eAppendix to 'RE: Selecting optimal subgroups for treatment using many covariate'

B.B.L. Penning de Vries; Rolf H.H. Groenwold; Alex Luedtke

In what follows, we denote by I the indicator function and by Y^a the counterfactual or potential outcome that would be realised if, possibly contrary to fact, A were set to a. Superscripts are reserved for assigned treatment levels rather than powers. For example, $Y^{I(S)}$ is the counterfactual outcome Y^1 if statement S is true and is Y^0 otherwise. We consider treatment assignment rules that map the vector X of covariate vector L and an error term ε to the value of 0 or 1. We generally require that ε be independent of $(Y^1 - Y^0, L)$ and uniformly distributed between 0 and 1, so that for fixed $p \in [0,1]$, $I(\varepsilon < p)$ takes the Bernoulli distribution with parameter p and, as such, behaves like an independent (unfair) coin toss.

Lemma 1. *Let $\mathcal{X}$ be the support of $X := (L, \varepsilon)$ and suppose that $(Y^1 - Y^0) \perp\!\!\!\perp \varepsilon | L$. If $\mathcal{X}_0 \subseteq \mathcal{X}_1 \subseteq \mathcal{X}$ such that $(L, \varepsilon) \in \mathcal{X}_1 \Rightarrow \mathbb{E}[Y^1 - Y^0|L] > 0$, then $\mathbb{E}\big[Y^{I(X\in\mathcal{X}_1)}\big] \geq \mathbb{E}\big[Y^{I(X\in\mathcal{X}_0)}\big]$. Also, for all $\mathcal{X}' \subseteq \mathcal{X}$, we have $\mathbb{E}\big[Y^{I(X\in\mathcal{X}' \wedge \mathbb{E}[Y^1-Y^0|L]>0)}\big] \geq \mathbb{E}\big[Y^{I(X\in\mathcal{X}')}\big]$.*

Proof. Define $\mathcal{X}_0$ and $\mathcal{X}_1$ as indicated above, so that

$$\begin{aligned}
\mathbb{E}\big[Y^{I(X\in\mathcal{X}_1)}\big] &= \mathbb{E}\big[Y^{I(X\in\mathcal{X}_0 \vee X\in\mathcal{X}_1\setminus\mathcal{X}_0)}\big] \\
&= \mathbb{E}[Y^0] + \mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_0 \vee X \in \mathcal{X}_1\setminus\mathcal{X}_0)] \\
&= \mathbb{E}[Y^0] + \mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_0)] + \mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_1\setminus\mathcal{X}_0)] \\
&= \mathbb{E}\big[Y^{I(X\in\mathcal{X}_0)}\big] + \mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_1\setminus\mathcal{X}_0)].
\end{aligned}$$

If $\Pr(X \in \mathcal{X}_1\setminus\mathcal{X}_0) > 0$, then

$$\begin{aligned}
\mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_1\setminus\mathcal{X}_0)] &= \mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1\setminus\mathcal{X}_0] \Pr(X \in \mathcal{X}_1\setminus\mathcal{X}_0) \\
&= \mathbb{E}\{\mathbb{E}[Y^1 - Y^0|L, \varepsilon]|X \in \mathcal{X}_1\setminus\mathcal{X}_0\} \Pr(X \in \mathcal{X}_1\setminus\mathcal{X}_0) \\
&= \mathbb{E}\{\mathbb{E}[Y^1 - Y^0|L]|X \in \mathcal{X}_1\setminus\mathcal{X}_0\} \Pr(X \in \mathcal{X}_1\setminus\mathcal{X}_0),
\end{aligned}$$

which is strictly positive, because the inner expectation is strictly positive on (any subset of) $\mathcal{X}_1$. Also, if $\Pr(X \in \mathcal{X}_1\setminus\mathcal{X}_0) = 0$, then $\mathbb{E}[(Y^1 - Y^0)I(X \in \mathcal{X}_1\setminus\mathcal{X}_0)] = 0$. In either case, $\mathbb{E}\big[Y^{I(X\in\mathcal{X}_1)}\big] \geq \mathbb{E}\big[Y^{I(X\in\mathcal{X}_0)}\big]$.

