Appendix for "Survivor-complier effects in the presence of selection on treatment, with application to a study of prompt ICU admission"

1 Results

1.1 Proof of Proposition 1

To prove non-identifiability of ψ, we only need to find an observed data distribution that can yield different values of the parameter ψ. We give a simple example here, but non-identifiability also follows from the fact that the bounds in Theorem 1 are sharp. Suppose there are no covariates ($\mathbf{X} = \emptyset$) and there are only two principal strata, indexed by variable $C \sim \text{Bernoulli}(0.5)$ with

$$C = 1 \iff (S^{(z=0)} = S^{(z=1)} = 1, A^{(z=1)} > A^{(z=0)}),$$

$$C = 0 \iff (S^{(z=0)} < S^{(z=1)}, A^{(z=1)} = 1).$$

This means there are only treatment-compliers and selection-compliers who take treatment. By unconfoundedness, $\mathbb{P}(Z = z, S = s, A = a) = \mathbb{P}(S^{(z)} = s, A^{(z)} = a) = \mathbb{1}(z = a = 1)/2$. Also suppose Y is binary and $Y^{(z=0)} = 0$ with probability one. Then $\mathbb{P}(Y = 1 \mid A, S, Z) = 0$ except when $A = S = Z = 1$, in which case

$$\mathbb{P}(Y = 1 \mid A = S = Z = 1) = \sum_{c=0}^{1} \mathbb{P}(Y^{(z=1)} = 1 \mid C = c)\mathbb{P}(C = c) = 0.5(\psi + \xi)$$

where $\xi = \mathbb{P}(Y^{(z=1)} = 1 \mid C = 0)$. Thus any choices of (ψ, ξ) with the same sum $(\psi + \xi)$ would yield the same observed data distribution; for example, both $(\psi, \xi) = (1, 0)$ and

$(\psi, \xi) = (0, 1)$ give $\mathbb{P}(Y = 1 \mid A = S = Z = 1) = 0.5$.

1.2 Proof of Proposition 2

We have

$$\begin{aligned}
\beta &= \mathbb{E}(Y^{(z=1)} - Y^{(z=0)} \mid S^{(z=1)} = S^{(z=0)} = 1) \\
&= \mathbb{E}(Y^{(z=1, A^{(z=1)})} - Y^{(z=0, A^{(z=0)})} \mid S^{(z=1)} = S^{(z=0)} = 1) \\
&= \mathbb{E}(Y^{(A^{(z=1)})} - Y^{(A^{(z=0)})} \mid S^{(z=1)} = S^{(z=0)} = 1) \\
&= \mathbb{E}\{(Y^{(a=1)} - Y^{(a=0)}) \mathbb{1}(A^{(z=1)} > A^{(z=0)}) \mid S^{(z=1)} = S^{(z=0)} = 1\} \\
&= \mathbb{E}(Y^{(a=1)} - Y^{(a=0)} \mid S^{(z=1)} = S^{(z=0)} = 1, A^{(z=1)} > A^{(z=0)}) \\
&\qquad \times \mathbb{P}(A^{(z=1)} > A^{(z=0)} \mid S^{(z=1)} = S^{(z=0)} = 1) \\
&= \psi\alpha / \mathbb{P}(S^{(z=1)} = S^{(z=0)} = 1)
\end{aligned}$$

where the first equality follows by definition, the second by the fact that $Y^{(z)} = Y^{(z, A^{(z)})}$ from Assumption 1 (consistency), the second by Assumption 5 (exclusion), the third by Assumption 6 (monotonicity), the fourth by iterated expectation and Assumption 3 (instrumentation), and the last by definition. Rearranging gives the desired result.

1.3 Proof of Theorem 1

To ease notation, in this section all potential outcomes are with respect to interventions on the instrument Z (not treatment A), so that we can write $Y^1 = Y^{(z=1)}$, for example.

1.3.1 Bounds on α

For the upper bound on α, clearly we have

$$\mathbb{P}(A^1 > A^0 \mid \mathbf{X}) \leq \mathbb{P}(A^1 > A^0 \mid \mathbf{X}) + \mathbb{P}(A^1 > S^0 \mid \mathbf{X}).$$

