Technical Appendix to "Beyond the Average: Distributional Causal Inference under Imperfect Compliance"

The Appendix is structured as follows. Section A provides a table summarizing the notation. Section B introduces some definitions. Section C presents all proofs. Section D discusses the construction of confidence intervals. Section E presents some additional experimental details.

A Summary of Notation

Table 1: Summary of Notation

X_i	pre-treatment covariates
S_i	stratum indicator
D_i	actual treatment received
Z_i	treatment assignment
Y_i	outcome variable
$Y_i(d)$	potential outcome for treatment group $d \in \{0,1\}$
$D_i(z)$	potential treatment choice under assignment $z \in \{0,1\}$
$p(s)$	proportion of stratum $s \in \mathcal{S}$
$\pi_z(s)$	treatment assignment probability for treatment group $z \in \{0,1\}$ in stratum $s \in \mathcal{S}$
n	sample size
$n_z(s)$	number of observations in treatment group $z \in \{0,1\}$ in stratum s
$n(s)$	number of observations in stratum $s \in \mathcal{S}$
$\widehat{p}(s)$	$n(s)/n$, proportion of stratum $s \in \mathcal{S}$ in the sample
$\widehat{\pi}_z(s)$	$n_z(s)/n(s)$, estimated treatment assignment probability for treatment group $z \in \{0,1\}$ in stratum $s \in \mathcal{S}$
$F_{Y(d)}(y)$	$\mathbb{E}[1\!\!1_{\{Y(d)\leq y\}}]$, potential outcome distribution function
$\mu_z(y,s,x)$	$\mathbb{E}[1\!\!1_{\{Y\leq y\}} \mid Z=z, S=s, X=x]$, conditional distribution function
$\eta_z(s,x)$	$\mathbb{E}[D \mid Z=z, S=s, X=x]$, conditional probability of treatment receipt
$[K]$	$\{1,\ldots,K\}$ for a positive integer K
$\|a\|$	$\sqrt{a^\top a}$, Euclidean norm of a vector $a=(a_1,\ldots,a_p)^\top \in \mathbb{R}^p$
$\|\cdot\|_{P,q}$	$L^q(P)$ norm
$\ell^\infty(\mathcal{Y})$	space of uniformly bounded functions mapping an arbitrary index set $\mathcal{Y}$ to the real line
$\rightsquigarrow$	convergence in distribution or law
$X_n = O_p(a_n)$	$\lim_{K\to\infty}\lim_{n\to\infty} P(\lvert X_n\rvert > Ka_n)=0$ for a sequence $a_n>0$
$X_n = o_p(a_n)$	$\sup_{K>0}\lim_{n\to\infty} P(\lvert X_n\rvert > Ka_n)=0$ for a sequence $a_n>0$
$x_n \lesssim y_n$	for sequences x_n and y_n in $\mathbb{R}$, $x_n \leq Ay_n$ for a constant A
$\lfloor b\rfloor$	$\max\{k\in\mathbb{Z} \mid k\leq b\}$, greatest integer less than or equal to b

B Definitions

Definition B.1 (Covering numbers)**.** The *covering number* $N(\mathcal{F},\epsilon,\|\cdot\|)$ is the minimal number of balls $\{g : \|g-f\| < \epsilon\}$ of radius ϵ needed to cover the set $\mathcal{F}$. The centers of the balls need not belong to $\mathcal{F}$, but they should have finite norms.

Definition B.2 (Envelope function)**.** An *envelope function* of a class $\mathcal{F}$ is any function $x \mapsto F(x)$ such that $|f(x)| \leq F(x)$ for every x and $f \in \mathcal{F}$.

Definition B.3 (VC-type class)**.** We say $\mathcal{F}$ is of VC-type with coefficients (α_n, v_n) if

$$\sup_Q N\left(\mathcal{F}, e_Q, \epsilon||F||_{Q,2}\right) \leq \left(\frac{\alpha_n}{\epsilon}\right)^{v_n}, \quad \forall\epsilon\in(0,1],$$

where $N(\cdot)$ denote the covering number, $e_Q(f,g)=||f-g||_{Q,2}$, and the supremum is taken over all finitely discrete probability measures Q.

C Proofs

C.1 Proof of Lemma 3.2

To prove Lemma 3.2, we introduce additional notation to categorize individuals based on their compliance type. Table 2 summarizes the four compliance types with respect to the potential treatment choices. We let $\mathcal{C}$ denote the compliance type, and $\mathcal{C} = c$ denote the compliers, i.e., those with $D(1) > D(0)$.

Table 2: Compliance types

$D(1)$	$D(0)$	type
0	0	never-takers
0	1	defiers
1	0	compliers
1	1	always-takers

Proof. Under the monotonicity assumption stated in Assumption 3.1(iv), we can identify the cumulative distribution functions of potential outcomes for the compliers conditional on S as follows:

$$F_{Y(1)}(y \mid S, \mathcal{C} = c) = \frac{\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 1, S] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 0, S]}{\mathbb{E}[D \mid Z = 1, S] - \mathbb{E}[D \mid Z = 0, S]}, \tag{1}$$

$$F_{Y(0)}(y \mid S, \mathcal{C} = c) = \frac{\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 1, S] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 0, S]}{\mathbb{E}[1 - D \mid Z = 1, S] - \mathbb{E}[1 - D \mid Z = 0, S]}. \tag{2}$$

We can then derive the unconditional CDF of the potential outcomes for the compliers by aggregating over the strata:

$$\begin{aligned}
F_{Y(1)}(y \mid \mathcal{C} = c) &= \sum_{s=1}^{S} P(S = s \mid \mathcal{C} = c) F_{Y(1)}(y \mid S = s, \mathcal{T} = c) \\
&= \sum_{s=1}^{S} \frac{P(\mathcal{C} = c \mid S = s)}{P(\mathcal{C} = c)} F_{Y(1)}(y \mid S = s, \mathcal{C} = c) \\
&= \frac{\sum_{s=1}^{S} p(s)(\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 1, S = s] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 0, S = s])}{\sum_{s=1}^{S} p(s)(\mathbb{E}[D \mid Z = 1, S = s] - \mathbb{E}[D \mid Z = 0, S = s])}.
\end{aligned}$$

The first equality holds by the law of total expectation. The second equality holds by the Bayes' law. The third equality follows from representation of the conditional distribution given in (1) and the fact that $P(\mathcal{C} = c \mid S = s) = \mathbb{E}[D \mid Z = 1, S = s] - \mathbb{E}[D \mid Z = 0, S = s]$. We can obtain similar expressions for $F_{Y(0)}(y \mid \mathcal{C} = c)$ using the representation given in (2) as follows:

$$F_{Y(0)}(y \mid \mathcal{C} = c) = \frac{\sum_{s=1}^{S} p(s)(\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 1, S = s] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 0, S = s])}{\sum_{s=1}^{S} p(s)(\mathbb{E}[1 - D \mid Z = 1, S = s] - \mathbb{E}[1 - D \mid Z = 0, S = s])}.$$

Then, the LDTE, the difference between the distribution functions is given by

$$\begin{aligned}
\beta(y) :&= F_{Y(1)}(y \mid \mathcal{C} = c) - F_{Y(0)}(y \mid \mathcal{C} = c) \\
&= \frac{\sum_{s=1}^{S} p(s)(\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 1, S = s] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot D \mid Z = 0, S = s])}{\sum_{s=1}^{S} p(s)(\mathbb{E}[D \mid Z = 1, S = s] - \mathbb{E}[D \mid Z = 0, S = s])} \\
&+ \frac{\sum_{s=1}^{S} p(s)(\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 1, S = s] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \cdot (1 - D) \mid Z = 0, S = s])}{\sum_{s=1}^{S} p(s)(\mathbb{E}[D \mid Z = 1, S = s] - \mathbb{E}[D \mid Z = 0, S = s])} \\
&= \frac{\sum_{s=1}^{S} p(s)(\mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \mid Z = 1, S = s] - \mathbb{E}[\mathbb{1}_{\{Y \leq y\}} \mid Z = 0, S = s])}{\sum_{s=1}^{S} p(s)(\mathbb{E}[D \mid Z = 1, S = s] - \mathbb{E}[D \mid Z = 0, S = s])}.
\end{aligned}$$

