pi\in\Pi} \widehat{\sigma}_\pi\right\} \\
&\subseteq \left\{\omega_{\pi^*} > \sup_{\pi\in\Pi} \left\{\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\} + 2\sup_{\pi\in\Pi} \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\}
\end{aligned}$$

$$\cup \left\{ \inf_{\pi \in \widehat{\Pi}_\beta} \omega_\pi < \inf_{\pi \in \widehat{\Pi}_\beta} \left\{ \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} - 2 \sup_{\pi \in \Pi} \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\}. \tag{S2.8}$$

In the remainder of this proof, we will show that the two events on the right-hand side each occur with probability no more than $\beta/2$. The result then follows by a union bound. Note that

$$\begin{aligned} \bigcap_{\pi \in \Pi} \left\{ \omega_\pi \le \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} &\subseteq \left\{ \omega_{\pi^*} \le \sup_{\pi \in \Pi} \left\{ \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \right\} \\ &\subseteq \left\{ \omega_{\pi^*} \le \sup_{\pi \in \Pi} \left\{ \widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} + 2 \sup_{\pi \in \Pi} \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\}, \end{aligned}$$

where the latter inclusion holds because $\sup[f+g] \le \sup f + \sup g$. So

$$\begin{aligned} &\liminf_{n \to \infty} \mathbb{P} \left(\omega_{\pi^*} - \sup_{\pi \in \Pi} \left\{ \widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \le 2 \sup_{\pi \in \Pi} \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right) \\ &\ge \liminf_{n \to \infty} \mathbb{P} \left(\bigcap_{\pi \in \Pi} \left\{ \omega_\pi \le \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \right) \ge 1 - \frac{\beta}{2}, \end{aligned}$$

where the last step follows from Lemma 4. Hence, the first event on the right-hand side of (S2.8) occurs with probability no more than probability $\beta/2$. We also have that

$$\begin{aligned} &\bigcap_{\pi \in \Pi} \left\{ \omega_\pi \ge \widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \subseteq \bigcap_{\pi \in \widehat{\Pi}_\beta} \left\{ \omega_\pi \ge \widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \\ &\subseteq \left\{ \inf_{\pi \in \widehat{\Pi}_\beta} \omega_\pi \ge \inf_{\pi \in \widehat{\Pi}_\beta} \left\{ \widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} \right\} \\ &\subseteq \left\{ \inf_{\pi \in \widehat{\Pi}_\beta} \omega_\pi \ge \inf_{\pi \in \widehat{\Pi}_\beta} \left\{ \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\} - 2 \sup_{\pi \in \widehat{\Pi}_\beta} \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \right\}, \end{aligned}$$

since $\inf[f-g] \geq \inf f - \sup g$. So

$$\liminf_{n\to\infty} \mathbb{P}\left(\inf_{\pi\in\widehat{\Pi}_\beta}\left\{\widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}}\right\} - 2\sup_{\pi\in\widehat{\Pi}_\beta}\frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \leq \inf_{\pi\in\widehat{\Pi}_\beta}\omega_\pi\right)$$
$$\geq \liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{\pi\in\Pi}\left\{\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi t_\beta}{n^{1/2}} \leq \omega_\pi\right\}\right) \geq 1 - \frac{\beta}{2},$$

where the last step follows from Lemma 4. Hence, the second event on the right-hand side of (S2.8) occurs with probability no more than probability $\beta/2$. □

Lemma 9. *In the setting of Lemma 3, under the event* $\{\Pi^* \subseteq \widehat{\Pi}_\beta\}$ *and* $\{\omega_{\pi^*} - \inf_{\pi\in\widehat{\Pi}_\beta}\omega_\pi \leq \frac{4t_\beta}{n^{1/2}}\sup_{\pi\in\Pi}\widehat{\sigma}_\pi\}$, *the width of the confidence interval for* ψ_0 *is* $O_p(n^{-1/2})$.

