A Additional discussion

A.1 Additional related work

Principal stratification and mediation analysis in causal inference [32] studies an optimal test-and-treat regime under a no-direct-effect assumption, that assigning a diagnostic test has no effect on outcomes except via propensity to treat, and studies semiparametric efficiency using Structural Nested-Mean Models. Though our exclusion restriction is also a no-direct-effect assumption, our optimal treatment regime is in the space of recommendations only as we do not have control over the final decision-maker, and we consider generally nonparametric models.

We briefly go into more detail about formal differences, due to our specific assumptions, that delineate the differences to mediation analysis. Namely, our conditional exclusion restriction implies that $Y_{1T_0} = Y_{T_0}$ and that $Y_{0T_1} = Y_{1T_1}$ (in mediation notation with $T_r = T(r)$ in our notation), so that so-called *net direct effects* are identically zero and the *net indirect effect* is the treatment effect (also called average encouragement effect here).

Human-in-the-loop in consequential domains. There is a great deal of interest in designing algorithms for the "human in the loop" and studying expertise and discretion in human oversight in consequential domains [14]. On the algorithmic side, recent work focuses on frameworks for learning to defer or human-algorithm collaboration. Our focus is *prior* to the design of these procedures for improved human-algorithm collaboration: we primarily hold fixed current human responsiveness to algorithmic recommendations. Therefore, our method can be helpful for optimizing local nudges. Incorporating these algorithmic design ideas would be interesting directions for future work.

Empirical literature on judicial discretion in the pretrial setting. Studying a slightly different substantive question, namely causal effects of pretrial decisions on later outcomes, a line of work uses individual judge decision-makers as a leniency instrumental variable for the treatment effect of (for example, EM) on pretrial outcomes [3, 2, 34]. And, judge IVs rely on quasi-random assignment of individual judges. We focus on the prescriptive question of optimal recommendation rules in view of patterns of judicial discretion, rather than the descriptive question of causal impacts of detention on downstream outcomes.

A number of works have emphasized the role of judicial discretion in pretrial risk assessments in particular [20, 15, 33]. In contrast to these works, we focus on studying decisions about electronic monitoring, which is an intermediate degree of decision lever to prevent FTA that nonetheless imposes costs. [23] study a randomized experiment of provision of the PSA and estimate (the sign of) principal causal effects, including potential group-conditional disparities. They are interested in a causal effect on the principal stratum of those marginal defendants who would not commit a new crime if recommended for detention. [9] study policy learning in the absence of positivity (since the PSA is a deterministic function of covariates) and consider a case study on determining optimal recommendation/detention decisions; however their observed outcomes are downstream of judicial decision-making. Relative to their approach, we handle lack of overlap via an exclusion restriction so that we only require ambiguity on *treatment responsivity models* rather than causal outcome models.

B Additional discussion on method

B.1 Additional discussion on constrained optimization

Feasibility program We can obtain upper/lower bounds on ϵ in order to obtain a feasible region for ϵ by solving the below optimization over maximal/minimal values of the constraint:

$$\overline{\epsilon}, \underline{\epsilon} \in \max_{\pi} / \min_{\pi} \mathbb{E}[T(\pi) \mid A = a] - \mathbb{E}[T(\pi) \mid A = b] \tag{3}$$

$$V_{\epsilon}^* = \max_{\pi} \{\mathbb{E}[c(\pi, T(\pi), Y(\pi))] \colon \mathbb{E}[T(\pi) \mid A = a] - \mathbb{E}[T(\pi) \mid A = b] \leq \epsilon\} \tag{4}$$

B.2 Additional discussion on Algorithm 2 (general algorithm)

B.2.1 Additional fairness constraints and examples in this framework

In this section we discuss additional fairness constraints and how to formulate them in the generic framework. Much of this discussion is quite similar to [1] (including in notation) and is included in this appendix for completeness only. We only additionally provide novel identification results for another fairness measure on causal policies in Appendix B.2.2, concrete discussion of the reduction to weighted classification, and provide concrete descriptions of the causal fairness constraints in the more general framework.

We first discuss how to impose the treatment parity constraint. This is similar to the demographic parity example in [1], with different coefficients, but included for completeness. (Instead, recommendation parity in $\mathbb{E}[\pi \mid A = a]$ is indeed nearly identical to demographic parity.)

Example 1 (Writing treatment parity in the general constrained classification framework.)**.** We write the constraint

$$\mathbb{E}[T(\pi) \mid A = a] - \mathbb{E}[T(\pi) \mid A = b] \tag{5}$$

in this framework as follows:

$$\mathbb{E}[T(\pi) \mid A = a] = \mathbb{E}[\pi_1(X)(p_{1|1}(X, A) - p_{1|0}(X, A)) + p_{1|0}(X, A) \mid A = a]$$

For each $u \in \mathcal{A}$ we enforce that

$$\textstyle\sum_{r \in \{0,1\}} \mathbb{E}\left[\pi_r(X) p_{1|r}(X, A) \mid A = u\right] = \sum_{r \in \{0,1\}} \mathbb{E}\left[\pi_r(X, A) p_{1|r}(X, A)\right]$$

We can write this in the generic notation given previously by letting $\mathcal{J} = \mathcal{A} \cup \{\circ\}$,

$$g_j(O, \pi(X); \eta) = \pi_1(X)(p_{1|1}(X, A) - p_{1|0}(X, A)) + p_{1|0}(X, A), \forall j.$$

We let the conditioning events $\mathcal{E}_a = \{A = a\}, \mathcal{E}_\circ = \{\text{True}\}$, i.e. conditioning on the latter is equivalent to evaluating the marginal expectation. Then we express Equation (5) as a set of equality constraints $h_a(\pi) = h_\circ(\pi)$, leading to pairs of inequality constraints,

$$\begin{Bmatrix} h_u(\pi) - h_\circ(\pi) \leq 0 \\ h_\circ(\pi) - h_u(\pi) \leq 0 \end{Bmatrix}_{u \in \mathcal{A}}$$

The corresponding coefficients of M over this enumeration over groups ($\mathcal{A}$) and epigraphical enforcement of equality ($\{+, -\}$) equation (1), gives $\mathcal{K} = \mathcal{A} \times \{+, -\}$ so that $M_{(a,+),a'} = \mathbf{1}\{a' = a\}, M_{(a,+),\star} = -1$, $M_{(a,-),a'} = -\mathbf{1}\{a' = a\}, M_{(a,-),\star} = 1$, and $\mathbf{d} = \mathbf{0}$. Further we can relax equality to small amounts of constraint relaxation by instead setting $d_k > 0$ for some (or all) k.

Next, we discuss a more complicated fairness measure. We first discuss identification and estimation before we also describe how to incorporate it in the generic framework.

B.2.2 Responder-dependent fairness measures

We consider a responder framework on outcomes (under our conditional exclusion restriction). Because the contribution to the AEE is indeed from the responder strata, this corresponds to additional estimation of the responder stratum.

We enumerate the four possible realizations of potential outcomes (given any fixed recommendation) as $(Y(0(r)), Y(1(r)) \in \{0,1\}^2$. We call units with $(Y(0(r)), Y(1(r))) = (0,1)$ responders, $(Y(0(r)), Y(1(r))) = (1,0)$ anti-responders, and $Y(0(r)) = Y(1(r))$ non-responders. Such a decomposition is general for the binary setting.

Assumption 8 (Binary outcomes, treatment)**.**

$$T, Y \in \{0,1\}$$

Assumption 9 (Monotonicity)**.**

$$Y(T(1)) \geq Y(T(0))$$

Importantly, the conditional exclusion restriction of Assumption 2 implies that responder status is independent of recommendation. Conditional on observables, whether a particular individual is a responder is independent of whether someone decides to treat them when recommended. In this way, we study responder status analogous to its use elsewhere in disparity assessment in algorithmic fairness [23, 28]. Importantly, this assumption implies that the conditioning event (of being a responder) is therefore independent of the policy π, so it can be handled in the same framework. s

We may consider reducing disparities in resource expenditure given responder status.