This completes the proof. $\square$

C.2 Proof of Theorem 5.2

To prove Theorem 5.2, we first introduce some additional notations. Let $\mathcal{D}_i := \{Y_i(1), Y_i(0), D_i(1), D_i(0), X_i\}$, $Y_i^1(y) := \mathbb{1}_{\{Y_i(D_i(1))\leq y\}}$, $Y_i^0(y) := \mathbb{1}_{\{Y_i(D_i(0))\leq y\}}$, $\tilde{Y}_i^1(y) := Y_i^1(y) - \mathbb{E}[Y_i^1(y)|S_i]$, $\tilde{Y}_i^0(y) := Y_i^0(y) - \mathbb{E}[Y_i^0(y)|S_i]$, $\tilde{X}_i := X_i - \mathbb{E}[X_i|S_i]$, $\tilde{D}_i(z) := D_i(z) - \mathbb{E}[D_i(z)|S_i]$, $\tilde{\mu}_z(y, S_i, X_i) := \mu_z(y, S_i, X_i) - \mathbb{E}[\mu_z(y, S_i, X_i)|S_i]$ and $\tilde{\eta}_z(S_i, X_i) := \eta_z(S_i, X_i) - \mathbb{E}[\eta_z(S_i, X_i)|S_i]$ for $z \in \{0, 1\}$.

Proof. Let

$$\begin{aligned}
B &:= \mathbb{E}[D(1) - D(0)],\\
T(y) &:= \mathbb{E}[(\mathbb{1}_{\{Y(1)\leq y\}} - \mathbb{1}_{\{Y(0)\leq y\}})(D(1) - D(0))],\\
\widehat{B} &:= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{Z_i\cdot(D_i - \widehat{\eta}_1(S_i, X_i))}{\widehat{\pi}(S_i)} - \frac{(1 - Z_i)\cdot(D_i - \widehat{\eta}_0(S_i, X_i))}{1 - \widehat{\pi}(S_i)} + \widehat{\eta}_1(S_i, X_i) - \widehat{\eta}_0(S_i, X_i)\right],
\end{aligned}$$

$$\begin{aligned}
\widehat{T}(y) := \frac{1}{n}\sum_{i=1}^{n}\Bigg[&\frac{Z_i\cdot(\mathbb{1}_{\{Y_i\leq y\}} - \widehat{\mu}_1(y, S_i, X_i))}{\widehat{\pi}(S_i)} - \frac{(1 - Z_i)\cdot(\mathbb{1}_{\{Y_i\leq y\}} - \widehat{\mu}_0(y, S_i, X_i))}{1 - \widehat{\pi}(S_i)}\\
&+ \widehat{\mu}_1(y, S_i, X_i) - \widehat{\mu}_0(y, S_i, X_i)\Bigg].
\end{aligned}$$

Then, we have

$$\begin{aligned}
\sqrt{n}\left(\widehat{\beta}(y) - \beta(y)\right) &= \sqrt{n}\left(\frac{\widehat{T}(y)}{\widehat{B}} - \frac{T(y)}{B}\right)\\
&= \frac{1}{\widehat{B}}\sqrt{n}\left(\widehat{T}(y) - T(y)\right) - \frac{T(y)}{\widehat{B}B}\sqrt{n}\left(\widehat{B} - B\right)\\
&= \frac{1}{\widehat{B}}\left[\sqrt{n}\left(\widehat{T}(y) - T(y)\right) - \beta(y)\sqrt{n}\left(\widehat{B} - B\right)\right]. \qquad (3)
\end{aligned}$$

Step 1. First, we start with the linear expansion of $\sqrt{n}\left(\widehat{T}(y) - T(y)\right)$.

$$\begin{aligned}
\sqrt{n}(\widehat{T}(y) - T(y)) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg[\frac{Z_i\cdot(\mathbb{1}_{\{Y_i\leq y\}} - \widehat{\mu}_1(y, S_i, X_i))}{\widehat{\pi}(S_i)} - \frac{(1 - Z_i)\cdot(\mathbb{1}_{\{Y_i\leq y\}} - \widehat{\mu}_0(y, S_i, X_i))}{1 - \widehat{\pi}(S_i)}\\
&\quad + \widehat{\mu}_1(y, S_i, X_i) - \widehat{\mu}_0(y, S_i, X_i)\Bigg] - \sqrt{n}T(y)\\
&= \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\widehat{\mu}_1(y, S_i, X_i) - \frac{Z_i\widehat{\mu}_1(y, S_i, X_i))}{\widehat{\pi}(S_i)}\right]}_{\equiv T_{n,1}}\\
&\quad + \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\frac{(1 - Z_i)\widehat{\mu}_0(y, S_i, X_i)}{1 - \widehat{\pi}(S_i)} - \widehat{\mu}_0(y, S_i, X_i)\right]}_{\equiv T_{n,2}}\\
&\quad + \underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i\cdot\mathbb{1}_{\{Y_i\leq y\}}}{\widehat{\pi}(S_i)} - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{(1 - Z_i)\cdot\mathbb{1}_{\{Y_i\leq y\}}}{1 - \widehat{\pi}(S_i)} - \sqrt{n}T(y)}_{\equiv T_{n,3}}. \qquad (4)
\end{aligned}$$

We first show that

$$
\begin{aligned}
\sqrt{n}(\widehat{T}(y) - T(y)) = &\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(1 - \frac{1}{\pi(S_i)}\right) \tilde{\mu}_1(y, S_i, X_i) - \tilde{\mu}_0(y, S_i, X_i) + \frac{\tilde{Y}_i^1(y)}{\pi(S_i)} \right] Z_i \right. \\
&+ \left. \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\left(\frac{1}{1 - \pi(S_i)} - 1 \right) \tilde{\mu}_0(y, S_i, X_i) + \tilde{\mu}_1(y, S_i, X_i) - \frac{\tilde{Y}_i^0}{1 - \pi(S_i)} \right] (1 - Z_i) \right\} \\
&+ \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\mathbb{E}[Y_i^1(y) - Y_i^0(y) | S_i] - \mathbb{E}[Y_i^1(y) - Y_i^0(y)] \Big) \right\} + R(y), \qquad (5)
\end{aligned}
$$

where $\sup_{y \in \mathcal{Y}} |R(y)| = o_p(1)$.

We start with the first term $T_{n,1}$ in (4).

$$
\begin{aligned}
T_{n,1} &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\widehat{\mu}_1(y, S_i, X_i) - \frac{Z_i \widehat{\mu}_1(y, S_i, X_i))}{\widehat{\pi}(S_i)} \right] \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - \widehat{\pi}(S_i)}{\widehat{\pi}(S_i)} \widehat{\mu}_1(y, S_i, X_i) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - \widehat{\pi}(S_i)}{\widehat{\pi}(S_i)} \Big[\widehat{\mu}_1(y, S_i, X_i) - \mu_1(y, S_i, X_i) + \mu_1(y, S_i, X_i) \Big] \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - \widehat{\pi}(S_i)}{\widehat{\pi}(S_i)} \delta_y(1, S_i, X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i}{\widehat{\pi}(S_i)} \mu_1(y, S_i, X_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mu_1(y, S_i, X_i) \\
&= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - \widehat{\pi}(S_i)}{\widehat{\pi}(S_i)} \delta_y(1, S_i, X_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i}{\widehat{\pi}(S_i)} \tilde{\mu}_1(y, S_i, X_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{\mu}_1(y, S_i, X_i) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(1 - \frac{1}{\pi(S_i)}\right) Z_i \tilde{\mu}_1(y, S_i, X_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (1 - Z_i) \tilde{\mu}_1(y, S_i, X_i) \\
&\quad + \underbrace{\frac{1}{\sqrt{n}} \sum_{s \in \mathcal{S}} \left(\frac{\widehat{\pi}(s) - \pi(s)}{\widehat{\pi}(s)\pi(s)} \right) \sum_{i=1}^{n} Z_i \tilde{\mu}_1(y, S_i, X_i) 1\{S_i = s\}}_{\equiv R_{1,1}(y)} - \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i - \widehat{\pi}(S_i)}{\widehat{\pi}(S_i)} \delta_y(1, S_i, X_i)}_{\equiv R_{1,2}(y)},
\end{aligned}
$$

where the second last equality holds because we have

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{Z_i}{\widehat{\pi}(S_i)} \mathbb{E}[\mu_1(y, S_i, X_i) \mid S_i] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}[\mu_1(y, S_i, X_i) \mid S_i].
$$

