Supplement to "Policy Learning with Asymmetric Counterfactual Utilities"

A Constrained optimization formulation

While we have arrived at the objective defined in Eqn (2) through a utility-based framework, we can also characterize this decision problem in the following constrained form,

$$\begin{aligned} \min_{\pi}\ & \Pr(Y(\pi(X)) < Y(1)) \\ \text{subject to}\ & \Pr(Y(\pi(X)) < Y(0)) \leq \delta, \\ & \mathbb{E}[\pi(X)] \leq B, \end{aligned} \tag{A.1}$$

where $\Pr(Y(\pi(X)) < Y(0)) = \Pr(Y(1) = 0, Y(0) = 1, \pi(X) = 1)$ and $\Pr(Y(\pi(X)) < Y(1)) = \Pr(Y(1) = 1, Y(0) = 0, \pi(X) = 0)$ represent the probabilities that policy π gives a harmful treatment or fails to give a useful treatment for a randomly selected member of the population, respectively.

In this formulation, the goal is to find a policy π that minimizes the expected proportion of *false negatives* — failing to give a useful treatment — subject to a constraint on the expected proportion of *false positives* — providing a harmful treatment — and the treatment *budget* — the proportion of units treated. Thus, the decision problem given in Eqn (A.1) allows the policy maker to explicitly state their preferences via the constraint on the number of false positives and the budget, rather than implicitly through the utility function in $R_{e_{01}}(\pi, \varpi)$. It is also possible to swap the constraints and the objective to minimize the proportion of false positives subject to a constraint on the proportion of false negatives. We can also interpret Eqn (A.1) through the lens of multiple testing, for each unit i we have a null hypothesis $H_{0i} : Y_i(1) < Y_i(0)$, i.e. that unit i is harmed by treatment. We can view the policy $\pi(X_i)$ as determining whether to reject H_{0i} or not. Then, the constraint on the proportion of false positives in Eqn (A.1) is a scaling of the false detection rate, where the budget constraint limits the number of rejections, and the objective is a measure of the average power under the alternative $H_{1i} : Y_i(1) > Y_i(0)$.

However, note that $\Pr(Y(\pi(X)) < Y(0)) = \mathbb{E}[\pi(X)e_{01}(X)]$ and $\Pr(Y(\pi(X)) < Y(1)) = \mathbb{E}[(1 - \pi(X))(\tau(x)+e_{01}(X))]$. Thus, we can view the expected utility loss $R_{e_{01}}(\pi, \varpi)$ for a constant comparison policy — either always or never providing treatment — as a Lagrangian relaxation of the decision problem defined in Eqn (A.1), where some choice of the false-positive constraint δ and budget B will correspond to a particular value of the utility ratio $\frac{u_g - u_l}{u_g}$ and cost ratio $\frac{c}{u_g}$. This is in contrast to the regret relative to the oracle policy that maximizes the true value, which involves unidentifiable terms in the relative weights on $\tau(x)$ and $e_{01}(x)$, so it cannot be written as a Lagrangian relaxation of Eqn (A.1).

B Connection to maximin policies

Under the maximin approach, we find a policy π that maximizes the worst-case expected utility. In this appendix we connect the minimax loss policies relative to never and always treating to maximin policies under particular choices of the utility. To do so, we need to specify the utilities under no treatment, $u(0;y_1,y_0)$. We consider two cases.

First, say that $u(0;y_1,y_0)=0$ for all principal strata y_1,y_0. In that case, the expected utility is

$$V(\pi)=\mathbb{E}\left[\pi(X)\left\{u_g\tau(X)+(u_g-u_l)e_{01}(X)-c\right\}\right]=-R_e(\pi,\pi_{\mathbb{O}}).$$

Therefore the maximin policy is equivalent to the minimax loss policy relative to never treating, $\pi^*_{\mathbb{O}}$.

Alternatively, say that the utility function under no treatment mirrors that under treatment, i.e.,

$$u(0;0,0)=u(0;1,1)=0, \qquad u(0;0,1)=u_l, \qquad u(0;1,0)=-u_g.$$

In this case, the expected utility is

$$V(\pi)=\mathbb{E}\left[(\pi(X)-1)\left\{u_g\tau(X)+(u_g-u_l)e_{01}(X)-c\right\}\right]-c=-R_e(\pi,\pi_{\mathbb{1}})-c.$$

So, the maximin policy is equivalent to the minimax loss policy relative to always treating, $\pi^*_{\mathbb{1}}$.

C Algorithms for learning minimax loss policies when estimating nuisance functions via empirical risk minimization

Algorithm 1 Estimated minimax policy $\hat{\pi}$ relative to the always-treat policy $\pi^{\mathbb{1}}$ (when $u_g \geq u_l$) and the never-treat policy $\pi^{\mathbb{O}}$ (when $u_g < u_l$)

Input: Policy classes Π and Δ_+

Output: Estimated minimax policy $\hat{\pi}$ relative to $\pi^{\mathbb{1}}$ or $\pi^{\mathbb{O}}$

1: Find $\hat{\delta}_+$ by solving

$$\min_{\delta\in\Delta_+} -\frac{1}{n}\sum_{i=1}^{n}\delta(X_i)\left\{\widehat{\Gamma}_1(X_i,D_i,Y_i)+\widehat{\Gamma}_0(X_i,D_i,Y_i)-1\right\}.$$

2: Compute weighting and cost functions

$$\hat{c}_1^{\varpi}(x)=u_g+\hat{\delta}_+(x)(u_l-u_g),\ \hat{c}_0^{\varpi}(x)=-u_l-\hat{\delta}_+(x)(u_g-u_l) \text{ and } \hat{c}^{\varpi}(x)=\hat{\delta}_+(x)(u_g-u_l).$$

3: Find a policy $\hat{\pi}\in\underset{\pi\in\Pi}{\arg\min}\,\hat{R}_{\sup}(\pi,\varpi)$.

Algorithm 2 Empirical minimax policy $\hat{\pi}$ relative to the oracle policy π^o

Input: Policy classes Π, Π', Δ_+, and Δ_τ

Output: Empirical minimax policy $\hat{\pi}$ relative to the oracle π^o

Find $\hat{\delta}_+$ by solving

$$\min_{\delta \in \Delta_+} -\frac{1}{n}\sum_{i=1}^{n} \delta(X_i)\left\{\widehat{\Gamma}_1(X_i, D_i, Y_i) + \widehat{\Gamma}_0(X_i, D_i, Y_i) - 1\right\}.$$

if $u_g \geq u_l$ **then**

  Find $\hat{\pi}_{\mathbb{1}}$ via Algorithm 1 with policy class Π'.

  Find $\hat{\pi}_{\mathbb{O}}$ by solving

$$\min_{\pi \in \Pi'} -\frac{1}{n}\sum_{i=1}^{n} \pi(X_i)\left[u_g\left\{\widehat{\Gamma}_1(X_i, D_i, Y_i) - \widehat{\Gamma}_0(X_i, D_i, Y_i)\right\} - c\right].$$

else

  Find $\hat{\pi}_{\mathbb{O}}$ via Algorithm 1 with policy class Π'.

  Find $\hat{\pi}_{\mathbb{1}}$ by solving

$$\min_{\pi \in \Pi'} -\frac{1}{n}\sum_{i=1}^{n} \pi(X_i)\left[u_g\left\{\widehat{\Gamma}_1(X_i, D_i, Y_i) - \widehat{\Gamma}_0(X_i, D_i, Y_i)\right\} - c\right].$$

end if

Find $\hat{\delta}_\tau$ by solving

$$\min_{\delta \in \Delta_\tau} -\frac{1}{n}\sum_{i=1}^{n} \delta(X_i)\left\{\widehat{\Gamma}_1(X_i, D_i, Y_i) - \widehat{\Gamma}_0(X_i, D_i, Y_i)\right\}.$$

Compute weighting and cost functions $\hat{c}_1^{\pi^o}(x), \hat{c}_0^{\pi^o}(x), \hat{c}^{\pi^o}(x)$ via Theorem 3.1.

Find the empirical minimax policy $\hat{\pi} \in \underset{\pi \in \Pi}{\arg\min} \hat{R}_{\sup}(\pi, \pi^o)$.

D Asymmetric utilities based on observed outcomes

Although it is possible to construct asymmetric utilities without relying on principal strata (Babii et al., 2021), doing so places additional restrictions on the structure of utilities. Consider the following utility function based on observed outcomes alone, $u(d, Y(d)) = u_{11} d Y_i(d) + u_{10} d\{1 - Y_i(d)\} + u_{01}(1 - d)Y(d) + u_{00}(1-d)\{1 - Y_i(d)\}$. This utility function includes the interaction between the decision and the observed outcome. Indeed, for a binary decision and outcome, this represents the most general utility that could be specified using the observed outcome.

Table D.1 summarizes the utility gain/loss for treating a unit that belongs to each principal stratum under this setting. With an interaction term, this utility has different utility gains/losses in principal strata $(Y(1) = 1, Y(0) = 0)$ and $(Y(1) = 0, Y(0) = 1)$, allowing for the asymmetry in the utilities as required by the Hippocratic principle. This utility, however, still places restrictions on its structure. In particular, it requires that the difference between the utility gains in principal strata $(Y(1) = 1, Y(0) = 1)$ and $(Y(1) = 0, Y(0) = 1)$ is the same as that between the utility losses in principal

	$Y_i(0) = 1$	$Y_i(0) = 0$
$Y_i(1) = 1$	Harmless $u_{11} - u_{01}$	Useful $u_{11} - u_{00}$
$Y_i(1) = 0$	Harmful $u_{10} - 2u_{01}$	Useless $u_{10} - u_{01} - u_{00}$

Table D.1: Asymmetric utilities gain/loss for treating a unit, $u(1, Y_i(1)) - u(0, Y_i(0))$ based on the observed outcomes for each of the principal strata. The utility function is given by $u(d, Y_i(d)) = u_{11} d Y_i(d) + u_{10} d\{1 - Y_i(d)\} + u_{01}(1-d)Y_i(d) + u_{00}(1-d)\{1 - Y_i(d)\}$. Each cell corresponds to the principal stratum defined by the values of the two potential outcomes, $Y_i(1)$ and $Y_i(0)$. Each entry represents the utility gain/loss of treatment assignment, relative to no treatment, for a unit that belongs to the corresponding principal stratum.

strata $(Y(1) = 1, Y(0) = 0)$ and $(Y(1) = 0, Y(0) = 0)$. Therefore, it might be violated if the difference between harmful and harmless decisions is much greater than that between useful and useless decisions. Thus, a fully general construction of asymmetric utilities requires the use of principal strata, and defining the utility function based on both potential outcomes, $u(d; Y(1), Y(0))$, with utility functions like the one above as a special case.

E Simulation study

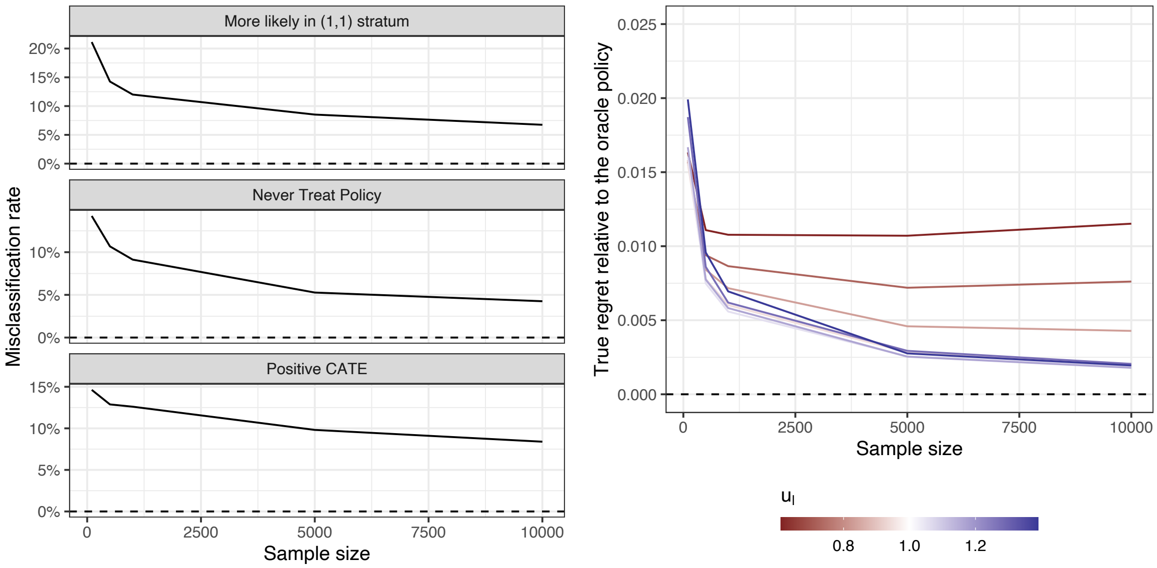
As the results in Section 4.2 show, the misclassification rates of the underlying nuisance classifiers are important in controlling the excess regret due to estimating the weighting and cost functions that make up the worst-case regret. Additionally, although the minimax policies we consider are designed to minimize the worst-case regret, in some cases it may be possible that the true, unidentifiable regret may also be small. To inspect how the misclassification rates and the true regret behave in finite samples as the sample size increases, we now conduct a brief simulation study, where we can know the true values of the principal scores $e_{y_1 y_0}(x)$.