Proof. We first show that $\sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] = \psi_0 + O_p(n^{-1/2})$. We know that

$$\sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] \leq \sup_{\pi\in\widehat{\Pi}_\beta}\psi_\pi + \sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi - \psi_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right].$$

We then show that $\sup_{\pi\in\Pi^*}\psi_\pi - \sup_{\pi\in\widehat{\Pi}_\beta}\psi_\pi = O_p(n^{-1/2})$. Consider some $\pi_1 \in \Pi^*$ and $\pi_2 \in \widehat{\Pi}_\beta$. Let $B_{1,0} = \{x \in \mathcal{X} : \pi_1(x) = 1, \pi_2(x) = 0\}$ and $B_{0,1} = \{x \in \mathcal{X} : \pi_1(x) = 0, \pi_2(x) = 1\}$. By the definition of Π^* we know that $\omega_{\pi_1} \geq \omega_{\pi_2}$, and

$$\omega_{\pi_1} - \omega_{\pi_2} = \int \mathbb{E}[Y^*|A = \pi_1(x), x]dP_0(x) - \int \mathbb{E}[Y^*|A = \pi_2(x), x]dP_0(x)$$

$$= \int_{B_{1,0}} q_{b,0}(x)dP_0(x) - \int_{B_{0,1}} q_{b,0}(x)dP_0(x).$$

Since $\pi_1 \in \Pi^*$ and Π^* contains unrestricted optimal policies by assumption, ω_{π_1} is largest among all $\pi \in \Pi$, which implies that for $x \in B_{1,0}$, $q_{b,0}(x) \geq 0$ and for $x \in B_{0,1}$, $q_{b,0}(x) \leq 0$. This gives us

$$\omega_{\pi_1} - \omega_{\pi_2} = \int_{B_{1,0}} |q_{b,0}(x)|dP_0(x) + \int_{B_{0,1}} |q_{b,0}(x)|dP_0(x).$$

On the other hand, on the event $\{\Pi^* \subseteq \widehat{\Pi}_\beta\}$, we have $\sup_{\pi \in \widehat{\Pi}_\beta} \psi_\pi \geq \sup_{\pi \in \Pi^*} \psi_\pi$, and

$$\begin{aligned}
|\psi_{\pi_2} - \psi_{\pi_1}| &= \left| \int \mathbb{E}[Y^\dagger | A = \pi_2(x), x] dP_0(x) - \int \mathbb{E}[Y^\dagger | A = \pi_1(x), x] dP_0(x) \right| \\
&= \left| \int_{B_{0,1}} s_{b,0}(x)dP_0(x) - \int_{B_{1,0}} s_{b,0}(x)dP_0(x) \right| \\
&\leq \int_{B_{1,0}} |s_{b,0}(x)|dP_0(x) + \int_{B_{0,1}} |s_{b,0}(x)|dP_0(x) \\
&\leq C \int_{B_{1,0}} |q_{b,0}(x)|dP_0(x) + C \int_{B_{0,1}} |q_{b,0}(x)|dP_0(x).
\end{aligned}$$

Therefore, $|\psi_{\pi_2} - \psi_{\pi_1}| \leq C(\omega_{\pi_1} - \omega_{\pi_2})$ for some $C < \infty$. Since this holds for any $\pi_1 \in \Pi^*$ and $\pi_2 \in \widehat{\Pi}_\beta$, we have $\sup_{\pi \in \widehat{\Pi}_\beta} \psi_\pi - \inf_{\pi \in \Pi^*} \psi_\pi \leq C(\sup_{\pi \in \Pi^*} \omega_\pi - \inf_{\pi \in \widehat{\Pi}_\beta} \omega_\pi)$. Under the event $\{\omega_{\pi^*} - \inf_{\pi \in \widehat{\Pi}_\beta} \omega_\pi \leq \frac{4t_\beta}{n^{1/2}} \sup_{\pi \in \Pi} \widehat{\sigma}_\pi\}$, we have that $\sup_{\pi \in \widehat{\Pi}_\beta} \psi_\pi - \inf_{\pi \in \Pi^*} \psi_\pi \leq C \frac{4t_\beta}{n^{1/2}} \sup_{\pi \in \Pi} \widehat{\sigma}_\pi$. Under Condition 9, we know that $\sup_{\pi \in \Pi} \widehat{\sigma}_\pi - \sup_{\pi \in \Pi} \sigma_\pi(P_0) = o_p(1)$, so