Now, consider the term $R_{1,2}(y)$:

$$
\begin{aligned}
&\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i-\widehat{\pi}(S_i)}{\widehat{\pi}(S_i)}\delta_y(1,S_i,X_i)\right| = \left|\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\frac{Z_i-\widehat{\pi}(s)}{\widehat{\pi}(s)}\delta_y(1,s,X_i)1\{S_i=s\}\right| \\
&= \frac{1}{\sqrt{n}}\left|\sum_{s\in\mathcal{S}}\frac{1}{\widehat{\pi}(s)}\sum_{i=1}^{n}Z_i\delta_y(1,s,X_i)1\{S_i=s\} - \sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\delta_y(1,s,X_i)1\{S_i=s\}\right| \\
&= \frac{1}{\sqrt{n}}\left|\sum_{s\in\mathcal{S}}\sum_{i\in I_1(s)}\delta_y(1,s,X_i)\frac{n(s)}{n_1(s)} - \sum_{s\in\mathcal{S}}\sum_{i\in I_0(s)\cup I_1(s)}\delta_y(1,s,X_i)\right| \\
&= \frac{1}{\sqrt{n}}\left|\sum_{s\in\mathcal{S}}\sum_{i\in I_1(s)}\delta_y(1,s,X_i)\frac{n_0(s)}{n_1(s)} - \sum_{s\in\mathcal{S}}\sum_{i\in I_0(s)}\delta_y(1,s,X_i)\right| \\
&= \frac{1}{\sqrt{n}}\left|\sum_{s\in\mathcal{S}}n_0(s)\left[\frac{\sum_{i\in I_1(s)}\delta_y(1,s,X_i)}{n_1(s)} - \frac{\sum_{i\in I_0(s)}\delta_y(1,s,X_i)}{n_0(s)}\right]\right| \\
&\leq \frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}n_0(s)\sup_{y\in\mathcal{Y}}\left|\frac{\sum_{i\in I_1(s)}\delta_y(1,s,X_i)}{n_1(s)} - \frac{\sum_{i\in I_0(s)}\delta_y(1,s,X_i)}{n_0(s)}\right| = o_p(1)
\end{aligned}
$$

where the last equality is due to Assumption 5.1 (i). Thus, we have

$$
\begin{aligned}
T_{n,1} &= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i}{\hat{\pi}(S_i)}\tilde{\mu}_1(y,S_i,X_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{\mu}_1(y,S_i,X_i) + R_1(y) \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{Z_i}{\hat{\pi}(S_i)}\tilde{\mu}_1(y,S_i,X_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\tilde{\mu}_1(y,S_i,X_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-Z_i)\tilde{\mu}_1(y,S_i,X_i) + R_1(y) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{1-\frac{1}{\hat{\pi}(S_i)}\right\}Z_i\tilde{\mu}_1(y,S_i,X_i) + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(1-Z_i)\tilde{\mu}_1(y,S_i,X_i) + R_1(y), \qquad (6)
\end{aligned}
$$

where $\sup_{y\in\mathcal{Y}}R_1(y)=o_p(1)$. Moreover, we note that

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{1}{\hat{\pi}(S_i)}\right)Z_i\tilde{\mu}_1(y,S_i,X_i) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(1-\frac{1}{\pi(S_i)}\right)Z_i\tilde{\mu}_1(y,S_i,X_i) \\
&\quad + \sum_{s\in\mathcal{S}}\left(\frac{1}{\pi(s)}-\frac{1}{\hat{\pi}(s)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\tilde{\mu}_1(y,s,X_i)1\{S_i=s\}.
\end{aligned}
$$

Note that under Assumption 3.1(i), conditional on $\{Z_i,S_i\}_{i=1}^n$, the distribution of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\tilde{\mu}_1(y,s,X_i)1\{S_i=s\}$$

is the same as the distribution of the same quantity where units are ordered by strata and then ordered by $Z_i=1$ first and $Z_i=0$ second within strata. To this end, define $N(s):=\sum_{i=1}^{n}\mathbb{1}_{\{S_i<s\}}$ and $F(s):=\mathbb{P}(S_i<s)$. Furthermore, independently for each $s\in\mathcal{S}$ and independently of $\{Z_i,S_i\}_{i=1}^n$, let $\{X_i^s:1\leq i\leq n\}$ be i.i.d with marginal distribution equal to the distribution of $X_i|S=s$. Define

$$\tilde{\mu}_z(y,s,X_i^s):=\mu_z(y,s,X_i^s)-\mathbb{E}[\mu_z(y,s,X_i^s)|S_i=s].$$

Then, we have, for $s\in\mathcal{S}$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\tilde{\mu}_1(y,s,X_i)1\{S_i=s\} \stackrel{d}{=} \frac{1}{\sqrt{n}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\mu}_1(y,s,X_i^s).$$

We also have

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\mu}_1(y,s,X_i^s)\right)^2\middle|\{Z_i,S_i\}_{i=1}^n\right]=\frac{n_1(s)}{n}\mathbb{E}\left[\tilde{\mu}_1^2(y,s,X_i^s)|\{S_i\}_{i=1}^n\right]$$
$$\leq\frac{n_1(s)}{n}E\left[\mu_1^2(y,s,X_i)|S_i=s\right]=O_p(1).$$

Hence, we have

$$\max_{s\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\tilde{\mu}_1(y,s,X_i^s)\right|=O_p(1).$$

Combining this with the facts that $\max_{s\in\mathcal{S}}|\hat{\pi}(s)-\pi(s)|=o_p(1)$ and $\min_{s\in\mathcal{S}}\pi(s)>c>0$ for some constant c, we have

$$\sum_{s\in\mathcal{S}}\left(\frac{1}{\pi(s)}-\frac{1}{\hat{\pi}(s)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i\tilde{\mu}_1(y,s,X_i)1\{S_i=s\}=o_p(1).$$

Therefore, we have

$$T_{n,1}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(1-\frac{1}{\pi(S_i)}\right)Z_i\tilde{\mu}_1(y,S_i,X_i)+\frac{1}{\sqrt{n}}\sum_{i=1}^n(1-Z_i)\tilde{\mu}_1(y,S_i,X_i)+o_p(1).$$

The linear expansion of $T_{n,2}$ can be established in the same manner. As for the third term $T_{n,3}$, first note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{Z_i\cdot 1\!\!1_{\{Y_i\leq y\}}}{\widehat{\pi}(S_i)}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{Z_i\cdot 1\!\!1_{\{Y_i(D_i(1))\leq y\}}}{\widehat{\pi}(S_i)}=:\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{Z_i\cdot Y_i^1(y)}{\widehat{\pi}(S_i)}$$
$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{(1-Z_i)1\!\!1_{\{Y_i\leq y\}}}{1-\widehat{\pi}(S_i)}=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{(1-Z_i)1\!\!1_{\{Y_i(D_i(0)\leq y\}}}{1-\widehat{\pi}(S_i)}=:\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{(1-Z_i)Y_i^0(y)}{1-\widehat{\pi}(S_i)}.$$