We first generate n 1-dimensional i.i.d. covariates $X_i \sim N(0, 2)$. We then construct log-linear principal scores as

$$e_{y_1 y_0}(x) = \frac{\exp\left(\alpha_{y_1 y_0} + x\beta_{y_1 y_0}\right)}{\sum_{y_1'=0}^{1} \sum_{y_0'=0}^{1} \exp\left(\alpha_{y_1' y_0'} + x\beta_{y_1' y_0'}\right)},$$

where $(\alpha_{00}, \alpha_{10}, \alpha_{01}, \alpha_{11}) = (.2, .15, 0, 0)$, $\beta_{y_1 y_0} \sim N(0, 40)$ for $(y_1, y_0) \in \{(0,0), (1,0), (0,1)\}$, and $\beta_{11} = 0$. We then jointly sample potential outcomes $\{Y_i(1), Y_i(0)\}$ according to the principal scores at covariate value X_i. In this simulation study, we consider a randomized control trial with binary treatment D_i sampled independently as Bernoulli random variables with probability one half.

For each value of sample size $n \in \{100, 500, 1000, 5000, 10000\}$, we draw 1,000 samples according to the above data generating process. In each simulation run, we find the minimax optimal policy with respect to the oracle following Algorithm 2 with zero cost $c = 0$, $u_g = 1$, and u_l varying between 0.6



(a) Misclassification rate for nuisance classifiers (b) Regret of minimax policy relative to oracle

Figure E.1: Performance of nuisance classifiers and the minimax optimal policy relative to the oracle across simulation runs. Panel (a) shows the misclassification rate for the nuisance classifiers $\hat{\delta}_+$ ("More likely in (1,1) stratum") and $\hat{\delta}_\tau$ ("Positive CATE"), as well as the minimax policy relative to the never-treat policy for $u_l = 0.833$. Panel (b) shows the true regret of the minimax optimal policy relative to the oracle, in the sample, for $u_g = 1$ and as u_l varies between 0.5 and 1.5.

and 1.4, where the value of u_l changes within each simulation run.

We use the IPW scoring function and restrict all policy classes to be the set of linear thresholds, solving the optimization problem exactly by direct search. Figure E.1a shows the average misclassification rate for the nuisance classifiers $\hat{\delta}_+$ and $\hat{\delta}_\tau$, as well as the misclassification rate for the the minimax policy relative to always treating for $u_l = 0.833$. As we expect, we see that these misclassification rates decrease as the sample size increases.

Figure E.1b shows the true regret of the minimax policy relative to the oracle as u_l varies. Since the oracle is the best possible policy, this regret is always positive. The regret does decrease along with the sample size, reflecting both the decrease in the nuisance misclassification rate and the decrease in the worst-case excess regret when the nuisance classifiers are known. Notice, however, that the regret does stop decreasing after a certain point, flattening out at a different level depending on the asymmetry in the utility function. In highly asymmetric settings where u_ℓ is small the regret is essentially flat. This is due to the fundamental identifiability problem, and even with infinite data we cannot guarantee that the true regret will be zero. In contrast, in the symmetric setting the regret continues to decrease as the sample size increases.

F Implementation details for application to RHC

F.1 Details on cross-fitting procedure

In the empirical application to Right Heart Catheterization in Section 5, we use a three-fold cross fitting procedure to estimate the nuisance functions. We then use the plug-in method to estimate the nuisance classifiers. Below we present this procedure step-by-step

1. Randomly split the data into three folds.
2. For each fold $k = 1, 2, 3$, estimate the outcome model $\hat{m}^{-k}(\cdot,\cdot)$ and $\hat{d}^{-k}(\cdot)$ on the two other folds via gradient boosted decision stumps.
3. For each unit i, denote $k[i]$ as the fold that it belongs to, then obtain estimates of the outcome model $\hat{m}^{-k[i]}(w, X_i)$, the propensity score $\hat{d}_w^{-k[i]}(X_i)$, and the IP weight $\hat{\gamma}_w^{-k[i]}(D_i, X_i)$.
4. Use these to construct cross-fit estimates of the DR scoring rule:

$$\widehat{\Gamma}_w^{-k[i]}(X_i, D_i, Y_i) = \hat{m}^{-k[i]}(w, X_i) + \{Y_i - \hat{m}^{-k[i]}(w, X_i)\}\hat{\gamma}_w^{-k[i]}(X_i, D_i),$$

   and cross-fit plug-in estimates of the classifiers

$$\begin{aligned}
&\hat{\delta}_+^{-k[i]}(X_i) = \mathbb{1}\{\hat{m}^{-k[i]}(1, X_i) + \hat{m}^{-k[i]}(0, X_i) \geq 1\},\\
&\hat{\delta}_\tau^{-k[i]}(X_i) = \mathbb{1}\{\hat{m}^{-k[i]}(1, X_i) - \hat{m}^{-k[i]}(0, X_i) \geq 0\},\\
&\hat{\pi}_{\mathbb{O}}^{-k[i]}(X_i) = \begin{cases} \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) - \hat{m}^{-k[i]}(0, X_i) \geq \frac{c}{u_g}\right\}, & u_g \geq u_l,\\ \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) \geq \frac{u_l}{u_g}\hat{m}^{-k[i]}(0, X_i) + \frac{c}{u_g}\right\}, & u_g < u_l \text{ and } \hat{\delta}_+^{-k[i]}(X_i) = 0,\\ \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) \geq \frac{u_g}{u_l}\hat{m}^{-k[i]}(0, X_i) + \frac{u_l - u_g + c}{u_l}\right\}, & u_g < u_l \text{ and } \hat{\delta}_+^{-k[i]}(X_i) = 1,\end{cases}\\
&\hat{\pi}_{\mathbb{1}}^{-k[i]}(X_i) = \begin{cases} \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) - \hat{m}^{-k[i]}(0, X_i) \geq \frac{c}{u_g}\right\}, & u_g < u_l,\\ \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) \geq \frac{u_l}{u_g}\hat{m}^{-k[i]}(0, X_i) + \frac{c}{u_g}\right\}, & u_g \geq u_l \text{ and } \hat{\delta}_+^{-k[i]}(X_i) = 0,\\ \mathbb{1}\left\{\hat{m}^{-k[i]}(1, X_i) \geq \frac{u_g}{u_l}\hat{m}^{-k[i]}(0, X_i) + \frac{u_l - u_g + c}{u_l}\right\}, & u_g \geq u_l \text{ and } \hat{\delta}_+^{-k[i]}(X_i) = 1.\end{cases}
\end{aligned}$$

   Then plug in the cross-fit classifiers into the formulas in Appendix H to create cross-fit estimates of $\hat{c}^{-k[i]\varpi}(X_i)$.
5. Solve Eqn (10) with the cross-fit estimates:

$$\hat{\pi} \in \underset{\pi \in \Pi}{\arg\min} -\frac{1}{n}\sum_{i=1}^{n} \pi(X_i)\left\{\hat{c}_1^{-k[i]\varpi}(X_i)\widehat{\Gamma}_1^{-k[i]}(X_i, D_i, Y_i) + \hat{c}_0^{-k[i]\varpi}(X_i)\widehat{\Gamma}_0^{-k[i]}(X_i, D_i, Y_i) + \hat{c}^{-k[i]\varpi}(X_i)\right\}.$$

F.2 Minimax loss policies using a subset of covariates

It is often that case that we wish to construct minimax loss decision rules that only use a subset of the covariates $\mathcal{V} \subset \mathcal{X}$. To consider this case, define $m_{\mathcal{V}}(w, v) \equiv \mathbb{E}[Y(w) \mid V = v]$ to be the expected potential outcome conditioned on the subset of covariates v. Applying Theorem 3.1 to this setting, we get that we can write the worst-case expected utility loss of π relative to ϖ as

$$R_{\text{sup}}(\pi, \varpi) = C - \mathbb{E}\left[\pi(X)\left\{c^{\varpi}_{1\mathcal{V}}(V)m_{\mathcal{V}}(1,V) + c^{\varpi}_{0\mathcal{V}}(V)m_{\mathcal{V}}(0,V) + c^{\varpi}_{\mathcal{V}}(V)\right\}\right],$$

where the weighting and cost functions $c^{\varpi}_{1\mathcal{V}}(\cdot), c^{\varpi}_{0\mathcal{V}}(\cdot), c^{\varpi}_{\mathcal{V}}(\cdot)$ depend on the nuisance classifiers given only the subset of the covariates V, i.e.

$$\begin{aligned}
\delta_{+\mathcal{V}}(v) &= \mathbb{1}\{m_{\mathcal{V}}(1,v) + m_{\mathcal{V}}(0,v) \geq 1\}, \\
\delta_{\tau\mathcal{V}}(v) &= \mathbb{1}\{m_{\mathcal{V}}(1,v) - m_{\mathcal{V}}(0,v) \geq 0\}, \\
\pi^*_{\mathbb{0}\mathcal{V}}(v) &= \begin{cases} \mathbb{1}\left\{m_{\mathcal{V}}(1,v) - m_{\mathcal{V}}(0,v) \geq \frac{c}{u_g}\right\}, & u_g \geq u_l, \\ \mathbb{1}\left\{m_{\mathcal{V}}(1,v) \geq \frac{u_l}{u_g} m_{\mathcal{V}}(0,v) + \frac{c}{u_g}\right\}, & u_g < u_l \text{ and } \delta_{+\mathcal{V}}(v) = 0, \\ \mathbb{1}\left\{m_{\mathcal{V}}(1,v) \geq \frac{u_g}{u_l} m_{\mathcal{V}}(0,v) + \frac{u_l - u_g + c}{u_l}\right\}, & u_g < u_l \text{ and } \delta_{+\mathcal{V}}(v) = 1, \end{cases} \\
\pi^*_{\mathbb{1}\mathcal{V}}(v) &= \begin{cases} \mathbb{1}\left\{m_{\mathcal{V}}(1,v) - m_{\mathcal{V}}(0,v) \geq \frac{c}{u_g}\right\}, & u_g < u_l, \\ \mathbb{1}\left\{m_{\mathcal{V}}(1,v) \geq \frac{u_l}{u_g} m_{\mathcal{V}}(0,v) + \frac{c}{u_g}\right\}, & u_g \geq u_l \text{ and } \delta_{+\mathcal{V}}(v) = 0, \\ \mathbb{1}\left\{m_{\mathcal{V}}(1,v) \geq \frac{u_g}{u_l} m_{\mathcal{V}}(0,v) + \frac{u_l - u_g + c}{u_l}\right\}, & u_g \geq u_l \text{ and } \delta_{+\mathcal{V}}(v) = 1. \end{cases}
\end{aligned}$$

However, note that in order to use observable data, we must account for confounding, since in general $m_{\mathcal{V}}(w,v) \neq \mathbb{E}(Y \mid V = v, W = w)$ when $\mathcal{V}$ is a subset of $\mathcal{X}$. We can however, still use the IPW or DR scoring functions since $m_{\mathcal{V}}(w,v) = \mathbb{E}[\Gamma_w(X,D,Y) \mid V = v]$. So we can write the worst-case expected utility loss in terms of the scoring functions—where we condition on X—and the nuisance classifiers only conditioning on the subset of covariates V:

$$R_{\text{sup}}(\pi, \varpi) = C - \mathbb{E}\left[\pi(V)\left\{c^{\varpi}_{1\mathcal{V}}(V)\Gamma_1(X,D,Y) + c^{\varpi}_{0\mathcal{V}}(V)\Gamma_0(X,D,Y) + c\mathcal{V}^{\varpi}(V)\right\}\right],$$

Constructing plug-in estimates of the nuisance classifiers involves estimating $m_{\mathcal{V}}(w,v) = \mathbb{E}[\Gamma_w(X,D,Y) \mid V = v]$, which we can do by regressing the estimated DR scores on the subset of the covariates V, a variant of the DR-learner (Kennedy, 2022).

Overall, this leads to the following steps:

1. Estimate the DR score $\widehat{\Gamma}_w(x,d,y)$ using *all covariates* X to account for confounding.

2. Estimate the expected potential outcomes given the subset of covariates V, $\hat{m}_{\mathcal{V}}(w,v)$ using the DR-learner and regressing the estimates $\widehat{\Gamma}_w(X_i,D_i,Y_i)$ on V.

3. Form plug in estimates of the nuisance classifiers, e.g. $\hat{\delta}_\tau(v) = \mathbb{1}\{\hat{m}_{\mathcal{V}}(1,v) - \hat{m}_{\mathcal{V}}(0,v)\}$ and $\hat{\delta}_+(v) = \mathbb{1}\{\hat{m}_{\mathcal{V}}(1,v) + \hat{m}_{\mathcal{V}}(0,v) - 1 \geq 0\}$.

4. Get plug-in estimates of the weighting and cost functions $\hat{c}^{\varpi}_{1\mathcal{V}}(V_i), \hat{c}^{\varpi}_{0\mathcal{V}}(V_i), \hat{c}^{\varpi}_{\mathcal{V}}(V_i)$, using the estimates of the nuisance classifiers.

