Fair and Robust Estimation of Heterogeneous Treatment Effects for Policy Learning

Kwangho Kim [1 2] **José R. Zubizarreta** [2 3]

Abstract

We propose a simple and general framework for nonparametric estimation of heterogeneous treatment effects under fairness constraints. Under standard regularity conditions, we show that the resulting estimators possess the double robustness property. We use this framework to characterize the trade-off between fairness and the maximum welfare achievable by the optimal policy. We evaluate the methods in a simulation study and illustrate them in a real-world case study.

1. Introduction

In today's data-centric world, an increasing number of decisions that affect people's lives are automatically made by machine learning models. Such decision-making systems are implemented in various domains, ranging from finance to healthcare. Considering the multifaceted implications of such decisions both at the individual and societal levels, it is crucial to ensure that the underlying models are not only accurate but fair. In this work, by *fairness* we mean that the model estimates are not biased so that they do not systematically benefit or harm a specific group of people. The need to address such algorithmic biases has given rise to a plethora of works on algorithmic fairness (e.g., see Barocas et al. (2019) for a review). However, despite the abundance of studies in this general area, comparatively little attention has been given to fairness in causal inference. In this work, we propose a novel framework for estimating heterogeneous treatment effects in a fair and robust manner, leveraging recent developments in counterfactual optimization (Kim et al., 2022a;b).

[1]Department of Statistics, Korea University, Seoul, South Korea [2]Department of Health Care Policy, Harvard Medical School, MA, USA [3]Departments of Biostatistics and Statistics, Harvard University, MA, USA. Correspondence to: José R. Zubizarreta <zubizarreta@hcp.med.harvard.edu>.

Proceedings of the 40th International Conference on Machine Learning, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

1.1. Related Work

Understanding of treatment-effect heterogeneity and identifying subgroups that respond in a similar way to a given treatment is of great importance across scientific domains. The most common target estimand to study heterogeneity of treatment effects is the conditional average treatment effect (CATE). Various methods have been proposed to obtain accurate estimates of and valid inferences for the CATE, with a special emphasis in recent years on incorporating flexible machine learning tools (e.g., Imai et al., 2013; van der Laan & Luedtke, 2014; Athey & Imbens, 2016; Wager & Athey, 2018; Künzel et al., 2017; Nie & Wager, 2017; Kennedy, 2020). See Jacob (2021) for a review.

Although CATE estimation and inference are of independent importance, these tasks are central also to better target interventions, for example by learning subgroup structures or optimal policies. Much of the earlier attempts on these areas involve postulating a parametric model for the CATE to find the subgroups that would benefit from the given treatment (e.g., Murphy, 2003; Brinkley et al., 2010; Henderson et al., 2010; Orellana et al., 2010). Optimal policies or treatment rules are often designed to identify subgroups with CATEs larger than a certain threshold of interest, so that investigators can select the most promising subgroups with certain efficacy or safety (e.g., Zhao et al., 2013; Schnell et al., 2016; Chen et al., 2017; Ballarini et al., 2018; Wang & Rudin, 2022). More robust approaches to find optimal policies based on the property of doubly robustness have been proposed in subsequent research (e.g., Zhang et al., 2012; 2013). In recent studies (Kallus, 2018; Kitagawa & Tetenov, 2018; Athey & Wager, 2021), flexible nonparametric approaches are discussed where an optimal policy is deployed from a pre-specified class that encodes problem-specific constraints (e.g., a budget or a capacity constraint).

However, this type of data-driven policy-making processes may result in discriminatory treatment of groups defined by sensitive features, such as gender or race. In order to mitigate these algorithmic biases, a wide array of fair estimation criteria have been developed by placing restrictions on the joint distribution of outcome variables and sensitive features (e.g., Barocas et al., 2019; Hardt et al., 2016). In some cases, such as risk assessment, counterfac-

tual fairness may be of interest where the fairness criteria depend on potential (or counterfactual) outcomes (e.g., Kusner et al., 2017; Nabi & Shpitser, 2018; Coston et al., 2020b; Mishler et al., 2021; Mishler & Kennedy, 2021). Recently, constraint-based frameworks have been proposed to flexibly incorporate such fairness criteria in classification (e.g., Zafar et al., 2019; Mishler & Kennedy, 2021). It is also known that fairness-accuracy trade-offs may exist because in some cases the most accurate models under consideration do not satisfy the desired fairness criterion (e.g., Kleinberg et al., 2016; Menon & Williamson, 2018; Obermeyer et al., 2019; Mishler et al., 2021).

To our knowledge, little work has been done at the intersection of algorithmic fairness and the estimation of heterogeneous treatment effects. Some significant studies have integrated aspects of algorithmic fairness and policy learning (Nabi et al., 2019; Viviano & Bradic, 2022). However, it is still unclear how to extend existing methods in algorithmic fairness to obtain an efficient and robust estimator for the CATE under general fairness criteria. Further, to the best of our knowledge, the trade-off between fairness and the maximum welfare achievable by the optimal policy has not been formally explored.

1.2. Contribution

Our method builds on promising literature at the intersection of algorithmic fairness, causal inference, and stochastic optimization, bridging the gap between algorithmic fairness and the analysis of heterogeneous treatment effects. At this intersection, our contribution is twofold. First, we propose a simple and general framework for nonparametric estimation of the CATE under general fairness constraints. We cast our estimator as a convex optimization problem that can be readily solved with off-the-shelf solvers, and show that the resulting estimators are doubly robust under standard regularity conditions. The proposed estimator can attain fast $\sqrt{n}$ rates with tractable inference even when incorporating flexible machine learning tools. Our proposed approach contributes to the existing works in terms of robustness, flexibility, and ease of implementation. Second, we characterize the trade-off between welfare and fairness, by analyzing the regret bounds relative to the optimal policy. To our knowledge, this is the first to quantify the cost of fairness in policy learning, which helps to understand, for example, how a desired level of fairness requires a social welfare compromise.

2. Setup and Framework

2.1. Heterogeneous Treatment Effects and Policy Learning

Consider an i.i.d. sample $(Z_1, ..., Z_n)$ of n tuples $Z = (Y, A, S, X) \sim \mathbb{P}$ for some distribution $\mathbb{P}$, outcome $Y \in \mathcal{Y}$, binary intervention $A \in \{0, 1\}$, sensitive feature $S \in \{0, 1\}$, and additional covariates $X \in \mathcal{X}$ for some compact subset $\mathcal{X}$. Here, we assume that larger values of Y denote better outcomes. We let $W = (S, X) \in \mathcal{W}$ represent the measured pre-intervention variables and let Y^a denote the potential outcome that would have been observed (possibly contrary to fact) under treatment or intervention $A = a$. Throughout, we assume the common causal identification assumptions of *consistency*, *no unmeasured confounding*, and *positivity* (e.g., Imbens & Rubin, 2015, Chapter 12). Under these assumptions, the CATE is defined and identified as

$$\tau(W) = \mathbb{E}(Y^1 - Y^0 \mid W) = \mu_1(W) - \mu_0(W),$$

where $\mu_a(W) = \mathbb{E}[Y \mid W, A = a], \forall a \in \{0, 1\}$. A treatment rule or *policy* g is defined by a mapping from the pre-treatment variables to the treatment: i.e., $g : \mathcal{W} \to \{0, 1\}$. The CATE function τ can be used to produce the optimal policies or to identify subgroups of interest. A policymaker often chooses a policy in such a way that the expected utility or *welfare* defined by

$$\mathcal{U}(g) = \mathbb{E}\left\{Y^1 g(W) + Y^0 (1 - g(W))\right\} \tag{1}$$

is maximized. It is straightforward to show that the optimal policy producing the largest value of $\mathcal{U}(g)$ is given by

$$g^*(W) = \mathbb{1}\left\{\tau(W) > 0\right\}. \tag{2}$$

g^* targets individuals that would have yielded the larger mean outcome had the treatment assigned[1]. More generally, however, one may aim to learn a targeting policy

$$g_{\mathcal{I}}(W) = \mathbb{1}\left\{\tau(W) \in \mathcal{I}\right\}, \tag{3}$$

where the interval $\mathcal{I}$ defines a subgroup of interest.

2.2. Motivating Example

As we illustrate here, ignoring fairness considerations in the CATE estimation problem may introduce serious biases and in turn result in unfair policies. Consider the following data-generating process

$$S \sim Bernoulli(0.5),\ [X_1, X_2]^\top \mid S \sim \mathcal{N}([0, 2S - 1]^\top, I_2),$$
$$\mathbb{P}(A = 1 \mid W) = expit\left(W^\top [1, 0, 0] + SX_1\right),$$
$$\mu_A(W) = AX_2^3/2 + f_\mu(W),$$

for some fixed function $f_\mu : \mathcal{W} \to \mathbb{R}$, where $expit$ and I_2 denote the inverse logit function and the 2×2 identity matrix. Then, $\tau(W) = X_2^3/2$ and $g^*(W) = \mathbb{1}(X_2 > 0)$. When we generate 100 observations from this model, serious fairness issues arise, as can be seen in Figure 1. For example, under g^* only less than 5% of individuals with $S = 0$ are treated,

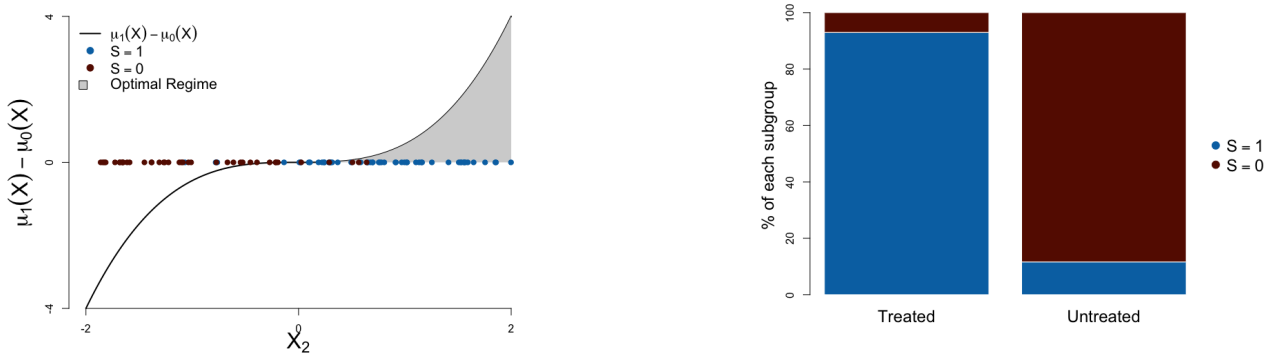

Figure 1. When the optimal policy is applied, only less than 7% of individuals with $S = 0$ are treated while more than 95% of individuals in the untreated group are $S = 0$.

while more than 90% of individuals in the untreated group are $S = 0$.