Then we have

$$\begin{aligned}T_{n,3}&=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{Z_i\cdot 1\!\!1_{\{Y_i\leq y\}}}{\widehat{\pi}(S_i)}-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{(1-Z_i)\cdot 1\!\!1_{\{Y_i\leq y\}}}{1-\widehat{\pi}(S_i)}-\sqrt{n}T(y)\\&=\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{\widehat{\pi}(S_i)}\tilde{Y}_i^1(y)Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1-Z_i}{1-\widehat{\pi}(S_i)}\tilde{Y}_i^0(y)\right\}\\&+\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{\widehat{\pi}(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1-Z_i}{1-\widehat{\pi}(S_i)}\mathbb{E}[Y_i^0(y)|S_i]-\sqrt{n}T(y)\right\}.\end{aligned}\tag{7}$$

First note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{\widehat{\pi}(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i=\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{1}{\pi(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\widehat{\pi}(S_i)-\pi(S_i)}{\widehat{\pi}(S_i)\pi(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i,$$

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\pi(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i &= \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\pi(s)}\mathbb{E}[Y_i^1(y)|S_i=s]Z_i1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\mathbb{E}[Y_i^1(y)|S_i=s]}{\pi(s)}(Z_i-\pi(s))1\{S_i=s\}\\
&+ \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\pi(s)}\mathbb{E}[Y_i^1(y)|S_i=s]\pi(s)1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\pi(s)\sqrt{n}}\sum_{i=1}^{n}(Z_i-\pi(s))1\{S_i=s\}\\
&+ \sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}}\sum_{i=1}^{n}1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\pi(s)\sqrt{n}}B_n(s)+\sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}}n(s),
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widehat{\pi}(S_i)-\pi(S_i)}{\widehat{\pi}(S_i)\pi(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i &= \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\widehat{\pi}(s)-\pi(s)}{\widehat{\pi}(s)\pi(s)}\mathbb{E}[Y_i^1(y)|S_i=s]Z_i1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{B_n(s)}{n(s)\widehat{\pi}(s)\pi(s)}\mathbb{E}[Y_i^1(y)|S_i=s]Z_i1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{B_n(s)\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}n(s)\widehat{\pi}(s)\pi(s)}\sum_{i=1}^{n}Z_i1\{S_i=s\}\\
&= \sum_{s\in\mathcal{S}}\frac{B_n(s)\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}n(s)\widehat{\pi}(s)\pi(s)}n_1(s)\\
&= \sum_{s\in\mathcal{S}}\frac{B_n(s)\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}\pi(s)}.
\end{aligned}
$$

Therefore, we have

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\widehat{\pi}(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i = \sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}}n(s).
$$

Similarly, we have

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1-Z_i}{1-\widehat{\pi}(S_i)}\mathbb{E}[Y_i^0(y)|S_i] = \sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^0(y)|S=s]}{\sqrt{n}}n(s)
$$

Then, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\widehat{\pi}(S_i)}\mathbb{E}[Y_i^1(y)|S_i]Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1-Z_i}{1-\widehat{\pi}(S_i)}\mathbb{E}[Y_i^0(y)|S_i]-\sqrt{n}T(y)\\
&=\sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^1(y)|S=s]}{\sqrt{n}}n(s)-\sum_{s\in\mathcal{S}}\frac{\mathbb{E}[Y^0(y)|S=s]}{\sqrt{n}}n(s)-\sqrt{n}T(y)\\
&=\sum_{s\in\mathcal{S}}\sqrt{n}\left(\frac{n(s)}{n}-p(s)\right)\mathbb{E}[Y^1(y)-Y^0(y)|S=s]+\sum_{s\in\mathcal{S}}\sqrt{n}p(s)\mathbb{E}[Y^1(y)-Y^0(y)|S=s]-\sqrt{n}T(y)\\
&=\sum_{s\in\mathcal{S}}\sqrt{n}\left(\frac{n(s)}{n}-p(s)\right)\mathbb{E}[Y^1(y)-Y^0(y)|S=s]+\sqrt{n}\mathbb{E}[Y^1(y)-Y^0(y)]-\sqrt{n}T(y)\\
&=\sum_{s\in\mathcal{S}}\frac{n(s)}{\sqrt{n}}\mathbb{E}[Y^1(y)-Y^0(y)|S=s]-\sqrt{n}\mathbb{E}[Y^1(y)-Y^0(y)]\\
&=\frac{1}{\sqrt{n}}\sum_{s\in\mathcal{S}}\sum_{i=1}^{n}\left(1\{S_i=s\}\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i=s]\right)-\sqrt{n}\mathbb{E}[Y^1(y)-Y^0(y)]\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i]-\sqrt{n}\mathbb{E}[Y^1(y)-Y^0(y)]\\
&=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i]-\mathbb{E}[Y_i^1(y)-Y_i^0(y)]\right). \qquad (8)
\end{aligned}
$$

Combining, we have

$$
\begin{aligned}
T_{n,3}=&\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\widehat{\pi}(S_i)}\tilde{Y}_i^1(y)Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1-Z_i}{1-\widehat{\pi}(S_i)}\tilde{Y}_i^0(y)\right\}\\
&+\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i]-\mathbb{E}[Y_i^1(y)-Y_i^0(y)]\right)\right\}\\
=&\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1}{\pi(S_i)}\tilde{Y}_i^1(y)Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{1-Z_i}{1-\pi(S_i)}\tilde{Y}_i^0(y)\right\}\\
&+\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i]-\mathbb{E}[Y_i^1(y)-Y_i^0(y)]\right)\right\}+o_p(1),
\end{aligned}
$$

where the second equality holds because

$$
\left(\frac{1}{\pi(s)}-\frac{1}{\widehat{\pi}(s)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Y}_i^1(y)Z_i1\{S_i=s\}=o_p(1)\quad\text{and}
$$

$$
\left(\frac{1}{\pi(s)}-\frac{1}{\widehat{\pi}(s)}\right)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{Y}_i^0(y)(1-Z_i)1\{S_i=s\}=o_p(1)
$$

due to the same argument used in the proofs of $T_{n,1}$.

Step 2. Using the same arguments, we can show that

$$
\begin{aligned}
\sqrt{n}(\widehat{B}-B)=&\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(1-\frac{1}{\pi(S_i)}\right)\tilde{\eta}_1(S_i,X_i)-\tilde{\eta}_0(S_i,X_i)+\frac{\tilde{D}_i(1)}{\pi(S_i)}\right]Z_i\right.\\
&\left.+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\left(\frac{1}{1-\pi(S_i)}-1\right)\tilde{\eta}_0(S_i,X_i)+\tilde{\eta}_1(S_i,X_i)-\frac{\tilde{D}_i(0)}{1-\pi(S_i)}\right](1-Z_i)\right\}\\
&+\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(\mathbb{E}[D_i(1)-D_i(0)|S_i]-\mathbb{E}[D_i(1)-D_i(0)]\Big)\right\}+o_p(1). \qquad (9)
\end{aligned}
$$

Step 3. Combining (5) and (9) into (3), we obtain the linear expansion for $\widehat{\beta}(y)$ as