$$\sup_{\pi \in \widehat{\Pi}_\beta} \psi_\pi - \inf_{\pi \in \Pi^*} \psi_\pi \leq C \frac{4t_\beta}{n^{1/2}} \sup_{\pi \in \Pi} \widehat{\sigma}_\pi$$

$$= C\frac{4t_\beta}{n^{1/2}}\left(\sup_{\pi\in\Pi}\sigma_\pi(P_0)+o_p(n^{-1/2})\right) = O_p(n^{-1/2}). \tag{S2.9}$$

Under the event $\{\Pi^* \subseteq \widehat{\Pi}_\beta\}$, we have $\sup_{\pi\in\widehat{\Pi}_\beta}\psi_\pi \geq \sup_{\pi\in\Pi^*}\psi_\pi \geq \inf_{\pi\in\Pi^*}\psi_\pi$, so we have $\sup_{\pi\in\Pi^*}\psi_\pi - \sup_{\pi\in\widehat{\Pi}_\beta}\psi_\pi = O_p(n^{-1/2})$. Also,

$$\begin{aligned}&\sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi - \psi_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right]\\ &\leq \sup_{\pi\in\Pi}\left[\widehat{\psi}_\pi - \psi_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] \leq \sup_{\pi\in\Pi}\left[\widehat{\psi}_\pi - \psi_\pi\right] + \sup_{\pi\in\Pi}\frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}.\end{aligned}$$

The first term is $O_p(n^{-1/2})$ under Condition 7. As for the second term, under Condition 9,

$$\sup_{\pi\in\Pi}\frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}} = \sup_{\pi\in\Pi}\frac{\kappa_\pi(P_0) z_{\alpha,\beta}}{n^{1/2}} + o_p(n^{-1/2}) = O_p(n^{-1/2}).$$

Therefore, $\sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi - \psi_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] = O_p(n^{-1/2})$ and so

$$\sup_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] = \psi_0^u + O_p(n^{-1/2})$$

as desired. By symmetry, $\inf_{\pi\in\widehat{\Pi}_\beta}\left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi z_{\alpha,\beta}}{n^{1/2}}\right] = \psi_0 - O_p(n^{-1/2})$ as well.

□

B.4 Proof of Theorem 3

Proof of Theorem 3. To establish this theorem, we show that

$$\liminf_n \mathbb{P}\left(\sup_{\pi\in\Pi^*}\psi_\pi \leq \sup_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right]\right) \geq 1-\alpha/2.$$

We can similarly get

$$\liminf_n \mathbb{P}\left(\inf_{\pi\in\Pi^*}\psi_\pi \geq \inf_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi - \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right]\right) \geq 1-\alpha/2.$$

Combining the two displays gives us the theorem statement. Note that

$$\begin{aligned}&\left\{\sup_{\pi\in\Pi^*}\psi_\pi \leq \sup_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right]\right\}\\ &\supseteq \left\{\sup_{\pi\in\Pi^*}\psi_\pi \leq \sup_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right], \Pi^*\subseteq\widehat{\Pi}^\dagger\right\}.\end{aligned}$$