As for the last statement, fix some $\mathcal{X}' \subseteq \mathcal{X}$, let $\mathcal{X}'' = \{X \subseteq \mathcal{X} : \mathbb{E}[Y^1 - Y^0|L] > 0\}$ and observe

$$\begin{aligned}
\mathbb{E}\big[Y^{I(X\in\mathcal{X}' \wedge \mathbb{E}[Y^1-Y^0|L]>0)}\big] &= \mathbb{E}\big[Y^{1-I(X\in\mathcal{X}\setminus\mathcal{X}' \vee X\in\mathcal{X}\setminus\mathcal{X}'')}\big] \\
&= \mathbb{E}[Y^1] + \mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}\setminus\mathcal{X}' \vee X \in \mathcal{X}\setminus\mathcal{X}'')] \\
&= \mathbb{E}[Y^1] + \mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}\setminus\mathcal{X}' \vee X \in (\mathcal{X}\setminus\mathcal{X}'')\setminus(\mathcal{X}\setminus\mathcal{X}'))] \\
&= \mathbb{E}[Y^1] + \mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}\setminus\mathcal{X}')] + \mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}'\setminus\mathcal{X}'')] \\
&= \mathbb{E}\big[Y^{X\in\mathcal{X}'}\big] + \mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}'\setminus\mathcal{X}'')]
\end{aligned}$$

with $\mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}'\setminus\mathcal{X}'')] = 0$ if $\Pr(X \in \mathcal{X}'\setminus\mathcal{X}'') = 0$ and, if $\Pr(X \in \mathcal{X}'\setminus\mathcal{X}'') > 0$,

$$\begin{aligned}
\mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}'\setminus\mathcal{X}'')] &= \mathbb{E}[Y^0 - Y^1|X \in \mathcal{X}'\setminus\mathcal{X}''] \Pr(X \in \mathcal{X}'\setminus\mathcal{X}'') \\
&= \mathbb{E}\{-\mathbb{E}[Y^1 - Y^0|L, \varepsilon]|X \in \mathcal{X}'\setminus\mathcal{X}''\} \Pr(X \in \mathcal{X}'\setminus\mathcal{X}'') \\
&= \mathbb{E}\{-\mathbb{E}[Y^1 - Y^0|L]|X \in \mathcal{X}'\setminus\mathcal{X}''\} \Pr(X \in \mathcal{X}'\setminus\mathcal{X}'').
\end{aligned}$$

Because the inner expectation is strictly negative on (any subset of) $\mathcal{X}\setminus\mathcal{X}''$, we have $\mathbb{E}[(Y^0 - Y^1)I(X \in \mathcal{X}'\setminus\mathcal{X}'')] > 0$ if $\Pr(X \in \mathcal{X}'\setminus\mathcal{X}'') > 0$. Hence, $\mathbb{E}\big[Y^{I(X\in\mathcal{X}' \wedge \mathbb{E}[Y^1-Y^0|L]>0)}\big] \geq \mathbb{E}\big[Y^{X\in\mathcal{X}'}\big]$, as desired. $\square$

Lemma 2. *Let $\mathcal{X}$ be the support of $X := (L, \varepsilon)$ and let Cost be a deterministic, positive function of L such that $\mathbb{E}[Cost(L)] \in \mathbb{R}$. For some positive real $\tau \leq \mathbb{E}[Cost(L)]$, define $\mathcal{G}$ to be the set of all deterministic functions $g : \mathcal{X} \to \{0, 1\}$ such that $\mathbb{E}[Cost(L)g(X)] = \tau$. Suppose that $\varepsilon \perp\!\!\!\perp (Y^1 - Y^0, L)$, that $\varepsilon \sim \mathrm{Uniform}[0, 1]$ and that $\mathbb{E}[Y^1 - Y^0|L]$ is defined almost surely. Let $h(L) = \mathbb{E}[Y^a - Y^0|L]/Cost(L)$ and define g^* such that*

$$g^*((L, \varepsilon)) = \begin{cases} 1 & \text{if } h(L) > k, \\ 1 & \text{if } h(L) = k \wedge \varepsilon < p, \\ 0 & \text{if } h(L) < k \end{cases}$$

for all $(L, \varepsilon) \in \mathcal{X}$, and let $k = -\infty$ denote that $h(L) > k$ is necessarily true. Then, there exist $k \in \mathbb{R} \cup \{-\infty\}$ and $p \in [0, 1]$ such that $g^* \in \mathcal{G}$.