5. Find the policy $\hat{\pi} : \mathcal{V} \to \{0, 1\}$ by solving

$$\min_{\pi \in \Pi} -\frac{1}{n} \sum_{i=1}^{n} \pi(V) \left\{ \hat{c}_{1\mathcal{V}}^{\varpi}(V) \widehat{\Gamma}_1(X, D, Y) + \hat{c}_{0\mathcal{V}}^{\varpi}(V) \widehat{\Gamma}_0(X, D, Y) + \hat{c}\mathcal{V}^{\varpi}(V) \right\}.$$

Finally, note that as in Section F.1 above, we can use cross-fit estimates here, where for each fold k, both $\widehat{\Gamma}_w^{-k}$ and $\hat{m}_{\mathcal{V}}^{-k}$ are fit on data not in fold k. In principle we could do a multi-stage cross-fitting procedure, where for each fold k, we further break up the fold into sub-folds and cross-fit $\hat{m}_{\mathcal{V}}^{-k}$ within the fold k. We opt to use a simpler cross-fitting procedure here, noting that it may impact the quality of the DR-learner estimate $\hat{m}_{\mathcal{V}}^{-k}$.

G Additional figures

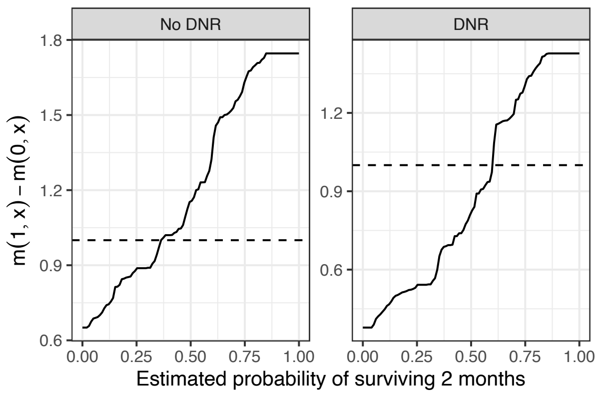


Figure G.1: Plug-in estimate of the decision rule $\hat{\delta}_+(v)$ to classify whether $m_{\mathcal{V}}(1, v) + m_{\mathcal{V}}(0, x) \geq 1$ using the estimated probability of survival and DNR status.

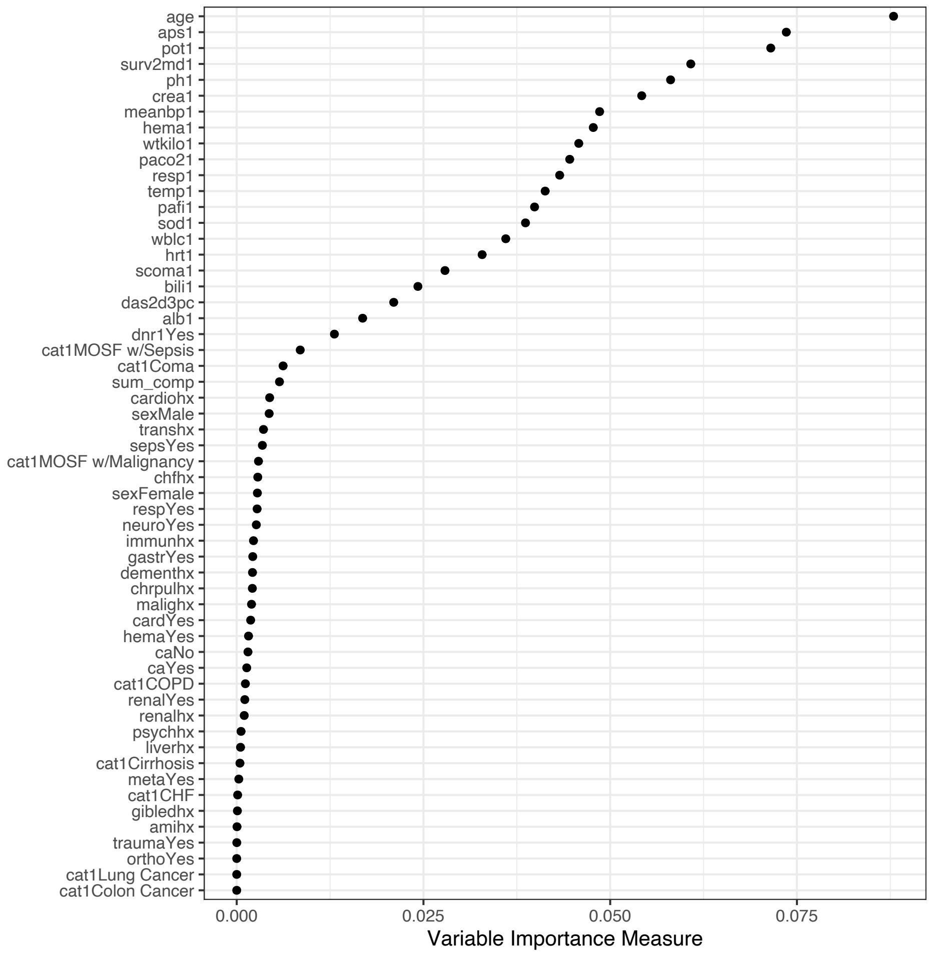


Figure G.2: Variable importance for estimated CATE.

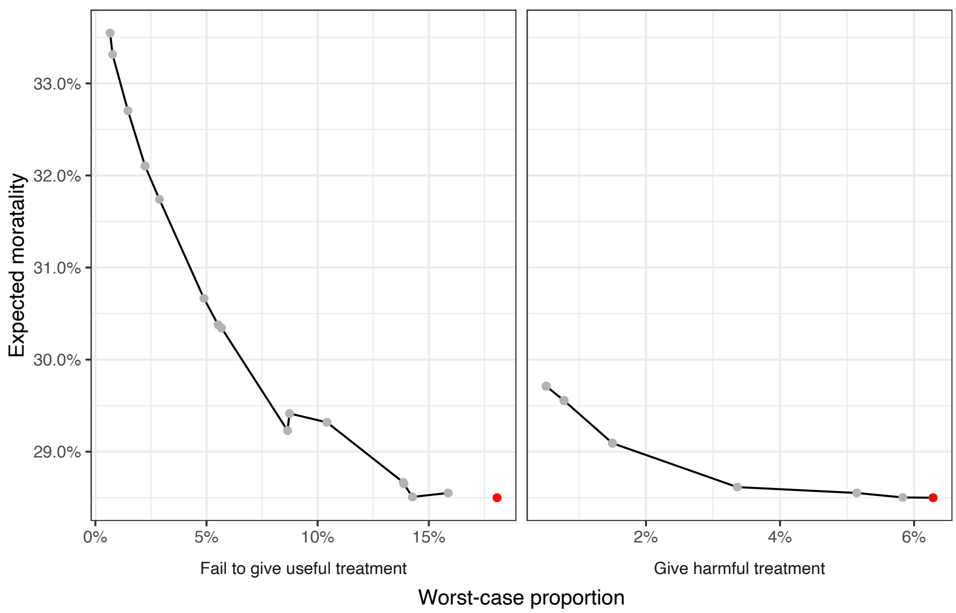


Figure G.3: Estimated expected mortality versus the worst case proportion of cases where the policy fails to give a useful treatment (left panel) and where the policy gives a harmful treatment (right panel), for linear minimax policies relative to never using RHC, as u_l varies in $[0.65, 1.2]$ and $u_g = 1$. The red point is the symmetric linear policy.

H Additional results

Full statement of Theorem 3.1 Let $\pi : \mathcal{X} \to \{0,1\}$ be a deterministic policy. For comparison policy $\varpi \in \{\pi^{\mathbb{0}}, \pi^{\mathbb{1}}\}$, the worst-case expected utility loss of π relative to ϖ is

$$\begin{aligned} R_{\sup}(\pi,\varpi) &= C - \mathbb{E}\left[\pi(X)\left\{c_1^{\varpi}(X)m(1,X) + c_0^{\varpi}(X)m(0,X) + c^{\varpi}(X)\right\}\right] \\ &= C - \mathbb{E}\left[\pi(X)\left\{c_1^{\varpi}(X)\Gamma_1(X,D,Y) + c_0^{\varpi}(X)\Gamma_0(X,D,Y) + c^{\varpi}(X)\right\}\right], \end{aligned} \tag{H.1}$$

where C is a constant that does not depend on π. For $u_g \geq u_l$,

$$\begin{aligned} c_1^{\pi^{\mathbb{0}}}(x) &= u_l + (u_g - u_l)\delta_\tau(x) & c_0^{\pi^{\mathbb{0}}}(x) &= -u_l - (u_g - u_l)\delta_\tau(x) & c^{\pi^{\mathbb{0}}}(x) &= -c, \\ c_1^{\pi^{\mathbb{1}}}(x) &= u_g + \delta_+(x)(u_l - u_g) & c_0^{\pi^{\mathbb{1}}}(x) &= -u_l - \delta_+(x)(u_g - u_l) & c^{\pi^{\mathbb{1}}}(x) &= \delta_+(x)(u_g - u_l) - c, \end{aligned}$$

and for $u_g < u_l$,

$$\begin{aligned} c_1^{\pi^{\mathbb{0}}}(x) &= u_g + \delta_+(x)(u_l - u_g) & c_0^{\pi^{\mathbb{0}}}(x) &= -u_l - \delta_+(x)(u_g - u_l) & c^{\pi^{\mathbb{0}}}(x) &= \delta_+(x)(u_g - u_l) - c, \\ c_1^{\pi^{\mathbb{1}}}(x) &= u_l + (u_g - u_l)\delta_\tau(x) & c_0^{\pi^{\mathbb{1}}}(x) &= -u_l - (u_g - u_l)\delta_\tau(x) & c^{\pi^{\mathbb{1}}}(x) &= -c. \end{aligned}$$

Define $\pi^*_{\mathbb{0}} \equiv \arg\min_\pi R_{\sup}(\pi, \pi^{\mathbb{0}})$ and $\pi^*_{\mathbb{1}} \equiv \arg\min_\pi R_{\sup}(\pi, \pi^{\mathbb{1}})$ as the minimax expected utility loss solutions relative to the never-treat policy and always-treat policy, respectively. The worst-case regret relative to the oracle policy π^o is of the form in Eqn (H.1) where for $u_g \geq u_l$,

$$\begin{aligned} \begin{pmatrix} c_1^{\pi^o}(x) \\ c_0^{\pi^o}(x) \\ c^{\pi^o}(x) \end{pmatrix} =& (1 - \pi^*_{\mathbb{1}}(x)) \begin{pmatrix} u_l + (u_g - u_l)\delta_\tau(x) \\ -u_l - (u_g - u_l)\delta_\tau(x) \\ -c \end{pmatrix} + \pi^*_{\mathbb{0}}(x) \begin{pmatrix} u_g - (u_g - u_l)\delta_+(x) \\ -u_l - (u_g - u_l)\delta_+(x) \\ (u_g - u_l)\delta_+(x) - c \end{pmatrix} \\ &+ (1 - \pi^*_{\mathbb{0}}(x))\pi^*_{\mathbb{1}}(x) \begin{pmatrix} u_l + u_g + (u_g - u_l)(\delta_\tau(x) - \delta_+(x)) \\ -2u_l - (u_g - u_l)(\delta_\tau(x) + \delta_+(x)) \\ (u_g - u_l)\delta_+(x) - 2c \end{pmatrix}, \end{aligned}$$

and for $u_g < u_l$,

$$\begin{aligned} \begin{pmatrix} c_1^{\pi^o}(x) \\ c_0^{\pi^o}(x) \\ c^{\pi^o}(x) \end{pmatrix} =& (1 - \pi^*_{\mathbb{1}}(x)) \begin{pmatrix} u_g - (u_g - u_l)\delta_+(x) \\ -u_l - (u_g - u_l)\delta_+(x) \\ (u_g - u_l)\delta_+(x) - c \end{pmatrix} + \pi^*_{\mathbb{0}}(x) \begin{pmatrix} u_l + (u_g - u_l)\delta_\tau(x) \\ -u_l - (u_g - u_l)\delta_\tau(x) \\ -c \end{pmatrix} \\ &+ (1 - \pi^*_{\mathbb{0}}(x))\pi^*_{\mathbb{1}}(x) \begin{pmatrix} u_l + u_g + (u_g - u_l)(\delta_\tau(x) - \delta_+(x)) \\ -2u_l - (u_g - u_l)(\delta_\tau(x) + \delta_+(x)) \\ (u_g - u_l)\delta_+(x) - 2c \end{pmatrix}. \end{aligned}$$

Corollary H.1 (Minimax regret relative to the always-treat policy)**.** If $u_g \geq u_l$, the minimax regret solution to Equation (5), $\pi^*_{\mathbb{1}} \equiv \arg\min_\pi R_{\sup}(\pi, \pi^{\mathbb{1}})$, is

$$\pi^*_{\mathbb{1}}(x) = \begin{cases} \mathbb{1}\left\{m(1,x) \geq \frac{u_l}{u_g} m(0,x) + \frac{c}{u_g}\right\}, & \delta_+(x) = 0, \\ \mathbb{1}\left\{m(1,x) \geq \frac{u_g}{u_l} m(0,x) + \frac{u_l - u_g + c}{u_l}\right\}, & \delta_+(x) = 1. \end{cases}$$

Otherwise, if $u_g < u_l$, it is given by the symmetric policy,

$$\pi^*_{\mathbb{1}}(x) = \mathbb{1}\left\{\tau(x) \geq \frac{c}{u_g}\right\} = \pi^{\text{symm}}(x).$$

Assumption H.1. There exists an $\alpha > 0$ and a constant C such that for any $t \geq 0$,

(a) $\Pr(|m(1, X) + m(0, X) - 1| \leq t) \leq Ct^\alpha$.