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{\beta}(y)-\beta(y)\right)\\
&=\frac{1}{\widehat{B}}\left[\sqrt{n}\left(\widehat{T}(y)-T(y)\right)-\beta(y)\sqrt{n}\left(\widehat{B}-B\right)\right]\\
&=\frac{1}{\widehat{B}}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_1(y,S_i,\mathcal{D}_i)Z_i-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_0(y,S_i,\mathcal{D}_i)(1-Z_i)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_s(y,S_i)\right]+I(y)
\end{aligned}
$$

where $\sup_{y\in\mathcal{Y}}|I(y)|=o_p(1)$ and

$$
\begin{aligned}
\phi_1(y,S_i,\mathcal{D}_i):=&\left[\left(1-\frac{1}{\pi(S_i)}\right)\tilde{\mu}_1(y,S_i,X_i)-\tilde{\mu}_0(y,S_i,X_i)+\frac{\tilde{Y}_i^1}{\pi(S_i)}\right]\\
&-\beta(y)\left[\left(1-\frac{1}{\pi(S_i)}\right)\tilde{\eta}_1(S_i,X_i)-\tilde{\eta}_0(S_i,X_i)+\frac{\tilde{D}_i(1)}{\pi(S_i)}\right],\\
\phi_0(y,S_i,\mathcal{D}_i):=&\left[\left(1-\frac{1}{1-\pi(S_i)}\right)\tilde{\mu}_0(y,S_i,X_i)-\tilde{\mu}_1(y,S_i,X_i)+\frac{\tilde{Y}_i^0}{1-\pi(S_i)}\right]\\
&-\beta(y)\left[\left(1-\frac{1}{1-\pi(S_i)}\right)\tilde{\eta}_0(S_i,X_i)-\tilde{\eta}_1(S_i,X_i)+\frac{\tilde{D}_i(0)}{1-\pi(S_i)}\right],\\
\phi_s(y,S_i):=&\Big(\mathbb{E}[Y_i^1(y)-Y_i^0(y)|S_i]-\mathbb{E}[Y_i^1(y)-Y_i^0(y)]\Big)-\beta(y)\Big(\mathbb{E}[D_i(1)-D_i(0)|S_i]-\mathbb{E}[D_i(1)-D_i(0)]\Big).
\end{aligned}
$$

Step 4. Denote

$$
\begin{aligned}
\varphi_{n,1}(y)&:=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_1(y,S_i,\mathcal{D}_i)Z_i+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_0(y,S_i,\mathcal{D}_i)(1-Z_i),\\
\varphi_{n,2}(y)&:=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\phi_s(y,S_i)
\end{aligned}
$$

Uniformly over $y\in\mathcal{Y}$, we show that

$$(\varphi_{n,1}(y),\varphi_{n,2}(y))\rightsquigarrow(\mathcal{G}_1(y),\mathcal{G}_2(y)),$$

where $(\mathcal{G}_1(y),\mathcal{G}_2(y))$ are two independent Gaussian processes with covariance kernels $\Omega_1(y,y')$ and $\Omega_2(y,y')$, respectively, such that

$$
\begin{aligned}
&\Omega_1(y,y')=\mathbb{E}[\pi(S_i)\phi_1(y,S_i,\mathcal{D}_i)\phi_1(y',S_i,\mathcal{D}_i)]+\mathbb{E}[(1-\pi(S_i))\phi_0(y,S_i,\mathcal{D}_i)\phi_0(y',S_i,\mathcal{D}_i)],\\
&\Omega_2(y,y')=\mathbb{E}[\phi_s(y,S_i)\phi_s(y',S_i)].
\end{aligned}
$$

The following argument follows the argument provided in the proof of Bugni et al. (2018, Lemma B.2). Note that under Assumption 3.1 (i), conditional on $\{Z_i,S_i\}_{i=1}^n$, the distribution of $\varphi_{n,1}(y)$ is the same as the distribution of the same quantity with units ordered by strata $s\in\mathcal{S}$ and then ordered by $Z_i=1$ first and $Z_i=0$ second within strata. Let $\{\mathcal{D}_i^s\}_{i=1}^n$ be a sequence of i.i.d. random variables with marginal distributions equal to the distribution of $\mathcal{D}_i|S_i=s$. Then we have

$$\varphi_{n,1}(y)|\{Z_i,S_i\}_{i=1}^n\stackrel{d}{=}\widetilde{\varphi}_{n,1}(y)|\{Z_i,S_i\}_{i=1}^n$$

where

$$\widetilde{\varphi}_{n,1}(y):=\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=N(s)+1}^{N(s)+n_1(s)}\phi_1(y,s,\mathcal{D}_i^s)+\sum_{s\in\mathcal{S}}\frac{1}{\sqrt{n}}\sum_{i=N(s)+n_1(s)+1}^{N(s)+n(s)}\phi_0(y,s,\mathcal{D}_i^s).$$

As $\varphi_{n,2}(y)$ is a function of $\{Z_i,S_i\}_{i=1}^n$, we have

$$(\varphi_{n,1}(y),\varphi_{n,2}(y))\stackrel{d}{=}(\widetilde{\varphi}_{n,1}(y),\varphi_{n,2}(y)).$$

Next, define

$$\varphi_{n,1}^{\star}(y) := \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=\lfloor nF(s)\rfloor+1}^{\lfloor n(F(s)+\pi(s)p(s)\rfloor} \phi_1(y,s,\mathcal{D}_i^s) + \sum_{s\in\mathcal{S}} \frac{1}{\sqrt{n}} \sum_{i=\lfloor n(F(s)+\pi(s)p(s)\rfloor+1}^{\lfloor n(F(s)+p(s))\rfloor} \phi_0(y,s,\mathcal{D}_i^s).$$

Note $\varphi_{n,1}^{\star}(y)$ is a function of $\{\mathcal{D}_i^s\}_{i\in[n],s\in\mathcal{S}}$, which is independent of $\{Z_i,S_i\}_{i=1}^n$ by construction. Therefore,

$$\varphi_{n,1}^{\star}(y) \perp\!\!\!\perp \varphi_{n,2}(y).$$

Note that

$$\frac{N(s)}{n} \xrightarrow{p} F(s), \quad \frac{n_1(s)}{n} \xrightarrow{p} \pi(s)p(s), \quad \text{and} \quad \frac{n(s)}{n} \xrightarrow{p} p(s).$$

Denote $\Gamma_{n,z}(s,t,y) := \sum_{i=1}^{\lfloor nt\rfloor} \phi_z(y,s,\mathcal{D}_i^s)/\sqrt{n}$ for $z\in\{0,1\}$ and $s\in\mathcal{S}$. In order to show that

$$\sup_{y\in\mathcal{Y}} |\widetilde{\varphi}_{n,1}(y) - \varphi_{n,1}^{\star}(y)| = o_p(1) \text{ and } \varphi_{n,1}^{\star}(y) \rightsquigarrow \mathcal{G}_1(y),$$

it suffices to show that (1) for $z\in\{0,1\}$ and $s\in\mathcal{S}$, the stochastic process $\{\Gamma_{n,z}(s,t,y) : t\in(0,1), y\in\mathcal{Y}\}$ is stochastically equicontinuous, and (2) $\varphi_{n,1}^{\star}(y) \rightsquigarrow \mathcal{G}_1(y)$ converges to $G_1(y)$ in finite dimension.