Now note that

$$\begin{aligned}\{A^1 > A^0\} \cup \{A^1 > S^0\} &= [\{A^1 > A^0\} \cup \{A^1 > S^0\} \cup \{A^1 = A^0 = 1\}] \cap \{A^1 = A^0 = 1\} \\ &= \{A^1 = 1\} \cap \{A^1 = A^0 = 1\} = \{A^1 = 1\} \cap \{A^0 = 1\}\end{aligned}$$

where the first equality follows from simple logic, the second by definition of $\{A^1 = 1\}$, and the third by monotonicity. Therefore

$$\begin{aligned}\mathbb{P}(A^1 > A^0 \mid \mathbf{X}) &\leq \mathbb{P}(A^1 = 1 \mid \mathbf{X}) - \mathbb{P}(A^0 = 1 \mid \mathbf{X}) \\ &= \mathbb{P}(A = 1 \mid \mathbf{X}, Z = 1) - \mathbb{P}(A = 1 \mid \mathbf{X}, Z = 0)\end{aligned}$$

where the equality follows by consistency, positivity, and unconfoundedness. Hence

$$\begin{aligned}\alpha &= \mathbb{E}\{\mathbb{P}(A^1 > A^0 \mid \mathbf{X})\} \\ &\leq \mathbb{E}\{\mathbb{P}(A = 1 \mid \mathbf{X}, Z = 1) - \mathbb{P}(A = 1 \mid \mathbf{X}, Z = 0)\} = \alpha_u.\end{aligned}$$

For the lower bound on α, we similarly have

$$\mathbb{P}(A^1 > A^0 \mid \mathbf{X}) \geq \{\mathbb{P}(A^1 > A^0 \mid \mathbf{X}) - \mathbb{P}(S^0 = A^1 = 0 \mid \mathbf{X})\}_+.$$

Now note that the right-hand side (before taking the positive part) is

$$\begin{aligned}\mathbb{P}(A^1 > A^0 \mid \mathbf{X}) + \mathbb{P}(A^0 = A^1 = 0 \mid \mathbf{X}) - \mathbb{P}(A^0 = A^1 = 0 \mid \mathbf{X}) - \mathbb{P}(S^0 = A^1 = 0 \mid \mathbf{X}) \\ = \mathbb{P}(A^0 = 0 \mid \mathbf{X}) - \mathbb{P}(A^1 = 0 \mid \mathbf{X})\end{aligned}$$

$$= \mathbb{P}(A = 0 \mid \mathbf{X}, Z = 0) - \mathbb{P}(A = 0 \mid \mathbf{X}, Z = 1)$$

where the first equality follows by the facts that $\{A^1 > A^0\} \cup \{A^0 = A^1 = 0\} = \{A^0 = 0\}$ since A is binary and

$$\{A^0 = A^1 = 0\} \cup \{S^0 = A^1 = 0\} = \{A^0 = 0\}$$

by monotonicity, and the second follows by consistency, positivity, and unconfoundedness. Hence

$$\alpha = \mathbb{E}\{\mathbb{P}(A^1 > A^0 \mid \mathbf{X})\} \geq \mathbb{E}[\{\mathbb{P}(A = 0 \mid \mathbf{X}, Z = 0) - \mathbb{P}(A \neq 1 \mid \mathbf{X}, Z = 1)\}_+] = \alpha_\ell.$$

1.3.2 Bounds on β

Before deriving the bounds on β, we first give a useful lemma.

Lemma 1. *Suppose $H = pF + qG$, where (H, F, G) are cumulative distribution functions for non-negative random variables (X_H, X_F, X_G), and $p > 0$ and $q > 0$ are weights with $p + q = 1$. Suppose the parent distribution H and weights p and q are known, but the component distributions F and G are unknown. Then sharp bounds on the mean under F are given by*

$$\int \frac{(p - H) \vee 0}{p} = \int (1 - F_\ell^*) \leq \int X_F \, dF \leq \int (1 - F_u^*) = \int \frac{(1 - H) \wedge p}{p}$$

for the bounding distributions $F_\ell^* = \left(\frac{H}{p}\right) \wedge 1$ and $F_u^* = \left(\frac{H-q}{p}\right) \vee 0$.

Proof. Since the random variables associated with the distributions (H, F, G) are non-negative, we can write expectations as integrated survival functions, as in $\int X_F \, dF = \int (1 - F)$. Therefore to show that the means under F_ℓ^* and F_u^* are valid bounds, we must show that $\int F \leq \int F_\ell^*$ and $\int F_u^* \leq \int F$.

For F_ℓ^* note that

$$\begin{aligned}
\int (F_\ell^* - F) &= \int \left(\frac{H}{p}\right) \wedge 1 - \int \left(\frac{H - qG}{p}\right) \\
&= \int_{H>p} (1 - F) + \int_{H \leq p} \frac{qG}{p} \geq 0
\end{aligned}$$

where the last inequality follows since $1 \geq F = (H - qG)/p$ and $qG/p \geq 0$, because (F, G, p, q) are all probabilities bounded between zero and one. Similarly, for F_u^* we have

$$\begin{aligned}
\int (F - F_u^*) &= \int \left(\frac{H - qG}{p}\right) - \int \left(\frac{H - q}{p}\right) \vee 0 \\
&= \int_{H>q} \frac{q(1 - G)}{p} + \int_{H \leq q} F \geq 0
\end{aligned}$$

where again the last inequality follows since (F, G, p, q) are all in $[0, 1]$.