(b) $\Pr(|m(1, X) - m(0, X)| \leq t) \leq Ct^\alpha$.

(c) For $u_g > u_l$ and c,

$$\Pr\left(|\{u_g - (u_g - u_l)\delta_+(X)\}m(1, X) - \{u_l + (u_g - u_l)\delta_+(X)\}m(0, X) + (u_g - u_l)\delta_+(X) - c| \leq t\right) \leq Ct^\alpha.$$

(d) For $u_g > u_l$ and c,

$$\Pr(|\{u_l + (u_g - u_l)\delta_\tau(X)\}\tau(X) - c| \leq t) \leq Ct^\alpha.$$

Theorem H.2. Let $u_g \geq u_l$, and define

$$\widehat{L}_b(x) = \{u_l + \hat{\delta}_\tau(x)(u_g - u_l)\}(\hat{m}(1, x) - \hat{m}(0, x)) - c,$$
$$\widehat{U}_b(x) = \{u_g - (u_g - u_l)\hat{\delta}_+(x)\}\hat{m}(1, x) - \{u_l + (u_g - u_l)\hat{\delta}_+(x)\}\hat{m}(0, x) + (u_g - u_l)\hat{\delta}_+(x) - c,$$

and let $\hat{\pi}^{\text{plug}}_{\mathbb{0}}(x) = \mathbb{1}\{\widehat{L}_b(x) \geq 0\}$ and $\hat{\pi}^{\text{plug}}_{\mathbb{1}}(x) = \mathbb{1}\{\widehat{U}_b(x) \geq 0\}$ be the plug-in estimates of the minimax optimal policies relative to never or always treating. Under Assumptions H.1(b) and H.1(d), the excess worst case regret for $\hat{\pi}^{\text{plug}}_{\mathbb{0}}$ relative to $\pi^*_{\mathbb{0}}$ is

$$R_{\sup}(\hat{\pi}^{\text{plug}}_{\mathbb{0}}, \pi^{\mathbb{0}}) - R_{\sup}(\pi^*_{\mathbb{0}}, \pi^{\mathbb{0}}) \leq 2^{1+\alpha}CU\|m - \hat{m}\|^{1+\alpha}_\infty,$$

where U is a constant depending on the utility values, α, and C. Under Assumptions 2 and H.1(c), the excess worst case regret for $\hat{\pi}^{\text{plug}}_{\mathbb{1}}$ relative to $\pi^*_{\mathbb{1}}$ is

$$R_{\sup}(\hat{\pi}^{\text{plug}}_{\mathbb{1}}, \pi^{\mathbb{1}}) - R_{\sup}(\pi^*_{\mathbb{0}}, \pi^{\mathbb{0}}) \leq 2^{1+\alpha}CU\|m - \hat{m}\|^{1+\alpha}_\infty,$$

where U is a constant depending on the utility values, α, and C.

Corollary H.3. Let $u_g \geq u_l$, $\hat{\pi}_o$ be a solution to Equation (10) with alternative policy $\varpi = \pi^o$ and with nuisance functions $\hat{m}$ and $\hat{d}$ fit on a separate sample and nuisance classifiers $\hat{\delta}_+(x) = \mathbb{1}\{m(1, x) + m(0, x) - 1 \geq 0\}, \hat{\delta}_\tau(m(1, x) - m(0, x) \geq 0), \hat{\pi}^{\text{plug}}_{\mathbb{0}}$, and $\hat{\pi}^{\text{plug}}_{\mathbb{1}}$, and let π^*_o be a solution to Equation (5), with alternative policy $\varpi = \pi^o$. Under the strict overlap condition in Assumption 1, the excess worst-case regret of $\hat{\pi}_o$ relative to π_o^* satisfies

$$R_{\sup}(\hat{\pi}_0, \pi^o) - R_{\sup}(\pi_o^*, \pi^o) \leq 2U_1 \times \left(\frac{6+\eta}{\eta} \times \left(2\mathcal{R}_n(\Pi) + \frac{t}{\sqrt{n}}\right) + \|\hat{m} - m\|_2 \|\hat{\gamma} - \gamma\|_2\right) + 2^{2+\alpha} C U_2 \|\hat{m} - m\|_\infty^{1+\alpha} + (u_g - u_l)\frac{t}{2\sqrt{n}},$$

with probability at least $1 - 2\exp\left(-\frac{t^2}{2}\right)$, where U_1 is a constant depending on the utility values, and U_2 is a constant depending on the utility values, α, and C.

Upper bounds on worst-case proportion of units given a harmful treatment or are failed to be given a useful treatment.

First, note that

$$\begin{aligned}
\Pr(Y(\pi(X)) < Y(0)) &= \Pr(\pi(X) = 1, Y(0) = 1, Y(1) = 0) = \mathbb{E}\left[\pi(X) e_{01}(X)\right], \\
\Pr(Y(\pi(X)) < Y(1)) &= \Pr(\pi(X) = 0, Y(0) = 0, Y(1) = 1) = \mathbb{E}\left[(1 - \pi(X))(\tau(X) + e_{01}(X))\right].
\end{aligned}$$

Plugging in the upper and lower bounds on $e_{01}(X)$ in Section 3, we get the following upper bounds:

$$\begin{aligned}
\sup_{e_{01}(x) \in [L(x), U(x)]} \Pr(Y(\pi(X)) < Y(0)) &= \mathbb{E}\left(\pi(X)\left[m(0, X) + \delta_+(X)\left\{1 - m(1, X) - m(0, X)\right\}\right]\right), \\
\sup_{e_{01}(x) \in [L(x), U(x)]} \Pr(Y(\pi(X)) < Y(1)) &= \mathbb{E}\left(\left\{1 - \pi(X)\right\}\left[m(1, X) + \delta_+(X)\left\{1 - m(1, X) - m(0, X)\right\}\right]\right).
\end{aligned}$$

I Continuous outcomes

Here we briefly consider extending our framework to the case with a binary decision $D \in \{0, 1\}$ but continuous potential outcomes $(Y(0), Y(1)) \in \mathbb{R}^2$. We define the utility function $u(d; y_1, y_0)$ as before and write the value function as

$$V(\pi) = \mathbb{E}\left[u(0; Y(1), Y(0)) + \pi(X) \times (u(1; Y(1), Y(0)) - u(0; Y(1), Y(0)))\right].$$

Defining $e_{y_1 y_0}(x)$ as the conditional joint density of the potential outcomes given $X = x$, the expected utility loss relative to ϖ is

$$V(\varpi) - V(\pi) = \mathbb{E}\left[\pi(X) \int_{y_1} \int_{y_0} (u(1; y_1, y_0) - u(0; y_1, y_0)) e_{y_1 y_0}(x) dy_0 dy_1\right].$$

With continuous outcomes, there are many potential ways to choose the utility function. One option is a utility function such that $u(1; y_1, y_0) - u(0; y_1, y_0) = y_1 - y_0 - u_\ell \mathbb{1}\{y_1 < y_0\}$. This is analogous to the utility function with binary outcomes, with an explicit utility gain/loss associated with a harmful $(Y(1) < Y(0))$ or useful $(Y(1) > Y(0))$ treatment. Defining the conditional probability of harm as $h(x) = \Pr(Y(1) < Y(0) \mid X = x)$, we can write the expected utility loss as

$$V(\varpi) - V(\pi) = \mathbb{E}\left[\pi(X)\{\tau(x) - u_\ell h(x)\}\right].$$

As in the binary case, we can use sharp bounds on the distribution of individual treatment effects (Fan and Park, 2010), $h(x) \in [L_h(x), U_h(x)]$, where

$$L_h(x) = \max\{\sup_y \{F_1(y \mid x) - F_0(y \mid x)\}, 0\},$$
$$U_h(x) = 1 + \min\{\inf_y \{F_1(y \mid x) - F_0(y \mid x)\}, 0\},$$

where $F_1(\cdot \mid x), F_0(\cdot \mid x)$ are the marginal CDFs conditional on $X = x$ for the potential outcomes under treatment and control, respectively. Now we can again define the minimax expected utility loss policy as the policy that solves

$$\min_\pi \max_{h(x) \in [L(x), U(x)]} \mathbb{E}\left[\pi(X)\{\tau(x) - u_l h(x)\}\right].$$

While this again leads to a point-identifiable objective, we note two ways in which this problem is more difficult than with binary outcomes. First, the upper and lower bounds on the probability of harm involves the conditional CDFs of $Y(1)$ and $Y(0)$. These can be more difficult to estimate than the conditional expected outcomes. Second, the bounds involve supremums and infimums over all $y \in \mathbb{R}$. This may require a more careful analysis and stronger assumptions in order to ensure that the default plug-in approach that we suggest for the binary outcome case will lead to reasonable guarantees on the excess expected utility loss.

J Proofs and derivations

J.1 Main results

Derivation of the expected utility loss First, notice that the expected utility of policy π is

$$V(\pi) = \mathbb{E}\left[\sum_{y_1=0}^{1}\sum_{y_0=0}^{1} e_{y_1 y_0}(X) u(0; y_1, y_0)\right] + \underbrace{\mathbb{E}\left[\sum_{y_1=0}^{1}\sum_{y_0=0}^{1} \pi(X) e_{y_1 y_0}(X)\left\{u(1; y_1, y_0) - u(0; y_1, y_0)\right\}\right]}_{(*)}.$$

The second term can be written as

$$\begin{aligned}
(*) &= \mathbb{E}\left[\pi(X)\left\{e_{10}(X)(u_g - c) - e_{01}(X)(u_l + c) - e_{00}(X)c - e_{11}(X)c\right\}\right] \\
&= \mathbb{E}\left[\pi(X))\left\{e_{10}(X)u_g - e_{01}(X)u_l - c(e_{10}(X) + e_{01}(X) + e_{00}(X) + e_{11}(X))\right\}\right] \\
&= \mathbb{E}\left[\pi(X)\left\{(\tau(X) + e_{01}(X))u_g - e_{01}(X)u_l - c\right\}\right] \\
&= \mathbb{E}\left[\pi(X)\left\{u_g \tau(X) + (u_g - u_l)e_{01}(X) - c\right\}\right],
\end{aligned}$$

where we have used the fact that $\tau(x) = e_{10}(x) - e_{01}(x)$. So the expected utility loss of policy π relative to policy ϖ is

$$V(\varpi) - V(\pi) = \mathbb{E}\left[(\varpi(X) - \pi(X))\left\{u_g \tau(X) + (u_g - u_l)e_{01}(X) - c\right\}\right].$$

Proof of Theorem 3.1. Define $b(x) = u_g\tau(x) + (u_g - u_l)e_{01}(X) - c$, and

$$L_b(x) = \min_{e(x)\in[L(x),U(x)]} \{u_g\tau(x) + (u_g - u_l)e_{01}(X) - c\},$$
$$U_b(x) = \max_{e(x)\in[L(x),U(x)]} \{u_g\tau(x) + (u_g - u_l)e_{01}(X) - c\}.$$

Note that the worst-case regret relative to the always and never treat policies are

$$R_{\text{sup}}(\pi, \pi^{\mathbb{O}}) = -\mathbb{E}[\pi(X)L_b(X)],$$
$$R_{\text{sup}}(\pi, \pi^{\mathbb{1}}) = \mathbb{E}[\{1-\pi(X)\}U_b(X)] = \mathbb{E}[U_b(X)] - \mathbb{E}[\pi(X)U_b(X)].$$

From this, we can find the unconstrained minimax regret policies

$$\pi^*_{\mathbb{O}} = \arg\min_{\pi} -\mathbb{E}[\pi(X)L_b(X)] = \mathbb{1}\{L_b(x) \geq 0\},$$
$$\pi^*_{\mathbb{1}} = \arg\min_{\pi} -\mathbb{E}[\pi(X)U_b(X)] = \mathbb{1}\{U_b(x) \geq 0\}.$$

Now, the oracle policy is $\pi^o(x) = \mathbb{1}\{b(x) \geq 0\}$. So if $L_b(x) \geq 0 \Leftrightarrow \pi^*_{\mathbb{O}}(x) = 1$ then $\pi^o(x) = 1$ for all possible values of the principal score $e_{01}(x)$. In this case,

$$\max_{e(x)\in[L(x),U(x)]} \{\pi^o(x) - \pi(x)\}b(x) = \{1-\pi(x)\}U_b(x).$$

Similarly, if $U_b(x) < 0 \Leftrightarrow \pi^*_{\mathbb{1}}(x) = 0$ then $\pi^o(x) = 0$, and

$$\max_{e(x)\in[L(x),U(x)]} \{\pi^o(x) - \pi(x)\}b(x) = -\pi(x)L_b(x).$$

Finally, if $L_b(x) < 0$ and $U_b(x) \geq 0$ (so $\pi^*_{\mathbb{O}}(x) = 0$ and $\pi^*_{\mathbb{1}}(x) = 1$), then the oracle policy can be either 0 or 1, $\pi^o(x) \in \{0, 1\}$. Therefore,

$$\max_{e(x)\in[L(x),U(x)]} \{\pi^o(x) - \pi(x)\}b(x) = \max\{(1-\pi(x))U_b(x), -\pi(x)L_b(x)\} = U_b(x) - \pi(x)\{U_b(x) + L_b(x)\}.$$