In this case, one would be inclined to believe that the sensitive feature S is the source of the heterogeneity. Also, this may lead to policies that discriminate against individuals belonging to the group $S = 0$. Hence, accurate estimation of τ alone may induce socially unacceptable consequences which can cause negative impacts on certain individuals. For example, if A represents the policing practice of stop-and-frisk program, it could be used as a recipe for discriminatory practice of stop-and-frisk toward specific ethnic groups (Gelman et al., 2007; White & Fradella, 2016). If A represents access to medical or health care resources, it could be used as an excuse to create a barrier to healthcare for people that are socially disadvantaged. Many recent studies advocate not only cost-effectiveness but also other ethical values for rationing limited health resources (e.g., Emanuel et al., 2020; Obermeyer et al., 2019). Unless properly addressed, the biases in estimated heterogeneity treatment effects can perpetuate and amplify unfair societal processes and outcomes. As a result, this can reduce the public's trust in data-driven systems (Schwartz et al., 2022).

2.3. Proposed Framework

In this section, we lay out a framework for estimating the CATE where we can maximize model fit while securing algorithmic fairness. Specifically, we aim to find a functional approximation of τ, defined by a projection onto a finite-dimensional parametric model subject to fairness constraints. Our target parameter can be reformulated as the solution to the following constrained stochastic optimization problem

$$\begin{aligned} \underset{\beta \in \mathbb{R}^k}{\text{minimize}} \quad & \mathcal{R}_{\text{MSE}}(\beta; \boldsymbol{b}) := \mathbb{E}\left[\left\{Y^1 - Y^0 - \beta^\top \boldsymbol{b}(W)\right\}^2\right] \\ \text{subject to} \quad & \left|\mathbb{E}\left\{\text{uf}_j(Z)\beta^\top \boldsymbol{b}(W)\right\}\right| \leq \delta_j,\ j \in J, \end{aligned} \tag{P}$$

for some $\delta_j \geq 0$ and $J = \{1, ..., m\}$. δ_j is a prespecified tolerance for the maximum acceptable level of unfairness in the j-th criterion. The solution of the above program corresponds to the coefficients of the best-fitting function of τ on the finite-dimensional model space spanned by the basis functions $\boldsymbol{b}(W) = [b_1(W), ..., b_k(W)]^\top$ subject to the m fairness constraints. Note that here we do not assume anything about the true functional relationship between $Y^1 - Y^0$ and W. This form of linear aggregation is widely used in nonparametric regression (e.g., Juditsky & Nemirovski, 2000; Tsybakov, 2003). Zhao (2019) discusses adaptive strategies that allow to consider very rich model family using $\boldsymbol{b}(\cdot)$.

We follow Mishler & Kennedy (2021) and use the *fairness function* $\text{uf}_j : \mathcal{Y}^2 \times \{0, 1\} \to \mathbb{R}$ to accommodate a broad range of (counterfactual) fairness measures. For example, the criterion of *independence*, arguably the most well-known criterion for fairness, can be applied by letting

$$\text{uf}_j(Z) = \frac{1 - S}{\mathbb{E}(1 - S)} - \frac{S}{\mathbb{E}(S)}. \tag{4}$$

This leads to

$$\left|\mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S = 0\right\} - \mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S = 1\right\}\right| \leq \delta_j,$$

thus requiring our fitted models to be marginally (approximately) independent of the sensitive feature with respect to the CATE. When we want to pursue independence for the group that benefits from the treatment, we may consider

$$\text{uf}_j(Z) = \frac{(1 - S)\mathbb{1}\{\tau(W) > 0\}}{\mathbb{E}\left[(1 - S)\mathbb{1}\{\tau(W) > 0\}\right]} - \frac{S\mathbb{1}\{\tau(W) > 0\}}{\mathbb{E}\left[S\mathbb{1}\{\tau(W) > 0\}\right]}, \tag{5}$$

[1]Here, the strict inequality follows from the convention (see, e.g., Zhang et al., 2012).

which leads to

$$\left|\mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S=0, \tau(W)>0\right\} - \mathbb{E}\left\{\beta^\top \boldsymbol{b}(W) \mid S=1, \tau(W)>0\right\}\right| \leq \delta_j.$$

This corresponds to the condition of *balance for the positive class* (Kleinberg et al., 2016), with $\tau(W)$ viewed as a risk score. One may also use other fairness criteria with appropriate fairness functions which may depend on Y^1, Y^0 as well (Mishler & Kennedy, 2021, Section 3).

Similar projection approaches have also been used in causal inference (e.g., Neugebauer & van der Laan, 2007; Semenova & Chernozhukov, 2021; Kennedy et al., 2021). There are several reasons why the above projection approach is preferred in our setting. First, as will be seen shortly, the coefficients β may be estimated with flexible nonparametric methods while achieving the parametric $\sqrt{n}$ rates and tractable inference due to the property of double robustness. It also aids interpretability: it allows practitioners to better grasp the heterogeneity in treatment effects and audit the resulting policies according to the specified level of unfairness. Finally, the optimal solution of (P) can be estimated by solving the optimization problem that approximates (P) using various off-the-shelf algorithms. We will discuss this in more detail in the following section.

Another helpful feature of this framework is that we can consider a general setting where only a subset of covariates $V \subseteq W$ can be used for predicting the counterfactual contrast $Y^1 - Y^0$. This allows for *runtime confounding*, where some factors used by decision-makers are recorded in the training data (used to construct nuisance estimates) but not available for prediction (see Kim et al. (2022a) for details; see also Coston et al. (2020a) and the references therein).

Remark 2.1. Following the convention in the fairness literature, in (P) we only place restrictions on the absolute difference between the first moments (i.e., the conditional means). However, one may consider restricting more comprehensive set of distributional features beyond the means by replacing the constraints with

$$\left|\mathbb{E}\left\{\mathrm{uf}_j(Z)\mathrm{d}\left(\beta^\top \boldsymbol{b}(W)\right)\right\}\right| \leq \delta_j,$$

where the function $\mathrm{d} : \mathbb{R} \to \mathbb{R}^p$ determines the distributional features which we want to constrain in each fairness criterion. For instance, one can match up to the p-th moment by setting $\mathrm{d}(x) = [x, \ldots, x^p]^\top$. This generalizes the conventional measure of fairness and allows to pursue stricter levels of fairness. However, this entails costs of much harder implementation (i.e., non-convex and non-linear optimization) and stronger regularity conditions for our theoretical results. We leave this extension for future work.

Notation. We briefly introduce some additional notation used in the rest of the paper. For any fixed vector v, we let $\|v\|_q$ denote the L_q-norm. Let $\mathbb{P}_n$ denote the empirical measure over $(Z_1, ..., Z_n)$. Given a sample operator h (e.g., an estimated function), we let $\mathbb{P}$ denote the conditional expectation over a new independent observation Z, as in $\mathbb{P}(h) = \mathbb{P}\{h(Z)\} = \int h(z) d\mathbb{P}(z)$ [2]. Then we use $\|h\|_{q,\mathbb{P}}$ to denote the $L_q(\mathbb{P})$ norm of h defined by $\|h\|_{q,\mathbb{P}} = \left[\int \|h(z)\|_q^q d\mathbb{P}(z)\right]^{\frac{1}{q}}$. Lastly, we let $\lesssim$ denote less than or equal to up to a nonnegative constant.

3. Estimation and Inference

(P) is not directly solvable so we need to find an *approximating program* of the "true" program (P). A complication arises since standard approaches to stochastic programming such as *stochastic approximation* (SA) and *sample average approximation* (SAA) (e.g., Nemirovski et al., 2009; Shapiro et al., 2014) are infeasible in our setting, because i) the relevant sample moments and stochastic (sub)gradients depend on unobserved counterfactuals, and ii) these approaches cannot incorporate efficient semiparametric estimators with cross-fitting (e.g., Newey & Robins, 2018). We therefore build our estimators on the recent developments by Kim et al. (2022a;b) where counterfactual components are estimated flexibly without any restrictions on our estimand or estimator.

We first describe simple estimators of each counterfactual component. Based on the identification assumptions, we obtain the following identity

$$\mathbb{E}[Y^a \boldsymbol{b}(W)] = \mathbb{E}\left[\mu_a(W)\boldsymbol{b}(W)\right] = \mathbb{E}\left[\frac{AY}{\pi_a(W)}\boldsymbol{b}(W)\right],$$

where $\pi_a(W) = \mathbb{P}[A = a \mid W]$ is the *propensity score*. Then one may estimate the counterfactual parameter $\mathbb{E}[Y^a \boldsymbol{b}(W)]$ using the *plug-in* (PI) estimator $\mathbb{P}_n\left\{\widehat{\mu}_a(W)\boldsymbol{b}(W)\right\}$ or the *inverse-probability-weighted* (IPW) estimator $\mathbb{P}_n\left\{AY/\widehat{\pi}_a(W)\boldsymbol{b}(W)\right\}$ depending on the quality of information to model the observational outcome or treatment process. Here $\widehat{\mu}_a$ and $\widehat{\pi}_a$ are some estimators of μ_a and π_a, respectively. Although widely used in practice, these estimators cannot attain $\sqrt{n}$ rates in general when nonparametric methods are used (Kennedy, 2016).

We provide more efficient influence-function-based semiparametric estimators for the counterfactual components in (P). Let φ_a denote the uncentered efficient influence function (EIF) for the parameter $\mathbb{E}[Y^a] = \mathbb{E}\left\{\mathbb{E}[Y \mid W, A = a]\right\}$, which is defined by

$$\varphi_a(Z;\eta) = \frac{\mathbb{1}(A=a)}{\pi_a(W)}\left\{Y - \mu_A(W)\right\} + \mu_a(W),$$

with a set of the nuisance components $\eta = \{\pi_a, \mu_a\}$ (Kennedy, 2017; 2022).