Stochastic equicontinuity We want to bound

$$\sup |\Gamma_{n,z}(s,t_2,y_2) - \Gamma_{n,z}(s,t_1,y_1)|,$$

where the supremum is taken over $0<t_1<t_2<t_1+\epsilon<1$ and $y_1<y_2<y_1+\epsilon$ such that $y_1, y_1+\epsilon\in\mathcal{Y}$. Note that,

$$\begin{aligned}\sup |\Gamma_{n,z}(s,t_2,y_2) - \Gamma_{n,z}(s,t_1,y_1)| \le & \sup_{0<t_1<t_2<t_1+\epsilon<1, y\in\mathcal{Y}} |\Gamma_{n,z}(s,t_2,y) - \Gamma_{n,z}(s,t_1,y)| \\ & + \sup_{t\in(0,1), y_1,y_2\in\mathcal{Y}, y_1<y_2<y_1+\epsilon} |\Gamma_{n,z}(s,t,y_2) - \Gamma_{n,z}(s,t,y_1)|.\end{aligned} \tag{10}$$

Then, for the first term of the RHS of (10), we have

$$\begin{aligned}&\mathbb{P}\left(\sup_{0<t_1<t_2<t_1+\epsilon<1, y\in\mathcal{Y}} |\Gamma_{n,z}(s,t_2,y) - \Gamma_{n,z}(s,t_1,y)| \ge \delta\right) \\ &= \mathbb{P}\left(\sup_{0<t_1<t_2<t_1+\epsilon<1, y\in\mathcal{Y}} |\sum_{i=\lfloor nt_1\rfloor+1}^{\lfloor nt_2\rfloor} \phi_z(y,s,\mathcal{D}_i^s)| \ge \sqrt{n}\delta\right) \\ &= \mathbb{P}\left(\sup_{0<t\le\epsilon, y\in\mathcal{Y}} |\sum_{i=1}^{\lfloor nt\rfloor} \phi_z(y,s,\mathcal{D}_i^s)| \ge \sqrt{n}\delta\right) \\ &\le \mathbb{P}\left(\max_{1\le k\le\lfloor n\epsilon\rfloor} \sup_{y\in\mathcal{Y}} |\sum_{i=1}^{k} \phi_z(y,s,\mathcal{D}_i^s)| \ge \sqrt{n}\delta\right) \\ &\le \frac{270\, \mathbb{E}[\sup_{y\in\mathcal{Y}} |\sum_{i=1}^{\lfloor n\epsilon\rfloor} \phi_z(y,s,\mathcal{D}_i^s)|]}{\sqrt{n}\delta} \\ &\lesssim \frac{\sqrt{n\epsilon}}{\sqrt{n}\delta} \lesssim \delta.\end{aligned}$$

The third equality holds due to Assumption 5.1 (ii) and by applying van der Vaart and Wellner (1996, Theorem 2.14.1). The last equality holds by taking $\epsilon = \delta^4$.

Similarly, for the second term on the RHS of (10), we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{t\in(0,1),y_1,y_2\in\mathcal{Y},y_1<y_2<y_1+\epsilon}|\Gamma_{n,z}(s,t,y_2)-\Gamma_{n,z}(s,t,y_1)|\geq\delta\right)\\
&=\mathbb{P}\left(\max_{1\leq k\leq n}\sup_{y_1,y_2\in\mathcal{Y},y_1<y_2<y_1+\epsilon}|\sum_{i=1}^{k}(\phi_z(y_2,s,\mathcal{D}_i^s)-\phi_z(y_1,s,\mathcal{D}_i^s))|\geq\sqrt{n}\delta\right)\\
&\leq\frac{270\,\mathbb{E}[\sup_{y_1,y_2\in\mathcal{Y},y_1<y_2<y_1+\epsilon}|\sum_{i=1}^{n}(\phi_z(y_2,s,\mathcal{D}_i^s)-\phi_z(y_1,s,\mathcal{D}_i^s))|]}{\sqrt{n}\delta}\\
&\lesssim\delta\sqrt{\log(\frac{C}{\delta^2})},
\end{aligned}$$

The last inequality holds due to Assumption 5.1 (ii) and by applying maximal inequality for bracketing in van der Vaart and Wellner (1996, Theorem 2.14.2). Note that $\delta\sqrt{\log(\frac{C}{\delta^2})}\to 0$ as $\delta\to 0$. This concludes the proof of stochastic equicontinuity.

Finite-dimensional convergence For a given y, by the triangular array central limit theorem,

$$\varphi_{n,1}^{\star}(y)\rightsquigarrow N(0,\Omega_1(y,y)),$$

where

$$\begin{aligned}
\Omega_1(y,y)=&\lim_{n\to\infty}\sum_{s\in\mathcal{S}}\frac{(\lfloor n(F(s)+\pi(s)p(s))\rfloor-\lfloor nF(s)\rfloor)}{n}\mathbb{E}[\phi_1^2(y,s,\mathcal{D}_i^s)]\\
&+\lim_{n\to\infty}\sum_{s\in\mathcal{S}}\frac{(\lfloor n(F(s)+p(s))\rfloor-\lfloor n(F(s)+p(s)\pi(s))\rfloor)}{n}\mathbb{E}[\phi_0^2(y,s,\mathcal{D}_i^s)]\\
&=\sum_{s\in\mathcal{S}}p(s)\mathbb{E}[\pi(s)\phi_1^2(y,S_i,\mathcal{D}_i)+(1-\pi(s))\phi_0^2(y,S_i,\mathcal{D}_i)|S_i=s]\\
&=\mathbb{E}[\pi(S_i)\phi_1^2(y,S_i,\mathcal{D}_i)]+\mathbb{E}[(1-\pi(S_i))\phi_0^2(y,S_i,\mathcal{D}_i)].
\end{aligned}$$

Lastly, finite dimensional convergence follows from the Cramér-Wold device. In particular, the covariance kernel is given by

$$\Omega_1(y,y')=\mathbb{E}[\pi(S_i)\phi_1(y,S_i,\mathcal{D}_i)\phi_1(y',S_i,\mathcal{D}_i)]+\mathbb{E}[(1-\pi(S_i))\phi_0(y,S_i,\mathcal{D}_i)\phi_0(y',S_i,\mathcal{D}_i)].$$

This concludes the proof of finite-dimensional convergence of $\varphi_{n,1}^{\star}(y)$.

Finally, since $\{\mu_z(y,s,x)(y):y\in\mathcal{Y}\}$ is of the VC-type with fixed coefficients (α,v) and a constant envelope function, $\{\phi_s(y,S_i):y\in\mathcal{Y}\}$ is a Donsker class and we have

$$\varphi_{n,2}(y)\rightsquigarrow\mathcal{G}_2(y),$$

where $\mathcal{G}_2(y)$ is a Gaussian process with covariance kernel

$$\Omega_2(y,y')=\mathbb{E}[\phi_s(y,S_i)\phi_s(y',S_i)].$$

This completes the proof of Step 4.

Step 5. Therefore, uniformly over $y\in\mathcal{Y}$, we have

$$\sqrt{n}\left(\widehat{\beta}(y)-\beta(y)\right)\rightsquigarrow\mathcal{G}(y),$$

where $\mathcal{G}(y)$ is a Gaussian process with covariance kernel

$$\begin{aligned}
\Omega(y,y')=\Big\{&\mathbb{E}[\pi(S_i)\phi_1(y,S_i,\mathcal{D}_i,)\phi_1(y',S_i,\mathcal{D}_i)]+\mathbb{E}[(1-\pi(S_i))\phi_0(y,S_i,\mathcal{D}_i)\phi_0(y',S_i,\mathcal{D}_i)]\\
&+\mathbb{E}[\phi_s(y,S_i)\phi_s(y',S_i)]\Big\}/\left\{\mathbb{E}[D(1)-D(0)]^2\right\}.
\end{aligned}$$

□

C.3 Proof of Theorem 5.3: Semiparametric Efficiency Bound

Proof. Part (a). We follow the approach used in Hahn (1998) and calculate the semiparametric efficiency bound of the LDTE, $\beta(y)$ for a given $y \in \mathcal{Y}$. First, we characterize the tangent space. To that end, the joint density of the observed variables (Y, D, Z, X, S) can be written as:

$$\begin{aligned} f(y,d,z,x,s) =& f(y \mid d,z,x,s) f(d \mid z,x,s) f(z \mid x,s) f(x \mid s) f(s) \\ =& f(y \mid d,z,x,s)\{\eta_z(x,s)^d \cdot (1-\eta_z(x,s))^{1-d}\}\{\pi(s)^z \cdot (1-\pi(s))^{1-z}\} f(x \mid s) f(s), \end{aligned}$$

where $\eta_z(x,s) := P(D=1|Z=z, X=x, S=s)$ and $\pi(s) = P(Z=1|X=x, S=s)$ for all $x \in \mathcal{X}$.