To show sharpness, we must give component distributions F and G that attain the bounds and can be mixed using (p, q) to form any known H, i.e., we can show that $pF_u^* + qG_\ell^* = H$ for $G_\ell^* = (H/q) \wedge 1$ (for F_ℓ^* we can simply reverse the role of F and G). This follows since

$$\begin{aligned}
pF_u^* + qG_\ell^* &= p\left\{\left(\frac{H - q}{p}\right) \vee 0\right\} + q\left\{\left(\frac{H}{q}\right) \wedge 1\right\} \\
&= \begin{cases} p\left(\frac{H-q}{p}\right) + q & \text{if } H > q \\ 0 + q\left(\frac{H}{q}\right) & \text{if } H \leq q \end{cases} \\
&= H.
\end{aligned}$$

$\square$

First note that

$$\begin{aligned}
\beta &= \mathbb{E}\{\mathbb{E}(Y^1 - Y^0 \mid \mathbf{X}, S^0 = S^1 = 1)\mathbb{P}(S^0 = S^1 = 1 \mid \mathbf{X})\}/\mathbb{P}(S^0 = S^1 = 1) \\
&= \mathbb{E}\{\mathbb{E}(Y^1 - Y^0 \mid \mathbf{X}, S^0 = 1)\mathbb{P}(S^0 = 1 \mid \mathbf{X})\}/\mathbb{P}(S^0 = 1) \\
&= \frac{\mathbb{E}[\{\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1) - \mathbb{E}(Y \mid \mathbf{X}, S = 1, Z = 0)\}\mathbb{P}(S = 1 \mid \mathbf{X}, Z = 0)]}{\mathbb{E}\{\mathbb{P}(S = 1 \mid \mathbf{X}, Z = 0)\}}
\end{aligned}$$

where the first equality follows by definition (and iterated expectation), the second by monotonicity, and the third by consistency, positivity, and unconfoundedness.

Now we use Lemma 1 to construct bounds for $\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1)$. Let

$$H = \mathbb{P}(Y^1 \leq y \mid \mathbf{X}, S^1 = 1)$$

$$F = \mathbb{P}(Y^1 \leq y \mid \mathbf{X}, S^0 = 1)$$

$$p = \mathbb{P}(S^0 = 1 \mid \mathbf{X}, S^1 = 1)$$

so that $\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1) = \int(1 - F) = 1 - F(0) = \overline{F}(0)$ since Y is binary. Also note that

$$\begin{aligned}
H &= \mathbb{P}(Y \leq y \mid \mathbf{X}, S = 1, Z = 1) \\
p &= \frac{\mathbb{P}(S^0 = 1 \mid \mathbf{X})}{\mathbb{P}(S^1 = 1 \mid \mathbf{X})} = \frac{\mathbb{P}(S = 1 \mid \mathbf{X}, Z = 0)}{\mathbb{P}(S = 1 \mid \mathbf{X}, Z = 1)} = \frac{\lambda_0(\mathbf{X})}{\lambda_1(\mathbf{X})}
\end{aligned}$$

where the equality for H and the second equality for p use consistency, positivity, and unconfoundedness, and the first equality for p uses monotonicity.

Therefore Lemma 1 gives

$$\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1) \geq \int \frac{(p - H) \vee 0}{p} = \int \frac{(\overline{H} - q) \vee 0}{p}$$

$$= \{\mu_1(\mathbf{X}) - \lambda_1(\mathbf{X}) + \lambda_0(\mathbf{X})\}_+/\lambda_0(\mathbf{X})$$

where the second equality comes from rearranging and using the fact that Y is binary so that $\int \overline{H} = \overline{H}(0) = \mathbb{E}(Y \mid \mathbf{X}, S = 1, Z = 1)$. Similarly we have

$$\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1) \leq \int \frac{(1-H) \wedge p}{p}$$
$$= \{\mu_1(\mathbf{X}) \wedge \lambda_0(\mathbf{X})\}/\lambda_0(\mathbf{X}).$$

Therefore, using the definitions from the main text and plugging in the results from Lemma 1 gives

$$\beta = \mathbb{E}\{\mathbb{E}(Y^1 \mid \mathbf{X}, S^0 = 1)\lambda_0(\mathbf{X}) - \mu_0(\mathbf{X})\}/\mathbb{E}\{\lambda_0(\mathbf{X})\}$$
$$\geq \mathbb{E}\{\{\mu_1(\mathbf{X}) - \lambda_1(\mathbf{X}) + \lambda_0(\mathbf{X})\}_+ - \mu_0(\mathbf{X})\}/\mathbb{E}\{\lambda_0(\mathbf{X})\} = \beta_\ell$$

and similarly

$$\beta \leq \mathbb{E}\{\{\mu_1(\mathbf{X}) \wedge \lambda_0(\mathbf{X})\} - \mu_0(\mathbf{X})\}/\mathbb{E}\{\lambda_0(\mathbf{X})\} = \beta_u$$