Since Π^* is P_0-Donsker following from Π being P_0-Donsker, Π^* is totally bounded in $L^2(P_0)$ Luedtke and Van Der Laan [2016]. Also, since $L^2(P_0)$ is complete, Π^* being closed in $L^2(P_0)$ implies that Π^* is complete in $L^2(P_0)$. So Π^* is compact in $L^2(P_0)$. Combining this with the fact that $\pi\mapsto\psi_\pi$ is continuous implies that there exists a $\pi^u\in\Pi^*$ such that $\psi_{\pi^u}=\sup_{\pi\in\Pi^*}\psi_\pi$. Combining this with the above, we see that

$$\begin{aligned}&\left\{\sup_{\pi\in\Pi^*}\psi_\pi \leq \sup_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right]\right\}\\ &\supseteq \left\{\psi_{\pi^u} \leq \sup_{\pi\in\widehat{\Pi}^\dagger}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u_\alpha^\dagger}{n^{1/2}}\right], \Pi^*\subseteq\widehat{\Pi}^\dagger\right\}\\ &\supseteq \left\{\psi_{\pi^u} \leq \widehat{\psi}_{\pi^u} + \frac{\widehat{\kappa}_{\pi^u} u_\alpha^\dagger}{n^{1/2}}, \Pi^*\subseteq\widehat{\Pi}^\dagger\right\}\\ &= \left\{\psi_{\pi^u} \leq \widehat{\psi}_{\pi^u} + \frac{\widehat{\kappa}_{\pi^u} u_\alpha^\dagger}{n^{1/2}}, \omega_{\pi'} < \sup_{\pi\in\Pi}\omega_\pi, \forall\pi'\in(\widehat{\Pi}^\dagger)^C\right\}.\end{aligned} \tag{S2.10}$$

Note that

$$\left\{\omega_{\pi'} < \sup_{\pi\in\Pi}\omega_\pi, \forall\pi'\in(\widehat{\Pi}^\dagger)^C\right\}^C = \left\{\exists\pi'\in(\widehat{\Pi}^\dagger)^C : \omega_{\pi'} = \sup_{\pi\in\Pi}\omega_\pi\right\}$$

$$\subseteq \left\{\exists \pi' \in (\widehat{\Pi}^\dagger)^C : \left[\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\alpha^\dagger}{n^{1/2}} + \sup_{\pi \in \Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}}\right]\right] > \sup_{\pi\in\Pi} \omega_\pi \right\}$$
$$= \left\{\exists \pi' \in (\widehat{\Pi}^\dagger)^C : \left[\omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\alpha^\dagger}{n^{1/2}}\right] > \sup_{\pi\in\Pi} \omega_\pi - \sup_{\pi \in \Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}}\right] \right\}, \tag{S2.11}$$

where the inclusion follows from the definition of $\widehat{\Pi}^\dagger$. Let $\mathcal{A}'$ denote the event

$$\left\{\sup_{\pi\in\Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}}\right] \leq \sup_{\pi\in\Pi} \omega_\pi \right\} \bigcap \left[\bigcap_{\pi\in\Pi} \left\{\omega_\pi \leq \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\alpha^\dagger}{n^{1/2}}\right\}\right].$$

Hence, (S2.11) shows that

$$\left\{\exists \pi' \in (\widehat{\Pi}^\dagger)^C : \omega_{\pi'} = \sup_{\pi\in\Pi} \omega_\pi \right\}$$
$$\subseteq \left[\left\{\exists \pi' \in (\widehat{\Pi}^\dagger)^C : \omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\alpha^\dagger}{n^{1/2}} > \sup_{\pi\in\Pi} \omega_\pi - \sup_{\pi \in \Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}}\right] \right\} \cap \mathcal{A}'\right] \cup \mathcal{A}'^C$$
$$\subseteq \left[\left\{\exists \pi' \in (\widehat{\Pi}^\dagger)^C : \omega_{\pi'} - \widehat{\omega}_{\pi'} - \frac{\widehat{\sigma}_{\pi'} t_\alpha^\dagger}{n^{1/2}} > 0 \right\} \cap \mathcal{A}'\right] \cup \mathcal{A}'^C$$
$$= \mathcal{A}'^C.$$