Consider a regular parametric submodel indexed by θ:

$$\begin{aligned} f(y,d,z,x,s;\theta) =& f^{11}(y \mid x,s;\theta)^{dz} f^{10}(y \mid x,s;\theta)^{d(1-z)} f^{01}(y \mid x,s;\theta)^{(1-d)z} f^{00}(y \mid x,s;\theta)^{(1-d)(1-z)} \\ & \{\eta_z(x,s;\theta)^d \cdot (1-\eta_z(x,s;\theta))^{1-d}\}\{\pi(s;\theta)^z \cdot (1-\pi(s;\theta))^{1-z}\} f(x \mid s;\theta) f(s;\theta), \end{aligned}$$

where $f^{dz}(y \mid x,s;\theta) := f(y \mid d,z,x,s;\theta)$. When the parameter takes the true value, $\theta = \theta_0$, $f(y,d,z,x,s;\theta_0) = f(y,d,z,x,s)$.

The corresponding score of $f(y,d,z,x,s;\theta)$ is given by

$$\begin{aligned} s(y,d,z,x,s;\theta) :=& \frac{\partial \ln f(y,d,z,x,s;\theta)}{\partial \theta} \\ &= dz\dot{f}^{11}(y \mid x,s;\theta) + d(1-z)\dot{f}^{10}(y \mid x,s;\theta) \\ &+ (1-d)z\dot{f}^{01}(y \mid x,s;\theta) + (1-d)(1-z)\dot{f}^{00}(y \mid x,s;\theta) \\ &+ \frac{d-\eta_z(x,s;\theta)}{1-\eta_z(x,s;\theta)}\dot{\eta}_z(x,s;\theta) + \frac{z-\pi(s;\theta)}{1-\pi(s;\theta)}\dot{\pi}(s;\theta) + \dot{f}(x,s;\theta) + \dot{f}(s;\theta), \end{aligned}$$

where $\dot{f}$ denotes a derivative of the log, i.e, $\dot{f}(x;\theta) = \frac{\partial \ln f(x;\theta)}{\partial \theta}$.

At the true value, the expectation of the score equals zero. The tangent space of the model is the set of functions that are mean zero and satisfy the additive structure of the score:

$$\mathscr{T} = \left\{ \begin{aligned} & dza^{11}(y \mid x,s) + d(1-z)a^{10}(y \mid x,s) \\ & + (1-d)za^{01}(y \mid x,s) + (1-d)(1-z)a^{00}(y \mid x,s) \\ & + (d-\eta_z(x,s))a_\eta(x,z,s) + (z-\pi(s))a_\pi(s) + a_x(x,s) + a_s(s) \end{aligned} \right\}, \tag{11}$$

where $a^{dz}(y|x,s)$, $a_x(x,s)$ and $a_s(s)$ are mean-zero functions and $a_\eta(x,z,s)$ and $a_\pi(s)$ are square-integrable functions.

The semiparametric variance bound of $\beta(y)$ is given by the variance of the projection of a function $\psi(Y,D,Z,X,S)$ onto the tangent space $\mathscr{T}$. This function must have mean zero, finite second order moment and satisfy the following condition for all regular parametric submodels:

$$\frac{\partial \beta(y;F_\theta)}{\partial \theta}\bigg|_{\theta=\theta_0} = \mathbb{E}[\psi(Y,D,Z,X,S) \cdot s(Y,D,Z,X,S)]\bigg|_{\theta=\theta_0}. \tag{12}$$

If ψ itself already lies in the tangent space, the variance bound is given by $\mathbb{E}[\psi^2]$.

Now, the LDTE is

$$\beta(y) = F_{Y(1)|\mathcal{C}=c}(y) - F_{Y(0)|\mathcal{C}=c}(y).$$

Following Lemma 3.2, it follows that

$$\begin{aligned} F_{Y(1)|\mathcal{C}=c}(y) &= \left\{ \iint (F_{Y|D=1,Z=1,X=x,S=s}(y) \cdot \eta_1(x,s) - F_{Y|D=1,Z=0,X=x,S=s}(y) \cdot \eta_0(x,s)) f(x|s) f(s) dx ds \right\} / P_C \\ F_{Y(0)|\mathcal{C}=c}(y) &= -\left\{ \iint (F_{Y|D=0,Z=1,X=x,S=s}(y) \cdot \eta_1(x,s) - F_{Y|D=0,Z=0,X=x,S=s}(y) \cdot \eta_0(x,s)) f(x|s) f(s) dx ds \right\} / P_C \end{aligned}$$

where $P_C = \iint (\eta_1(x,s) - \eta_0(x,s)) f(x|s) f(s) dx ds$.

We first need to calculate the derivative evaluated at true θ_0:

$$\frac{\partial \beta(y; F_\theta)}{\partial \theta}|_{\theta=\theta_0} = \frac{\partial}{\partial \theta} F_{Y(1)|\mathcal{C}=c}(y;\theta_0) - \frac{\partial}{\partial \theta} F_{Y(0)|\mathcal{C}=c}(y;\theta_0).$$

We have,

$$\begin{aligned}
&\frac{\partial}{\partial \theta} F_{Y(1)|\mathcal{C}=c}(y;\theta_0)\\
&= \frac{1}{P_C}\frac{\partial}{\partial \theta}\left\{\iint (F_{Y|D=1,Z=1,X=x,S=s}(y)\cdot \eta_1(x,s) - F_{Y|D=1,Z=0,X=x,S=s}(y)\cdot \eta_0(x,s)) f(x|s)f(s)dxds\right\}\\
&- \left\{\iint (F_{Y|D=1,Z=1,X=x,S=s}(y)\cdot \eta_1(x,s) - F_{Y|D=1,Z=0,X=x,S=s}(y)\cdot \eta_0(x,s)) f(x|s)f(s)dxds\right\}\frac{\partial P_C(\theta_0)}{\partial \theta}.
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
&\frac{\partial}{\partial \theta} F_{Y(0)|\mathcal{C}=c}(y;\theta_0)\\
&= -\frac{1}{P_C}\frac{\partial}{\partial \theta}\left\{\iint (F_{Y|D=0,Z=1,X=x,S=s}(y)\cdot \eta_1(x,s) - F_{Y|D=0,Z=0,X=x,S=s}(y)\cdot \eta_0(x,s)) f(x|s)f(s)dxds\right\}\\
&+ \left\{\iint (F_{Y|D=0,Z=1,X=x,S=s}(y)\cdot \eta_1(x,s) - F_{Y|D=0,Z=0,X=x,S=s}(y)\cdot \eta_0(x,s)) f(x|s)f(s)dxds\right\}\frac{\partial P_C(\theta_0)}{\partial \theta}.
\end{aligned}$$

We choose $\psi(Y,D,Z,X,S)$ as

$$\begin{aligned}
&\psi(Y,D,Z,X,S)\\
&=\left\{\frac{Z}{\pi_1(S)}\cdot\big(1\!\!1_{\{Y\le y\}} - \mu_1(y,S,X)\big) - \frac{1-Z}{\pi_0(S)}\cdot\big(1\!\!1_{\{Y\le y\}} - \mu_0(y,S,X)\big) + \mu_1(y,S,X) - \mu_0(y,S,X)\right\}/\\
&\left\{\frac{Z}{\pi_1(S)}\cdot(D-\eta_1(S,X)) - \frac{1-Z}{\pi_0(S)}\cdot(D-\eta_0(S,X)) + \eta_1(S,X) - \eta_0(S,X)\right\} - \beta(y).
\end{aligned}$$

Then, notice that ψ satisfies (12) and that ψ lies in the tangent space $\mathscr{T}$ given in (11). Since ψ lies in the tangent space, the variance bound is given by the expected square of ψ:

$$\begin{aligned}
\Omega(y) &:= \mathbb{E}\big[\psi(Y,D,Z,X,S)^2\big]\\
&= \mathbb{E}\Bigg[\Bigg(\left\{\frac{Z}{\pi_1(S)}\cdot\big(1\!\!1_{\{Y\le y\}} - \mu_1(y,S,X)\big) - \frac{1-Z}{\pi_0(S)}\cdot\big(1\!\!1_{\{Y\le y\}} - \mu_0(y,S,X)\big) + \mu_1(y,S,X) - \mu_0(y,S,X)\right\}/\\
&\qquad\left\{\frac{Z}{\pi_1(S)}\cdot(D-\eta_1(S,X)) - \frac{1-Z}{\pi_0(S)}\cdot(D-\eta_0(S,X)) + \eta_1(S,X) - \eta_0(S,X)\right\} - \beta(y)\Bigg)^2\Bigg]
\end{aligned}$$

This concludes the proof of part (a).