1.4 Proof of Theorem 2

Theorem 2 follows from the bounds on α and β given in Theorem 1, along with the expression for ψ given in Proposition 2. In particular, note that the bounds on β imply

$$\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\} \leq \beta \mathbb{E}\{\lambda_0(\mathbf{X})\} \leq \beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}$$

where we used the fact that

$$\mathbb{P}(S^0 = S^1 = 1) = \mathbb{P}(S^0 = 1) = \mathbb{E}\{\mathbb{P}(S = 1 \mid \mathbf{X}, Z = 0)\} > 0$$

where the first equality comes from monotonicity, and the second from consistency, positivity, and unconfoundedness.

Now we consider three cases depending on whether the above bounds on the numerator are positive or negative. All three results follow from the fact that $c > 0$ and $0 < \alpha_\ell \leq \alpha \leq \alpha_u \leq 1$, we have $c/\alpha_u \leq c/\alpha_\ell$ and $-c/\alpha_u \geq -c/\alpha_\ell$. If both numerator bounds are non-negative so that $0 \leq \beta_\ell$ then

$$\frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_u} \leq \frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \psi \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_\ell}.$$

Similarly if both numerator bounds are zero or negative so that $\beta_u \leq 0$ then

$$\frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_\ell} \leq \frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \psi \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_u}.$$

Finally if the lower numerator bound is non-positive and the upper is non-negative so that $\beta_\ell \leq 0 \leq \beta_u$ then it follows that

$$\frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_\ell} \leq \frac{\beta_\ell \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \psi \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha} \leq \frac{\beta_u \mathbb{E}\{\lambda_0(\mathbf{X})\}}{\alpha_\ell}.$$

This yields the desired result.

1.5 Proof of Theorem 3

1.5.1 Efficient influence function for $\psi^*(\boldsymbol{\delta})$

First we prove two lemmas describing efficient influence functions for bounds on (α, β).

Lemma 2. *Under Condition 1 the efficient influence functions for the bounds (α_ℓ, α_u) on* *the survivor-complier proportion are given by* $\varphi_{\ell/u}^{(\alpha)} - \alpha_{\ell/u}$ *for*

$$\varphi_{\ell}^{(\alpha)} = \mathbb{1}\Big\{\theta_0(0 \mid \mathbf{X}) > \theta_1(0 \mid \mathbf{X})\Big\}\Big[\phi_1\{\mathbb{1}(A \neq 0)\} - \phi_0\{\mathbb{1}(A \neq 0)\}\Big]$$

$$\varphi_{u}^{(\alpha)} = \phi_1\{\mathbb{1}(A = 1)\} - \phi_0\{\mathbb{1}(A = 1)\}.$$

Proof. The fact that $\varphi_u^{(\alpha)} - \alpha_u$ is the efficient influence function for α_u is well-known, since α_u is mathematically equivalent to the $\mathbf{X}$-adjusted average treatment effect of Z on $\mathbb{1}(A = 1)$. This result has been previously discussed by Robins *et al.* (1994), Hahn (1998), and Scharfstein *et al.* (1999) among others. Thus we need only consider the parameter α_ℓ.

The lower bound α_ℓ is more delicate, in particular since it is a non-smooth function of the nuisance functions $\theta_z(0 \mid \mathbf{X})$. Nonetheless we can adapt a result from the optimal treatment regime literature to give conditions under which it is pathwise differentiable (i.e., has an influence function) and regularly estimable at $\sqrt{n}$ rates.

Letting $\gamma_1(\mathbf{X}) = \theta_1(0 \mid \mathbf{X}) - \theta_0(0 \mid \mathbf{X})$, note that

$$\alpha_\ell = \mathbb{E}\{\gamma_1(\mathbf{X})_+\} = \mathbb{E}[\gamma_1(\mathbf{X})\mathbb{1}\{\gamma_1(\mathbf{X}) > 0\}].$$

Now we can use the same logic as in Theorem 2 and Lemma 2 of van der Laan & Luedtke (2014) to show that the influence function for α_ℓ under Condition 1 is the same as that treating the indicator $\mathbb{1}\{\gamma_1(\mathbf{X}) > 0\}$ in α_ℓ as known. Hence under Condition 1 the non-smoothness of α_ℓ is inconsequential.