For each $\pi \in \Pi$, we define $\widehat{B}_{n,\pi} := n^{1/2} \frac{\widehat{\omega}_\pi - \omega_\pi}{\widehat{\sigma}_\pi}$ and $\widetilde{B}_{n,\pi} := n^{1/2} \frac{\widehat{\psi}_\pi - \psi_\pi}{\widehat{\kappa}_\pi}$. Then starting from (S2.11), we have

$$\left\{\omega_{\pi'} < \sup_{\pi\in\Pi} \omega_\pi, \forall \pi' \in (\widehat{\Pi}^\dagger)^C \right\} \supseteq \mathcal{A}'$$
$$= \left\{\sup_{\pi\in\Pi} \left[\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}}\right] < \sup_{\pi\in\Pi} \omega_\pi \right\} \bigcap \left[\bigcap_{\pi\in\Pi} \left\{\omega_\pi \leq \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\alpha^\dagger}{n^{1/2}}\right\}\right]$$
$$\supseteq \bigcap_{\pi\in\Pi} \left\{\widehat{\omega}_\pi - \frac{\widehat{\sigma}_\pi s_\alpha^\dagger}{n^{1/2}} < \omega_\pi < \widehat{\omega}_\pi + \frac{\widehat{\sigma}_\pi t_\alpha^\dagger}{n^{1/2}}\right\}$$

$$= \bigcap_{\pi\in\Pi} \left\{ -t^\dagger_\alpha < n^{1/2}\frac{\widehat{\omega}_\pi - \omega_\pi}{\widehat{\sigma}_\pi} < s^\dagger_\alpha \right\}$$

$$= \bigcap_{\pi\in\Pi} \{ -t^\dagger_\alpha < B_{n,\pi} < s^\dagger_\alpha \}$$

$$= \{ -t^\dagger_\alpha \le \inf_{\pi\in\Pi} B_{n,\pi} \} \cap \{ \sup_{\pi\in\Pi} B_{n,\pi} \le s^\dagger_\alpha \}.$$

Using the above to study the event on the right-hand side of (S2.10) shows that

$$\begin{aligned} &\left\{ \psi_{\pi^u} \le \widehat{\psi}_{\pi^u} + \frac{\widehat{\kappa}_{\pi^u} u^\dagger_\alpha}{n^{1/2}}, \omega_{\pi'} < \sup_{\pi\in\Pi} \omega_\pi, \forall \pi' \in (\widehat{\Pi}^\dagger)^C \right\} \\ &\supseteq \left\{ \psi_{\pi^u} < \widehat{\psi}_{\pi^u} + \frac{\widehat{\kappa}_{\pi^u} u^\dagger_\alpha}{n^{1/2}}, -t^\dagger_\alpha \le \inf_{\pi\in\Pi} B_{n,\pi}, \sup_{\pi\in\Pi} B_{n,\pi} \le s^\dagger_\alpha \right\} \\ &= \left\{ \widetilde{B}_{n,\pi^u} > -u^\dagger_\alpha, -t^\dagger_\alpha \le \inf_{\pi\in\Pi} B_{n,\pi}, \sup_{\pi\in\Pi} B_{n,\pi} \le s^\dagger_\alpha \right\}. \end{aligned} \tag{S2.12}$$

We know that the choices $(s^\dagger_\alpha, t^\dagger_\alpha, u^\dagger_\alpha)$ satisfy that

$$\inf_{\pi\in\Pi} \mathbb{P} \left\{ \inf_{f\in\mathcal{F}} \mathbb{G}f \ge -t^\dagger_\alpha, \sup_{f\in\mathcal{F}} \mathbb{G}f \le s^\dagger_\alpha, \mathbb{G}\tilde{f}_\pi \ge -u^\dagger_\alpha \right\} \ge 1 - \alpha/2. \tag{S2.13}$$