Putting together the pieces, the worst-case regret relative to the oracle is

$$\begin{aligned} R_{\text{sup}}(\pi, \pi^o) = \mathbb{E}([\pi^*_{\mathbb{O}}(X) + \{1-\pi^*_{\mathbb{O}}(X)\}\pi^*_{\mathbb{1}}(X)]U_b(X)) \\ - \mathbb{E}[\pi(X)\{\pi^*_{\mathbb{O}}(X)U_b(X) + (1-\pi^*_{\mathbb{1}}(X))L_b(X) + (1-\pi^*_{\mathbb{O}}(X))\pi^*_{\mathbb{1}}(X)(U_b(X) + L_b(X))\}], \end{aligned}$$

and the unconstrained minimizer is

$$\pi^*_o = \arg\min_{\pi} R_{\text{sup}}(\pi, \pi^o) = \begin{cases} \pi^*_{\mathbb{1}}(x), & \pi^*_{\mathbb{O}}(x) = 1, \\ \pi^*_{\mathbb{O}}(x), & \pi^*_{\mathbb{1}}(x) = 0, \\ \mathbb{1}\{U_b(x) \geq -L_b(x)\}, & \pi^*_{\mathbb{O}}(x) = 0, \pi^*_{\mathbb{1}}(x) = 1. \end{cases}$$

Now notice that $\pi^*_{\mathbb{O}}(x) = 1 \Leftrightarrow L_b(x) \geq 0 \Rightarrow U_b(x) \geq 0 \Leftrightarrow \pi^*_{\mathbb{1}}(x) = 1$, and $\pi^*_{\mathbb{1}}(x) = 0 \Leftrightarrow U_b(x) < 0 \Rightarrow$

$L_b(x) < 0 \Leftrightarrow \pi^*_{\mathbb{O}}(x) = 0$, so we can simplify this to

$$\pi^*_o = \arg\min_\pi R_{\sup}(\pi, \pi^o) = \begin{cases} 1, & \pi^*_{\mathbb{O}}(x) = 1, \\ 0, & \pi^*_{\mathbb{1}}(x) = 0, \\ \mathbb{1}\{U_b(x) \geq -L_b(x)\}, & \pi^*_{\mathbb{O}}(x) = 0, \pi^*_{\mathbb{1}}(x) = 1. \end{cases}$$

To complete the proof, we need to compute $L_b(x)$ and $U_b(x)$. First, we begin with the case where $u_g \geq u_l$. In this case,

$$L_b(x) = \{u_l + (u_g - u_l)\delta_\tau(x)\}\tau(x) - c = \{u_l + (u_g - u_l)\delta_\tau(x)\}m(1,x) - \{u_l + (u_g - u_l)\delta_\tau(x)\}m(0,x) - c,$$
$$U_b(x) = \{u_g - (u_g - u_l)\delta_+(x)\}m(1,x) - \{u_l + (u_g - u_l)\delta_+(x)\}m(0,x) + (u_g - u_l)\delta_+(x) - c.$$

This gives the form of the worst-case regret relative to $\pi^{\mathbb{1}}$ and $\pi^{\mathbb{O}}$. For the worst-case regret relative to the oracle, we collect terms to get

$$\begin{pmatrix} c_1^{\pi^o}(x) \\ c_0^{\pi^o}(x) \\ c^{\pi^o}(x) \end{pmatrix} = \begin{cases} (u_l + (u_g - u_l)\delta_\tau(x), -u_l - (u_g - u_l)\delta_\tau(x), -c), & \pi^*_{\mathbb{1}}(x) = 0, \\ (u_g - (u_g - u_l)\delta_+(x), -u_l - (u_g - u_l)\delta_+(x), (u_g - u_l)\delta_+(x) - c), & \pi^*_{\mathbb{O}}(x) = 1, \\ (u_l + u_g + (u_g - u_l)(\delta_\tau(x) - \delta_+(x)), -2u_l - (u_g - u_l)(\delta_\tau(x) + \delta_+(x)), (u_g - u_l)\delta_+(x) - 2c), & \pi^*_{\mathbb{O}}(x) \neq \pi^*_{\mathbb{1}}(x). \end{cases}$$

Now for the case where $u_g < u_l$, the lower and upper bounds switch:

$$L_b(x) = \{u_g - (u_g - u_l)\delta_+(x)\}m(1,x) - \{u_l + (u_g - u_l)\delta_+(x)\}m(0,x) + (u_g - u_l)\delta_+(x) - c,$$
$$U_b(x) = \{u_l + (u_g - u_l)\delta_\tau(x)\}\tau(x) - c = \{u_l + (u_g - u_l)\delta_\tau(x)\}m(1,x) - \{u_l + (u_g - u_l)\delta_\tau(x)\}m(0,x) - c.$$

For the worst-case regret relative to the oracle, we collect terms to get

$$\begin{pmatrix} c_1^{\pi^o}(x) \\ c_0^{\pi^o}(x) \\ c^{\pi^o}(x) \end{pmatrix} = \begin{cases} (u_g - (u_g - u_l)\delta_+(x), -u_l - (u_g - u_l)\delta_+(x), (u_g - u_l)\delta_+(x) - c), & \pi^*_{\mathbb{1}}(x) = 0, \\ (u_l + (u_g - u_l)\delta_\tau(x), -u_l - (u_g - u_l)\delta_\tau(x), -c), & \pi^*_{\mathbb{O}}(x) = 1, \\ (u_l + u_g + (u_g - u_l)(\delta_\tau(x) - \delta_+(x)), -2u_l - (u_g - u_l)(\delta_\tau(x) + \delta_+(x)), (u_g - u_l)\delta_+(x) - 2c), & \pi^*_{\mathbb{O}}(x) \neq \pi^*_{\mathbb{1}}(x). \end{cases}$$

□

For the Proofs of Theorems 4.1 H.2 and 4.2, we prove the result for the case where $u_g \geq u_l$. The case where $u_g < u_l$ follows in the same way, with $\pi_{\mathbb{O}}$ taking the place for $\pi_{\mathbb{1}}$

Proof of Theorem 4.1. This follows directly from combining Lemmas J.1 and J.2 below via the union bound. □

Proof of Theorem 4.2. This follows directly from combining Lemmas J.1 and J.3 below via the union bound. □

Proof of Corollary 4.3. This follows by combining Theorem 4.1 and Lemma J.4 below. □

Proof of Theorem H.2. This follows from Lemmas J.4 and J.5 below. □

Proof of Corollary H.3. This follows by combining Theorem 4.2, Theorem H.2, and Lemma J.4 below. □

J.2 Auxiliary lemmas

Lemma J.1. Let $\hat{\pi}$ be a solution to Equation (10) with nuisance functions $\hat{m}$ and $\hat{d}$ fit on a separate sample, and let π^* be a solution to Equation (5). Under the strict overlap condition in Assumption 1, the excess worst-case regret between $\hat{\pi}$ and π^* is bounded by

$$R_{\text{sup}}(\hat{\pi}, \varpi) - R_{\text{sup}}(\pi^*, \varpi) \leq U \times \left\{ \frac{6+\eta}{\eta} \times \left(2\mathcal{R}_n(\Pi) + \frac{t}{\sqrt{n}} \right) + \sum_{w=0,1} \|\hat{\gamma}_w - \gamma_w\|_2 \, \|\hat{m}(w,\cdot) - m(w,\cdot)\|_2 \right\}$$
$$+ \sup_{\pi \in \Pi} \left| \tilde{R}(\pi, \hat{c}(\cdot), m(\cdot); \varpi) - \tilde{R}(\pi, c(\cdot), m(\cdot); \varpi) \right|,$$

with probability at least $1 - \exp\left(-\frac{t^2}{2}\right)$, where

$$\tilde{R}_{\text{sup}}(\pi, c(\cdot), m(\cdot); \varpi) = \frac{1}{n} \sum_{i=1}^{n} \pi(X_i) \left\{ c_1(X_i) m(1, X_i) + c_0(X_i) m(0, X_i) + c(X_i) \right\}.$$

and U is a constant depending on the utility values.

Proof of Lemma J.1. First, note that the excess regret can be decomposed into

$$R_{\text{sup}}(\hat{\pi}, \varpi) - R_{\text{sup}}(\pi^*, \varpi) = R_{\text{sup}}(\hat{\pi}, \varpi) - \hat{R}_{\text{sup}}(\hat{\pi}, \varpi) + \underbrace{\hat{R}_{\text{sup}}(\hat{\pi}, \varpi) - \hat{R}_{\text{sup}}(\pi^*, \varpi)}_{\leq 0} + \hat{R}_{\text{sup}}(\pi^*, \varpi) - R_{\text{sup}}(\pi^*, \varpi)$$
$$\leq 2 \sup_{\pi \in \Pi} |\hat{R}_{\text{sup}}(\pi, \varpi) - R_{\text{sup}}(\pi, \varpi)|,$$

where we have used that $\hat{\pi}$ minimizes $\hat{R}_{\text{sup}}(\pi^*, \varpi)$.

We further decompose $\hat{R}_{\text{sup}}(\pi, \varpi) - R_{\text{sup}}(\pi, \varpi)$ into

$$\begin{aligned} \hat{R}_{\text{sup}}(\pi, \varpi) - R_{\text{sup}}(\pi, \varpi) &= \hat{R}_{\text{sup}}(\pi, \varpi) - \tilde{R}(\pi, \hat{c}(\cdot), m(\cdot); \varpi) && \text{(a)} \\ &\quad + \tilde{R}(\pi, c(\cdot), m(\cdot); \varpi) - R_{\text{sup}}(\pi, \varpi) && \text{(b)} \\ &\quad + \tilde{R}(\pi, \hat{c}(\cdot), m(\cdot); \varpi) - \tilde{R}(\pi, c(\cdot), m(\cdot); \varpi) \end{aligned}$$

We will now control terms (a) and (b), following closely the proof of Lemma 4 in Athey and Wager

(2021). First note that we have the decompositions

$$\begin{aligned}\hat{\Gamma}_1(X,D,Y)-m(1,X)&=\hat{m}(1,X)-m(1,X)+\frac{D}{\hat{d}(X)}\{Y-\hat{m}(1,X)\}\\&=\{\hat{m}(1,X)-m(1,X)\}\times\left(1-\frac{D}{d(X)}\right)+\frac{D}{\hat{d}(X)}\{Y-m(1,X)\}\\&\quad+\left(\frac{D}{\hat{d}(X)}-\frac{D}{d(X)}\right)\times\{m(1,X)-\hat{m}(1,X)\}\end{aligned}$$

and

$$\begin{aligned}\hat{\Gamma}_0(X,D,Y)-m(0,X)&=\hat{m}(0,X)-m(0,X)+\frac{1-D}{1-\hat{d}(X)}\{Y-\hat{m}(0,X)\}\\&=\{\hat{m}(0,X)-m(0,X)\}\times\left(1-\frac{1-D}{1-d(X)}\right)+\frac{1-D}{1-\hat{d}(X)}\{Y-m(0,X)\}\\&\quad+\left(\frac{1-D}{1-\hat{d}(X)}-\frac{1-D}{1-d(X)}\right)\times\{m(0,X)-\hat{m}(0,X)\}.\end{aligned}$$

With this, we can compute the expectation of term (a):

$$\begin{aligned}\mathbb{E}[(a)]&=\mathbb{E}\left[\pi(X)\left(\hat{c}_1(X)\left\{\hat{\Gamma}_1(X,D,Y)-m(1,X)\right\}+\hat{c}_0(X)\left\{\hat{\Gamma}_0(X,D,Y)-m(0,X)\right\}\right)\right]\\&=\mathbb{E}\left[\pi(X)\hat{c}_1(X)\left(\frac{D}{\hat{d}(X)}-\frac{D}{d(X)}\right)\times(m(1,X)-\hat{m}(1,X))\right]\\&\quad+\mathbb{E}\left[\pi(X)\hat{c}_0(X)\left(\frac{1-D}{1-\hat{d}(X)}-\frac{1-D}{1-d(X)}\right)\times(m(0,X)-\hat{m}(0,X))\right],\end{aligned}$$

where we have used the fact that

$$\begin{aligned}\mathbb{E}\left[\pi(X)\hat{c}_1(X)\left(\hat{m}(1,X)-m(1,X)\right)\times\left(1-\frac{D}{d(X)}\right)\right]&=0,\\\mathbb{E}\left[\pi(X)\hat{c}_1(X)\frac{D}{\hat{d}(X)}\{Y-m(1,X)\}\right]&=0,\\\mathbb{E}\left[\pi(X)\hat{c}_0(X)\{\hat{m}(0,X)-m(0,X)\}\times\left(1-\frac{1-D}{1-d(X)}\right)\right]&=0,\\\mathbb{E}\left[\pi(X)\hat{c}_0(X)\frac{1-D}{1-\hat{d}(X)}\{Y-m(0,X)\}\right]&=0,\end{aligned}$$

because $\hat{c}$, $\hat{m}$, and $\hat{d}$ come from a different sample.