Proof. If $\tau = \mathbb{E}[Cost(L)]$, then letting $k = -\infty$ and $p = 0$ gives the result. So assume that $\tau < \mathbb{E}[Cost(L)]$.

Now, let

$$f : k \mapsto \mathbb{E}\big[Cost(L)I(h(L) \geq k)\big]$$

and $K = \{k \in \mathbb{R} : f(k) < \tau\}$.

Note that f is upper semi-continuous (which can be seen to hold because f is left continuous with right limits and monotonically non-increasing). Since upper semi-continuity of f implies $\{x \in \mathbb{R} : f(x) < y\}$ is open for every $y \in \mathbb{R}$, we see that $\mathbb{R}\backslash K$ is closed.

To see that $\mathbb{R}\backslash K$ is nonempty, note that, by the dominated convergence theorem, $\lim_{k\to-\infty} f(k) = \mathbb{E}[Cost(L)] > \tau$. Hence, there exists $k_0 > -\infty$ such that $f(k_0) \geq \tau$, which in turn implies that $\mathbb{R}\backslash K$ is non-empty. Moreover, $\lim_{k\to\infty} f(k) = 0 < \tau$, and so there exists a k_1 such that $f(k_1) < \tau$. Hence, $\mathbb{R}\backslash K$ is bounded above.

Since $\mathbb{R}\backslash K$ is closed, non-empty, and bounded above, we see that $k := \sup \mathbb{R}\backslash K$ belongs to $\mathbb{R}\backslash K$, which implies that $f(k) \geq \tau$. The proof is complete if we can show that there exists a $p \in [0, 1]$ such that $\tau = \mathbb{E}[Cost(L)g^*((L, \varepsilon))]$, where we note that g^* depends on the choice of p. To see that this is the case, first note that

$$\begin{aligned}
\mathbb{E}[Cost(L)g^*((L, \varepsilon))] &= \mathbb{E}\big[Cost(L)I(h(L) > k)\big] + p\mathbb{E}\big[Cost(L)I(h(L) = k)\big] \\
&= (1 - p)\mathbb{E}\big[Cost(L)I(h(L) > k)\big] + pf(k) \\
&= (1 - p) \lim_{k' \downarrow k} f(k') + pf(k).
\end{aligned}$$

Now, for any $k' \geq k$, it holds that $k' \in K$, implying that $f(k') < \tau$. Hence, $\lim_{k'\downarrow k} f(k') \leq \tau$. Combining this fact with the fact that $f(k) \geq \tau$, we see that there exists a $p \in [0, 1]$ such that $(1 - p) \lim_{k'\downarrow k} f(k') + pf(k) = \tau$. This completes the proof. □

Remark. The constraint $\tau \leq \mathbb{E}[Cost(L)]$ in Lemma 2 is weaker than, and so may me replaced with, $\tau \leq \mathbb{E}\big[Cost(L)I(\mathbb{E}[Y^1 - Y^0|L] > 0)\big]$.

Theorem. *Consider some positive real τ. In the setting of Lemma 2, except with $\mathcal{G}$ defined to be the set of all deterministic functions $g : \mathcal{X} \to \{0, 1\}$ such that $\mathbb{E}[Cost(L)g(X)] \leq \tau$, (i) there exist $k \in (0, \infty)$ and $p \in [0, 1]$ such that $g^* \in \mathcal{G}$ and (ii)*

$$g^* \in \operatorname*{arg\,max}_{g \in \mathcal{G}} \mathbb{E}\big[Y^{g(X)}\big].$$

Proof. Since $Y^{g(X)} = Y^0 + (Y^1 - Y^0)g(X)$ by consistency, we have

$$\begin{aligned}
\mathbb{E}\big[Y^{g(X)}\big] &= \mathbb{E}\big[Y^0 + (Y^1 - Y^0)g(X)\big] \\
&= \mathbb{E}\big[Y^0\big] + \mathbb{E}\big[(Y^1 - Y^0)g(X)\big] \\
&= \mathbb{E}\big[Y^0\big] + \mathbb{E}\big\{\mathbb{E}\big[(Y^1 - Y^0)g(X)|g(X)\big]\big\} \\
&= \mathbb{E}\big[Y^0\big] + \mathbb{E}\big[Y^1 - Y^0|g(X) = 1\big]\mathbb{E}[g(X)] \\
&= \mathbb{E}\big[Y^0\big] + \frac{\mathbb{E}\big[Y^1 - Y^0|g(X) = 1\big]}{\mathbb{E}[Cost(L)|g(X) = 1]}\mathbb{E}[Cost(L)g(X)].
\end{aligned}$$