We may be interested in the probability of receiving treatment assignment given responder status.

Example 2 (Fair treatment expenditure given responder status)**.**

$$\mathbb{E}[T(\pi) \mid Y(1(R)) > Y(0(R)), A = a] - \mathbb{E}[T(\pi) \mid Y(1(R)) > Y(0(R)), A = b] \leq \epsilon$$

We can obtain identification via regression adjustment:

Proposition 7 (Identification of treatment expenditure given responder status)**.** Assume Assumptions 8 and 9.

$$P(T(\pi) = 1 \mid A = a, Y(1(\pi)) > Y(0(\pi))) = \frac{\sum_r \mathbb{E}[\pi_r(X) p_{1|r}(X,A)(\mu_1(X,A) - \mu_0(X,A)) \mid A = a])}{\mathbb{E}[(\mu_1(X,A) - \mu_0(X,A)) \mid A = a]}$$

Therefore this can be expressed in the general framework.

Example 3 (Writing treatment responder-conditional parity in the general constrained classification framework.)**.** For each $u \in \mathcal{A}$ we enforce that

$$\frac{\sum_r \mathbb{E}[\pi_r(X) p_{1|r}(X,A)(\mu_1(X,A) - \mu_0(X,A)) | A = u])}{\mathbb{E}[(\mu_1(X,A) - \mu_0(X,A)) | A = u]} = \frac{\sum_r \mathbb{E}[\pi_r(X) p_{1|r}(X,A)(\mu_1(X,A) - \mu_0(X,A))])}{\mathbb{E}[(\mu_1(X,A) - \mu_0(X,A))]}$$

We can write this in the generic notation given previously by letting $\mathcal{J} = \mathcal{A} \cup \{\circ\}$,

$$g_j(O, \pi(X); \eta) = \frac{\{\pi_1(X)(p_{1|1}(X,A) - p_{1|0}(X,A)) + p_{1|0}(X,A)\}(\mu_1(X,A) - \mu_0(X,A))}{\mathbb{E}[(\mu_1(X,A) - \mu_0(X,A)) \mid A = a]}, \forall j.$$

Let $\mathcal{E}_a^j = \{A = a_j\}, \mathcal{E}_\circ = \{\text{True}\}$, and we express Equation (5) as a set of equality constraints of the above moment $h_a(\pi) = h_\circ(\pi)$, leading to pairs of inequality constraints,

$$\left\{ \begin{matrix} h_u(\pi) - h_\circ(\pi) \leq 0 \\ h_\circ(\pi) - h_u(\pi) \leq 0 \end{matrix} \right\}_{u \in \mathcal{A}}$$

The corresponding coefficients of M proceed analogously as for treatment parity.

B.2.3 Best-response oracles

Best-responding classifier π, given λ: $\text{BEST}_\pi(\lambda)$ The best-response oracle, given a particular λ value, optimizes the Lagrangian given π:

$$\begin{aligned} L(\pi, \lambda) &= \hat{V}(\pi) + \lambda^\top (M\hat{h}(\pi) - \hat{d}) \\ &= \hat{V}(\pi) - \lambda^\top \hat{d} + \sum_{k,j} \frac{M_{k,j}\lambda_k}{p_j} \mathbb{E}_n \left[g_j(O,\pi) 1 \left\{ O \in \mathcal{E}_j \right\} \right]. \end{aligned}$$

Best-responding Lagrange multiplier λ, given π: $\mathrm{BEST}_{\boldsymbol{\lambda}}(Q)$ is the best response of the Λ player. It can be chosen to be either 0 or put all the mass on the most violated constraint. Let $\gamma(Q) := Mh(Q)$ denote the constraint values, then $\mathrm{BEST}_{\boldsymbol{\lambda}}(Q)$ returns $\begin{cases} \mathbf{0} & \text{if } \widehat{\gamma}(Q) \leq \widehat{\mathbf{c}} \\ B\mathbf{e}_{k^*} & \text{otherwise, where } k^* = \arg\max_k \left[\widehat{\gamma}_k(Q) - \widehat{c}_k\right] \end{cases}$

B.2.4 Weighted classification reduction

There is a well-known reduction of optimizing the zero-one loss for policy learning to weighted classification. A cost-sensitive classification problem is

$$\underset{\pi_1}{\arg\min} \sum_{i=1}^{n} \pi_1\left(X_i\right) C_i^1 + \left(1 - \pi_1\left(X_i\right)\right) C_i^0$$

The weighted classification error is $\sum_{i=1}^{n} W_i 1\left\{h\left(X_i\right) \neq Y_i\right\}$ which is an equivalent formulation if $W_i = \left|C_i^0 - C_i^1\right|$ and $Y_i = 1\left\{C_i^0 \geq C_i^1\right\}$.

The reduction to weighted classification is particularly helpful since taking the Lagrangian will introduce datapoint-dependent penalties that can be interpreted as additional weights. We can consider the centered regret $J(\pi) = \mathbb{E}[Y(\pi)] - \frac{1}{2}\mathbb{E}[\mathbb{E}[Y \mid R = 1, X] + \mathbb{E}[Y \mid R = 0, X]]$. Then

$$J(\theta) = J(\mathrm{sgn}(g_\theta(\cdot))) = \mathbb{E}[\mathrm{sgn}(g_\theta(X))\{\psi\}]$$

where ψ can be one of, where $\mu_r^R(X) = \mathbb{E}[Y \mid R = r, X]$,

$$\psi_{DM} = (p_{1|1}(X) - p_{1|0}(X))(\mu_1(X) - \mu_0(X)), \psi_{IPW} = \frac{RY}{e_R(X)}, \psi_{DR} = \psi_{DM} + \psi_{IPW} + \frac{R\mu^R(X)}{e_R(X)}$$

We can apply the standard reduction to cost-sensitive classification since $\psi_i \,\mathrm{sgn}(g_\theta(X_i)) = |\psi_i|\,(1 - 2\mathbb{I}\left[\mathrm{sgn}(g_\theta(X_i)) \neq \mathrm{sgn}(\psi_i)\right])$. Then we can use surrogate losses for the zero-one loss,

$$L(\theta) = \mathbb{E}[|\psi|\,\ell(g_\theta(X), \mathrm{sgn}(\psi))]$$

Although many functional forms for $\ell(\cdot)$ are Fisher-consistent, the logistic (cross-entropy) loss will be particularly relevant: $l(g, s) = 2\log(1 + \exp(g)) - (s + 1)g$.

Example 4 (Treatment parity, continued (weighted classification reduction))**.** The cost-sensitive reduction for a vector of Lagrange multipliers can be deduced by applying the weighted classification reduction to the Lagrangian:

$$L(\beta) = \mathbb{E}\left[|\tilde{\psi}^\lambda| \ell\left(g_\beta(X), \mathrm{sgn}(\tilde{\psi}^\lambda)\right)\right], \qquad \text{where } \tilde{\psi}^\lambda = \psi + \frac{\lambda_A}{p_A}(p_{1|1} - p_{1|0}) - \sum_{a \in \mathcal{A}} \lambda_a.$$

where $p_a := \hat{P}(A = a)$ and $\lambda_a := \lambda_{(a,+)} - \lambda_{(a,-)}$, effectively replacing two non-negative Lagrange multipliers by a single multiplier, which can be either positive or negative.

Example 5 (Responder-conditional treatment parity, continued)**.** The Lagrangian is $L(\beta) = \mathbb{E}\left[|\tilde{\psi}^\lambda| \ell\left(g_\beta(X), \mathrm{sgn}(\tilde{\psi}^\lambda)\right)\right]$ with weights:

$$\tilde{\psi}^\lambda = \psi + \frac{\lambda_A}{p_A} \frac{(p_{1|1} - p_{1|0})(\mu_1 - \mu_0)}{\mathbb{E}_n[(\mu_1(X, A) - \mu_0(X, A)) \mid A = a]} - \sum_{a \in \mathcal{A}} \lambda_a.$$

where $p_a := \hat{P}(A = a)$ and $\lambda_a := \lambda_{(a,+)} - \lambda_{(a,-)}$.