[2] When h is a fixed operator, $\mathbb{P}$ and $\mathbb{E}$ are used interchangeably.

Following Robins et al. (2008), Zheng & Van Der Laan (2010), Chernozhukov et al. (2017), Newey & Robins (2018), and Kennedy (2020), we propose to use *sample splitting* (or *cross fitting*) to allow for arbitrarily complex nuisance estimators $\widehat{\eta}$. Specifically, we split the data into K disjoint groups, each with size n/K approximately, by drawing variables $(B_1, ..., B_n)$ independent of the data, with $B_i = b$ indicating that subject i was split into group $b \in \{1, ..., K\}$. Then the semiparametric estimator for $\mathbb{E}[Y^a \boldsymbol{b}(W)]$ based on the EIF and sample splitting are given by

$$\frac{1}{K}\sum_{b=1}^{K} \mathbb{P}_n^b \left\{\varphi_a(Z; \widehat{\eta}_{-b})\boldsymbol{b}(W)\right\} \equiv \mathbb{P}_n \left\{\varphi_a(Z; \widehat{\eta}_{-B})\boldsymbol{b}(W)\right\},$$

where we let $\mathbb{P}_n^b$ denote empirical averages only over the set of units in group b $\{i : B_i = b\}$ and let $\widehat{\eta}_{-b}$ denote the nuisance estimator constructed only using those units $\{i : B_i \neq b\}$. This suggests

$$\mathbb{P}_n \left\{\varphi_1(Z; \widehat{\eta}_{-B}) - \varphi_0(Z; \widehat{\eta}_{-B})\boldsymbol{b}(W)\right\} \tag{6}$$

as our estimator for $\mathbb{E}[(Y^1 - Y^0)\boldsymbol{b}(W)] = \mathbb{E}[\tau(W)\boldsymbol{b}(W)]$. Under weak regularity conditions, this sample-splitting-based semiparametric estimator attains the efficiency bound with the double robustness property, and thus allow us to employ flexible machine learning methods while achieving the $\sqrt{n}$-rate of convergence and valid inference (Kennedy, 2017)[3].

Development of an efficient estimator for $\mathbb{E}\left\{\mathrm{uf}_j(Z)\boldsymbol{b}(W)\right\}$ depends on the form of the fairness function uf_j. Here, we provide a few illustrative cases. First, for the fairness function (4) corresponding to the criterion of independence, one may simply use the following sample-average type estimator:

$$\mathbb{P}_n \left\{\widehat{\mathrm{uf}}_j \boldsymbol{b}(W)\right\} = \mathbb{P}_n \left[\left\{\frac{(1-S)}{\mathbb{P}_n(1-S)} - \frac{S}{\mathbb{P}_n(S)}\right\}\boldsymbol{b}(W)\right],$$

which is naturally $\sqrt{n}$-consistent without any need for nuisance estimation.

Next, for the fairness functions involving a non-smooth component as in (5), we can employ some standard techniques in the nonparametric literature. One of them is the margin condition. In case of (5), we can restrict the probability that the two outcome regression functions get too close to each other, by imposing the following margin condition:

Definition 3.1 (Margin Condition)**.** For any margin exponent $\alpha > 0$ and for all $t \geq 0$,

$$\mathbb{P}(|\mu_1(W) - \mu_0(W)| \leq t) \lesssim t^{\alpha}. \tag{7}$$

[3]If one is willing to rely on appropriate empirical process conditions (e.g., Donsker-type or low entropy conditions (Van der Vaart, 2000)), then η can be estimated on the same sample without sample splitting. However this would limit the flexibility of the nuisance estimators.

The above margin condition is analogous to that used in Kitagawa & Tetenov (2018); Luedtke & van der Laan (2016a;b); Kennedy et al. (2019; 2020) as well as other problems involving estimation of non-smooth functionals such as classification (Audibert & Tsybakov, 2007), couneterfactual density estimation (Kim et al., 2018), and clustering (Levrard, 2018). This margin condition allows us to use the following plug-in type estimator with sample splitting:

$$\mathbb{P}_n \left\{\widehat{\mathrm{uf}}_j \boldsymbol{b}(W)\right\} = \frac{1}{K}\sum_{b=1}^{K} \mathbb{P}_n^b \Bigg(\boldsymbol{b}(W) \times \left[\frac{(1-S)\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W) - \widehat{\mu}_{0,-b}(W) > 0\right\}}{\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left[(1-S)\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W) - \widehat{\mu}_{0,-b}(W) > 0\right\}\right]} - \frac{S\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W) - \widehat{\mu}_{0,-b}(W) > 0\right\}}{\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left[S\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W) - \widehat{\mu}_{0,-b}(W) > 0\right\}\right]}\right]\Bigg).$$

The above estimator attains $\sqrt{n}$-consistency and asymptotic normality if $\mathbb{P}\left[\mathbb{1}\{\widehat{\mu}_1 - \widehat{\mu}_0 > 0\} \neq \mathbb{1}\{\mu_1 - \mu_0 > 0\}\right] = o_{\mathbb{P}}(1)$ and $\max_a \|\widehat{\mu}_a - \mu_a\|_{\infty,\mathbb{P}}^{\alpha} = o_{\mathbb{P}}(n^{-\frac{1}{2}})$ (more details can be found in Appendix C). Alternatively, one may use smoothing approximation techniques such as kernel or polynomial smoothing (e.g., Kim et al., 2018; Kennedy, 2020). We do not discuss each of them here as it is beyond the scope of this paper.

Further, if our fairness function involves the potential outcomes, we can utilize estimators similar to those employed for our objective function. Specifically, if the fairness function $\mathrm{uf}_j(Y^0, Y^1, W)$ is a smooth function of Y^1, Y^0, then one may use the following EIF-based semiparametric estimator

$$\begin{aligned}&\mathbb{P}_n \left\{\widehat{\mathrm{uf}}_j \boldsymbol{b}(W)\right\} \\ &= \mathbb{P}_n \left\{\mathrm{uf}_j\left(\varphi_0(Z; \widehat{\eta}_{-B}), \varphi_1(Z; \widehat{\eta}_{-B}), W\right)\boldsymbol{b}(W)\right\},\end{aligned}$$

which is asymptotically normal and efficient according to the same logic as used for (6).

Consequently, our approximating program can be found as the following convex quadratic program (QP):

$$\begin{aligned}\underset{\beta \in \mathbb{R}^k}{\text{minimize}} \quad & \frac{1}{2}\beta^{\top}\mathbb{P}_n\left\{\boldsymbol{b}(W)\boldsymbol{b}(W)^{\top}\right\}\beta \\ & - \beta^{\top}\mathbb{P}_n\left[\left\{\varphi_1(Z; \widehat{\eta}_{-B}) - \varphi_0(Z; \widehat{\eta}_{-B})\right\}\boldsymbol{b}(W)\right] \\ \text{subject to} \quad & \left|\beta^{\top}\mathbb{P}_n\left\{\widehat{\mathrm{uf}}_j \boldsymbol{b}(W)\right\}\right| \leq \delta_j,\ j \in J.\end{aligned} \tag{$\widehat{\mathsf{P}}$}$$

The above QP can be readily solved using off-the-shelf solvers. Let $\widehat{\beta}$ be an optimal solution to ($\widehat{\mathsf{P}}$). Our proposed estimator for τ is then given by $\widehat{\tau}(W) = \widehat{\beta}^{\top}\boldsymbol{b}(W)$.

Next, we introduce the following assumptions on our basis expansion and nuisance estimation.

(A1) $\mathbb{E}[\boldsymbol{b}(W)\boldsymbol{b}(W)^\top]$ is positive definite

(A2) $\mathbb{P}(\widehat{\pi}_a \in [\epsilon, 1-\epsilon]) = 1$ for some $\epsilon > 0$

(A3) $\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(1)$ or $\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(1)$

(A4) $\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}}\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} = o_{\mathbb{P}}(n^{-\frac{1}{2}})$

Assumption (A1), requiring that our basis functions are never perfectly collinear, ensures that the quadratic growth condition holds at the optimal solution in (P). This could be replaced with a weaker technical condition (see Kim et al., 2022b, or Appendix B). It also guarantees uniqueness of the optimal solution to (P). Assumptions (A2) - (A4) are conditions regarding our nuisance estimation, and commonly used in semiparametric estimation in the causal inference literature (e.g., Kennedy, 2016; 2022). We also need the following assumptions to ensure that the stochastic components in the constraints are consistent at fast rates and attain asymptotic normality.

(A5) $\mathbb{P}_n\left\{\widehat{\mathrm{uf}}_j\boldsymbol{b}(W)\right\} - \mathbb{E}\left\{\mathrm{uf}_j\boldsymbol{b}(W)\right\} = O_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$

(A6) $\sqrt{n}\left[\mathbb{P}_n\left\{\widehat{\mathrm{uf}}_j\boldsymbol{b}(W)\right\} - \mathbb{E}\left\{\mathrm{uf}_j\boldsymbol{b}(W)\right\}\right]$ converges in distribution to a normal random variable with zero mean and finite variance.

In the following theorem, we provide the large-sample properties of our proposed estimator.

Theorem 3.2. *Let β^* and $\widehat{\beta}$ denote the optimal solutions to* (P) *and* ($\widehat{\mathsf{P}}$)*, respectively. If Assumptions (A1) - (A3), and (A5) hold, then*

$$\|\widehat{\beta} - \beta^*\|_2 = O_{\mathbb{P}}\left(\max_a \|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}}\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} + n^{-\frac{1}{2}}\right).$$

If we additionally assume (A4) and (A6), and that the Linear Independence Constraint Qualification (LICQ) and Strict Complementarity (SC) hold at β^*, then $\sqrt{n}(\widehat{\beta} - \beta^*)$ converges in distribution to a normal random variable with zero mean and finite variance.