Lemma 1 suggests choosing among all $g \in \mathcal{G}$ such that $\mathbb{E}[Cost(L)g(X)] = \min\{\tau, \mathbb{E}[Cost(L)I(\mathbb{E}[Y^1 - Y^0|L] > 0)]\}$. Let $\mathcal{G}'$ be the set of all such g. Since $\mathbb{E}[Y^0]$ and $\mathbb{E}[Cost(L)g(X)]$ are invariant under changes in $g \in \mathcal{G}'$,

$$\arg\max_{g\in\mathcal{G}} \mathbb{E}[Y^{g(X)}] \supseteq \arg\max_{g\in\mathcal{G}'} \frac{\mathbb{E}[Y^1 - Y^0|g(X) = 1]}{\mathbb{E}[Cost(L)|g(X) = 1]}.$$

Part (i) now follows from Lemma 2. In the remainder of this proof, we show that (ii) holds also. It suffices to show that

$$g^* \in \arg\max_{g\in\mathcal{G}'} \frac{\mathbb{E}[Y^1 - Y^0|g(X) = 1]}{\mathbb{E}[Cost(L)|g(X) = 1]}.$$

To show that the above expression is true, consider first any non-empty $\mathcal{X}_0, \mathcal{X}_1 \subseteq \mathcal{X}$ such that $\mathbb{E}[Cost(L)I(X \in \mathcal{X}_0)] = \mathbb{E}[Cost(L)I(X \in \mathcal{X}_1)] = \tau'$ for some $\tau' \in \mathbb{R}_+$. It holds that

$$\begin{aligned}
\tau' &= \mathbb{E}[Cost(L)I(X \in \mathcal{X}_0)] \\
&= \mathbb{E}[Cost(L)I(X \in \mathcal{X}_0 \cap \mathcal{X}_1) + Cost(L)I(X \in \mathcal{X}_0\backslash\mathcal{X}_1)] \\
&= \mathbb{E}[Cost(L)|X \in \mathcal{X}_0 \cap \mathcal{X}_1] \Pr(X \in \mathcal{X}_0 \cap \mathcal{X}_1) + \mathbb{E}[Cost(L)|X \in \mathcal{X}_0\backslash\mathcal{X}_1] \Pr(X \in \mathcal{X}_0\backslash\mathcal{X}_1)
\end{aligned}$$

and, similarly,

$$\tau' = \mathbb{E}[Cost(L)|X \in \mathcal{X}_0 \cap \mathcal{X}_1] \Pr(X \in \mathcal{X}_0 \cap \mathcal{X}_1) + \mathbb{E}[Cost(L)|X \in \mathcal{X}_1\backslash\mathcal{X}_0] \Pr(X \in \mathcal{X}_1\backslash\mathcal{X}_0),$$

so that $\mathbb{E}[Cost(L)|X \in \mathcal{X}_0\backslash\mathcal{X}_1] \Pr(X \in \mathcal{X}_0\backslash\mathcal{X}_1) = \mathbb{E}[Cost(L)|X \in \mathcal{X}_1\backslash\mathcal{X}_0] \Pr(X \in \mathcal{X}_1\backslash\mathcal{X}_0)$. Therefore, for there exist $a \in \mathbb{R}$ and $b, c \in \mathbb{R}_+ \cup \{0\}$ such that $b + c \neq 0$ and for all $i \in \{0, 1\}$,

$$\begin{aligned}
&\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_i]} \\
&\quad= \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i \cap \mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_{1-i}|X \in \mathcal{X}_i) + \mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}] \Pr(X \notin \mathcal{X}_{1-i}|X \in \mathcal{X}_i)}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_i \cap \mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_{1-i}|X \in \mathcal{X}_i) + \mathbb{E}[Cost(L)|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}] \Pr(X \notin \mathcal{X}_{1-i}|X \in \mathcal{X}_i)} \\
&\quad= \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i \cap \mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_{1-i} \cap \mathcal{X}_i) + \mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i})}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_i \cap \mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_{1-i} \cap \mathcal{X}_i) + \mathbb{E}[Cost(L)|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}] \Pr(X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i})} \\
&\quad= \frac{a + \mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}]\mathbb{E}[Cost(L)|X \in \mathcal{X}_i\backslash\mathcal{X}_{1-i}]^{-1}b}{c + b}.
\end{aligned}$$