B.3 Proofs

Proof of Proposition 7.

$$
\begin{aligned}
&P(T(\pi) = 1 \mid A = a, Y(1(\pi)) > Y(0(\pi))) \\
&= \frac{P(T(\pi) = 1, Y(1(r)) > Y(0(r)) \mid A = a)}{P(Y(1(\pi)) > Y(0(\pi)) \mid A = a)} && \text{by Bayes' rule} \\
&= \frac{P(T(\pi) = 1, Y(1) > Y(0) \mid A = a)}{P(Y(1) > Y(0) \mid A = a)} && \text{by Assumption 2} \\
&= \frac{\sum_r \mathbb{E}[\mathbb{E}[\pi_r(X)\mathbb{I}\left[T(r) = 1\right]\mathbb{I}\left[Y(1) > Y(0)\right] \mid A = a, X]])}{P(Y(1) > Y(0) \mid A = a)} && \text{by iter. exp} \\
&\frac{\sum_r \mathbb{E}[\pi_r(X)p_{1|r}(X, A)(\mu_1(X, A) - \mu_0(X, A)) \mid A = a])}{\mathbb{E}[(\mu_1(X, A) - \mu_0(X, A)) \mid A = a]} && \text{by Proposition 1}
\end{aligned}
$$

$\square$

C Proofs

C.1 Proofs for generalization under unconstrained policies

Proposition 8 (Policy value generalization)**.** Assume the nuisance models $\eta = [p_{1|0}, p_{1|1}, \mu_1, \mu_0, e_r(X)]^\top, \eta \in H$ are consistent and well-specified with finite VC-dimension V_η over the product function class H. We provide a proof for the general case, including doubly-robust estimators, which applies to the statement of Proposition 8 by taking $\eta = [p_{1|0}, p_{1|1}, \mu_1, \mu_0]$.

Let $\Pi = \{\mathbb{I}\{\mathbb{E}[L(\lambda, X, A; \eta) \mid X] > 0 \colon \lambda \in \mathbb{R}; \eta \in \mathcal{F}\}$.

$$\sup_{\pi\in\Pi,\lambda\in\mathbb{R}} |(\mathbb{E}_n[\pi L(\lambda, X, A)] - \mathbb{E}[\pi L(\lambda, X, A)])| = O_p(n^{-\frac{1}{2}})$$

The generalization bound allows deducing risk bounds on the out-of-sample value:

Corollary 2.

$$\mathbb{E}[L(\hat{\lambda}, X, A)_+] \leq \mathbb{E}[L(\lambda^*, X, A)_+] + O_p(n^{-\frac{1}{2}})$$

Proof of Proposition 8. We study a general Lagrangian, which takes as input pseudo-outcomes $\psi^{t|r}(O;\eta), \psi^{y|t}(O;\eta), \psi^{1|0,\Delta A}$ where each satisfies that

$$\begin{aligned}
\mathbb{E}[\psi^{t|r}(O;\eta) \mid X, A] &= p_{1|1}(X, A) - p_{1|0}(X, A) \\
\mathbb{E}[\psi^{y|t}(O;\eta) \mid X, A] &= \tau(X, A) \\
\mathbb{E}[\psi^{1|0,\Delta A} \mid X] &= p_{1|0}(X, a) - p_{1|0}(X, b)
\end{aligned}$$

We make high-level stability assumptions on pseudooutcomes ψ relative to the nuisance functions η (these are satisfied by standard estimators that we will consider):

Assumption 10. $\psi^{t|r}, \psi^{y|t}, \psi^{1|0,\Delta A}$ respectively are Lipschitz contractions with respect to η and bounded

We study a generalized Lagrangian of an optimization problem that took these pseudooutcome estimates as inputs:

$$L(\lambda, X, A;\eta) = \psi_{t|r}(O;\eta)\left\{\psi_{y|t}(O;\eta) + \frac{\lambda}{p(A)}(\mathbb{I}[A=a] - \mathbb{I}[A=b])\right\} + \lambda(\psi^{1|0,\Delta A}(O;\eta))$$

We will show that

$$\sup_{\pi\in\Pi,\lambda\in\mathbb{R}} |(\mathbb{E}_n[\pi L(\lambda, X, A)] - \mathbb{E}[\pi L(\lambda, X, A)])| = O_p(n^{-\frac{1}{2}})$$

which, by applying the generalization bound twice gives that

$$\mathbb{E}_n[\pi L(\lambda, X, A)] = \mathbb{E}[\pi L(\lambda, X, A)]) + O_p(n^{-\frac{1}{2}})$$

Write Lagrangian as

$$\max_\pi \min_\lambda = \min_\lambda \max_\pi = \min_\lambda \mathbb{E}[L(O, \lambda;\eta)_+]$$

Suppose the Rademacher complexity of η_k is given by $\mathcal{R}(H_k)$, so that [7, Thm. 12] gives that the Rademacher complexity of the product nuisance class H is therefore $\sum_k \mathcal{R}(H_k)$. The main result follows by applying vector-valued extensions of Lipschitz contraction of Rademacher complexity given in [36]. Suppose that $\psi^{t|r}, \psi^{y|t}, \psi^{1|0,\Delta A}$ are Lipschitz with constants $C^L_{t|r}, C^L_{y|t}, C^L_{1|0,\Delta A}$.

We establish VC-properties of

$$\mathcal{F}_{L_1}(O_{1:n}) = \{(g_\eta(O_1), g_\eta(O_i), \ldots g_\eta(O_n)) \colon \eta \in H\}, \text{ where } g_\eta(O) = \psi_{t|r}(O;\eta)\psi_{y|t}(O;\eta)$$

$$\mathcal{F}_{L_2}(O_{1:n}) = \{(h_\eta(O_1), h_\eta(O_i), \ldots h_\eta(O_n)) \colon \eta \in H\}, \text{ where } h_\eta(O) = \psi_{t|r}(O;\eta)\frac{\lambda}{p(A)}(\mathbb{I}[A=a] - \mathbb{I}[A=b])$$

$$\mathcal{F}_{L_3}(O_{1:n}) = \{(m_\eta(O_1), m_\eta(O_i), \ldots m_\eta(O_n)) \colon \eta \in H\}, \text{ where } m_\eta(O) = \lambda(\psi^{1|0,\Delta A}(O;\eta))$$

and the function class for the truncated Lagrangian,

$$\mathcal{F}_{L_+} = \{\{(g_\eta(O_i) + h_\eta(O_i) + m_\eta(O_i))_+\}_{1:n} \colon g \in \mathcal{F}_{L_1}(O_{1:n}), h \in \mathcal{F}_{L_2}(O_{1:n}), m \in \mathcal{F}_{L_3}(O_{1:n}), \eta \in H\}$$

[36, Corollary 4] (and discussion of product function classes) gives the following: Let $\mathcal{X}$ be any set, $(x_1, \ldots, x_n) \in \mathcal{X}^n$, let F be a class of functions $f : \mathcal{X} \to \ell_2$ and let $h_i : \ell_2 \to \mathbb{R}$ have Lipschitz norm L. Then

$$\mathbb{E} \sup_{\eta \in H} \sum_i \epsilon_i \psi_i \left(\eta\left(O_i\right)\right) \leq \sqrt{2} L \mathbb{E} \sup_{\eta \in H} \sum_{i,k} \epsilon_{ik} \eta\left(O_i\right) \leq \sqrt{2} L \sum_k \mathbb{E} \sup_{\eta_k \in H_k} \sum_i \epsilon_i \eta_k\left(O_i\right) \qquad (6)$$

where ϵ_{ik} is an independent doubly indexed Rademacher sequence and $f_k\left(x_i\right)$ is the k-th component of $f\left(x_i\right)$.