The expectation of term (b) is

$$\mathbb{E}[(b)] = \mathbb{E}[\pi(X)\{c_1(X)m(1,X) + c_0(X)m(0,X) + c(X)\}] - R_{\text{sup}}(\pi, \varpi) = 0.$$

Now define a function $f : \mathcal{X} \to \mathbb{R}$ as

$$\begin{aligned} f_\pi(x,d,y) \quad \equiv \quad & \pi(x)\left[\hat{c}_1(x)\left\{\hat{\Gamma}_1(x,d,y) - m(1,x)\right\} + \hat{c}_0(x)\left\{\hat{\Gamma}_0(x,d,y) - m(0,x)\right\}\right] \\ & + \pi(x)\{c_1(x)m(1,x) + c_0(x)m(0,x) + c(x)\} \end{aligned}$$

and the function class $\mathcal{F} \equiv \{f_\pi \mid \pi \in \Pi\}$ as the set of all functions f as we vary π in Π.

With this notation, we can write the sum of terms (a) and (b) as

$$(a) + (b) = \frac{1}{n}\sum_{i=1}^{n} f_\pi(X_i, D_i, Y_i) - R_{\text{sup}}(\pi, \varpi),$$

and from above the expectation of $f_\pi(X_i, D_i, Y_i)$ is

$$\begin{aligned} \mathbb{E}[f_\pi(X,D,Y)] = R_{\text{sup}}(\pi,\varpi) & + \mathbb{E}\left[\pi(X)\hat{c}_1(X)\left(\frac{D}{\hat{d}(X)} - \frac{D}{d(X)}\right) \times (m(1,X) - \hat{m}(1,X))\right] \\ & + \mathbb{E}\left[\pi(X)\hat{c}_0(X)\left(\frac{1-D}{1-\hat{d}(X)} - \frac{1-D}{1-d(X)}\right) \times (m(0,X) - \hat{m}(0,X))\right]. \end{aligned}$$

Putting together the pieces, we can write

$$\begin{aligned} |(\text{a}) + (\text{b})| &= \left|\frac{1}{n}\sum_{i=1}^{n} f_\pi(X_i,D_i,Y_i) - R_{\text{sup}}(\pi,\varpi)\right| \\ &= \left|\frac{1}{n}\sum_{i=1}^{n} f_\pi(X_i,D_i,Y_i) - \mathbb{E}[f_\pi(X,D,Y)] + \mathbb{E}[f_\pi(X,D,Y)] - R_{\text{sup}}(\pi,\varpi)\right| \\ &\leq \left|\frac{1}{n}\sum_{i=1}^{n} f_\pi(X_i,D_i,Y_i) - \mathbb{E}[f_\pi(X,D,Y)]\right| \\ &\quad + \left|\mathbb{E}\left[\pi(X)\hat{c}_1(X)\left(\frac{D}{\hat{d}(X)} - \frac{D}{d(X)}\right) \times (m(1,X) - \hat{m}(1,X))\right]\right| \\ &\quad + \left|\mathbb{E}\left[\pi(X)\hat{c}_0(X)\left(\frac{1-D}{1-\hat{d}(X)} - \frac{1-D}{1-d(X)}\right) \times (m(0,X) - \hat{m}(0,X))\right]\right|. \end{aligned}$$

Now notice that for $\varpi \in \{\pi^{\mathbb{0}}, \pi^{\mathbb{1}}, \pi^{o}\}$, $|c_1(x)m(1,x) + c_0(x)m(0,x) + c(x)|$, $|c_1(x)|$, and $|c_0(x)|$ are bounded by some constant U depending on the utilities. From the decompositions above, by the strict overlap condition in Assumption 1, and because $Y_i \in \{0,1\}$,

$$
\begin{aligned}
\left|\hat{\Gamma}_1(X_i, D_i, Y_i) - m(1,x)\right| \leq & \left|\{\hat{m}(1,X_i) - m(1,X_i)\} \times \left(1 - \frac{D_i}{d(X_i)}\right)\right| \\
& + \left|\frac{D_i}{\hat{d}(X_i)} \times \{Y_i - m(1,X_i)\}\right| \\
& + \left|\left(\frac{D_i}{d(X_i)} - \frac{D_i}{\hat{d}(X_i)}\right) \times \{\hat{m}(1,X_i) - m(1,X_i)\}\right| \\
\leq & \frac{1-\eta}{\eta}\|\hat{m} - m\|_\infty + \frac{1}{\eta} + \left\|\frac{1}{d} - \frac{1}{\hat{d}}\right\|_\infty \|\hat{m} - m\|_\infty \\
\leq & \frac{1-\eta}{\eta} + \frac{1}{\eta} + \frac{1}{\eta} - \frac{1}{1-\eta} \\
\leq & \frac{3}{\eta}.
\end{aligned}
$$

Similarly,

$$
\left|\hat{\Gamma}_0(X_i, D_i, Y_i) - m(0,x)\right| \leq \frac{1-\eta}{\eta}\|\hat{m} - m\|_\infty + \frac{1}{\eta} + \left\|\frac{1}{1-d} - \frac{1}{1-\hat{d}}\right\|_\infty \|\hat{m} - m\|_\infty \leq \frac{3}{\eta}.
$$

This combines to give that for any x, d, y,

$$
|f_\pi(x,d,y)| \leq U \times \frac{6+\eta}{\eta}.
$$

This also shows that the Rademacher complexity of $\mathcal{F}$ is:

$$
\mathcal{R}_n(\mathcal{F}) = 2U \times \frac{6+\eta}{\eta} \times \mathcal{R}_n(\Pi).
$$

So by Wainwright (2019) Theorem 4.2, for any $n \geq 1$ and $t \geq 0$,

$$
\sup_{f \in \mathcal{F}} \left|\frac{1}{n}\sum_{i=1}^n f(X_i) - \mathbb{E}[f(X)]\right| \leq 2U \times \frac{6+\eta}{\eta} \times \left(2\mathcal{R}_n(\Pi) + \frac{t}{\sqrt{n}}\right),
$$

with probability at least $1 - \exp\left(-\frac{t^2}{2}\right)$.

Finally, notice that by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&\left|\mathbb{E}\left[\pi(X)\hat{c}_1(X)\left(\frac{D}{\hat{d}(X)} - \frac{D}{d(X)}\right) \times (m(1,X) - \hat{m}(1,X))\right]\right| \\
&\qquad \leq U\sqrt{\mathbb{E}\left[\left(\frac{D}{\hat{d}(X)} - \frac{D}{d(X)}\right)^2\right]\mathbb{E}\left[(m(1,X) - \hat{m}(1,X))^2\right]},
\end{aligned}
$$

and

$$\left|\mathbb{E}\left[\pi(X)\hat{c}_0(X)\left(\frac{1-D}{1-\hat{d}(X)}-\frac{1-D}{1-d(X)}\right)\times(m(0,X)-\hat{m}(0,X))\right]\right|$$
$$\leq U\sqrt{\mathbb{E}\left[\left(\frac{1-D}{1-\hat{d}(X)}-\frac{1-D}{1-d(X)}\right)^2\right]\mathbb{E}\left[(m(0,X)-\hat{m}(0,X))^2\right]}.$$

Combining these two bounds gives the result. □

Lemma J.2. For $u_g \geq u_l$,

$$\sup_{\pi\in\Pi}\left|\tilde{R}_{\text{sup}}(\pi,\hat{c},m;\pi^*_{\mathbb{1}})-\tilde{R}_{\text{sup}}(\pi,c,m,\pi^*_{\mathbb{1}})\right| \leq (u_g-u_l)\times\left(R_+(\hat{\delta}_+)+\frac{t}{2\sqrt{n}}\right),$$

with probability at least $1-e^{-\frac{t^2}{2}}$.

Proof of Lemma J.2. First we have the bound,

$$\begin{aligned}\tilde{R}_{\text{sup}}(\pi,\hat{c},m;\pi^*_{\mathbb{1}})-\tilde{R}_{\text{sup}}(\pi,c,m,\pi^*_{\mathbb{1}}) &= \frac{u_g-u_l}{n}\sum_{i=1}^n \pi(X_i)\left\{\hat{\delta}_+(X_i)-\delta_+(X_i)\right\}\{m(1,X_i)+m(0,X_i)-1\}\\ &\leq \frac{u_g-u_l}{n}\sum_{i=1}^n \mathbb{1}\left\{\hat{\delta}_+(X_i)\neq\delta_+(X_i)\right\}|m(1,X_i)+m(0,X_i)-1|.\end{aligned}$$

Now note that

$$\mathbb{E}\left[\mathbb{1}\left\{\hat{\delta}_+(X_i)\neq\delta_+(X_i)\right\}|m(1,X_i)+m(0,X_i)-1|\right]=R_+(\hat{\delta}_+)$$

For each i, since $\mathbb{1}\left\{\hat{\delta}_+(X_i)\neq\delta_+(X_i)\right\}|m(1,X_i)+m(0,X_i)-1|$ is bounded between 0 and 1, it is sub-Gaussian with scale parameter 1. Furthermore, they are independent across $i=1,\ldots,n$, so by the Hoeffding bound,

$$\Pr\left(\frac{1}{n}\sum_{i=1}^n \mathbb{1}\left\{\hat{\delta}_+(X_i)\neq\delta_+(X_i)\right\}|m(1,X_i)+m(0,X_i)-1| \leq R_+(\hat{\delta}_+)+\frac{t}{\sqrt{n}}\right)\geq 1-\exp\left(-2t^2\right).$$

Combining this with the deterministic bound above gives the result.

□

Lemma J.3. For $u_g \geq u_l$,

$$
\begin{aligned}
& \sup_{\pi \in \Pi} \left| \tilde{R}_{\text{sup}}(\pi, \hat{c}, m; \pi_o^*) - \tilde{R}_{\text{sup}}(\pi, c, m, \pi_o^*) \right| \\
\leq \;& 2 \times \left\{ R_{\text{sup}}(\hat{\pi}_{\mathbb{1}}, \pi^{\mathbb{1}}) - R_{\text{sup}}(\pi_{\mathbb{1}}^*, \pi^{\mathbb{1}}) \right\} + 2 \times \left\{ R_{\text{sup}}(\hat{\pi}_{\mathbb{O}}, \pi^{\mathbb{O}}) - R_{\text{sup}}(\pi_{\mathbb{O}}^*, \pi^{\mathbb{O}}) \right\} \\
& + (u_g - u_l) \times \left(R_+(\hat{\delta}_+) + R_\tau(\hat{\delta}_\tau) + \frac{t}{2\sqrt{n}} \right),
\end{aligned}
$$

with probability at least $1 - 2e^{-\frac{t^2}{2}}$.

Proof of Lemma J.3. Define

$$
\begin{aligned}
\check{L}_b(x) &= \{u_l + (u_g - u_l)\hat{\delta}_\tau(x)\} m(1, x) - \{u_l + (u_g - u_l)\hat{\delta}_\tau(x)\} m(0, x) - c, \\
\check{U}_b(x) &= \{u_g - (u_g - u_l)\hat{\delta}_+(x)\} m(1, x) - \{u_l + (u_g - u_l)\hat{\delta}_+(x)\} m(0, x) + (u_g - u_l)\hat{\delta}_+(x) - c, \\
Q(x) &= \pi_{\mathbb{O}}^*(x) U_b(x) + (1 - \pi_{\mathbb{1}}^*(x)) L_b(x) + (1 - \pi_{\mathbb{O}}^*(x)) \pi_{\mathbb{1}}^*(x)(U_b(x) + L_b(x)), \\
\tilde{Q}(x) &= \hat{\pi}_{\mathbb{O}}(x) U_b(x) + (1 - \hat{\pi}_{\mathbb{1}}(x)) L_b(x) + (1 - \hat{\pi}_{\mathbb{O}}(x)) \hat{\pi}_{\mathbb{1}}(x)(U_b(x) + L_b(x)), \\
\check{Q}(x) &= \hat{\pi}_{\mathbb{O}}(x) \check{U}_b(x) + (1 - \hat{\pi}_{\mathbb{1}}(x)) \check{L}_b(x) + (1 - \hat{\pi}_{\mathbb{O}}(x)) \hat{\pi}_{\mathbb{1}}(x)(\check{U}_b(x) + \check{L}_b(x)).
\end{aligned}
$$

With these definitions, we can write

$$
\begin{aligned}
\left| \tilde{R}_{\text{sup}}(\pi, \hat{c}, m; \pi_o^*) - \tilde{R}_{\text{sup}}(\pi, c, m, \pi_o^*) \right| &= \left| \frac{1}{n} \sum_{i=1}^n \pi(X) \{ Q(X_i) - \check{Q}(X_i) \} \right| \\
&= \left| \frac{1}{n} \sum_{i=1}^n \pi(X) \{ Q(X_i) - \tilde{Q}(X_i) \} + \frac{1}{n} \sum_{i=1}^n \pi(X) \{ \tilde{Q}(X_i) - \check{Q}(X_i) \} \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n \left| Q(X_i) - \tilde{Q}(X_i) \right| + \frac{1}{n} \sum_{i=1}^n \left| \tilde{Q}(X_i) - \check{Q}(X_i) \right|.
\end{aligned}
$$