Specifically, letting $\{\mathbb{P}_\epsilon : \epsilon \in \mathbb{R}\}$ denote a smooth parametric submodel passing through $\mathbb{P}$ at $\epsilon = 0$ (e.g., $d\mathbb{P}_\epsilon = (1 + \epsilon h)d\mathbb{P}$ for bounded mean-zero $h = h(\mathbf{O})$), we have (suppressing the subscript on γ_1 for simplicity)

$$\begin{aligned}\alpha_\ell(\mathbb{P}_\epsilon) - \alpha_\ell(\mathbb{P}) &= \int \gamma_\epsilon \mathbb{1}(\gamma_\epsilon > 0) \ d\mathbb{P}_\epsilon - \int \gamma \mathbb{1}(\gamma > 0) \ d\mathbb{P} \\ &= \int \mathbb{1}(\gamma > 0)(\gamma_\epsilon \ d\mathbb{P}_\epsilon - \gamma \ d\mathbb{P}) + \int \gamma_\epsilon \Big\{ \mathbb{1}(\gamma_\epsilon > 0) - \mathbb{1}(\gamma > 0) \Big\} \ d\mathbb{P}_\epsilon.\end{aligned}$$

The first term, after dividing by ϵ and letting $\epsilon \to 0$, is the pathwise derivative for the case where the indicator $\mathbb{1}(\gamma_1 > 0)$ is known. This corresponds to the influence function $\varphi_\ell^{(\alpha)}$ given in the statement of Lemma 2, which follows by standard chain rule arguments as in Hahn (1998) and elsewhere. Now, again following the same logic as in Lemma 2 of van der Laan & Luedtke (2014), we will show that the second term is $o(|\epsilon|)$ under Condition 1, and so does not contribute to the influence function.

In absolute value, the second term is bounded above by

$$\begin{aligned}\int |\gamma_\epsilon| \Big| \mathbb{1}(\gamma_\epsilon > 0) - \mathbb{1}(\gamma > 0) \Big| \ d\mathbb{P}_\epsilon &\leq \int |\gamma_\epsilon| \mathbb{1}(|\gamma| < |\gamma_\epsilon - \gamma|) \ d\mathbb{P}_\epsilon \\ &\leq \int (|\gamma| + C|\epsilon|) \mathbb{1}(|\gamma| < C|\epsilon|) \ d\mathbb{P}_\epsilon \lesssim |\epsilon|(1 + |\epsilon|) \int \mathbb{1}(|\gamma| < C|\epsilon|) \ d\mathbb{P} \\ &= |\epsilon|(1 + |\epsilon|)\mathbb{P}(0 < |\gamma| < C|\epsilon|) = o(|\epsilon|)\end{aligned}$$

where the second bound follows since $|\gamma| = |\gamma_\epsilon - \gamma| - |\gamma_\epsilon|$ whenever γ_ϵ and γ have different signs, the third and the fourth by the submodel construction and boundedness of γ, and the fifth by Condition 1. Hence this term does not contribute to the pathwise derivative, and the influence function for α_ℓ is the same as if the indicator $\mathbb{1}(\gamma_1 > 0)$ was known. $\square$

Lemma 3. *Under Condition 2 the efficient influence functions for the bounds (β_ℓ, β_u) on* *the survivor intention-to-treat effect are given by* $\mathbb{E}\{\phi_0(S)\}^{-1}\{\varphi_{\ell/u}^{(\beta)} - \beta_{\ell/u}\phi_0(S)\}$ *for*

$$\varphi_\ell^{(\beta)} = \mathbb{1}\left\{\frac{\mu_1(\mathbf{X})}{\lambda_1(\mathbf{X}) - \lambda_0(\mathbf{X})} > 1\right\}\Big[\phi_1\{S(Y-1)\} - \phi_0(S)\Big] - \phi_0(SY)$$
$$\varphi_u^{(\beta)} = \mathbb{1}\left\{\frac{\mu_1(\mathbf{X})}{\lambda_0(\mathbf{X})} > 1\right\}\Big\{\phi_0(S) - \phi_1(SY)\Big\} + \phi_1(SY) - \phi_0(SY).$$

Proof. This proof is similar to that of Lemma 2. The efficient influence function for the denominator $\mathbb{E}\{\lambda_0(\mathbf{X})\}$ of the bounds on β is straightforward, as this parameter is mathematically equivalent to the marginal mean of an outcome missing at random (Robins *et al.* 1994; Hahn 1998; Scharfstein *et al.* 1999). Specifically the influence function is given by $\phi_0(S) - \mathbb{E}\{\lambda_0(\mathbf{X})\}$. The same goes for the subtracted numerator term $\mathbb{E}\{\mu_0(\mathbf{X})\}$, which similarly has influence function $\phi_0(SY) - \mathbb{E}\{\mu_0(\mathbf{X})\}$.