Applying Equation (6) to each of the component classes $\mathcal{F}_{L_1}(O_{1:n}), \mathcal{F}_{L_2}(O_{1:n}), \mathcal{F}_{L_3}(O_{1:n})$, and Lipschitz contraction [7, Thm. 12.4] of the positive part function $\mathcal{F}_{L_+}$, we obtain the bound

$$\sup_{\lambda, \eta} |\mathbb{E}_n[L(O, \lambda; \eta)_+] - \mathbb{E}[L(O, \lambda; \eta)_+]| \leq \sqrt{2}(C^L_{t|r} C^L_{y|t} + C^L_{t|r} B_{p_a} B + B C^L_{1|0,\Delta A}) \sum_k \mathcal{R}(H_k)$$

□

Proposition 9 (Threshold solutions)**.** Define

$$L(\lambda, X, A) = (p_{1|1}(X, A) - p_{1|0}(X, A)) \left\{\tau(X, A) + \frac{\lambda}{p(A)}(\mathbb{I}\left[A = a\right] - \mathbb{I}\left[A = b\right])\right\} + \lambda(p_{1|0}(X, a) - p_{1|0}(X, b))$$

$$\lambda^* \in \arg\min_\lambda \; \mathbb{E}[L(\lambda, X, A)_+], \quad \pi^*(x, u) = \mathbb{I}\{L(\lambda^*, X, u) > 0\}$$

If instead $d(x)$ is a function of covariates x only,

$$\lambda^* \in \arg\min_\lambda \; \mathbb{E}[\mathbb{E}[L(\lambda, X, A) \mid X]_+], \quad \pi^*(x) = \mathbb{I}\{\mathbb{E}[L(\lambda^*, X, A) \mid X] > 0\}$$

Proof of Proposition 9. The characterization follows by strong duality in infinite-dimensional linear programming [42]. Strict feasibility can be satisfied by, e.g. solving eq. (3) to set ranges for ϵ. □

C.2 Proofs for robust characterization

Proof of Proposition 5.

$$
\begin{aligned}
V(\pi) &= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\pi_r(X)\mathbb{E}[c_{rt}(Y(t))\mathbb{I}\left[T(r)=t\right] \mid R=r,X]] \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\pi_r(X)\mathbb{E}[c_{rt}(Y(t)) \mid R=r,X]P(T(r)=t \mid R=r,X)] && \text{unconf.} \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\pi_r(X)\mathbb{E}[c_{rt}(Y(t)) \mid X]P(T(r)=t \mid R=r,X)] && \text{Assumption 2 (ER)} \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}\left[\pi_r(X)\mathbb{E}\left[c_{rt}(Y(t))\frac{\mathbb{I}\left[T(r)=t\right]}{p_t(X)} \mid X\right]P(T(r)=t \mid R=r,X)\right] && \text{unconf.} \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}\left[\pi_r(X)\left\{\mathbb{E}\left[c_{rt}(Y(t))\frac{\mathbb{I}\left[T(r)=t\right]}{p_t(X)} + \left(1-\frac{T}{p_t(X)}\right)\mu_t(X) \mid X\right]p_{t|r}(X)\right\}\right] && \text{control variate} \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}\left[\pi_r(X)\left\{\left\{c_{rt}(Y(t))\frac{\mathbb{I}\left[T(r)=t\right]}{p_t(X)} + \left(1-\frac{T}{p_t(X)}\right)\mu_t(X)\right\}p_{t|r}(X)\right\}\right] && \text{(LOTE)}
\end{aligned}
\tag{7}
$$

where $p_t(X) = P(T=t \mid X)$ (marginally over R in the observational data) and (LOTE) is an abbreviation for the law of total expectation. $\square$

Proof of Lemma 1.

$$
\begin{aligned}
\overline{V}_{no}(\pi) &:= \max_{q_{tr}(X)\in\mathcal{U}} \left\{ \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\pi_r(X)\mu_t(X)q_{tr}(X)\mathbb{I}\left[X\in\mathcal{X}^{\text{no}}\right]]\right\} \\
&= \max_{q_{tr}(X)\in\mathcal{U}} \left\{ \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\pi_r(X)\mathbb{E}[Y \mid T=t,X]q_{tr}(X)\mathbb{I}\left[X\in\mathcal{X}^{\text{no}}\right]]\right\}
\end{aligned}
$$

Note the objective function can be reparametrized under a surjection of $q_{t|r}(X)$ to its marginalization, i.e. marginal expectation over a $\{T=t\}$ partition (equivalently $\{T=t, A=a\}$ partition for a fairness-constrained setting).

Define

$$\beta_{t|r}(a) := \mathbb{E}[q_{t|r}(X,A) \mid T=t, A=a], \beta_{t|r} := \mathbb{E}[q_{t|r}(X,A) \mid T=t]$$

Therefore we may reparametrize $\overline{V}_{no}(\pi)$ as an optimization over constant coefficients (bounded by B):

$$
\begin{aligned}
&= \max \left\{ \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\{c_t\beta_{t|r}\}\pi_r(X)\mathbb{E}[Y \mid T=t,X]\mathbb{I}\left[X\in\mathcal{X}^{\text{no}}\right]]\colon \underline{B}\leq c_1\leq\overline{B}, c_0 = 1-c_1\right\} \\
&= \max \left\{ \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[\{c_t\beta_{t|r}\}\mathbb{E}[Y\pi_r(X) \mid T=t]\mathbb{I}\left[X\in\mathcal{X}^{\text{no}}\right]]\colon \underline{B}\leq c_1\leq\overline{B}, c_0 = 1-c_1\right\} && \text{LOTE} \\
&= \sum_{t\in\mathcal{T},r\in\{0,1\}} \mathbb{E}[c^*_{rt}\beta_{t|r}\mathbb{E}[Y\pi_r(X) \mid T=t]\mathbb{I}\left[X\in\mathcal{X}^{\text{no}}\right]]
\end{aligned}
$$

$$
\text{where } c^*_{rt} = \begin{cases} \overline{B}\mathbb{I}\left[t=1\right] + \underline{B}\mathbb{I}\left[t=0\right] & \text{if } \mathbb{E}[Y\pi_r(X) \mid T=t] \geq 0 \\ \overline{B}\mathbb{I}\left[t=0\right] + \underline{B}\mathbb{I}\left[t=1\right] & \text{if } \mathbb{E}[Y\pi_r(X) \mid T=t] < 0 \end{cases}
$$

$\square$

Proof of proposition 6.

$$\max_{\pi} \mathbb{E}[c(\pi, T(\pi), Y(\pi))\mathbb{I}\left[X \notin \mathcal{X}^{\text{no}}\right]] + \mathbb{E}[c(\pi, T(\pi), Y(\pi))\mathbb{I}\left[X \in \mathcal{X}^{\text{no}}\right]] \tag{8}$$

$$\mathbb{E}[T(\pi)\mathbb{I}\left[X \notin \mathcal{X}^{\text{no}}\right] \mid A = a] - \mathbb{E}[T(\pi)\mathbb{I}\left[X \notin \mathcal{X}^{\text{no}}\right] \mid A = b] \tag{9}$$

$$+ \mathbb{E}[T(\pi)\mathbb{I}\left[X \in \mathcal{X}^{\text{no}}\right] \mid A = a] - \mathbb{E}[T(\pi)\mathbb{I}\left[X \in \mathcal{X}^{\text{no}}\right] \mid A = b] \leq \epsilon, \forall q_{r1} \in \mathcal{U} \tag{10}$$

Define

$$g_r(x, u) = (\mu_{r1}(x, u) - \mu_{r0}(x, u))$$

then we can rewrite this further and apply the standard epigraph transformation:

$$\max t$$

$$t - \int_{x \in \mathcal{X}^{\text{no}}} \sum_{u \in \{a,b\}} \sum_{r \in \{0,1\}} \{g_r(x, u)\pi_r(x, u)f(x, u)\}q_{r1}(x, u)\}dx \leq V_{ov}(\pi) + \mathbb{E}[\mu_0], \forall q_{r1} \in \mathcal{U}$$

$$\int_{x \in \mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x, a)q_{r1}(x, a)) - f(x \mid b)(\sum_r \pi_r(x, b)q_{r1}(x, b))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon, \forall q_{r1} \in \mathcal{U}$$