The above result immediately follows by Theorems 3.1 and 3.2 of Kim et al. (2022b), and gives conditions under which $\widehat{\beta}$ is $\sqrt{n}$-consistent and asymptotically normal. Thus, valid confidence intervals and hypothesis tests can be constructed via the bootstrap. The doubly robust second-order term $\|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}}\|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}}$ will be small if either π_a or μ_a are estimated accurately. In nonparametric modeling, this second-order error substantially lowers the bar for the nuisance estimator convergence rate, which allows much more flexible methods to be employed while still achieving $\sqrt{n}$ rates; for example, it suffices that both nuisance functions are estimated consistently at $n^{\frac{1}{4}}$ rates (Kennedy, 2016). LICQ and SC are regularity conditions commonly found in the optimization literature (e.g., Still, 2018; Shapiro et al., 2014); see Appendix B for the formal definitions. Here we use weaker assumptions than the standard results in stochastic optimization (Shapiro, 1993) due to the linear constraints (Kim et al., 2022b). We also remark that, while we choose the mean squared error $\mathcal{R}_{\mathrm{MSE}}$ as our default risk function, other risk functions, even those with regularization terms, may be used in our framework if they satisfy the necessary regularity conditions (see Kim et al., 2022b, Section 3).

4. Regret Bounds and Fairness-Welfare Tradeoff

In this section, we analyze the regret upper bounds and discuss the implications of using the proposed CATE estimator under fairness criteria for the estimation of the optimal policy g^* defined in (2). One may estimate g^* by

$$\widehat{g}(W) = \mathbb{1}\left\{\widehat{\beta}^\top\boldsymbol{b}(W) > 0\right\}, \tag{8}$$

where $\widehat{\beta}^\top\boldsymbol{b}$ is our estimates of τ obtained by solving ($\widehat{\mathsf{P}}$). Note that since $\widehat{g}$ is a non-smooth function of $\widehat{\beta}^\top\boldsymbol{b}$, one should not in general expect to achieve the same level of fairness for $\widehat{g}$ as for $\widehat{\beta}^\top\boldsymbol{b}$ (see Appendix A for more details). Following the convention in the literature, we evaluate the performance of the above estimated policy in terms of the welfare loss or *regret* relative to the maximum achievable welfare $\mathcal{U}(g^*)$, i.e., $\mathcal{U}(g^*) - \mathcal{U}(\widehat{g})$.

We shall use the margin condition (3.1) for our analysis, which helps us to avoid expensive classification errors for $\widehat{g}$ when τ is close to 0. In the following lemma, we adapt the comparison results in Audibert & Tsybakov (2007) and provide two useful inequalities between the regrets and the L_q risks of the proposed CATE estimator.

Lemma 4.1. *Assume that the margin condition* (7) *holds with margin exponent $0 < \alpha < \infty$. Define the risk score $\Delta \equiv \Delta(W) = \tau(W)$ and $\widehat{\Delta} \equiv \widehat{\Delta}(W) = \widehat{\beta}^\top\boldsymbol{b}(W)$. Then*

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}}^{\alpha+1}.$$

Further, for any $1 \le q < \infty$,

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}}.$$

Using this lemma, the following theorem provides upper bounds of the regret for our proposed estimator $\widehat{g}$ in (8). These results are asymptotic in the sample size n.

Theorem 4.2. *Assume (A1) - (A3), (A5) and that the margin condition* (7) *holds with margin exponent $0 < \alpha < \infty$. Let*

$$\widetilde{\beta} = \underset{\beta\in\mathbb{R}^k}{\arg\min} \quad \mathbb{E}\left\{\left(Y^1 - Y^0 - \beta^\top\boldsymbol{b}(W)\right)^2\right\}, \tag{9}$$

and define the remainder terms

$$R_{1,n} = O_{\mathbb{P}}\left(\max_a \|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}} \|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} + n^{-\frac{1}{2}}\right),$$
$$R_2 = O\Big(\Big\|\sum_j \sqrt{\lambda_j} u f_j(Z)\Big\|_{2,\mathbb{P}} \Big\|\boldsymbol{b}(W)\Big\|_{2,\mathbb{P}}\Big),$$

where $\lambda_j \geq 0$ is the Lagrange multiplier associated with the j-th fairness constraint in (P)*. Then*

$$\text{(i) } \mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \Big\|\tau(W) - \widetilde{\beta}^\top \boldsymbol{b}(W)\Big\|_{\infty,\mathbb{P}}^{1+\alpha} + R_{1,n}^{1+\alpha} + R_2^{1+\alpha},$$

$$\text{(ii) } \mathbb{P}\left\{\widehat{g}(W) \neq g^*(W)\right\} \lesssim \Big\|\tau(W) - \widetilde{\beta}^\top \boldsymbol{b}(W)\Big\|_{\infty,\mathbb{P}}^{\alpha} + R_{1,n}^{\alpha} + R_2^{\alpha},$$

$$\text{(iii) } \mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \Big\|\tau(W) - \widetilde{\beta}^\top \boldsymbol{b}(W)\Big\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}} + R_{1,n}^{\frac{q(1+\alpha)}{q+\alpha}} + R_2^{\frac{q(1+\alpha)}{q+\alpha}}, \forall 1 \leq q < \infty.$$

In (ii), $\Pr\left\{\widehat{g}(W) \neq g^*(W)\right\}$ is the probability that $\widehat{g}$ differs from the true optimal policy g^* over a new observation. Theorem 4.2 shows that the regret bounds depend on the levels of both the nuisance estimation accuracy and fairness we wish to achieve in the CATE estimation problem.

Specifically, each bound listed in Theorem 4.2 consists of three terms. The first term is an unavoidable modeling error minimized through least square estimation, which will vanish if $\mu_a(\cdot)$ lies in the function space spanned by the basis functions $\boldsymbol{b}(\cdot)$.

The second term, $R_{1,n}$, is essentially the doubly robust second-order term that appears in Theorem 3.2. $R_{1,n}$ converges to zero at $\sqrt{n}$ rates even when π_a and μ_a are flexibly estimated at slower than $\sqrt{n}$ rates.

The third term, R_2, has important implications. It indicates the cumulative unfairness in the CATE with respect to the sensitive features, measured by a series of the fairness functions. If we use small values of the tolerance level δ_j so that the optimum β^* is constrained by the j-th fairness constraint (i.e., the j-th constraint is active), then the corresponding Lagrange multiplier, λ_j, is positive. On the contrary, if we loosen the standard by using large values of δ_j so that the j-th fairness constraint does not constrain β^*, λ_j is set to zero. Therefore, our attempts toward making optimal policies more fair may lead to an additional welfare loss (regret) relative to the universally maximum possible welfare $\mathcal{U}(g^*)$. In other words, there is a trade-off between fairness in the optimal policy and the maximum achievable welfare.

In short, Theorem 4.2 states that, while the proposed approach can significantly reduce the cost of estimating the nuisance components, there is still a cost associated with the fairness constraints in order to estimate the optimal policy at the required level of fairness.

Finally, we remark that a comparable analysis for $g_{\mathcal{I}}$ in (3) can also be performed with a slightly different margin condition, where the probability is restricted on the edge of the interval $\mathcal{I}$.

5. Experiments

5.1. Simulation study

In this section, we conduct a simulation study to illustrate the theoretical properties and finite-sample performance of the proposed estimators. We generate synthetic data according to the process described in Section 2.2 where we set

$$f_\mu(W) = \log(SX_1^2 + 10) + \exp(-SX_2/5) + SX_1,$$
$$Y^a = \mu_a(W) + \epsilon,\ \epsilon \sim N(0,1), \quad Y = Y^1 A + Y^0(1-A).$$

In the following experiments, we consider the independence criterion with the fairness function defined in (4). $\boldsymbol{b}(W)$ consists of the polynomial terms W, W^2, W^3 and $\{W_j W_k W_s\}_{j,k,s}$ to form the third-order Taylor expansion. All nuisance functions are estimated using the cross-validation super learner ensemble implemented in the `SuperLearner` R package to combine generalized additive models, adaptive regression splines, and random forests.

We first assess the ability of the proposed estimators to mitigate unfairness in the estimated CATE. To this end, we generate a sample of size 2000 observations and estimate $\widehat{\tau}$ using ($\widehat{\text{P}}$) with $K = 2$ splits, under $\delta = \infty$ (i.e., with no fairness constraints) and $\delta = 0$. We then compare the densities of $\widehat{\tau}$ for each value of δ, and show how individuals belonging to distinct groups are distributed in terms of S in Figure (2). In Figure 3, we compute the proportions of $S = 1$ to $S = 0$ for each decision of $\widehat{g}(W) = \mathbb{1}\left\{\widehat{\tau}(W) > 0\right\}$ under $\delta = \infty$ and $\delta = 0$.

In Figure 2, without the fairness constraints, we observe a pronounced violation of the independence criterion in $\widehat{\tau}$; individuals belonging to the group $S = 1$ ($S = 0$) are mostly distributed in the regime $\widehat{\tau} > 0$ ($\widehat{\tau} < 0$). This eventually leads to disproportionate (unfair) policies (top of Figure 3) where individuals with $S = 1$ are to be treated. However, when the fairness constraint is applied with $\delta = 0$, this issue is largely resolved. In Figure 2, we observe that the conditional sample means of $\widehat{\tau}$ given $S = 0$ and $S = 1$ are nearly identical, and many individuals belonging to the $S = 1$ ($S = 0$) group are shifted to the left (right) compared to the case $\delta = \infty$. This produces policies where the treated and untreated groups are more balanced along the sensitive

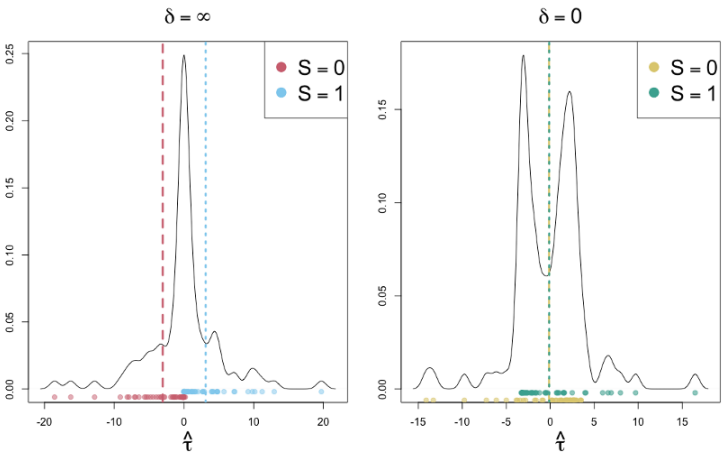

Figure 2. Densities of $\widehat{\tau}(W)$ for $\delta = \infty$ and $\delta = 0$. In each figure, two dashed vertical lines correspond to $\mathbb{P}_n\{\widehat{\tau} \mid S = 0\}$ and $\mathbb{P}_n\{\widehat{\tau} \mid S = 1\}$.