This readily shows that

$$\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0]} > \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1]} \Leftrightarrow \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0\backslash\mathcal{X}_1]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0\backslash\mathcal{X}_1]} > \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1\backslash\mathcal{X}_0]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1\backslash\mathcal{X}_0]} \tag{1}$$

for any non-empty $\mathcal{X}_0, \mathcal{X}_1 \subseteq \mathcal{X}$ such that $\mathbb{E}[Cost(L)I(X \in \mathcal{X}_0)] = \mathbb{E}[Cost(L)I(X \in \mathcal{X}_1)] = \tau'$ for some $\tau' \in \mathbb{R}_+$.

Let $\mathcal{X}_0 = \{X \in \mathcal{X} : g^*(X) = 1\}$. Suppose, by way of contradiction, that there exists $\mathcal{X}_1$ such that $\mathbb{E}[Cost(L)I(X \in \mathcal{X}_0)] = \mathbb{E}[Cost(L)I(X \in \mathcal{X}_1)]$ and

$$\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0]} < \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1]},$$

so that, by (1),

$$\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0\backslash\mathcal{X}_1]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0\backslash\mathcal{X}_1]} < \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1\backslash\mathcal{X}_0]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1\backslash\mathcal{X}_0]}. \tag{2}$$

Sets $\mathcal{X}_0 \backslash \mathcal{X}_1$ and $\mathcal{X}_1 \backslash \mathcal{X}_0$ are disjoint and $\mathbb{E}[Cost(L)I(X \in \mathcal{X}_0 \backslash \mathcal{X}_1)] = \mathbb{E}[Cost(L)I(X \in \mathcal{X}_1 \backslash \mathcal{X}_0)]$. In addition, for all non-empty subsets $\mathcal{X}_0' \subseteq \mathcal{X}_0 \backslash \mathcal{X}_1$ and $\mathcal{X}_1' \subseteq \mathcal{X}_1 \backslash \mathcal{X}_0$, we have, by construction of $\mathcal{X}_0$ and disjointedness, that

$$\inf \left\{ \frac{\mathbb{E}[Y^1 - Y^0|L]}{Cost(L)} : X \in \mathcal{X}_0' \right\} \geq \sup \left\{ \frac{\mathbb{E}[Y^1 - Y^0|L]}{Cost(L)} : X \in \mathcal{X}_1' \right\}. \tag{3}$$

Let $f(L) = \mathbb{E}[Y^1 - Y^0|L]$ and $g(L) = Cost(L)$, so that $h(L) = f(L)/g(L)$, and observe that

$$\begin{aligned}
\frac{\mathbb{E}[f(L)|X \in \mathcal{X}_0']}{\mathbb{E}[g(L)|X \in \mathcal{X}_0']} &= \mathbb{E}\left[\frac{f(L)}{g(L)} \frac{g(L)}{\mathbb{E}[g(L)|X \in \mathcal{X}_0']} \middle| X \in \mathcal{X}_0' \right] \\
&\geq \mathbb{E}\left[\inf \left\{ \frac{f(L)}{g(L)} : X \in \mathcal{X}_0' \right\} \frac{g(L)}{\mathbb{E}[g(L)|X \in \mathcal{X}_0']} \middle| X \in \mathcal{X}_0' \right] \\
&= \inf \left\{ \frac{f(L)}{g(L)} : X \in \mathcal{X}_0' \right\} \mathbb{E}\left[\frac{g(L)}{\mathbb{E}[g(L)|X \in \mathcal{X}_0']} \middle| X \in \mathcal{X}_0' \right] \\
&= \inf \left\{ h(L) : X \in \mathcal{X}_0' \right\}. \qquad (4)
\end{aligned}$$

Similarly, we have

$$\frac{\mathbb{E}[f(L)|X \in \mathcal{X}_1']}{\mathbb{E}[g(L)|X \in \mathcal{X}_1']} \leq \sup \left\{ h(L) : X \in \mathcal{X}_1' \right\}. \tag{5}$$