Project the uncertainty set onto the direct product of uncertainty sets:

$$\max t$$

$$t - \int_{x \in \mathcal{X}^{\text{no}}} \sum_{u \in \{a,b\}} \sum_{r \in \{0,1\}} \{g_r(x, u)\pi_r(x, u)f(x, u)\}q_{r1}(x, u)\}dx \leq V_{ov}(\pi) + \mathbb{E}[\mu_0], \forall q_{r1} \in \mathcal{U}_\infty$$

$$\int_{x \in \mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x, a)q_{r1}(x, a)) - f(x \mid b)(\sum_r \pi_r(x, b)q_{r1}(x, b))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon, \forall q_{r1} \in \mathcal{U}_\in$$

Clearly robust feasibility of the resource parity constraint over the interval is obtained by the highest/lowest bounds for groups a, b, respectively:

$$\max t$$

$$t - \int_{x \in \mathcal{X}^{\text{no}}} \sum_{u \in \{a,b\}} \sum_{r \in \{0,1\}} \{g_r(x, u)\pi_r(x, u)f(x, u)\}q_{r1}(x, u)\}dx \leq V_{ov}(\pi) + \mathbb{E}[\mu_0], \forall q_{r1} \in \mathcal{U}_\infty$$

$$\int_{x \in \mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x, a)\overline{B}_r(x, a)) - f(x \mid b)(\sum_r \pi_r(x, b)\underline{B}_r(x, u))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon$$

We define

$$\delta_{r1}(x, u) = \frac{2(q_{r1}(x, u) - \underline{B}_r(x, u))}{\overline{B}_r(x, u) - \underline{B}_r(x, u)} - (\overline{B}_r(x, u) - \underline{B}_r(x, u)),$$

then

$$\{\underline{B}_r(x, u) \leq q_{r1}(x, u) \leq \overline{B}_r(x, u)\} \implies \{\|\delta_{r1}(x, u)\|_\infty \leq 1\}$$

and

$$q_{r1}(x, u) = \underline{B}_r(x, u) + \frac{1}{2}(\overline{B}_r(x, u) - \underline{B}_r(x, u))(\delta_{r1}(x, u) + 1).$$

For brevity we denote $\Delta B = (\overline{B}_r(x, u) - \underline{B}_r(x, u))$, so

$$\max t$$

$$t + \max_{\substack{\|\delta_{r1}(x,u)\|_\infty \leq 1 \\ r \in \{0,1\}, u \in \{a,b\}}} \left\{ - \int_{x \in \mathcal{X}^{\text{no}}} \sum_{u \in \{a,b\}} \sum_{r \in \{0,1\}} \{g_r(x, u)\pi_r(x, u)f(x, u)\}\frac{1}{2}\Delta B(x, u)\delta_{r1}(x, u)dx \right\} - c_1(\pi) \leq V_{ov}(\pi) + \mathbb{E}$$

$$\int_{x \in \mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x, a)\overline{B}_r(x, a)) - f(x \mid b)(\sum_r \pi_r(x, b)\underline{B}_r(x, u))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon,$$

where

$$c_1(\pi) = \int_{x\in\mathcal{X}^{\text{no}}} \sum_{u\in\{a,b\}} \sum_{r\in\{0,1\}} \{g_r(x,u)\pi_r(x,u)f(x,u)\}(\underline{B}_r(x,u) + \frac{1}{2}(\overline{B}_r(x,u) - \underline{B}_r(x,u)))dx$$

This is equivalent to:

$$\max t$$

$$t + \int_{x\in\mathcal{X}^{\text{no}}} \sum_{u\in\{a,b\}} \sum_{r\in\{0,1\}} |-g_r(x,u)\pi_r(x,u)f(x,u)| \frac{1}{2}\Delta B(x,u)dx - c_1(\pi) \leq V_{ov}(\pi) + \mathbb{E}[\mu_0]$$

$$\int_{x\in\mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x,a)\overline{B}_r(x,a)) - f(x \mid b)(\sum_r \pi_r(x,b)\underline{B}_r(x,u))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon$$

Undoing the epigraph transformation, we obtain:

$$\max V_{ov}(\pi) + \mathbb{E}[\mu_0] + c_1(\pi) - \int_{x\in\mathcal{X}^{\text{no}}} \sum_{u\in\{a,b\}} \sum_{r\in\{0,1\}} |-g_r(x,u)\pi_r(x,u)f(x,u)| \frac{1}{2}\Delta B(x,u)dx$$

$$\int_{x\in\mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x,a)\overline{B}_r(x,a)) - f(x \mid b)(\sum_r \pi_r(x,b)\underline{B}_r(x,u))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon$$

and simplifying the absolute value:

$$\max V_{ov}(\pi) + \mathbb{E}[\mu_0] + c_1(\pi) - \int_{x\in\mathcal{X}^{\text{no}}} \sum_{u\in\{a,b\}} \sum_{r\in\{0,1\}} |g_r(x,u)\pi_r(x,u)f(x,u)| \frac{1}{2}\Delta B(x,u)dx$$

$$\int_{x\in\mathcal{X}^{\text{no}}} \{f(x \mid a)(\sum_r \pi_r(x,a)\overline{B}_r(x,a)) - f(x \mid b)(\sum_r \pi_r(x,b)\underline{B}_r(x,u))\} + \mathbb{E}[\Delta_{ov}T(\pi)] \leq \epsilon$$

□

C.3 Proofs for general fairness-constrained policy optimization algorithm and analysis

We begin with some notation that will simplify some statemetns. Define, for observation tuples $O \sim (X, A, R, T, Y)$, the value estimate $v(Q;\eta)$ given some pseudo-outcome $\psi(O;\eta)$ dependent on observation information and nuisance functions η. (We often suppress notation of η for brevity). We let estimators sub/super-scripted by 1 denote estimators from the first dataset.

$$\begin{aligned}
v^{(\cdot)}(Q) &= \mathbb{E}_{\pi\sim Q}[\pi\psi_{(\cdot)} \mid O], \text{ for } (\cdot) \in \{\emptyset, \mathrm{DR}\} \\
V^{(\cdot)}(Q) &= \mathbb{E}[v^{(\cdot)}(Q)] \\
\hat{V}_1^{(\cdot)}(Q) &= \mathbb{E}_{n_1}[v^{(\cdot)}(Q)] \\
g_j(O;Q) &= \mathbb{E}_{\pi\sim Q}[g_j(O;\pi) \mid O, \mathcal{E}_j] \\
h_j(Q) &= \mathbb{E}[g_j(O;Q) \mid \mathcal{E}_j] \\
\hat{h}_j^1(Q) &= \mathbb{E}_{n_1}[g_j(O;Q) \mid \mathcal{E}_j]
\end{aligned}$$

C.3.1 Preliminaries: results from other works used without proof

Theorem 3 (Thm. 3, [1] (saddle point generalization bound (non-localized))). *Let $\rho := \max_h \|M\hat{\mu}(h) - \hat{c}\|_\infty$. Let Assumption 1 hold for $C' \geq 2C + 2 + \sqrt{\ln(4/\delta)/2}$, where $\delta > 0$. Let $Q^\star$ minimize $V(Q)$ subject to $M\mu(Q) \leq c$. Then Algorithm 1 with $\nu \propto n^{-\alpha}, B \propto n^{\alpha}$ and $\eta \propto \rho^{-2}n^{-2\alpha}$ terminates in $O\left(\rho^2 n^{4\alpha}\ln|\mathcal{K}|\right)$ iterations and returns $\hat{Q}$. If $np_j^\star \geq 8\log(2/\delta)$ for all j, then with probability at least $1-(|\mathcal{J}|+1)\delta$ then for all k $\hat{Q}$ satisfies:*

$$\begin{aligned}
V(\hat{Q}) &\leq V\left(Q^\star\right) + \widetilde{O}\left(n^{-\alpha}\right) \\
\gamma_k(\widehat{Q}) &\leq c_k + \frac{1+2\nu}{B} + \sum_{j\in\mathcal{J}} |M_{k,j}|\, \widetilde{O}\left(\left(np_j^\star\right)^{-\alpha}\right)
\end{aligned}$$

The proof of [1, Thm. 3] is modular in invoking Rademacher complexity bounds on the objective function and constraint moments, so that invoking standard Rademacher complexity bounds for off-policy evaluation/learning [5, 45] yields the above statement for $V(\pi)$ (and analogously, randomized policies by [7, Thm. 12.2] giving stability for convex hulls of policy classes).