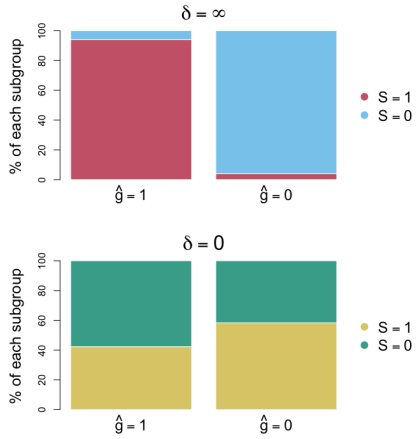

Figure 3. Proportions of $S = 1$ to $S = 0$ for each decision of $\widehat{g}$.

feature, giving individuals with $S = 0$ more chances to be treated (bottom of Figure 3).

To illustrate the theoretical findings in Section 4, which state that the fairness in optimal policies comes at a price, we estimate g^* across different values of δ ranging between 0 and 4. In our simulation settings, $\delta = 4$ is equivalent to having no constraints ($\delta = \infty$) as the fairness constraint becomes inactive. This time, we also estimate the welfare $\mathcal{U}(g^*)$ and the unfairness of our policy $\mathbb{E}\{\text{uf}(Z)g^*(W)\}$ on a separate independent sample of the same size. We compare our methods to the doubly robust score approach (WA21) of Wager & Athey (2018), and the m-hybrid (KT18-m) and e-hybrid (KT18-e) rules by Kitagawa & Tetenov (2018). We do not incorporate our fairness constraint for these approaches since it is not addressed in the original works. The results across 500 simulations are presented in Figure 4.

The results illustrate the predictions of our theory. In Figure 4, the estimated welfare gradually decreases as we apply stricter tolerance levels (δ). The figure also shows that one may reduce unfairness without any noticeable decrease in welfare up to a certain point. To get the result of the bottom of Figure (3), we have to sacrifice the overall welfare by 0.13. This illustrates the trade-off between fairness and social welfare. Further in the simulation, we see that the cost of reducing the same amount of unfairness grows rapidly as we continue to decrease δ. Without the fairness constraints, the proposed method achieves a comparable performance to the doubly robust approach of Athey & Wager (2021).

Finally, to illustrate the double robustness property in Theorems 3.2 and 4.2, we consider three estimators: the proposed estimator and the previous two where the counterfactual component is estimated by the PI and IPW estimators. In

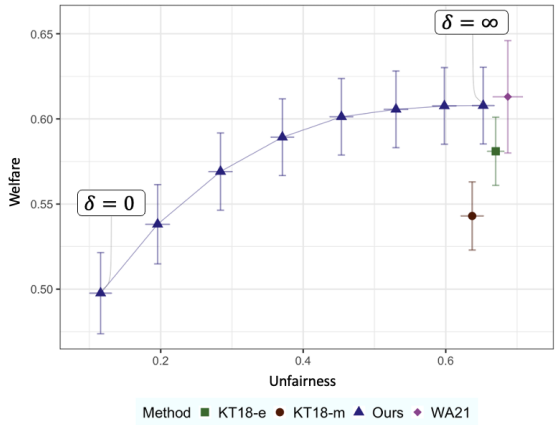

Figure 4. The curve visualizes how much welfare is required to sacrifice in order to achieve a desired level of fairness.

each simulated dataset, again $\widehat{g}$ is estimated on samples of size 2000 and 10000 across the same values of δ as above, and the regret and the unfairness are estimated on new, separate samples. The results across 500 simulations are presented in Figure 5. Again we observe the trade-off between fairness and regret. However, the regret of the proposed method decays faster than the other methods as the sample size grows, due to the double robustness property (the second-order product of nuisance errors).

5.2. Case study

We next illustrate our methods on the COMPAS dataset originally gathered to assess the risk of recidivism (Angwin et al., 2016). Following Mishler et al. (2021), we let A represent pretrial release, with $A = 1$ if defendants are released and $A = 0$ if they are incarcerated. Our sensitive

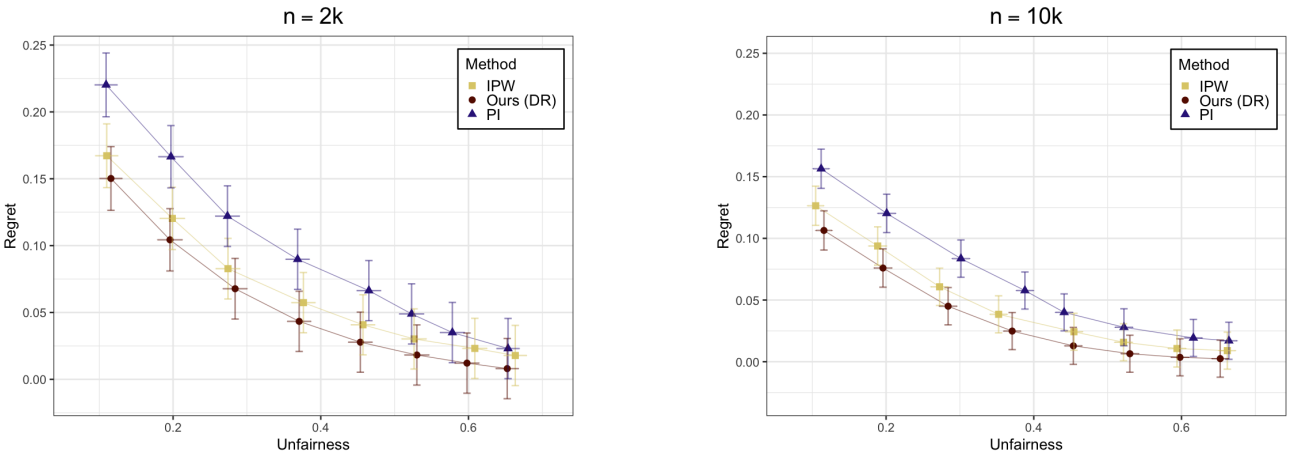

Figure 5. Regret and unfairness curves of the three estimators for $n = 2k$ (left) and $n = 10k$ (right).

Method	Recidivism Risk	IDP	PB
KT18-e	0.33	0.28	0.15
KT18-m	0.28	0.31	0.16
WA21	0.25	0.33	0.17
Ours ($\delta = \infty$)	0.27	0.32	0.16
Ours ($\delta = 0$)	0.34	**0.09**	**0.07**

Table 1. The estimated risk of recidivism and unfairness.

feature represents race, restricted to defendants who are Caucasian ($n = 2013$) and African-American ($n = 3175$), coded as $S = 0$ and $S = 1$, respectively. The outcome of interest, Y, is rearrest within two years. We use three covariates: age, sex, and number of prior arrests. We are interested in estimating the policy that minimizes the risk of recidivism, so in this case our target optimal policy is $g^*(W) = \mathbb{1}\{\mu_1(W) < \mu_0(W)\}$ where the corresponding welfare defines the risk of recidivism. We use roughly two-thirds of the data to estimate $\widehat{g}$, and the rest to estimate the welfare and unfairness using the same setting as in the preceding subsection. Here, we consider two fairness criteria simultaneously: independence (4) (IDP) and positive balance (5) (PB).

In Table 1, we compare our estimators with $\delta = 0$ and $\delta = \infty$ to the three other approaches used in Figure 4. As expected, without the fairness constraint, while producing similar risks, all the methods show large disparities; in fact, the estimated optimal policies suggest incarcerating more African-American defendants. However, when the proposed estimator is used with $\delta = 0$, both disparities are jointly reduced by a significant margin with a small impact on the risk. This illustrates the effectiveness of the proposed methodology.

6. Discussion

We propose a new framework for fair and robust estimation of heterogeneous treatment effects and discuss its implications for policy learning. Our method is easily implementable and allows practitioners to flexibly incorporate various fairness constraints to meet the desired level of fairness. This affords new opportunities to leverage recent advances in algorithmic fairness for robust estimation of heterogeneous treatment effects. The proposed CATE estimator enables us to design targeting policies in a fairer fashion. Importantly, our theoretical results in Section 4 help us to understand the trade-off between fairness and welfare.

There are some caveats to mention, and ways in which our work could be improved. Our method is intended to attain a desired level of fairness in CATE estimation, not in targeting policies. Different methods are required to estimate a particular targeting policy that satisfies a specific fairness criterion. In a forthcoming paper, we develop a novel nonparametric estimator for this kind of fair optimal policies. Further, although the setup we consider here is widely used, in future work it is desirable to consider extensions to time-varying treatments, instrumental variables, and mediation. Finally, it is also of interest to apply the proposed methods to various real-world problems, to bring new insights into fair heterogeneous treatment effect estimation and policy learning.

7. Acknowledgments

This work was completed while Kwangho Kim was a research associate at Harvard Medical School. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea governement (MSIT)(No. NRF-2022M3J6A1063595).

References

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. Machine bias. In *Ethics of Data and Analytics*, pp. 254–264. Auerbach Publications, 2016.

Athey, S. and Imbens, G. Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences*, 113(27):7353–7360, 2016.

Athey, S. and Wager, S. Policy learning with observational data. *Econometrica*, 89(1):133–161, 2021.

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

Ballarini, N. M., Rosenkranz, G. K., Jaki, T., König, F., and Posch, M. Subgroup identification in clinical trials via the predicted individual treatment effect. *PloS one*, 13 (10):e0205971, 2018.

Barocas, S., Hardt, M., and Narayanan, A. *Fairness and Machine Learning*. fairmlbook.org, 2019. `http://www.fairmlbook.org`.

Brinkley, J., Tsiatis, A., and Anstrom, K. J. A generalized estimator of the attributable benefit of an optimal treatment regime. *Biometrics*, 66(2):512–522, 2010.

Chen, S., Tian, L., Cai, T., and Yu, M. A general statistical framework for subgroup identification and comparative treatment scoring. *Biometrics*, 73(4):1199–1209, 2017.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., and Newey, W. Double/debiased/neyman machine learning of treatment effects. *American Economic Review*, 107(5):261–65, May 2017.