Working with the first term:

$$
\begin{aligned}
Q(x) - \tilde{Q}(x) &= (\hat{\pi}_{\mathbb{1}}(x) - \pi_{\mathbb{1}}^*(x)) U_b(x) - (\hat{\pi}_{\mathbb{1}}(x)\hat{\pi}_{\mathbb{O}}(x) - \pi_{\mathbb{1}}^*(x)\pi_{\mathbb{O}}^*(x))(U_b(x) + L_b(x)) + (\hat{\pi}_{\mathbb{O}}(x) - \pi_{\mathbb{O}}^*(x)) U_b(x) \\
&= (\hat{\pi}_{\mathbb{O}}(x) - \pi_{\mathbb{O}}^*(x)) \times (-L_b(x)\pi_{\mathbb{1}}^*(x) + (1 - \pi_{\mathbb{1}}^*(x)) U_b(x)) \qquad (*) \\
&\quad + (\hat{\pi}_{\mathbb{1}}(x) - \pi_{\mathbb{1}}^*(x)) \times (-L_b(x)\hat{\pi}_{\mathbb{O}}(x) + (1 - \hat{\pi}_{\mathbb{O}}(x)) U_b(x)) \qquad (**)
\end{aligned}
$$

Notice that $\pi_{\mathbb{1}}^*(x) = 0 \Leftrightarrow U_b(x) \leq 0$, since $L_b(x) \leq U_b(x)$, this implies that when $\pi_{\mathbb{1}}^*(x) = 0$, $|U_b(x)| \leq |L_b(x)|$. Therefore,

$$
|(*)| \leq \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \neq \pi_{\mathbb{O}}^*(x)\} |L_b(x)|.
$$

Similarly, if $\pi^*_{\mathbb{O}}(x) = 1$, then $0 \le L_b(x) \le U_b(x)$, so $|L_b(x)| \le |U_b(x)|$. So,

$$\begin{aligned}
|(**)| &\le \mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}\mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) = \hat{\pi}_{\mathbb{O}}(x)\}|-L_b(x)\pi^*_{\mathbb{O}}(x) + (1-\pi^*_{\mathbb{O}}(x))U_b(x)| \\
&\qquad + \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|-\hat{\pi}_{\mathbb{O}}(x)L_b(x) + (1-\hat{\pi}_{\mathbb{O}}(x))U_b(x)| \\
&\le \mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}\mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) = \hat{\pi}_{\mathbb{O}}(x)\}|U_b(x)| \\
&\qquad + \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|L_b(x)| + \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|U_b(x)| \\
&\le \mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|U_b(x)| + \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}|L_b(x)| + \mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|U_b(x)| \\
&\le 2\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|U_b(x)| + \mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}|L_b(x)|.
\end{aligned}$$

Putting together the pieces, we get that

$$|Q(x) - \tilde{Q}(x)| \le 2\mathbb{1}\{\hat{\pi}_{\mathbb{O}}(x) \ne \pi^*_{\mathbb{O}}(x)\}|L_b(x)| + 2\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(x) \ne \pi^*_{\mathbb{1}}(x)\}|U_b(x)|.$$

So the expectation is bounded by two regret terms:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}\left|Q(X_i) - \tilde{Q}(X_i)\right|\right] &\le 2\mathbb{E}\left[\mathbb{1}\{\hat{\pi}_{\mathbb{O}}(X) \ne \pi^*_{\mathbb{O}}(X)\}|L_b(X)|\right] + 2\mathbb{E}\left[\mathbb{1}\{\hat{\pi}_{\mathbb{1}}(X) \ne \pi^*_{\mathbb{1}}(X)\}|U_b(X)|\right] \\
&= 2 \times \{R_{\text{sup}}(\hat{\pi}_{\mathbb{1}}, \pi^{\mathbb{1}}) - R_{\text{sup}}(\pi^*_{\mathbb{1}}, \pi^{\mathbb{1}})\} + 2 \times \{R_{\text{sup}}(\hat{\pi}_{\mathbb{O}}, \pi^{\mathbb{O}}) - R_{\text{sup}}(\pi^*_{\mathbb{O}}, \pi^{\mathbb{O}})\}.
\end{aligned}$$

Next, $\left|Q(X_i) - \tilde{Q}(X_i)\right|$ is bounded between 0 and $u_g - u_l$, so by the Hoeffding bound it concentrates around its expectation:

$$\begin{aligned}
&\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|Q(X_i) - \tilde{Q}(X_i)\right| \le 2\{R_{\text{sup}}(\hat{\pi}_{\mathbb{1}}, \pi^{\mathbb{1}}) - R_{\text{sup}}(\pi^*_{\mathbb{1}}, \pi^{\mathbb{1}})\} + 2\{R_{\text{sup}}(\hat{\pi}_{\mathbb{O}}, \pi^{\mathbb{O}}) - R_{\text{sup}}(\pi^*_{\mathbb{O}}, \pi^{\mathbb{O}})\} + \frac{t}{\sqrt{n}}\right) \\
\ge \quad & 1 - \exp\left(-\frac{2t^2}{(u_g - u_l)^2}\right).
\end{aligned}$$

Now for the second term:

$$\begin{aligned}
|\tilde{Q}(x) - \check{Q}(x)| &= \left|(L_b(x) - \check{L}_b(x))(1 - \hat{\pi}_{\mathbb{1}} + (1-\hat{\pi}_{\mathbb{O}})\hat{\pi}_{\mathbb{1}}) + (U_b(x) - \check{U}_b(x))(\hat{\pi}_{\mathbb{O}} + (1-\hat{\pi}_{\mathbb{O}})\hat{\pi}_{\mathbb{1}})\right| \\
&\le |L_b(x) - \check{L}_b(x)| + |U_b(x) - \check{U}_b(x)|.
\end{aligned}$$

To re-write this, notice that

$$\begin{aligned}
|L_b(x) - \check{L}_b(x)| &= (u_g - u_l)\mathbb{1}\{\hat{\delta}_\tau(x) \ne \delta_\tau(x)\}|m(1,x) - m(0,x)|, \\
|U_b(x) - \check{U}_b(x)| &= (u_g - u_l)\mathbb{1}\{\hat{\delta}_+(x) \ne \delta_+(x)\}|m(1,x) + m(0,x) - 1|.
\end{aligned}$$

So,

$$\frac{|\tilde{Q}(x) - \check{Q}(x)|}{u_g - u_l} \le \mathbb{1}\{\hat{\delta}_\tau(x) \ne \delta_\tau(x)\}|m(1,x) - m(0,x)| + \mathbb{1}\{\hat{\delta}_+(x) \ne \delta_+(x)\}|m(1,x) + m(0,x) - 1|.$$

Taking the expectation, we see that it is bounded by:

$$\frac{1}{u_g - u_l}\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{Q}(X_i) - \check{Q}(X_i)\right| \leq \mathbb{E}\left[\mathbb{1}\{\hat{\delta}_\tau(x) \neq \delta_\tau(X)\}|m(1,X) - m(0,X)|\right]$$
$$+ \mathbb{E}\left[\mathbb{1}\{\hat{\delta}_+(X) \neq \delta_+(x)\}|m(1,X) + m(0,X) - 1|\right]$$
$$= R_+(\hat{\delta}_+) + R_\tau(\hat{\delta}_\tau).$$

Again noting that $|\tilde{Q}(X_i) - \check{Q}(X_i)|$ is bounded between 0 and $u_g - u_l$, and applying the Hoeffding inequality gives

$$\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left|\tilde{Q}(X_i) - \check{Q}(X_i)\right| \leq (u_g - u_l) \times \left(R_+(\hat{\delta}_+) + R_\tau(\hat{\delta}_\tau) + \frac{t}{\sqrt{n}}\right)\right) \geq 1 - \exp\left(-2t^2\right).$$

Combining these two bounds via the union bound gives the result.

□

Lemma J.4. Let $\hat{\delta}_+(x) = \mathbb{1}\{\hat{m}(1,x) + \hat{m}(0,x) - 1 \geq 0\}$ and $\hat{\delta}_\tau(x) = \mathbb{1}\{\hat{m}(1,x) - \hat{m}(0,x)\}$. Under Assumption 2,

$$R_+(\hat{\delta}_+) \leq 2^{1+\alpha}C\|\hat{m} - m\|_\infty^{1+\alpha},$$
$$\Pr(\hat{\delta}_+(X) \neq \delta_+(X)) \leq 2^\alpha C\|\hat{m} - m\|_\infty^\alpha.$$

Under Assumption H.1(b),

$$R_\tau(\hat{\delta}_\tau) \leq 2^{1+\alpha}C\|\hat{m} - m\|_\infty^{1+\alpha},$$
$$\Pr(\hat{\delta}_\tau(X) \neq \delta_\tau(X)) \leq 2^\alpha C\|\hat{m} - m\|_\infty^\alpha.$$

Proof of Lemma J.4. This Lemma directly follows Lemma 5.1 in Audibert and Tsybakov (2007). Note that if $\hat{\delta}_+(x) \neq \delta_+(x)$, then the error is greater than the margin, i.e.,

$$|\hat{m}(1,x) - m(1,x) + \hat{m}(0,x) - m(0,x)| \geq |m(1,X) + m(0,X) - 1|$$

So,

$$\Pr(\hat{\delta}_+(X) \neq \delta_+(X)) \leq \Pr(|\hat{m}(1,X) - m(1,X) + \hat{m}(0,X) - m(0,X)| \geq |m(1,X) + m(0,X) - 1|)$$
$$\leq C(\|\hat{m}(1,\cdot) - m(1,\cdot)\|_\infty + \|\hat{m}(0,\cdot) - m(0,\cdot)\|_\infty)^\alpha.$$

By a similar argument,

$$\begin{aligned}
R_+(\hat{\delta}_+) - R_+(\delta_+) &= \mathbb{E}\left[\mathbb{1}\left\{\hat{\delta}_+(X) \neq \delta_+(X)\right\}|m(1,X)+m(0,X)-1|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{|\hat{m}(1,X)-m(1,X)+\hat{m}(0,X)-m(0,X)| \geq |m(1,X)+m(0,X)-1|\right\}\right. \\
&\qquad \left.\times\ |m(1,X)+m(0,X)-1|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{|\hat{m}(1,X)-m(1,X)+\hat{m}(0,X)-m(0,X)| \geq |m(1,X)+m(0,X)-1|\right\}\right. \\
&\qquad \left.\times\ |m(1,X)-\hat{m}(1,X)+m(0,X)-\hat{m}(0,X)|\right] \\
&\leq (\|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty + \|\hat{m}(0,\cdot)-m(0,\cdot)\|_\infty) \\
&\qquad \times \Pr(|\hat{m}(1,X)-m(1,X)+\hat{m}(0,X)-m(0,X)| \geq |m(1,X)+m(0,X)-1|) \\
&\leq C(\|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty + \|\hat{m}(0,\cdot)-m(0,\cdot)\|_\infty)^{1+\alpha}.
\end{aligned}$$

Similarly, if $\hat{\delta}_\tau(x) \neq \delta_\tau(x)$, then

$$|m(1,x)-\hat{m}(1,x)-m(0,x)+\hat{m}(0,x)| \geq |m(1,x)-m(0,x)|.$$

By the same argument as above,

$$\Pr(\hat{\delta}_\tau(X) \neq \delta_\tau(X)) \leq C(\|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty + \|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty)^\alpha,$$

and

$$\begin{aligned}
R_\tau(\hat{\delta}_\tau) - R_\tau(\delta_\tau) &= \mathbb{E}\left[\mathbb{1}\left\{\hat{\delta}_\tau(X_i) \neq \delta_\tau(X_i)\right\}|m(1,X_i)-m(0,X_i)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{|m(1,X)-\hat{m}(1,X)-m(0,X)+\hat{m}(0,X)| \geq |m(1,X)-m(0,X)|\right\}\right. \\
&\qquad \left.\times\ |m(1,X)-m(0,X)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\left\{|m(1,X)-\hat{m}(1,X)-m(0,X)+\hat{m}(0,X)| \geq |m(1,X)-m(0,X)|\right\}\right. \\
&\qquad \left.\times\ |m(1,X)-\hat{m}(1,X)-m(0,X)+\hat{m}(0,X)|\right] \\
&\leq (\|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty + \|\hat{m}(0,\cdot)-m(0,\cdot)\|_\infty) \\
&\qquad \times \Pr(|m(1,X)-\hat{m}(1,X)-m(0,X)+\hat{m}(0,X)| \geq |m(1,X)-m(0,X)|) \\
&\leq C(\|\hat{m}(1,\cdot)-m(1,\cdot)\|_\infty + \|\hat{m}(0,\cdot)-m(0,\cdot)\|_\infty)^{1+\alpha}.
\end{aligned}$$