Lemma 2 (Lemma 4, [17]). *Consider a function class $\mathcal{F}: \mathcal{X} \to \mathbb{R}^d$, with $\sup_{f\in\mathcal{F}} \|f\|_{\infty,2} \leq 1$ and pick any $f^* \in \mathcal{F}$. Assume that $v(\pi)$ is L-Lipschitz in its first argument with respect to the ℓ_2 norm and let:*

$$Z_n(r) = \sup_{Q\in\mathcal{Q}} \{|\mathbb{E}_n[\hat{v}(Q) - \hat{v}(Q^*)] - \mathbb{E}[v(Q) - v(Q^*)]| : \mathbb{E}[(\mathbb{E}_{\pi\sim Q}[v(\pi)] - \mathbb{E}_{\pi\sim Q^*}[v(\pi)])^2]^{\frac{1}{2}} \leq r\}$$

Then for some universal constants c_1, c_2 :

$$\Pr\left[Z_n(r) \geq 16L\sum_{t=1}^{d} \mathcal{R}\left(r, \operatorname{conv}(\Pi_t) - Q_t^*\right) + u\right] \leq c_1 \exp\left\{-\frac{c_2 n u^2}{L^2 r^2 + 2Lu}\right\}$$

Moreover, if δ_n is any solution to the inequalities:

$$\forall t \in \{1,\ldots,d\} : \mathcal{R}\left(\delta; \operatorname{star}\left(\operatorname{conv}(\Pi_t) - Q_t^*\right)\right) \leq \delta^2$$

then for each $r \geq \delta_n$:

$$P\left(Z_n(r) \geq 16Ldr\delta_n + u\right) \leq c_1 \exp\left\{-\frac{c_2 n u^2}{L^2 r^2 + 2Lu}\right\}$$

Lemma 3 (Concentration of conditional moments ([1, 47]). *For any $j \in \mathcal{J}$, with probability at least $1-\delta$, for all Q,*

$$\left|\widehat{h}_j(Q) - h_j(Q)\right| \leq 2\mathcal{R}_{n_j}(\mathcal{H}) + \frac{2}{\sqrt{n_j}} + \sqrt{\frac{\ln(2/\delta)}{2n_j}}$$

If $np_j^\star \geq 8\log(2/\delta)$, then with probability at least $1-\delta$, for all Q,

$$\left|\widehat{h}_j(Q) - h_j(Q)\right| \leq 2\mathcal{R}_{np_j^\star/2}(\mathcal{H}) + 2\sqrt{\frac{2}{np_j^\star}} + \sqrt{\frac{\ln(4/\delta)}{np_j^\star}}$$

Lemma 4 (Orthogonality (analogous to [12] (Lemma 8), others)). *Suppose the nuisance estimates satisfy a mean-squared-error bound*

$$\max_{l}\{\mathbb{E}[(\hat{\eta}_l-\eta_l)^2]\}_{l\in[L]}\coloneqq\chi_n^2$$

Then w.p. $1-\delta$ *over the randomness of the policy sample,*

$$V(Q_0)-V(\hat{Q})\leq O(R_{n,\delta}+\chi_n^2)$$

C.4 Adapted lemmas

In this subsection we collect results similar to those that have appeared previously, but that require substantial additional argumentation in our specific saddle point setting.

Lemma 5 (Feasible vs. oracle nuisances in low-variance regret slices ([12], Lemma 9)). *Consider the setting of Corollary 7. Suppose that the mean squared error of the nuisance estimates is upper bounded w.p.* $1-\delta$ *by* $h_{n,\delta}^2$ *and suppose* $h_{n,\delta}^2\leq\epsilon_n$. *Then:*

$$V_2^0=\sup_{\pi,\pi'\in\Pi_*\left(\epsilon_n+2h_{n,\delta}^2\right)}\operatorname{Var}\left(v_{DR}^0(x;\pi)-v_{DR}^0(x;\pi')\right)$$

Then $V_2\leq V_2^0+O\left(h_{n,\delta}\right)$.

C.5 Proof of Theorem 1

Proof of Theorem 1. We first study the meta-algorithm with "oracle" nuisance functions $\eta=\eta_0$.

Define

$$\Pi_2\left(\epsilon_n\right)=\left\{\pi\in\Pi:\mathbb{E}_{n_1}[v(Q;\eta_0)-v(\hat{Q}_1;\eta_0)]\leq\epsilon_n,\mathbb{E}_{n_1}\left[g_j(O;\pi,\eta_0)-g_j\left(O;\hat{\pi}_1,\eta_0\right)\mid\mathcal{E}_j\right]\leq\epsilon_n,j\in\hat{\mathcal{I}}_1\right\}$$
$$\mathcal{Q}_2\left(\epsilon_n\right)=\left\{Q\in\Delta\left(\Pi_2(\epsilon_n)\right)\right\}$$
$$\mathcal{Q}^*\left(\epsilon_n\right)=\left\{Q\in\Delta(\Pi):\mathbb{E}[(v(Q;\eta_0)-v(Q^*;\eta_0)]\leq\epsilon_n,\mathbb{E}[g_j(O;Q,\eta_0)\mid\mathcal{E}_j]-\mathbb{E}[g_j(O;Q^*,\eta_0)\mid\mathcal{E}_j]\leq\epsilon_n\right\}$$

In the following, we suppress notational dependence on η_0.

Note that $\hat{Q}_1\in\mathcal{Q}_2\left(\epsilon_n\right)$.

Step 1: First we argue that w.p. $1-\delta/6$, $Q^*\in\mathcal{Q}_2$.

Invoking Theorem 3 on the output of the first stage of the algorithm, yields that with probability $1-\frac{\delta}{6}$ over the randomness in $\mathcal{D}_1$, by choice of $\epsilon_n=\bar{O}(n^{-\alpha})$),

$$\begin{aligned}&V(\hat{Q}_1)\leq V(Q^*)+\epsilon_n/2\\&\gamma_k(\hat{Q}_1)\leq d_k+\sum_{j\in\mathcal{J}}|M_{k,j}|\,\widetilde{O}\left((np_j^*)^{-\alpha}\right)\leq d_k+\epsilon_n/2\quad\text{for all }k\end{aligned}$$

Further, by Lemma 2,

$$\begin{aligned}&\sup_{Q\in\mathcal{Q}}|\mathbb{E}_{n_1}[(v(Q)-v(Q^*))]-\mathbb{E}[(v(Q)-v(Q^*))]|\leq\epsilon_n/2\\&\sup_{Q\in\mathcal{Q}}|\mathbb{E}_{n_1}[(g(O;Q)-g(O;Q^*))]-\mathbb{E}[(g(O;Q)-g(O;Q^*))]|\leq\epsilon_n/2\end{aligned}$$

Therefore, with high probability on the good event, $Q^*\in\mathcal{Q}_2$.