Coston, A., Kennedy, E., and Chouldechova, A. Counterfactual predictions under runtime confounding. In *Advances in neural information processing systems*, volume 33, pp. 4150–4162, 2020a.

Coston, A., Mishler, A., Kennedy, E. H., and Chouldechova, A. Counterfactual risk assessments, evaluation, and fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 582–593, 2020b.

Emanuel, E. J., Persad, G., Upshur, R., Thome, B., Parker, M., Glickman, A., Zhang, C., Boyle, C., Smith, M., and Phillips, J. P. Fair allocation of scarce medical resources in the time of covid-19, 2020.

Gelman, A., Fagan, J., and Kiss, A. An analysis of the new york city police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American statistical association*, 102(479):813–823, 2007.

Hardt, M., Price, E., and Srebro, N. Equality of opportunity in supervised learning. *Advances in neural information processing systems*, 29, 2016.

Henderson, R., Ansell, P., and Alshibani, D. Regret-regression for optimal dynamic treatment regimes. *Biometrics*, 66(4):1192–1201, 2010.

Imai, K., Ratkovic, M., et al. Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470, 2013.

Imbens, G. W. and Rubin, D. B. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

Jacob, D. Cate meets ml: Conditional average treatment effect and machine learning. *Digital Finance*, 3(2):99–148, 2021.

Juditsky, A. and Nemirovski, A. Functional aggregation for nonparametric regression. *The Annals of Statistics*, 28(3): 681–712, 2000.

Kallus, N. Balanced policy evaluation and learning. *Advances in neural information processing systems*, 31, 2018.

Kennedy, E. H. Semiparametric theory and empirical processes in causal inference. In *Statistical causal inferences and their applications in public health research*, pp. 141–167. Springer, 2016.

Kennedy, E. H. Semiparametric theory. *arXiv preprint arXiv:1709.06418*, 2017.

Kennedy, E. H. Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*, 2020.

Kennedy, E. H. Semiparametric doubly robust targeted double machine learning: a review. *arXiv preprint arXiv:2203.06469*, 2022.

Kennedy, E. H., Harris, S., and Keele, L. J. Survivor-complier effects in the presence of selection on treatment, with application to a study of prompt icu admission. *Journal of the American Statistical Association*, 114(525): 93–104, 2019.

Kennedy, E. H., Balakrishnan, S., and G'Sell, M. Sharp instruments for classifying compliers and generalizing causal effects. *The Annals of Statistics*, 48(4):2008–2030, 2020.

Kennedy, E. H., Balakrishnan, S., and Wasserman, L. Semiparametric counterfactual density estimation. *arXiv preprint arXiv:2102.12034*, 2021.

Kim, K., Kim, J., and Kennedy, E. H. Causal effects based on distributional distances. *arXiv preprint arXiv:1806.02935*, 2018.

Kim, K., Kennedy, E., and Zubizarreta, J. R. Doubly robust counterfactual classification. *Advances in Neural Information Processing Systems*, 36, 2022a.

Kim, K., Mishler, A., and Zubizarreta, J. R. Counterfactual mean-variance optimization. *arXiv preprint arXiv:2209.09538*, 2022b.

Kitagawa, T. and Tetenov, A. Who should be treated? empirical welfare maximization methods for treatment choice. *Econometrica*, 86(2):591–616, 2018.

Kleinberg, J., Mullainathan, S., and Raghavan, M. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016.

Künzel, S. R., Sekhon, J. S., Bickel, P. J., and Yu, B. Metalearners for estimating heterogeneous treatment effects using machine learning. *arXiv preprint arXiv:1706.03461*, 2017.

Kusner, M. J., Loftus, J. R., Russell, C., and Silva, R. Counterfactual fairness. *arXiv preprint arXiv:1703.06856*, 2017.

Levrard, C. Quantization/clustering: when and why does k-means work? *Journal de la société française de statistique*, 159(1):1–26, 2018.

Luedtke, A. R. and van der Laan, M. J. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016a.

Luedtke, A. R. and van der Laan, M. J. Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of statistics*, 44(2):713, 2016b.

Menon, A. K. and Williamson, R. C. The cost of fairness in binary classification. In *Conference on Fairness, Accountability and Transparency*, pp. 107–118. PMLR, 2018.

Mishler, A. and Kennedy, E. Fade: Fair double ensemble learning for observable and counterfactual outcomes. *arXiv preprint arXiv:2109.00173*, 2021.

Mishler, A., Kennedy, E. H., and Chouldechova, A. Fairness in risk assessment instruments: Post-processing to achieve counterfactual equalized odds. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 386–400, 2021.

Murphy, S. A. Optimal dynamic treatment regimes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):331–355, 2003.

Nabi, R. and Shpitser, I. Fair inference on outcomes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Nabi, R., Malinsky, D., and Shpitser, I. Learning optimal fair policies. In *International Conference on Machine Learning*, pp. 4674–4682. PMLR, 2019.

Nemirovski, A., Juditsky, A., Lan, G., and Shapiro, A. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on optimization*, 19(4):1574–1609, 2009.

Neugebauer, R. and van der Laan, M. Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference*, 137(2):419–434, 2007.

Newey, W. K. and Robins, J. R. Cross-fitting and fast remainder rates for semiparametric estimation. *arXiv preprint arXiv:1801.09138*, 2018.

Nie, X. and Wager, S. Quasi-oracle estimation of heterogeneous treatment effects. *arXiv preprint arXiv:1712.04912*, 2017.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, 2019.

Orellana, L., Rotnitzky, A., and Robins, J. M. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part i: main content. *The international journal of biostatistics*, 6(2), 2010.

Robins, J., Li, L., Tchetgen, E., and van der Vaart, A. Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, pp. 335–421. Institute of Mathematical Statistics, 2008.

Schnell, P. M., Tang, Q., Offen, W. W., and Carlin, B. P. A bayesian credible subgroups approach to identifying patient subgroups with positive treatment effects. *Biometrics*, 72(4):1026–1036, 2016.

Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., Hall, P., et al. Towards a standard for identifying and managing bias in artificial intelligence. 2022.

Semenova, V. and Chernozhukov, V. Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal*, 24(2):264–289, 2021.

Shapiro, A. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research*, 18(4):829–845, 1993.

Shapiro, A., Dentcheva, D., and Ruszczyński, A. *Lectures on stochastic programming: modeling and theory*. SIAM, 2014.

Still, G. Lectures on parametric optimization: An introduction. *Optimization Online*, 2018.

Tsybakov, A. B. Optimal rates of aggregation. In *Learning theory and kernel machines*, pp. 303–313. Springer, 2003.

van der Laan, M. J. and Luedtke, A. R. Targeted learning of an optimal dynamic treatment, and statistical inference for its mean outcome. 2014.

Van der Vaart, A. W. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.

Viviano, D. and Bradic, J. Fair policy targeting. *Journal of the American Statistical Association*, pp. 1–37, 2022.

Wager, S. and Athey, S. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

Wang, T. and Rudin, C. Causal rule sets for identifying subgroups with enhanced treatment effects. *INFORMS Journal on Computing*, 34(3):1626–1643, 2022.

White, M. D. and Fradella, H. F. *Stop and frisk: The use and abuse of a controversial policing tactic*. NYU Press, 2016.

Zafar, M. B., Valera, I., Gomez-Rodriguez, M., and Gummadi, K. P. Fairness constraints: A flexible approach for fair classification. *The Journal of Machine Learning Research*, 20(1):2737–2778, 2019.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. A robust method for estimating optimal treatment regimes. *Biometrics*, 68(4):1010–1018, 2012.

Zhang, B., Tsiatis, A. A., Laber, E. B., and Davidian, M. Robust estimation of optimal dynamic treatment regimes for sequential treatment decisions. *Biometrika*, 100(3): 681–694, 2013.

Zhao, L., Tian, L., Cai, T., Claggett, B., and Wei, L.-J. Effectively selecting a target population for a future comparative study. *Journal of the American Statistical Association*, 108(502):527–539, 2013.

Zhao, Q. Covariate balancing propensity score by tailored loss functions. *The Annals of Statistics*, 47(2):965–993, 2019.

Zheng, W. and Van Der Laan, M. J. Asymptotic theory for cross-valiyeard targeted maximum likelihood estimation. *Working Paper 273*, 2010.

APPENDIX

A. Fairness in Optimal Policy

Even though a certain fairness criterion is met for a prediction $f(W)$ that we are concerned with, we might not be able to achieve the same level of fairness for $\mathbb{1}\{f(W) > 0\}$. To illustrate this, consider the criterion of independence and assume that for some $\sigma > 0$,

$$f(W) \mid S = 0 \sim N(-\delta, \sigma^2)$$
$$f(W) \mid S = 1 \sim N(\delta, \sigma^2).$$

The above setup naturally satisfies the independence fairness constraint:

$$|\mathbb{E}\{f(W) \mid S = 0\} - \mathbb{E}\{f(W) \mid S = 1\}\|_{2,\mathbb{P}} \leq \delta.$$

However we get

$$\begin{aligned}&\left|\mathbb{E}\left[\mathbb{1}\{f(W) > 0\} \mid S = 1\right] - \mathbb{E}\left[\mathbb{1}\left\{\beta^\top \boldsymbol{b}(W) > 0\right\} \mid S = 0\right]\right\|_{2,\mathbb{P}} \\ &= \frac{1}{2}\left\{\text{erf}\left(\frac{\delta}{\sigma}\right) - \text{erf}\left(-\frac{\delta}{\sigma}\right)\right\}.\end{aligned}$$

Hence, when σ is large so there is sufficient overlap between the two conditional distributions, RHS of the above equation vanishes as $\delta \to 0$ and thereby the criterion of independence will likely to hold for $\mathbb{1}\{f(W) > 0\}$ as well. On the other extreme, when σ is very small so there is no overlap, then RHS remains as $1/2$ and the same fairness criterion would not hold at all for $\mathbb{1}\{f(W) > 0\}$.

This shows that the following fairness condition alone

$$\left|\mathbb{E}\left\{\text{uf}(Z)\beta^\top \boldsymbol{b}(W)\right\}\right| \leq \delta$$

does not guarantee that the same condition holds for $\mathbb{1}\left\{\beta^\top \boldsymbol{b}(W) > 0\right\}$ in general.