Now consider the non-smooth terms in the numerators. Letting $\gamma_2(\mathbf{X}) = \mu_1(\mathbf{X}) + \lambda_0(\mathbf{X}) - \lambda_1(\mathbf{X})$, we have that the non-smooth term in the numerator of β_ℓ is

$$\mathbb{E}\{\gamma_2(\mathbf{X})_+\} = \mathbb{E}[\gamma_2(\mathbf{X})\mathbb{1}\{\gamma_2(\mathbf{X}) > 0\}]$$

and so can be analyzed exactly as in Lemma 2, except replacing Condition 1 with

$$\mathbb{P}\{\mu_1(\mathbf{X}) = \lambda_1(\mathbf{X}) - \lambda_0(\mathbf{X})\} = 0$$

as in Condition 2, to ensure that $\gamma_2 = \mu_1 + \lambda_0 - \lambda_1$ does not have a point mass at zero. Therefore the influence function for this parameter is the same as that treating the indicator $\mathbb{1}(\mu_1 > \lambda_1 - \lambda_0)$ as known, which is given in the statement of Lemma 3.

Similarly, now letting $\gamma_3(\mathbf{X}) = \mu_1(\mathbf{X}) - \lambda_0(\mathbf{X})$, the non-smooth term in the numerator of

β_u is given by

$$\begin{aligned}\mathbb{E}\{\mu_1(\mathbf{X}) \wedge \lambda_0(\mathbf{X})\} &= \mathbb{E}[\mu_1(\mathbf{X})\mathbb{1}\{\lambda_0(\mathbf{X}) \geq \mu_1(\mathbf{X})\} + \lambda_0(\mathbf{X})\mathbb{1}\{\mu_1(\mathbf{X}) > \lambda_0(\mathbf{X})\}] \\ &= \mathbb{E}[\mu_1(\mathbf{X}) + \{\lambda_0(\mathbf{X}) - \mu_1(\mathbf{X})\}\mathbb{1}\{\mu_1(\mathbf{X}) > \lambda_0(\mathbf{X})\}] \\ &= \mathbb{E}[\mu_1(\mathbf{X}) - \gamma_3(\mathbf{X})\mathbb{1}\{\gamma_3(\mathbf{X}) > 0\}].\end{aligned}$$

Again the first term above can be analyzed with standard techniques, as it is mathematically equivalent to the marginal mean of an outcome missing at random. The second term is exactly the same as in the previous two examples in Lemma 2 and the first part of this Lemma 3. The second part of Condition 2, that

$$\mathbb{P}\{\mu_1(\mathbf{X}) = \lambda_0(\mathbf{X})\} = 0,$$

again ensures that $\gamma_3 = \mu_1 - \lambda_0$ does not have a point mass at zero. Therefore the influence function is the same as that treating the indicator $\mathbb{1}(\gamma_3 > 0) = \mathbb{1}(\mu_1 > \lambda_0)$ as known.

Now that we have the influence functions for each the components making up the lower and upper bounds of β, the final result of Lemma 3 follows after combining these influence functions using the chain rule, and rearranging. □

That the efficient influence function of $\psi^*(\boldsymbol{\delta})$ is given as in Theorem 3, i.e.,

$$\varphi(\boldsymbol{\eta}) = \{\varphi^{(\beta)}(\delta_2) - \psi^*(\boldsymbol{\delta})\varphi^{(\alpha)}(\delta_1)\}/\mathbb{E}\{\varphi^{(\alpha)}(\delta_1)\}$$

with $\varphi^{(\alpha)}(\delta_1) = \delta_1\varphi_u^{(\alpha)} + (1-\delta_1)\varphi_\ell^{(\alpha)}$ and $\varphi^{(\beta)}(\delta_2) = \delta_2\varphi_u^{(\beta)} + (1-\delta_2)\varphi_\ell^{(\beta)}$ now follows directly from Lemmas 2–3 together with the chain rule.

1.5.2 Asymptotic results for $\widehat{\psi}^*(\boldsymbol{\delta})$

To ease notation, in this subsection we drop the dependence of all quantities on $\boldsymbol{\delta}$, and write

$$\hat{\psi} = \widehat{\psi}^*(\boldsymbol{\delta}) \ , \ \varphi_\alpha = \varphi^{(\alpha)}(\delta_1) \ , \ \varphi_\beta = \varphi^{(\beta)}(\delta_2).$$

By definition, the estimator $\hat{\psi}$ solves the efficient influence function estimating equation, i.e., $\hat{\psi} = \mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\}/\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}$. Therefore we have

$$\begin{aligned}
\hat{\psi} - \psi &= \frac{\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\}}{\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}} - \frac{\mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\}}{\mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}} = \frac{\mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\}\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}}{\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}\mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}} \\
&= \frac{\mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}[\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\}] - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\}[\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}]}{\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}\mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}} \\
&= \mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}^{-1}\left(\Big[\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\}\Big] - \psi\Big[\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\}\Big]\right)
\end{aligned}$$