Taken together, (3), (4) and (5) imply

$$\frac{\mathbb{E}\{\mathbb{E}[Y^1 - Y^0|L]|X \in \mathcal{X}_0'\}}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0']} \geq \frac{\mathbb{E}\{\mathbb{E}[Y^1 - Y^0|L]|X \in \mathcal{X}_1'\}}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1']},$$

which, by assumption that $(Y^1 - Y^0, L) \perp\!\!\!\perp \varepsilon$ (and, in turn, $(Y^1 - Y^0) \perp\!\!\!\perp \varepsilon|L$ by weak union), implies

$$\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0']}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0']} \geq \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1']}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1']}.$$

In particular, this implies

$$\frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_0 \backslash \mathcal{X}_1]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_0 \backslash \mathcal{X}_1]} \geq \frac{\mathbb{E}[Y^1 - Y^0|X \in \mathcal{X}_1 \backslash \mathcal{X}_0]}{\mathbb{E}[Cost(L)|X \in \mathcal{X}_1 \backslash \mathcal{X}_0]}.$$

However, in view of (2), this poses a contradiction. Hence, for all $g \in \mathcal{G}'$, we have

$$\frac{\mathbb{E}[Y^1 - Y^0|g^*(X) = 1]}{\mathbb{E}[Cost(L)|g^*(X) = 1]} \geq \frac{\mathbb{E}[Y^1 - Y^0|g(X) = 1]}{\mathbb{E}[Cost(L)|g(X) = 1]},$$

so that $g^* \in \arg\max_{g \in \mathcal{G}} \mathbb{E}[Y^{g(X)}]$, as desired. □

The counterexample to the following proposition suggests that the a greedy approach need not optimize mean potential outcomes with multiple treatment levels and cost or resource constraints.

Proposition. *Let $\mathcal{A}$ be a finite set that includes 0 and denote by $\mathcal{X}$ the support of $X := (L, \varepsilon)$. For $a \in \mathcal{A} \backslash \{0\}$, let $Cost_a$ be a deterministic, positive function of L such that $\mathbb{E}[Cost_a(L)] \in \mathbb{R}$. Let I denote the indicator function and define $\mathcal{G}$ to be the set of all deterministic functions $g : \mathcal{X} \to \mathcal{A}$ such that $\mathbb{E}[Cost_a(L)I(g(X) = a)] = \tau_a$ for all $a \in \mathcal{A} \backslash \{0\}$ and some positive reals $\tau_a \leq \mathbb{E}[Cost_a(L)]$. Suppose $(Y^1 - Y^0) \perp\!\!\!\perp \varepsilon|L$, $\mathbb{E}[Y^1 - Y^0|L] \in \mathbb{R}$ and $\varepsilon|L \sim \text{Uniform}[0, 1]$. Let $h_a(L) = \mathbb{E}[Y^a - Y^0|L]/Cost_a(L)$ for all $a \in \mathcal{A} \backslash \{0\}$ and define g^* such that*

$$g^*((L, \varepsilon)) = \begin{cases} \min \left\{ \underset{a \in \mathcal{A} \backslash \{0\} : \mathcal{P}(a, L)}{\arg\max} \; h_a(L) \right\} & \text{if } \mathcal{P}(a, L) \text{ for some } a \in \mathcal{A} \backslash \{0\}, \\ 0 & \text{otherwise} \end{cases}$$

for all $(L,\varepsilon)\in\mathcal{X}$ *and where* $\mathcal{P}(a,L)$ *is true if and only if* $h_a(L)>k_a\vee[h_a(L)=k_a\wedge\varepsilon<p]$*. Then, (i) there exist* $k_a\in\mathbb{R}\cup\{-\infty\}$ *and* $p_a\in[0,1]$ *for* $a\in\mathcal{A}\backslash\{0\}$ *such that* $g^*\in\mathcal{G}$ *and (ii)*

$$g^*\in\operatorname*{arg\,max}_{g\in\mathcal{G}}\mathbb{E}\big[Y^{g(X)}\big].$$