However, this issue can be greatly alleviated by employing the fairness function such as (5) under the margin condition. In this case, the unfairness would differ at most by the order of $\delta + \left\|\tau(W) - \beta^\top \boldsymbol{b}(W)\right\|_{2,\mathbb{P}}$.

B. Formal Definitions of the Regularity Conditions

Let $\mathcal{C}_{\text{fair}} := \left\{\beta \;\middle|\; \left|\mathbb{E}\left\{\text{uf}_j(Z)\beta^\top \boldsymbol{b}(W)\right\}\right| \leq \delta_j,\, j \in J\right\}$. For simplicity, define a set of inequality constraints $\mathcal{C} = \{\beta \mid g_j(\beta) \leq 0, 1 \leq j \leq M\}$ such that $\mathcal{C} = \{\mathcal{C}_{\text{fair}}, \mathcal{C}_{\text{lin}}\}$, where $\mathcal{C}_{\text{lin}}$ is a set of extra linear constraints Then for a feasible point $\bar{\beta} \in \mathcal{C}$, we define the active index set.

Definition B.1 (Active set)**.** We define the active index set J_0 by

$$J_0(\bar{\beta}) = \{1 \leq j \leq M \mid g_j(\bar{\beta}) = 0\}.$$

In what follows, we define LICQ and SC with respect to (P).

Definition B.2 (LICQ)**.** *Linear independence constraint qualification* (LICQ) is satisfied at $\bar{\beta} \in \mathcal{S}$ if the vectors $\nabla_\beta g_j(\bar{\beta})$, $j \in J_0(\bar{\beta})$ are linearly independent.

Definition B.3 (SC)**.** Let $L(\beta, \gamma)$ be the Lagrangian. *Strict Complementarity* (SC) is satisfied at $\bar{\beta} \in \mathcal{S}$ if, with multipliers $\bar{\gamma}_j \geq 0$, $j \in J_0(\bar{\beta})$, the Karush-Kuhn-Tucker (KKT) condition

$$\nabla_\beta L(\bar{\beta}, \bar{\gamma}) := \nabla_\beta \mathcal{L}(\bar{\beta}) + \sum_{j \in J_0(\bar{\beta})} \bar{\gamma}_j \nabla_\beta g_j(\bar{\beta}) = 0,$$

is satisfied such that $\bar{\gamma}_j > 0, \forall j \in J_0(\bar{\beta})$.

LICQ is arguably one of the most widely-used constraint qualifications that admit the first-order necessary conditions. SC means that if the j-th inequality constraint is active then the corresponding dual variable is strictly positive, so exactly one of them is zero for each $1 \leq j \leq m$. SC is widely used in the optimization literature, particularly in the context of parametric optimization (e.g., Shapiro et al., 2014).

Next, we give the second-order condition that can replace Assumption (A1).

(A1') For each optimal solution β^* in (P),

$$\varsigma^\top \mathbb{E}[\boldsymbol{b}(W)\boldsymbol{b}(W)^\top]\varsigma > 0, \quad \forall \varsigma \in \{\varsigma \in \mathbb{R}^k \mid \boldsymbol{b}(W)^\top \varsigma \leq 0, \mathrm{uf}_j(Z)\boldsymbol{b}(W)^\top \varsigma \leq 0, j \in J_0(\beta^*)\} \setminus \{0\}.$$

Here, $J_0(\beta^*)$ is the active index set for β^*. Assumption (A1') implies that β^* is locally isolated, and is weaker than Assumption (A1) which we used in the main text for simplicity. This guarantees that the quadratic growth condition holds at each β^* (Shapiro et al., 2014).

C. Estimation of Non-smooth Functionals under the Margin Condition

We first consider the fairness function which involves the non-smooth component of the form $\mathbb{1}\{\vartheta(W) > 0\}$ for some ϑ. We first provide the following useful lemma.

Lemma C.1. *For any functionals $\vartheta, \xi : \mathcal{Z} \to \mathbb{R}$, $\|\xi\|_{\infty,\mathbb{P}} < \infty$ and their estimates $\widehat{\xi}, \widehat{\vartheta}$, let $\psi = \mathbb{P}\{\phi_\vartheta\} \equiv \mathbb{P}\{\xi(Z)\mathbb{1}\{\vartheta(W) > 0\}\}$ and consider the plug-in estimator $\widehat{\psi} = \mathbb{P}_n\left\{\widehat{\phi}_{\vartheta_{-B}}\right\} \equiv \frac{1}{K}\sum_{b=1}^K \mathbb{P}_n^b\left\{\widehat{\phi}_{\vartheta_{-b}}\right\} = \frac{1}{K}\sum_{b=1}^K \mathbb{P}_n^b\left\{\widehat{\xi}(Z)\mathbb{1}\{\widehat{\vartheta}_{-b}(W) > 0\}\right\}$. Assume that*

1. *For any margin exponent $\alpha > 0$ and for all $t \geq 0$, the following margin condition holds:*

$$\mathbb{P}(|\vartheta(W)| \leq t) \lesssim t^\alpha,$$

2. $\mathbb{P}\left[\mathbb{1}\{\widehat{\vartheta}(W) > 0\} \neq \mathbb{1}\{\vartheta(W) > 0\}\right] + \|\widehat{\xi} - \xi\|_{2,\mathbb{P}} = o_{\mathbb{P}}(1)$.

Then,

$$\widehat{\psi} - \psi = O_{\mathbb{P}}\left(\|\widehat{\xi} - \xi\|_{1,\mathbb{P}} + \|\widehat{\vartheta} - \vartheta\|_{\infty,\mathbb{P}}^\alpha + \frac{1}{\sqrt{n}}\right).$$

If we further assume

3 $\|\widehat{\xi} - \xi\|_{1,\mathbb{P}} + \|\widehat{\vartheta} - \vartheta\|_{\infty,\mathbb{P}}^\alpha = o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$,

then

$$\sqrt{n}\left(\widehat{\psi} - \psi\right) \xrightarrow{d} N\left(0, \mathit{var}(\phi)\right).$$

Proof. We can write $\widehat{\psi} - \psi$ as

$$(\mathbb{P}_n - \mathbb{P})\,\phi_\vartheta + (\mathbb{P}_n - \mathbb{P})\left(\widehat{\phi}_{\vartheta_{-B}} - \phi_\vartheta\right) + \mathbb{P}\left(\widehat{\phi}_{\vartheta_{-B}} - \phi_\vartheta\right)$$

For the second term, noting that $(\mathbb{P}_n - \mathbb{P})\left(\widehat{\phi}_{\vartheta_{-B}} - \phi_\vartheta\right) = \sum_{b=1}^K (\mathbb{P}_n - \mathbb{P})\left\{(\widehat{\phi}_{\vartheta_{-b}} - \phi_\vartheta)(\mathbb{1}(B = b)\right\}$ and $n \lesssim \frac{n}{K}$, we have

$$\begin{aligned}
\left\|(\mathbb{P}_n - \mathbb{P})\left\{(\widehat{\phi}_{\vartheta_{-b}} - \phi_\vartheta)(\mathbb{1}(B = b)\right\}\right\|_{2,\mathbb{P}} &\leq \left\|\widehat{\phi}_{\vartheta_{-b}} - \phi_\vartheta\right\|_{2,\mathbb{P}} \lesssim \left\|\widehat{\phi}_\vartheta - \phi_\vartheta\right\|_{2,\mathbb{P}} \\
&= \left\|(\widehat{\xi} - \xi)\mathbb{1}\{\widehat{\vartheta} > 0\} + \xi\left(\mathbb{1}\{\widehat{\vartheta} > 0\} - \mathbb{1}\{\vartheta > 0\}\right)\right\|_{2,\mathbb{P}} \\
&\lesssim \mathbb{P}\left[\mathbb{1}\{\widehat{\vartheta}(W) > 0\} \neq \mathbb{1}\{\vartheta(W) > 0\}\right] + \|\widehat{\xi} - \xi\|_{2,\mathbb{P}}.
\end{aligned}$$

Hence, if $\mathbb{P}\left[\mathbb{1}\{\widehat{\vartheta}(W)>0\}\neq\mathbb{1}\{\vartheta(W)>0\}\right]+\|\widehat{\xi}-\xi\|_{2,\mathbb{P}}=o_{\mathbb{P}}(1)$, then the the second term is $o_{\mathbb{P}}(\frac{1}{\sqrt{n}})$ by Kennedy et al. (2020, Lemma 1).

The third term can be rewritten as

$$\mathbb{P}\left\{(\widehat{\xi}-\xi)\mathbb{1}(\widehat{\vartheta}_{-B}>0)\right\}+\mathbb{P}\left[\xi\left\{\mathbb{1}(\widehat{\vartheta}_{-B}>0)-\mathbb{1}(\vartheta>0)\right\}\right].$$

The first term in the above display can be simply bounded as

$$\mathbb{P}\left\{(\widehat{\xi}-\xi)\mathbb{1}(\widehat{\vartheta}_{-B}>0)\right\}\lesssim\|\widehat{\xi}-\xi\|_{1,\mathbb{P}}.$$

For the second term, using the similar logic as before, we have that

$$\begin{aligned}\mathbb{P}\left[\xi\left\{\mathbb{1}(\widehat{\vartheta}_{-B}>0)-\mathbb{1}(\vartheta>0)\right\}\right]&\lesssim\mathbb{P}\left|\xi\left\{\mathbb{1}(\widehat{\vartheta}>0)-\mathbb{1}(\vartheta>0)\right\}\right|\\&\leq\|\xi\|_{\infty,\mathbb{P}}\mathbb{P}\left|\left\{\mathbb{1}(\widehat{\vartheta}>0)-\mathbb{1}(\vartheta>0)\right\}\right|\\&\lesssim\mathbb{P}\left(|\vartheta|\leq|\widehat{\vartheta}-\vartheta|\right)\\&\leq\|\widehat{\vartheta}-\vartheta\|_{\infty,\mathbb{P}}^{\alpha},\end{aligned}$$

where the last inequality follows by the margin condition. Putting the two pieces together, we conclude

$$\mathbb{P}\left(\widehat{\phi}_{\vartheta_{-B}}-\phi_{\vartheta}\right)\lesssim\|\widehat{\xi}-\xi\|_{1,\mathbb{P}}+\|\widehat{\vartheta}-\vartheta\|_{\infty,\mathbb{P}}^{\alpha}$$

If this term is $o_{\mathbb{P}}\left(n^{-\frac{1}{2}}\right)$ (Condition 3), then the $\sqrt{n}$-consistency and asymptotic normality follows by the central limit theorem and Slutsky's theorem. □

Lemma C.1 gives corresponding convergence rates, as well as conditions under which $\widehat{\psi}$ is asymptotically normal and efficient.