Now we will analyze the two terms in parentheses. For the β part we have

$$\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\} = (\mathbb{P}_n - \mathbb{P})\{\varphi_\beta(\hat{\boldsymbol{\eta}}) - \varphi_\beta(\boldsymbol{\eta})\} + (\mathbb{P}_n - \mathbb{P})\varphi_\beta(\boldsymbol{\eta}) - \mathbb{P}\{\varphi_\beta(\hat{\boldsymbol{\eta}}) - \varphi_\beta(\boldsymbol{\eta})\}$$

The first term is a centered empirical process and is $o_{\mathbb{P}}(1/\sqrt{n})$ by, for example, Lemma 19.24 in van der Vaart (2000) since φ_β lies in a Donsker class and $\|\varphi_\beta(\hat{\boldsymbol{\eta}}) - \varphi_\beta(\boldsymbol{\eta})\|^2 = o_{\mathbb{P}}(1)$ by the assumptions of Theorem 3. With sample splitting this term will be $o_{\mathbb{P}}(1/\sqrt{n})$ under only the consistency assumption, without requiring any Donsker conditions. The second term is asymptotically normal after scaling by $\sqrt{n}$, by the central limit theorem.

The third term captures the effect of nuisance estimation, and for it we have

$$\mathbb{P}\{\varphi_\beta(\hat{\boldsymbol{\eta}}) - \varphi_\beta(\boldsymbol{\eta})\} = \delta_2\mathbb{P}\{\varphi_{\beta,u}(\hat{\boldsymbol{\eta}}) - \varphi_{\beta,u}(\boldsymbol{\eta})\} + (1-\delta_2)\mathbb{P}\{\varphi_{\beta,\ell}(\hat{\boldsymbol{\eta}}) - \varphi_{\beta,\ell}(\boldsymbol{\eta})\}.$$

First consider the first term on the right hand side, referring to the definition of φ_β in Lemma 3. Repeated iterated expectation shows that

$$\mathbb{P}\{\hat{\phi}_z(SY) - \phi_z(SY)\} \lesssim \|\hat{\pi}_1 - \pi_1\| \|\hat{\mu}_z - \mu_z\|.$$

Let $\gamma_3 = \mu_1 - \lambda_0$ as before and also let $\phi = \phi_0(S) - \phi_1(SY)$. Then the same arguments, combined with the result from Theorem 3 of van der Laan & Luedtke (2014), show that

$$\begin{aligned}\mathbb{P}\{\mathbb{1}(\hat{\gamma}_3 > 0)\hat{\phi} - \mathbb{1}(\gamma_3 > 0)\phi\} &= \mathbb{P}\{\mathbb{1}(\hat{\gamma}_3 > 0)(\hat{\phi} - \phi) + \{\mathbb{1}(\hat{\gamma}_3 > 0) - \mathbb{1}(\gamma_3 > 0)\}\phi\} \\ &\lesssim \|\hat{\pi}_1 - \pi_1\| \Big(\|\hat{\mu}_1 - \mu_1\| + \|\hat{\lambda}_0 - \lambda_0\|\Big) + \|\hat{\gamma}_3 - \gamma_3\| \sqrt{\mathbb{P}(\gamma_3 < |\hat{\gamma}_3 - \gamma_3|)}.\end{aligned}$$

Therefore, combining the above results, we have that $\mathbb{P}\{\varphi_{\beta,u}(\hat{\boldsymbol{\eta}}) - \varphi_{\beta,u}(\boldsymbol{\eta})\}$ is bounded above (up to constants) by

$$\|\hat{\pi}_1 - \pi_1\| \Big(\max_z \|\hat{\mu}_z - \mu_z\| + \|\hat{\lambda}_0 - \lambda_0\|\Big) + \|\hat{\gamma}_3 - \gamma_3\| \sqrt{\mathbb{P}(\gamma_3 < |\hat{\gamma}_3 - \gamma_3|)}$$

and this is $o_{\mathbb{P}}(1/\sqrt{n})$ by Assumptions 3–4 of Theorem 3.

The same logic shows that, for $\gamma_2 = \mu_1 - (\lambda_1 - \lambda_0)$,

$$\begin{aligned}\mathbb{P}\{\varphi_{\beta,\ell}(\hat{\boldsymbol{\eta}}) - \varphi_{\beta,\ell}(\boldsymbol{\eta})\} \lesssim \|\hat{\pi}_1 - \pi_1\| &\Big(\max_z \|\hat{\lambda}_z - \lambda_z\| + \max_z \|\hat{\mu}_z - \mu_z\|\Big) \\ &+ \|\hat{\gamma}_2 - \gamma_2\| \sqrt{\mathbb{P}(\gamma_2 < |\hat{\gamma}_2 - \gamma_2|)},\end{aligned}$$

and this is also $o_{\mathbb{P}}(1/\sqrt{n})$ by Assumptions 3–4 of Theorem 3.