Note that by Condition 7, we have $\sup_{\pi\in\Pi} \left[n^{1/2}\frac{\widehat{\omega}_\pi - \omega_\pi}{\widehat{\sigma}_\pi} - \mathbb{G}_n f_\pi \right] = o_p(1)$ and also $\frac{\widehat{\psi}_{\pi^u} - \psi_{\pi^u}}{\widehat{\kappa}_{\pi^u}} - \mathbb{G}_n \tilde{f}_{\pi^u} = o_p(1)$. Since $\sup_{f\in\mathcal{F}} \mathbb{G}_n f \rightsquigarrow \sup_{f\in\mathcal{F}} \mathbb{G}f$, $\inf_{f\in\mathcal{F}} \mathbb{G}_n f \rightsquigarrow \inf_{f\in\mathcal{F}} \mathbb{G}f$, and for each $\pi \in \Pi$, $\widehat{\sigma}_\pi$ is a consistent estimator of σ_π, by Slutsky Theorem, we have $\sup_{\pi\in\Pi} B_{n,\pi} \rightsquigarrow \sup_{f\in\mathcal{F}} \mathbb{G}f$ and $\inf_{\pi\in\Pi} B_{n,\pi} \rightsquigarrow \inf_{f\in\mathcal{F}} \mathbb{G}f$. Also, since for each $\tilde{f} \in \tilde{\mathcal{F}}$, $\mathbb{G}_n \tilde{f} \rightsquigarrow \mathbb{G}\tilde{f}$ and $\widehat{\sigma}_\pi$ is a consistent estimator of σ_π, we similarly have $\widetilde{B}_{n,\pi^u} \rightsquigarrow \mathbb{G}\tilde{f}_{\pi^u}$. Combining

(S2.10), (S2.12), and (S2.13), we have

$$
\begin{aligned}
&\liminf_n \mathbb{P}\left(\sup_{\pi\in\Pi^*}\psi_\pi < \sup_{\pi\in\widehat{\Pi}}\left[\widehat{\psi}_\pi + \frac{\widehat{\kappa}_\pi u^\dagger_\alpha}{n^{1/2}}\right]\right)\\
&\geq \liminf_n \mathbb{P}\left(\widetilde{B}_{n,\pi^u} < u^\dagger_\alpha, -s^\dagger_\alpha < \inf_{\pi\in\Pi} B_{n,\pi}, \sup_{\pi\in\Pi} B_{n,\pi} < t^\dagger_\alpha\right)\\
&\to \mathbb{P}\left(\mathbb{G}\tilde{f}_{\pi^u} < u^\dagger_\alpha, -s^\dagger_\alpha < \inf_{f\in\mathcal{F}}\mathbb{G}f, \sup_{f\in\mathcal{F}}\mathbb{G}f < t^\dagger_\alpha\right)\\
&\geq \inf_{\pi\in\Pi}\mathbb{P}\left(\mathbb{G}\tilde{f}_{\pi} < u^\dagger_\alpha, -s^\dagger_\alpha < \inf_{f\in\mathcal{F}}\mathbb{G}f, \sup_{f\in\mathcal{F}}\mathbb{G}f < t^\dagger_\alpha\right) = 1-\alpha/2.
\end{aligned}
$$

□

C. Additional simulation results

C.1 A 1D simulation with large sample size

In this section, we run the same 1D instance described in Section 4.1 but with larger sample size. Table 2 provides coverages and confidence interval widths with a larger sample size of 5000. In the non-unique setting, since there are multiple optimal policies for the primary outcome, $[\psi_0^\ell, \psi_0^u]$ will be an interval with some length. In our setting, we can see from the lower-left plot of Figure 3 that the length of $[\psi_0^\ell, \psi_0^u]$ is about 0.5, so any valid confidence interval for $[\psi_0^\ell, \psi_0^u]$ must have at least that length. Comparing the widths in Table 1 and 2, we can see that both the union bounding method and the joint method produce confidence intervals approaching that

	coverage				width				
	union	joint	one-step	os-split	union	joint	one-step	os-split	oracle
non-unique	1.000	1.000	0.000	0.000	1.091	1.061	0.061	0.096	0.561
unique non-margin	0.981	0.986	0.810	0.734	0.036	0.035	0.017	0.027	0.016
unique margin	0.983	0.989	0.946	0.949	0.040	0.036	0.023	0.037	0.023

Table 2: Coverages and widths of $[\psi_0^\ell, \psi_0^u]$ with sample size $n = 5000$.

limit. In the setting where Ω-optimal policy is unique, the widths of the confidence intervals for all methods approach zero as n goes to infinity.