□

Lemma J.5. Let $u_g \geq u_l$. Define

$$\begin{aligned}
\widehat{L}_b(x) &= \{u_l + \hat{\delta}_\tau(x)(u_g-u_l)\}\{\hat{m}(1,x)-\hat{m}(0,x)\} - c, \\
\widehat{U}_b(x) &= \{u_g - (u_g-u_l)\hat{\delta}_+(x)\}\hat{m}(1,x) - \{u_l + (u_g-u_l)\hat{\delta}_+(x)\}\hat{m}(0,x) + (u_g-u_l)\hat{\delta}_+(x) - c.
\end{aligned}$$

and let $\hat{\pi}_{\mathbb{0}}^{\text{plug}}(x) = \mathbb{1}\{\widehat{L}_b(x) \geq 0\}$ and $\hat{\pi}_{\mathbb{1}}^{\text{plug}}(x) = \mathbb{1}\{\widehat{U}_b(x) \geq 0\}$ be the plug-in estimates of the minimax optimal policies relative to never or always treating. Under Assumption H.1(d), the excess worst case regret for $\hat{\pi}_{\mathbb{0}}^{\text{plug}}$ relative to $\pi_{\mathbb{0}}^{*}$ is

$$R_{\sup}(\hat{\pi}_{\mathbb{0}}^{\text{plug}}, \pi^{\mathbb{0}}) - R_{\sup}(\pi_{\mathbb{0}}^{*}, \pi^{\mathbb{0}}) \leq u_g^{\alpha} C (2\|m - \hat{m}\|_{\infty})^{1+\alpha} + 2u_g C \|m - \hat{m}\|_{\infty} \Pr\left(\hat{\delta}_{\tau}(X) \neq \delta_{\tau}(X)\right) + (u_g - u_l) R_{\tau}(\hat{\delta}_{\tau}).$$

Under Assumption H.1(c), the excess worst case regret for $\hat{\pi}_{\mathbb{1}}^{\text{plug}}$ relative to $\pi_{\mathbb{1}}^{*}$ is

$$R_{\sup}(\hat{\pi}_{\mathbb{1}}^{\text{plug}}, \pi^{\mathbb{1}}) - R_{\sup}(\pi_{\mathbb{0}}^{*}, \pi^{\mathbb{0}}) \leq u_g^{\alpha} C (2\|m - \hat{m}\|_{\infty})^{1+\alpha} + 2u_g C \|m - \hat{m}\|_{\infty} \Pr\left(\hat{\delta}_{+}(X) \neq \delta_{+}(X)\right) + (u_g - u_l) R_{+}(\hat{\delta}_{+}).$$

Proof of Lemma J.5. First, as in the proof of Lemma J.4, note that $\hat{\pi}_{\mathbb{0}}^{\text{plug}}(x) \neq \pi_{\mathbb{0}}^{*}(x)$ implies that $|L_b(x) - \widehat{L}_b(x)| \geq |L_b(x)|$. Now, if $\hat{\delta}_{\tau}(x) = \delta_{\tau}(x)$, then

$$\begin{aligned}
|L_b(x) - \widehat{L}_b(x)| &= |((1 - \delta_{\tau}(x))u_l + \delta_{\tau}(x)u_g)(m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x))| \\
&\leq u_g |m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)|,
\end{aligned}$$

because $|(1 - \delta_{\tau}(x))u_l + \delta_{\tau}(x)u_g| = |u_l + (u_g - u_l)\delta_{\tau}(X)| \leq \max\{u_g, u_l\} \leq u_g$ in the case where $u_g \geq u_l$. If $\hat{\delta}_{\tau}(x) \neq \delta_{\tau}(x)$ and $\delta_{\tau}(x) = 1$, we have that

$$\begin{aligned}
|L_b(x) - \widehat{L}_b(x)| &= |u_l(m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)) + (u_g - u_l)(m(1,x) - m(0,x))| \\
&\leq u_l |m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)| + (u_g - u_l)|m(1,x) - m(0,x)| \\
&\leq u_g |m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)| + (u_g - u_l)|m(1,x) - m(0,x)|.
\end{aligned}$$

Similarly, if $\hat{\delta}_{\tau}(x) \neq \delta_{\tau}(x)$ and $\delta_{\tau}(x) = 0$,

$$\begin{aligned}
|L_b(x) - \widehat{L}_b(x)| &= |u_g(m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)) - (u_g - u_l)(m(1,x) - m(0,x))| \\
&\leq u_g |m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)| + (u_g - u_l)|m(1,x) - m(0,x)|.
\end{aligned}$$

Putting together the pieces, we get that

$$\begin{aligned}
R_{\sup}(\hat{\pi}_{\mathbb{0}}^{\text{plug}}, \pi^{\mathbb{0}}) - R_{\sup}(\pi_{\mathbb{0}}^{*}, \pi^{\mathbb{0}}) &= \mathbb{E}\left[\mathbb{1}\{\hat{\pi}_{\mathbb{0}}^{\text{plug}} \neq \pi_{\mathbb{0}}^{*}\}|L_b(x)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\{|L_b(X) - \widehat{L}_b(X)| \geq |L_b(X)|\}|L(X)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\{|L_b(X) - \widehat{L}_b(X)| \geq |L_b(X)|\}|L_b(X) - \widehat{L}_b(X)|\right] \\
&= \mathbb{E}\left[\mathbb{1}\{|L_b(X) - \widehat{L}_b(X)| \geq |L_b(X)|\}|L_b(X) - \widehat{L}_b(X)|\mathbb{1}\{\hat{\delta}_{\tau}(X) = \delta_{\tau}(X)\}\right] \quad (*) \\
&\quad + \mathbb{E}\left[\mathbb{1}\{|L_b(X) - \widehat{L}_b(X)| \geq |L_b(X)|\}|L_b(X) - \widehat{L}_b(X)|\mathbb{1}\{\hat{\delta}_{\tau}(X) \neq \delta_{\tau}(X)\}\right]. \quad (**)
\end{aligned}$$

By Hölder's inequality and the margin condition (Assumption H.1(d)), the first term is

$$\begin{aligned}
(*) &\le \mathbb{E}\left[\mathbb{1}\{|((1-\delta_\tau(x))u_l + \delta_\tau(X)u_g)(m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x))| \ge |L(X)|\}\right. \\
&\qquad \left.\times\, |m(1,X) - \hat{m}(1,X) - m(0,X) + \hat{m}(0,X)|\right] \\
&\le \mathbb{E}\left[\mathbb{1}\{u_g|m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)| \ge |L(X)|\}\right. \\
&\qquad \left.\times\, |m(1,X) - \hat{m}(1,X) - m(0,X) + \hat{m}(0,X)|\right] \\
&\le \mathbb{E}\left[\mathbb{1}\{u_g|m(1,x) - \hat{m}(1,x) - m(0,x) + \hat{m}(0,x)| \ge |L(X)|\}\right] \times 2\|m - \hat{m}\|_\infty \\
&\le C u_g^\alpha (2\|m - \hat{m}\|_\infty)^{1+\alpha}.
\end{aligned}$$

Similarly, we can bound the second term as

$$\begin{aligned}
(**) &\le \mathbb{E}\left[|L_b(X) - \widehat{L}_b(X)|\mathbb{1}\{\hat{\delta}_\tau(X) \ne \delta_\tau(X)\}\right] \\
&\le \mathbb{E}\left[u_g|m(1,X) - \hat{m}(1,X) - m(0,X) + \hat{m}(0,X)|\mathbb{1}\{\hat{\delta}_\tau(X) \ne \delta_\tau(X)\}\right] \\
&\qquad + (u_g - u_l)\mathbb{E}\left[|m(1,X) - m(0,X)|\mathbb{1}\{\hat{\delta}_\tau(X) \ne \delta_\tau(X)\}\right] \\
&\le u_g 2C\|m - \hat{m}\|_\infty \Pr\left(\hat{\delta}_\tau(X) \ne \delta_\tau(X)\right) + (u_g - u_l)R_\tau(\hat{\delta}_\tau).
\end{aligned}$$

Combining these two terms gives the first result.

Now, also note that $\hat{\pi}_{\mathbb{1}}^{\text{plug}}(x) \ne \pi_{\mathbb{1}}^*(x)$ implies that $|U_b(x) - \widehat{U}_b(x)| \ge |U_b(x)|$. We again break this error term into cases depending on $\hat{\delta}_+(x)$ and $\delta_+(x)$. First, if $\hat{\delta}_+(x) = \delta_+(x)$, then

$$\begin{aligned}
|U_b(x) - \widehat{U}_b(x)| &= \begin{cases} |u_g(m(1,x) - \hat{m}(1,x)) - u_l(m(0,x) - \hat{m}(0,x))|, & \delta_+(x) = 0 \\ |u_l(m(1,x) - \hat{m}(1,x)) - u_g(m(0,x) - \hat{m}(0,x))|, & \delta_+(x) = 1 \end{cases} \\
&\le u_g|m(1,x) - \hat{m}(1,x)| + u_g|m(0,x) - \hat{m}(0,x)|.
\end{aligned}$$

If $\hat{\delta}_+(x) \ne \delta_+(x)$

$$\begin{aligned}
|U_b(x) - \widehat{U}_b(x)| &= \begin{cases} |u_g(m(1,x) - \hat{m}(1,x)) - u_l(m(0,x) - \hat{m}(0,x)) + (u_g - u_l)(m(1,x) + m(0,x) - 1)|, & \delta_+(x) = 0 \\ |u_l(m(1,x) - \hat{m}(1,x)) - u_g(m(0,x) - \hat{m}(0,x))| - (u_g - u_l)(m(1,x) + m(0,x) - 1), & \delta_+(x) = 1 \end{cases} \\
&\le u_g|m(1,x) - \hat{m}(1,x)| + u_g|m(0,x) - \hat{m}(0,x)| + (u_g - u_l)|m(1,x) + m(0,x) - 1|.
\end{aligned}$$

Mirroring the decomposition above, we have that

$$
\begin{aligned}
R_{\sup}(\hat{\pi}^{\text{plug}}_{\mathbb{1}}, \pi^{\mathbb{1}}) - R_{\sup}(\pi^*_{\mathbb{1}}, \pi^{\mathbb{1}}) &= \mathbb{E}\left[\mathbb{1}\{\hat{\pi}^{\text{plug}}_{\mathbb{1}} \neq \pi^*_{\mathbb{1}}\}|U_b(x)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\{|U_b(X) - \widehat{U}_b(X)| \geq |U_b(X)|\}|U_b(X)|\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\{|U_b(X) - \widehat{U}_b(X)| \geq |U_b(X)|\}|U_b(X) - \widehat{U}_b(X)|\right] \\
&= \mathbb{E}\left[\mathbb{1}\{|U_b(X) - \widehat{U}_b(X)| \geq |U_b(X)|\}|U_b(X) - \widehat{U}_b(X)|\mathbb{1}\{\hat{\delta}_+(X) = \delta_+(X)\}\right] \\
&\quad + \mathbb{E}\left[\mathbb{1}\{|U_b(X) - \widehat{U}_b(X)| \geq |U_b(X)|\}|U_b(X) - \widehat{U}_b(X)|\mathbb{1}\{\hat{\delta}_+(X) \neq \delta_+(X)\}\right] \\
&\leq \mathbb{E}\left[\mathbb{1}\{u_g|m(1,X) - \hat{m}(1,X)| + u_g|m(0,X) - \hat{m}(0,X)| \geq |U_b(X)|\}\right. \\
&\qquad \left. \times (u_g|m(1,X) - \hat{m}(1,X)| + u_g|m(0,X) - \hat{m}(0,X)|)\right] \\
&\quad + \mathbb{E}\left[u_g|m(1,x) - \hat{m}(1,x)|\mathbb{1}\{\hat{\delta}_+(X) \neq \delta_+(X)\}\right] \\
&\quad + \mathbb{E}\left[u_g|m(0,x) - \hat{m}(0,x)|\mathbb{1}\{\hat{\delta}_+(X) \neq \delta_+(X)\}\right] \\
&\quad + \mathbb{E}\left[(u_g - u_l)|m(1,x) + m(0,x) - 1|\mathbb{1}\{\hat{\delta}_+(X) \neq \delta_+(X)\}\right] \\
&\leq u_g^\alpha C(2\|m - \hat{m}\|_\infty)^{1+\alpha} + u_g C 2\|m - \hat{m}\|_\infty P(\hat{\delta}_+(X) \neq \delta_+(X)) \\
&\quad + (u_g - u_l) R_+(\hat{\delta}_+).
\end{aligned}
$$

□

References

Athey, S. and Wager, S. (2021). Policy Learning With Observational Data. *Econometrica*, 89(1):133–161.

Audibert, J. Y. and Tsybakov, A. B. (2007). Fast learning rates for plug-in classifiers. *Annals of Statistics*, 35(2):608–633.

Babii, A., Chen, X., Ghysels, E., and Kumar, R. (2021). Binary Choice with Asymmetric Loss in a Data-Rich Environment: Theory and an Application to Racial Justice.

Fan, Y. and Park, S. S. (2010). Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, 26(3):931–951.

Kennedy, E. H. (2022). Towards optimal doubly robust estimation of heterogeneous causal effects.

Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.