Counterexample. Let $\mathcal{A}=\{0,1,2\}$ and suppose L is binary with $\Pr(L=1)=1/2$. Suppose also that $Cost_a(L)=1$ and that $\tau_a=1/4$ for all $a\in\mathcal{A}\backslash\{0\}$. Suppose further that

$$\mathbb{E}\big[Y^a|L\big]=\begin{cases}0 & \text{if } a=0,\\ 5 & \text{if } a=1\wedge L=0,\\ 4 & \text{if } a=1\wedge L=1,\\ 4 & \text{if } a=2\wedge L=0,\\ 1 & \text{if } a=2\wedge L=1,\end{cases}\quad\text{so that}\quad h_a(L)=\begin{cases}5 & \text{if } a=1\wedge L=0,\\ 4 & \text{if } a=1\wedge L=1,\\ 4 & \text{if } a=2\wedge L=0,\\ 1 & \text{if } a=2\wedge L=1.\end{cases}$$

Suppose now that $g^*\in\mathcal{G}$. Then, $k_1=5$, $k_2=1$ and $p_1=p_2=1/2$. Indeed, if $k_1>5$, then $\mathcal{P}(1,L)$ is false for all L and, so, $\mathbb{E}[g^*(X)=1]=0\neq\tau_1$. If $k_1<5$, then $\mathcal{P}(1,L)$ is true for all L and $\mathbb{E}[g^*(X)=1]=\mathbb{E}[g^*(X)=1|L=0]/2+\mathbb{E}[g^*(X)=1|L=1]/2=1\neq\tau_1$. If $k_1=5$, then $\mathcal{P}(1,L)$ is true if and only if $L=0$ and $\varepsilon<p$, so $\mathbb{E}[g^*(X)=1]=\Pr(L=0,\varepsilon<p)=\Pr(L=0)\Pr(\varepsilon<p)=p/2$ and $p/2=\tau_1=1/4$ if and only if $p=1/2$. Similar arguments establish that $k_2=1$ and $p_2=1/2$ if $g^*\in\mathcal{G}$.

Hence,

$$\begin{aligned}\mathbb{E}\big[Y^{g^*(X)}\big]&=\mathbb{E}\big[Y^0+(Y^1-Y^0)I(g^*(X)=1)+(Y^2-Y^0)I(g^*(X)=2)\big]\\&=\mathbb{E}\big[Y^0\big]+\mathbb{E}\big[Y^1-Y^0|g^*(X)=1\big]\tau_1+\mathbb{E}\big[Y^2-Y^0|g^*(X)=2\big]\tau_2\\&=\mathbb{E}\big[Y^1-Y^0|L=0,\varepsilon<1/2\big]\tau_1+\mathbb{E}\big[Y^2-Y^0|L=1,\varepsilon<1/2\big]\tau_2\\&=\mathbb{E}\big[Y^1-Y^0|L=0\big]\tau_1+\mathbb{E}\big[Y^2-Y^0|L=1\big]\tau_2\\&=5/4+1/4=1.5.\end{aligned}$$

Now, define $\tilde{g}:\mathcal{X}\to\mathcal{A}$ such that

$$\tilde{g}((L,\varepsilon))=\begin{cases}1 & \text{if } L=1\wedge\varepsilon<1/2,\\ 2 & \text{if } L=0\wedge\varepsilon<1/2,\\ 0 & \text{otherwise,}\end{cases}$$

so that $\mathbb{E}[\tilde{g}(X)=1]=\tau_1$ and $\mathbb{E}[\tilde{g}(X)=2]=\tau_2$. But

$$\begin{aligned}\mathbb{E}\big[Y^{\tilde{g}(X)}\big]&=\mathbb{E}\big[Y^0\big]+\mathbb{E}\big[Y^1-Y^0|\tilde{g}(X)=1\big]\tau_1+\mathbb{E}\big[Y^2-Y^0|\tilde{g}(X)=2\big]\tau_2\\&=\mathbb{E}\big[Y^1-Y^0|L=1\big]\tau_1+\mathbb{E}\big[Y^2-Y^0|L=0\big]\tau_2\\&=4/4+4/4=2.\end{aligned}$$

Hence, $\mathbb{E}\big[Y^{\tilde{g}(X)}\big]>\mathbb{E}\big[Y^{g^*(X)}\big]$ and $\tilde{g}\in\mathcal{G}$ and, so, $g^*\notin\operatorname*{arg\,max}_{g\in\mathcal{G}}\mathbb{E}\big[Y^{g(X)}\big]$. □