C.2 A 3D simulation

We also added a scenario where we have a 3D policy and the optimal policy is unique. The policy class is a restricted tree class, denoted as $\Pi = \{x \mapsto \mathbf{1}\{x_1 \geq a_1, x_2 \geq a_2, x_3 \geq a_3\} : a_1, a_2, a_3 \in [-1, 1]\}\}$. The optimal policy is $\pi^*(x) = \mathbf{1}\{x_1 \geq 0, x_2 \geq 0, x_3 \geq 0\}$ so it lies in the tree class Π. We compare the outcome interval from three approaches: union, joint, and one-step. The method os-split provides a wider interval while having a worse coverage than one-step in 1D simulation results, so we drop it from the simulation. For each scenario, we consider a sample size n of 500. We again use 1000 multiplier bootstrap replicates to estimate the supremum and infimum. In this scenario, instead of generating a fine grid and computing the maximum over the grid, we use the nlopt package to numerically approximate the maximum. We let $\alpha = 0.05$ and use 500 Monte Carlo repli-

	coverage			width			
	union	joint	one-step	union	joint	one-step	oracle
3D margin	0.970	0.948	0.940	0.199	0.186	0.124	0.124
3D non-margin	1.000	0.988	0.594	0.185	0.175	0.092	0.092

Table 3: Coverage and width for 3D policy class with sample size $n = 500$.

cations to compute the coverage and approximate the average confidence interval widths. Table 3 shows the results. The joint methods achieves slightly shorter widths in this setting (5-6%), and the results are otherwise similar to those from Section 4.1.

C.3 Linear classes

To demonstrate the benefit of the joint method and the flexibility of our method in high-dimensional scenarios, we consider another scenario where the policy class is linear, taking the form $\Pi_\theta = \{x \mapsto \mathbf{1}\{x^\top\theta \geq 0\}\}$. We consider a high-dimensional sparse linear setting where $\theta^* = [0.1, 0.2, ..., 0.5, 0, ..., 0] \in \mathbb{R}^{10}$. We again compare the outcome interval from three approaches: union, joint, and one-step and use 1000 multiplier bootstrap replicates to estimate the supremum and infimum. To approximate the maximum more accurately and avoid the issue of getting a local optimum, we use the differential evolution method in the scipy package. We let $\alpha = 0.05$ and

	coverage			width			
	union	joint	one-step	union	joint	one-step	oracle
linear margin	1.000	0.989	0.990	0.107	0.076	0.038	0.037
linear non-margin	1.000	0.998	0.618	0.183	0.161	0.062	0.050

Table 4: Coverage and width for sparse linear policy class with sample size $n = 1000$.

use 500 Monte Carlo replications to compute the coverage and approximate the average confidence interval widths. Table 4 shows the results. We can see that when the margin condition is not satisfied, the one-step estimator only achieves coverage of 0.618, while both the union bounding and the joint methods achieve valid coverages. Also, the joint method generally has a smaller confidence interval width than the union bounding method (29% decrease when the margin condition is satisfied and 12% when the margin condition is not satisfied).

We also consider the true parameter vector $\theta^* = [0.1, 0.2, \cdots, 1] \in \mathbb{R}^{10}$ and we run the same set of simulations as described in Section C.3. Table 5 shows the results. The joint methods achieve much shorter widths (almost 20% decrease in margin and 17% decrease in non-margin setting) in this setting, and the results are otherwise similar to those from earlier sections.

	coverage			width			
	union	joint	one-step	union	joint	one-step	oracle
linear margin	1.000	0.985	0.980	0.112	0.090	0.057	0.042
linear non-margin	0.995	0.964	0.895	0.102	0.085	0.062	0.043

Table 5: Coverage and width for linear policy class with sample size $n = 1000$.