Next, for part (b), under Assumption 5.1, the regression-adjusted estimator defined in Algorithm 1 satisfies the following asymptotic distribution for any given $y \in \mathcal{Y}$:

$$\sqrt{n}\big(\widehat{\beta}(y) - \beta(y)\big) \rightsquigarrow \mathcal{N}(0, \Omega(y)),$$

where $\Omega(y)$ is the semiparametric efficiency bound derived in part (a). This completes the proof of part (b).

□

D Inference

We consider two approaches to estimate the standard errors and construct confidence intervals for the regression-adjusted LDTE, $\widehat{\beta}(y)$, at a given threshold $y \in \mathcal{Y}$. Using the asymptotic distribution derived in Theorem 5.2, we can construct a $(1-\alpha) \times 100\%$ confidence interval for $\widehat{\beta}(y)$ based on a consistent estimator:

$$\left\{\widehat{\beta}(y) \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\widehat{\Omega}^2(y)}/\sqrt{n}\right\},$$

where Φ is the standard normal distribution function. For a 95% confidence interval, $\Phi^{-1}(1-\alpha/2) = 1.96$. The consistent estimator $\widehat{\Omega}^2(y)$ is given by

$$\widehat{\Omega}^2(y) := \frac{\frac{1}{n}\sum_{i=1}^n \left[Z_i \widehat{\phi}_1^2(y, S_i, \mathcal{D}_i) + (1-Z_i)\widehat{\phi}_0^2(y, S_i, \mathcal{D}_i) + \widehat{\phi}_s^2(y, S_i)\right]}{\left(\frac{1}{n}\sum_{i=1}^n (\Xi_{1,i}^D - \Xi_{0,i}^D)\right)^2}, \quad \text{where}$$

$$\begin{aligned}
\widehat{\phi}_1(y, s, \mathcal{D}_i) &:= \tilde{\phi}_1(y, s, \mathcal{D}_i) - \frac{1}{n_1(s)} \sum_{j \in I_1(s)} \tilde{\phi}_1(y, s, \mathcal{D}_j), \\
\widehat{\phi}_0(y, s, \mathcal{D}_i) &:= \tilde{\phi}_0(y, s, \mathcal{D}_i) - \frac{1}{n_0(s)} \sum_{j \in I_0(s)} \tilde{\phi}_0(y, s, \mathcal{D}_j), \\
\widehat{\phi}_s(y, s) &:= \frac{1}{n_1(s)} \sum_{i \in I_1(s)} (1\!\!1_{\{Y_i \le y\}} - \widehat{\beta}(y) D_i) - \frac{1}{n_0(s)} \sum_{i \in I_0(s)} (1\!\!1_{\{Y_i \le y\}} - \widehat{\beta}(y) D_i), \\
\tilde{\phi}_1(y, s, \mathcal{D}_i) &:= \left[\left(1 - \frac{1}{\widehat{\pi}_1(s)}\right) \widehat{\mu}_1(y, s, X_i) - \widehat{\mu}_0(y, s, X_i) + \frac{1\!\!1_{\{Y_i \le y\}}}{\widehat{\pi}_1(s)}\right] \\
&\quad - \widehat{\beta}(y) \left[\left(1 - \frac{1}{\widehat{\pi}_1(s)}\right) \widehat{\eta}_1(s, X_i) - \widehat{\eta}_0(s, X_i) + \frac{D_i}{\widehat{\pi}_1(s)}\right], \quad \text{and} \\
\tilde{\phi}_0(y, s, \mathcal{D}_i) &:= \left[\left(\frac{1}{\widehat{\pi}_0(s)} - 1\right) \widehat{\mu}_0(y, s, X_i) + \widehat{\mu}_1(y, s, X_i) - \frac{1\!\!1_{\{Y_i \le y\}}}{\widehat{\pi}_0(s)}\right] \\
&\quad - \widehat{\beta}(y) \left[\left(\frac{1}{\widehat{\pi}_0(s)} - 1\right) \widehat{\eta}_0(s, X_i) + \widehat{\eta}_1(s, X_i) - \frac{D_i}{\widehat{\pi}_0(s)}\right].
\end{aligned}$$

Second, an alternative method for inference is empirical bootstrap. The procedure is summarized in Algorithm 1.

Algorithm 1 Bootstrap confidence intervals for regression-adjusted LDTE

Input: Original sample $\{(Y_i, D_i, Z_i, S_i, X_i)\}_{i=1}^n$
Output: $(1-\alpha) \times 100\%$ confidence intervals for the regression-adjusted LDTE

1. For each bootstrap iteration $b = 1, \ldots, B$:

2. Draw a bootstrap sample of size n with replacement: $\{(Y_i^b, D_i^b, Z_i^b, S_i^b, X_i^b)\}_{i=1}^n$ from $\{(Y_i, D_i, Z_i, S_i, X_i)\}_{i=1}^n$

3. Compute regression-adjusted LDTE $\widehat{\beta}(y)$ given the conditional distribution estimator based on the original sample

4. Calculate standard errors $\widehat{\Sigma}(y)$ as the standard deviation of the bootstrapped LDTEs $\{\widehat{\beta}(y)\}_{b=1}^B$,

5. Construct the confidence band:

$$\left\{\widehat{\beta}(y) \pm \Phi^{-1}(1-\alpha/2) \times \widehat{\Sigma}(y) : y \in \mathcal{Y}\right\},$$

where Φ is the standard normal distribution function.

E Additional experimental details

All experiments are run on a Macbook Pro with 36 GB memory and the Apple M3 Pro chip. The code is publicly available at [TBA later].

Table 3: Pre-treatment covariates included in regression adjustment in Oregon Health Insurance Experiment

Variable
Number of ED visits pre-randomization
Number of ED visits resulting in a hospitalization, pre-randomization
Number of Outpatient ED visits, pre-randomization
Number of weekday daytime ED visits, pre-randomization
Number of weekend or nighttime ED visits, pre-randomization
Number of emergent, non-preventable ED visits, pre-randomization
Number of emergent, preventable ED visits, pre-randomization
Number of primary care treatable ED visits, pre-randomization
Number of non-emergent ED visits, pre-randomization
Number of unclassified ED visits, pre-randomization
Number of ED visits for chronic conditions, pre-randomization
Number of ED visits for injury, pre-randomization
Number of ED visits for skin conditions, pre-randomization
Number of ED visits for abdominal pain, pre-randomization
Number of ED visits for back pain, pre-randomization
Number of ED visits for chest pain, pre-randomization
Number of ED visits for headache, pre-randomization
Number of ED visits for mood disorders, pre-randomization
Number of ED visits for psych conditions/substance abuse, pre-randomization
Number of ED visits for a high uninsured volume hospital, pre-randomization
Number of ED visits for a low uninsured volume hospital, pre-randomization
Sum of total charges, pre-randomization
Age
Gender
Health (last 12 months)
Education (highest completed)

References

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113(524):1784–1796.

Hahn, J. (1998). On the role of the propensity score in efficient semiparametric estimation of average treatment effects. *Econometrica*, pages 315–331.

van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Science & Business Media.