Step 2: Again invoking Theorem 3, this time on the output of the second stage of the algorithm with function space Π_2 (hence implicitly $\mathcal{Q}_2$), and conditioning on the "good event" that $Q^*\in\mathcal{Q}_2$, we obtain the bound that with probability $\geq 1-\delta/3$ over the randomness of the second sample $\mathcal{D}_2$,

$$\begin{aligned}&V(\hat{Q}_2)\leq V(Q^*)+\epsilon_n/2\\&\gamma_k(\hat{Q}_2)\leq\gamma_k(Q^*)+\epsilon_n/2\end{aligned}$$

Step 3: empirical small-regret slices relate to population small-regret slices, and variance bounds

We show that if $Q \in \mathcal{Q}_2$, then with high probability $Q \in \mathcal{Q}_2^0$ (defined on small population value- and constraint-regret slices relative to $\hat{Q}_1$ rather than small empirical regret slices)

$$\mathcal{Q}_2^0 = \{Q \in \mathrm{conv}(\Pi) \colon \left|V(Q) - V(\hat{Q}_1)\right| \leq \epsilon_n/2, \mathbb{E}[g_j(O;Q) - g_j(O;\hat{Q}_1)) \mid \mathcal{E}_j] \leq \epsilon_n, \forall j\}$$

so that w.h.p. $\mathcal{Q}_2 \subseteq \mathcal{Q}_2^0$.

Note that for $Q \in \mathcal{Q}$, w.h.p. $1-\delta/6$ over the first sample, we have that

$$\begin{aligned}
&\sup_{Q\in\mathcal{Q}} \left|\mathbb{E}_n[v(Q) - v(\hat{Q}_1)] - \mathbb{E}[v(Q) - v(\hat{Q}_1)]\right| \leq 2 \sup_{Q\in\mathcal{Q}} \left|\mathbb{E}_n[v(Q)] - \mathbb{E}[v(Q)]\right| \leq \epsilon, \\
&\sup_{Q\in\mathcal{Q}} \left|\mathbb{E}_{n_1}[g_j(O;Q) - g_j(O;\hat{Q}_1) \mid \mathcal{E}_j] - \mathbb{E}[g_j(O;Q) - g_j(O;\hat{Q}_1) \mid \mathcal{E}_j]\right| \\
&\qquad \leq 2 \sup_{Q\in\mathcal{Q}} \left|\mathbb{E}_{n_1}[g_j(O;Q) \mid \mathcal{E}_j] - \mathbb{E}[g_j(O;Q) \mid \mathcal{E}_j]\right| \leq \epsilon, \forall j
\end{aligned}$$

The second bound follows from [7, Theorem 12.2] (equivalence of Rademacher complexity over convex hull of the policy class) and linearity of the policy value and constraint estimators in π, and hence Q.

On the other hand since Q_1 achieves low policy regret, the triangle inequality implies that we can contain the true policy by increasing the error radius. That is, for all $Q \in \mathcal{Q}_2$, with high probability $\geq 1-\delta/3$:

$$\begin{aligned}
&|\mathbb{E}[(v(Q) - v(Q^*))]| \leq \left|\mathbb{E}[(v(Q) - v(\hat{Q}_1))]\right| + \left|\mathbb{E}[(v(\hat{Q}_1) - v(Q^*))]\right| \leq \epsilon_n \\
&|\mathbb{E}[g_j(O;Q) - g_j(O;Q^*) \mid \mathcal{E}_j]| \leq \left|\mathbb{E}[g_j(O;Q) - g_j(O;\hat{Q}_1) \mid \mathcal{E}_j]\right| + \left|\mathbb{E}[g_j(O;\hat{Q}_1) - g_j(O;Q^*) \mid \mathcal{E}_j]\right| \leq \epsilon_n
\end{aligned}$$

Define the space of distributions over policies that achieve value and constraint regret in the population of at most ϵ_n :

$$\mathcal{Q}_*(\epsilon_n) = \{Q \in \mathcal{Q} \colon V(Q) - V(Q^*) \leq \epsilon_n, \;\; \mathbb{E}[g_j(O;Q) - g_j(O;Q^*) \mid \mathcal{E}_j] \leq \epsilon_n, \forall j\},$$

so that on that high-probability event,

$$\mathcal{Q}_2^0(\epsilon_n) \subseteq \mathcal{Q}_*(\epsilon_n). \tag{11}$$

Then on that event with probability $\geq 1-\delta/3$,

$$\begin{aligned}
r_2^2 &= \sup_{Q\in\mathcal{Q}_2} \mathbb{E}[(v(Q) - v(Q^*))^2] \leq \sup_{Q\in\mathcal{Q}^*(\epsilon_n)} \mathbb{E}[(v(Q) - v(Q^*))^2] \\
&= \sup_{Q\in\mathcal{Q}^*(\epsilon_n)} \mathrm{Var}(v(Q) - v(Q^*)) + \mathbb{E}[(v(Q) - v(Q^*))]^2 \\
&\leq \sup_{Q\in\mathcal{Q}^*(\epsilon_n)} \mathrm{Var}(v(Q) - v(Q^*)) + \epsilon_n^2
\end{aligned}$$

Therefore:

$$r_2 \leq \sqrt{\sup_{Q\in\mathcal{Q}_*(\epsilon_n)} \mathrm{Var}\left(v(Q) - v(Q^*)\right)} + 2\epsilon_n = \sqrt{V_2} + 2\epsilon_n$$

Combining this with the local Rademacher complexity bound, we obtain that:

$$\mathbb{E}[v(\hat{Q}_2) - v(Q^*)] = O\left(\kappa\left(\sqrt{V_2} + 2\epsilon_n, \mathcal{Q}_*\left(\epsilon_n\right)\right) + \sqrt{\frac{V_2 \log(3/\delta)}{n}}\right)$$

These same arguments apply for the variance of the constraints

$$V_2^j = \sup\left\{\mathrm{Var}\left(g_j(O;Q) - g_j\left(O;Q'\right)\right) : Q, Q' \in \mathcal{Q}_*(\tilde{\epsilon}_n)\right\}$$

$\square$

C.6 Proofs of auxiliary/adapted lemmas

Proof of Lemma 5. The proof is analogous to that of [12, Lemma 9] except for the step of establishing that $\pi_* \in \mathcal{Q}^0_{\epsilon_n + O(\chi^2_{n,\delta})}$: in our case we must establish relationships between saddlepoints under estimated vs. true nuisances. We show an analogous version below.

Define the saddle points to the following problems (with estimated vs. true nuisances):

$$\begin{aligned}
(Q^*_{0,0}, \lambda^*_{0,0}) &\in \arg\min_Q \max_\lambda \mathbb{E}[v_{DR}(Q;\eta_0)] + \lambda^\top(\gamma_{DR}(Q;\eta_0) - d) \coloneqq L(Q,\lambda;\eta_0,\eta_0) \coloneqq L(Q,\lambda), \\
(Q^*_{\eta,0}, \lambda^*_{\eta,0}) &\in \arg\min_Q \max_\lambda \mathbb{E}[v_{DR}(Q;\eta)] + \lambda^\top(\gamma_{DR}(Q;\eta_0) - d), \\
(Q^*, \lambda^*) &\in \arg\min_Q \max_\lambda \mathbb{E}[v_{DR}(Q;\eta)] + \lambda^\top(\gamma_{DR}(Q;\eta) - d).
\end{aligned}$$

We have that:

$$\begin{aligned}
&\mathbb{E}[v_{DR}(Q^*)] \leq L(Q^*,\lambda^*;\eta,\eta) + \nu \\
&\leq L(Q^*,\lambda^*;\eta,\eta_0) + \nu + \chi^2_{n,\delta} \\
&\leq L(Q^*,\lambda^*;\eta,\eta_0) + \nu + \chi^2_{n,\delta} && \text{by Lemma 4} \\
&\leq L(Q^*,\lambda^*_{\eta,0};\eta,\eta_0) + \nu + \chi^2_{n,\delta} && \text{by saddlepoint prop.} \\
&\leq L(Q^*_{\eta,0},\lambda^*_{\eta,0};\eta,\eta_0) + \left| L(Q^*_{\eta,0},\lambda^*_{\eta,0};\eta,\eta_0) - L(Q^*,\lambda^*_{\eta,0};\eta,\eta_0) \right| + \nu + \chi^2_{n,\delta} && \text{triangle ineq.} \\
&\leq L(Q^*_{\eta,0},\lambda^*_{\eta,0};\eta,\eta_0) + \epsilon_n + \nu + \chi^2_{n,\delta} && \text{assuming } \epsilon_n \geq \chi^2_{n,\delta} \\
&\leq \mathbb{E}[v_{DR}(Q^*_{\eta,0};\eta)] + \epsilon_n + 2\nu + \chi^2_{n,\delta} && \text{apx. complementary slackness} \\
&\leq \mathbb{E}[v_{DR}(Q^*_{0,0};\eta)] + \epsilon_n + 2\nu + \chi^2_{n,\delta} && \text{suboptimality}
\end{aligned}$$

Hence

$$\mathbb{E}[v_{DR}(Q^*;\eta)] - \mathbb{E}[v_{DR}(Q^*_{0,0};\eta)] \leq \epsilon_n + 2\nu + \chi^2_{n,\delta}.$$

We generally assume that the saddlepoint suboptimality ν is of lower order than ϵ_n (since it is under our computational control).