D. Multiplier bootstrap

In practice, we use multiplier bootstrap to estimate the quantiles described in Section 3 and we provide the pseudocodes of the algorithms below. Algorithm 1 estimates t_β defined just above Lemma 2. Algorithm 2 estimates the quantiles described in (3.7). In this algorithm, we take $s_\alpha^\dagger = t_\alpha^\dagger$ for simplicity and estimate the best $(t_\alpha^\dagger, u_\alpha^\dagger)$ given samples. Both algorithms approximate suprema and infima over sets indexed by $\pi \in \Pi$ by maxima and minima over π belonging to a grid approximation of Π.

Algorithm 1 Multiplier bootstrap

Input: samples $\{(x_i, a_i, y_i)\}_{i=1}^n$, policy set Π, bootstrap sample size B, confidence level β

Take a grid estimate $\{\pi_1, \cdots, \pi_K\}$ of Π

for each $k \in [K]$, compute normalized one-step estimates $\{o_i^{(\pi_k)}\}_{i=1}^n$ using collected samples $\{(x_i, a_i, y_i)\}_{i=1}^n$

for $j = 1, \cdots, B$ **do**

    get multiplier bootstrap samples ϵ_{ij} for $i = 1, \cdots, n$ and $k = 1, \cdots, K$

    compute $n^{-1/2} \sum_{i=1}^n \epsilon_{ij} o_i^{(\pi_k)}$ and denote the result as $f_{\pi_k}^{(j)}$

end for

compute $\max_{k \in [K]} f_{\pi_k}^{(j)}$ for each j and denote the resulting dataset as $\{t_i\}_{i=1}^B$

Output: $(1 - \beta)$-th quantile of $\{t_i\}_{i=1}^B$

Algorithm 2 Multiplier bootstrap for joint probability

Input: samples $\{(x_i, a_i, y_i, z_i)\}_{i=1}^n$, policy set Π, bootstrap sample size B, confidence level α

Take a grid estimate $\{\pi_1, \cdots, \pi_K\}$ of Π

for $k \in [K]$ **do**

  compute normalized one-step estimates $\{o_i^{(\pi_k)}\}_{i=1}^n$ using collected samples $\{(x_i, a_i, y_i)\}_{i=1}^n$

  compute normalized one-step estimates $\{\tilde{o}_i^{(\pi_k)}\}_{i=1}^n$ using collected samples $\{(x_i, a_i, z_i)\}_{i=1}^n$

end for

for $j = 1, \cdots, B$ **do**

  get multiplier bootstrap samples $\epsilon_{ik}^{(j)}$ for $i = 1, \cdots, n$ and $k = 1, \cdots, K$

  compute $n^{-1/2} \sum_{i=1}^n \epsilon_{ik}^{(j)} o_i^{(\pi_k)}$ and denote the result as $f_{\pi_k}^{(j)}$

  compute $n^{-1/2} \sum_{i=1}^n \epsilon_{ik}^{(j)} \tilde{o}_i^{(\pi_k)}$ and denote the result as $\tilde{f}_{\pi_k}^{(j)}$

end for

compute $\max_{k \in [K]} f_{\pi_k}^{(j)}$ for each j and denote the results as $\{s_j\}_{j=1}^B$

compute probability $\mathbb{P}(\max_{k \in [K]} f_{\pi_k} \le t, \tilde{f}_{\pi_k} \le u)$ for each $k = 1, \cdots, K$ using the B samples

Output: pairs (t, u) such that $\min_{k \in [K]} \mathbb{P}(\max_{k \in [K]} f_{\pi_k} \le t, \tilde{f}_{\pi_k} \le u) = 1 - \alpha$.