Therefore $\mathbb{P}\{\varphi_\beta(\hat{\boldsymbol{\eta}}) - \varphi_\beta(\boldsymbol{\eta})\} = o_{\mathbb{P}}(1/\sqrt{n})$, which implies

$$\mathbb{P}_n\{\varphi_\beta(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\beta(\boldsymbol{\eta})\} = (\mathbb{P}_n - \mathbb{P})\varphi_\beta(\boldsymbol{\eta}) + o_{\mathbb{P}}(1/\sqrt{n}).$$

Similarly for the α part of the earlier decomposition we have

$$\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\} = (\mathbb{P}_n - \mathbb{P})\{\varphi_\alpha(\hat{\boldsymbol{\eta}}) - \varphi_\alpha(\boldsymbol{\eta})\} + (\mathbb{P}_n - \mathbb{P})\varphi_\alpha(\boldsymbol{\eta}) - \mathbb{P}\{\varphi_\alpha(\hat{\boldsymbol{\eta}}) - \varphi_\alpha(\boldsymbol{\eta})\}$$

The first term is again centered empirical process and is $o_{\mathbb{P}}(1/\sqrt{n})$ since φ_α lies in a Donsker class and $\|\varphi_\alpha(\hat{\boldsymbol{\eta}}) - \varphi_\alpha(\boldsymbol{\eta})\|^2 = o_{\mathbb{P}}(1)$ by the assumptions of Theorem 3. The second term is asymptotically normal after scaling by $\sqrt{n}$, by the central limit theorem.

As with the β part of the decomposition, we have

$$\mathbb{P}\{\varphi_\alpha(\hat{\boldsymbol{\eta}}) - \varphi_\alpha(\boldsymbol{\eta})\} = \delta_1\mathbb{P}\{\varphi_{\alpha,u}(\hat{\boldsymbol{\eta}}) - \varphi_{\alpha,u}(\boldsymbol{\eta})\} + (1-\delta_1)\mathbb{P}\{\varphi_{\alpha,\ell}(\hat{\boldsymbol{\eta}}) - \varphi_{\alpha,\ell}(\boldsymbol{\eta})\}.$$

Standard iterated expectation arguments show that

$$\mathbb{P}\{\varphi_{\alpha,u}(\hat{\boldsymbol{\eta}}) - \varphi_{\alpha,u}(\boldsymbol{\eta})\} \lesssim \|\hat{\pi}_1 - \pi_1\|\Big(\max_z \|\hat{\theta}_{z1} - \theta_{z1}\|\Big)$$

and this is $o_{\mathbb{P}}(1/\sqrt{n})$ by Assumptions 3–4 of Theorem 3. As with β, for $\gamma_1 = \theta_{10} - \theta_{00}$ we have

$$\mathbb{P}\{\varphi_{\alpha,\ell}(\hat{\boldsymbol{\eta}}) - \varphi_{\alpha,\ell}(\boldsymbol{\eta})\} \lesssim \|\hat{\pi}_1 - \pi_1\|\Big(\max_z \|\hat{\theta}_{z0} - \theta_{z0}\|\Big) + \|\hat{\gamma}_1 - \gamma_1\|\sqrt{\mathbb{P}(\gamma_1 < |\hat{\gamma}_1 - \gamma_1|)},$$

and this is also $o_{\mathbb{P}}(1/\sqrt{n})$ by Assumptions 3–4 of Theorem 3.

Therefore $\mathbb{P}\{\varphi_\alpha(\hat{\boldsymbol{\eta}}) - \varphi_\alpha(\boldsymbol{\eta})\} = o_{\mathbb{P}}(1/\sqrt{n})$, which implies

$$\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\} = (\mathbb{P}_n - \mathbb{P})\varphi_\alpha(\boldsymbol{\eta}) + o_{\mathbb{P}}(1/\sqrt{n}).$$

Hence

$$\hat{\psi} - \psi = \mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\}^{-1}\Big[(\mathbb{P}_n - \mathbb{P})\Big\{\varphi_\beta(\boldsymbol{\eta}) - \psi\varphi_\alpha(\boldsymbol{\eta})\Big\}\Big] + o_{\mathbb{P}}(1/\sqrt{n})$$

which yields the result of the theorem, after applying the continuous mapping theorem and Slutsky's theorem (noting that $\mathbb{P}_n\{\varphi_\alpha(\hat{\boldsymbol{\eta}})\} - \mathbb{P}\{\varphi_\alpha(\boldsymbol{\eta})\} = o_{\mathbb{P}}(1)$ by the above results).