Applying Lemma 4 gives;

$$V(Q^*) - V(Q^*_{0,0}) \leq \epsilon_n + 2\nu + 2\chi^2_{n,\delta}.$$

Define policy classes with respect to small-population regret slices (with a nuisance-estimation enlarged radius):

$$\mathcal{Q}^0(\epsilon) = \{Q \in \Delta(\Pi) \colon V(Q^*_0) - V(Q) \leq \epsilon, \gamma(Q^*_0) - \gamma(Q) \leq \epsilon\}$$

Then we have that

$$V_2^{obj} \leq \sup_{Q \in \mathcal{Q}^0(\epsilon_n)} \mathrm{Var}(v_{DR}(O;\pi) - v_{DR}(O;\pi^*)),$$

where we have shown that $\pi^* \in \mathcal{Q}^0(\epsilon + 2\nu + 2\chi^2_{n,\delta})$.

Following the result of the argumentation in [12, Lemma 9] from here on out gives the result. □

D Case Studies

D.1 Oregon Health Insurance Study

The Oregon Health Insurance Study [16] is an important study on the causal effect of expanding public health insurance on healthcare utilization, outcomes, and other outcomes. It is based on a randomized controlled trial made possible by resource limitations, which enabled the use of a randomized lottery to expand Medicaid eligibility for low-income uninsured adults. Outcomes of interest included health care utilization, financial hardship, health, and labor market outcomes and political participation.

Because the Oregon Health Insurance Study expanded access to *enroll* in Medicaid, a social safety net program, the effective treatment policy is in the space of *encouragement* to enroll in insurance (via access to Medicaid) rather than direct enrollment. This encouragement structure is shared by many other interventions in social services that may invest in nudges to individuals to enroll, tailored assistance, outreach, etc., but typically do not automatically enroll or automatically initiate transfers. Indeed this so-called *administrative burden* of requiring eligible individuals to undergo a costly enrollment process, rather than automatically enrolling all eligible individuals, is a common policy design lever in social safety net programs. Therefore we expect many beneficial interventions in this consequential domain to have this encouragement structure.

We preprocess the data by partially running the Stata replication file, obtaining a processed data file as input, and then selecting a subset of covariates. These covariates include household information that affected stratified lottery probabilities, socioeconomic demographics, medical status and other health information.

In the notation of our framework, the setup of the optimal/fair encouragement policy design question is as follows:

- X covariates (baseline household information, socioeconomic demographics, health information)
- A race (non-white/white), or gender (female/male)

  These protected attributes were binarized.
- R encouragement: lottery status of expanded eligibility (i.e. invitation to enroll when individual was previously ineligible to enroll)
- T: whether the individual is enrolled in insurance ever

  Note that for $R = 1$ this can be either Medicaid or private insurance while for $R = 0$ this is still well-defined as this can be private insurance.
- Y: number of doctor visits

  This outcome was used as a measure of healthcare utilization. Overall, the study found statistically significant effects on healthcare utilization. An implicit assumption is that increased healthcare utilization leads to better health outcomes.

We subsetted the data to include complete cases only (i.e. without missing covariates). We learned propensity and treatment propensity models via logistic regression for each group, and used gradient-boosted regression for the outcome model. We first include results for regression adjustment identification.

In Figure 3 we plot descriptive statistics. We include histograms of the treatment responsivity lifts $p_{1|1a}(x, a) - p_{1|0a}(x, a)$. We see some differences in distributions of responsivity by gender and race. We then regress treatment responsivity on the outcome-model estimate of τ. We find substantially more heterogeneity in treatment responsivity by race than by gender: whites are substantially more likely to take up insurance when made eligible, conditional on the same expected treatment effect heterogeneity in increase in healthcare utilization. (This is broadly consistent with health policy discussions regarding mistrust of the healthcare system).

Next we consider imposing treatment parity constraints on an unconstrained optimal policy (defined on these estimates). In Figure 4 we display the objective value, and $\mathbb{E}[T(\pi) \mid A = a]$, for gender and race, respectively, enumerated over values of the constraint. We use costs of 2 for the number of doctors visits and 1 for enrollment in Medicaid (so $\mathbb{E}[T(\pi) \mid A = a]$ is on the scale of probability of

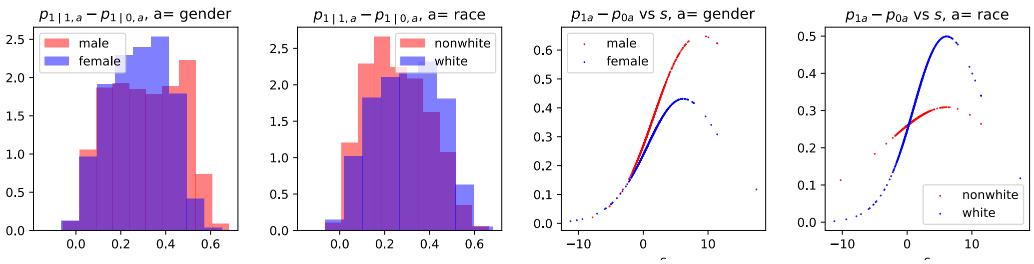

Figure 3: Distribution of lift in treatment probabilities $p_{1|1,a} - p_{1|0,a} = P(T = 1 \mid R = 1, A = a, X) - P(T = 1 \mid R = 0, A = a, X)$, and plot of $p_{1|1,a} - p_{1|0,a}$ vs. τ.

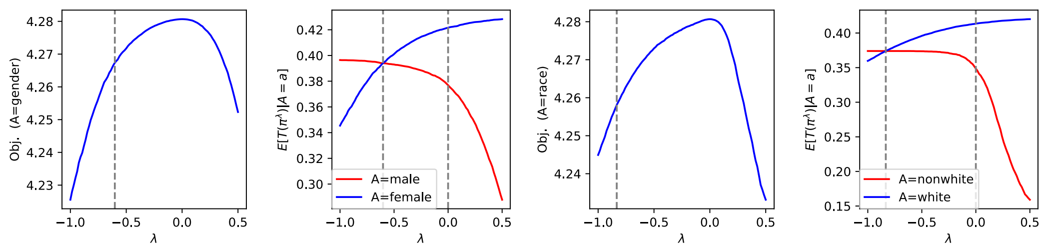

Figure 4: Policy value $V(\pi^\lambda)$, treatment value $\mathbb{E}[T(\pi^\lambda) \mid A = a]$, for A = race, gender.

enrollment). These costs were chosen arbitrarily. Finding optimal policies that improve disparities in group-conditional access can be done with relatively little impact to the overall objective value. These group-conditional access disparities can be reduced from 4 percentage points (0.04) for gender and about 6 percentage points (0.06) for race at a cost of 0.01 or 0.02 in objective value (twice the number of doctors' visits). On the other hand, relative improvements/compromises in access value for the "advantaged group" show different tradeoffs. Plotting the tradeoff curve for race shows that, consistent with the large differences in treatment responsivity we see for whites, improving access for blacks. Looking at this disparity curve given λ however, we can also see that small values of λ can have relatively large improvements in access for blacks before these improvements saturate, and larger λ values lead to smaller increases in access for blacks vs. larger decreases in access for whites.