Now consider the unfairness estimand $\mathbb{E}\left\{\text{uf}_j(Z)\boldsymbol{b}(W)\right\}$ with the fairness function (5)

$$\text{uf}_j(Z)=\frac{(1-S)\mathbb{1}\left\{\tau(W)>0\right\}}{\mathbb{E}\left[(1-S)\mathbb{1}\left\{\tau(W)>0\right\}\right]}-\frac{S\mathbb{1}\left\{\tau(W)>0\right\}}{\mathbb{E}\left[S\mathbb{1}\left\{\tau(W)>0\right\}\right]}.$$

By Lemma C.1 and the continuous mapping theorem, one can deduce that the estimator

$$\begin{aligned}&\mathbb{P}_n\left\{\widehat{\text{uf}}_j\boldsymbol{b}(W)\right\}\\&=\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\Bigg(\Bigg[\frac{(1-S)\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W)-\widehat{\mu}_{0,-b}(W)>0\right\}}{\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left[(1-S)\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W)-\widehat{\mu}_{0,-b}(W)>0\right\}\right]}\\&\quad-\frac{S\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W)-\widehat{\mu}_{0,-b}(W)>0\right\}}{\frac{1}{K}\sum_{b=1}^{K}\mathbb{P}_n^b\left[S\mathbb{1}\left\{\widehat{\mu}_{1,-b}(W)-\widehat{\mu}_{0,-b}(W)>0\right\}\right]}\Bigg]\boldsymbol{b}(W)\Bigg)\end{aligned}$$

attains the $\sqrt{n}$ rate of convergence if

$$\begin{aligned}&\max\|\widehat{\mu}_a-\mu_a\|_{\infty,\mathbb{P}}^{\alpha}=O_{\mathbb{P}}(n^{-\frac{1}{2}}),\\&\mathbb{P}\left[\mathbb{1}\{\widehat{\tau}(W)>0\}\neq\mathbb{1}\{\tau(W)>0\}\right]=o_{\mathbb{P}}(1),\end{aligned}$$

and is asymptotically normal and efficient if it additionally holds that

$$\max\|\widehat{\mu}_a-\mu_a\|_{\infty,\mathbb{P}}^{\alpha}=o_{\mathbb{P}}(n^{-\frac{1}{2}}).$$

Importantly, Lemma C.1 could also be used for constructing efficient estimators for target functionals involving a non-smooth component of the more general form $\mathbb{1}\left\{\tau(W)\in\mathcal{I}\right\}$ for some interval (or union of intervals) $\mathcal{I}$ on $\mathbb{R}$, as long as $\mathbb{1}\left\{\tau(W)\in\mathcal{I}\right\}$ can be expressed as a smooth function of $\left\{\mathbb{1}\left\{\kappa_{1,j}\tau(W)+\kappa_{2,j}>0\right\}\right\}_j$ for some constants $\kappa_{1,j},\kappa_{2,j}\in\mathbb{R}$.

D. Proofs

D.1. Proof of Theorem 3.2

One may rewrite (P) as the following nonlinear program with linear constraints

$$\begin{aligned} \underset{\beta \in \mathbb{R}^k}{\text{minimize}} \quad & f(\beta, T) \\ \text{subject to} \quad & C\beta \leq \delta, \end{aligned}$$

where

$$\begin{aligned} \delta &= [\delta_1, \ldots, \delta_m]^\top, \\ T &= \{\mathbb{E}[\boldsymbol{b}(W)\boldsymbol{b}(W)^\top]_{i,j}, \mathbb{E}[(Y^1 - Y^0)\boldsymbol{b}(W)]_i\}_{i,j=1}^k \in \mathbb{R}^{k(k+1)}, \\ C &= [\mathbb{E}\{\text{uf}_1(Z)\boldsymbol{b}(W)\}, \ldots, \mathbb{E}\{\text{uf}_m(Z)\boldsymbol{b}(W)\}]^\top \in \mathbb{R}^{m \times k}. \end{aligned}$$

According to our approximating program ($\widehat{\mathsf{P}}$) we have that,

$$\begin{aligned} \widehat{T} &= \{\mathbb{P}_n[\boldsymbol{b}(W)\boldsymbol{b}(W)^\top]_{i,j}, \mathbb{P}_n\left\{\left(\varphi_1(Z; \widehat{\vartheta}_{-B}) - \varphi_0(Z; \widehat{\vartheta}_{-B})\right)\boldsymbol{b}(W)\right\}_i\}_{i,j=1}^k, \\ \widehat{C} &= [\mathbb{P}_n\left\{\widehat{\text{uf}}_1\boldsymbol{b}(W)\right\}, \ldots, \mathbb{P}_n\left\{\widehat{\text{uf}}_m\boldsymbol{b}(W)\right\}]^\top. \end{aligned}$$

Given the conditions (A1) - (A3), (A5), by Lemma A.1 of (Kim et al., 2022a), it follows that

$$\begin{aligned} \|\widehat{T} - T\|_2 &= O_{\mathbb{P}}\left(\max \|\widehat{\pi} - \pi\|_{2,\mathbb{P}} \|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} + n^{-\frac{1}{2}}\right), \\ \|\widehat{C} - C\|_F &= O_{\mathbb{P}}\left(\max \|\widehat{\pi} - \pi\|_{2,\mathbb{P}} \|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} + n^{-\frac{1}{2}}\right). \end{aligned}$$

Further, under the conditions (A1), Theorem 3.1 of (Kim et al., 2022b) gives that

$$\|\widehat{\beta} - \beta^*\|_2 = O_{\mathbb{P}}\left(\max \|\widehat{\pi} - \pi\|_{2,\mathbb{P}} \|\widehat{\mu}_a - \mu_a\|_{2,\mathbb{P}} + n^{-\frac{1}{2}}\right).$$

If we further assume the conditions (A4),(A6) so that every element in $\widehat{T}$, $\widehat{C}$ is indeed a $\sqrt{n}$-consistent and asymptotically normal estimator for the corresponding element in T, C, then the asymptotic normality with $\sqrt{n}$ rates can be obtained by Theorem 3.2 of (Kim et al., 2022b), under the necessary regularity conditions described in Theorem 3.2.

D.2. Proof of Lemma 4.1

Proof. The proof mimics the proofs of Lemma 5.1 and Lemma 5.2 in (Audibert & Tsybakov, 2007). To show the first inequality, note that

$$\begin{aligned} \mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) &= \mathbb{P}\left[\Delta\left(\mathbb{1}\{\Delta > 0\} - \mathbb{1}\left\{\widehat{\Delta} > 0\right\}\right)\right] \\ &\leq \mathbb{P}\left[|\Delta|\left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right)\right] \\ &\leq \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}} \mathbb{P}\left\{|\Delta| \leq \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}}\right\} \\ &\lesssim \left\|\widehat{\Delta} - \Delta\right\|_{\infty,\mathbb{P}}^{\alpha+1}, \end{aligned}$$

where the first inequality follows by Lemma 1 of (Kennedy et al., 2020) and the last by the margin condition.

Next, for any $t > 0$ we have

$$\begin{aligned}
\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) &\leq \mathbb{P}\left[|\Delta| \left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right) \mathbb{1}\left\{|\Delta| \leq t\right\}\right] \\
&\quad + \mathbb{P}\left[|\Delta| \left(\mathbb{1}\left\{|\Delta| \leq \left|\widehat{\Delta} - \Delta\right|\right\}\right) \mathbb{1}\left\{|\Delta| > t\right\}\right] \\
&\leq \mathbb{P}\left[\left|\widehat{\Delta} - \Delta\right| \mathbb{1}\left\{|\Delta| \leq t\right\}\right] + \mathbb{P}\left[\left|\widehat{\Delta} - \Delta\right| \mathbb{1}\left\{\left|\widehat{\Delta} - \Delta\right| > t\right\}\right] \\
&\leq \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} \Pr\left\{|\Delta| \leq t\right\}^{\frac{q-1}{q}} + \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} \left(\frac{\mathbb{P}|\widehat{\Delta} - \Delta|^q}{t^q}\right)^{\frac{q-1}{q}} \\
&\lesssim \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}} t^{\frac{q-1}{q}} + \frac{\left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^q}{t^{q-1}},
\end{aligned}$$

where the third inequality follows by the Hölder and Markov inequalities and the last by the margin condition. Now, the RHS in the last display is minimized when $t = O\left(\left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^{\frac{q}{q+\alpha}}\right)$, yielding

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\|\widehat{\Delta} - \Delta\right\|_{q,\mathbb{P}}^{\frac{q(1+\alpha)}{q+\alpha}}.$$

□

D.3. Proof of Theorem 4.2

Proof. By the first inequality in Lemma 4.1, we have

$$\mathcal{U}(g^*) - \mathcal{U}(\widehat{g}) \lesssim \left\|\tau(W) - \widehat{\beta}^\top \boldsymbol{b}(W)\right\|_{\infty,\mathbb{P}}^{\alpha+1}$$

Recall that β^* and $\widetilde{\beta}$ are optimal solutions to (P) and (9), respectively. Then $\forall a$, by the triangle and Cauchy–Schwarz inequalities,

$$\left|\tau(W) - \widehat{\beta}^\top \boldsymbol{b}(W)\right\|_{2,\mathbb{P}} \leq \left|\tau(W) - \widetilde{\beta}^\top \boldsymbol{b}(W)\right\|_{2,\mathbb{P}} + \|\boldsymbol{b}(W)\|_2 \left\{\|\widetilde{\beta} - \beta^*\|_2 + \|\beta^* - \widehat{\beta}\|_2\right\}.$$

Next, for some $\lambda \in \mathbb{R}^m$ consider the following Lagrange form of (P).

$$\underset{\beta \in \mathcal{B}}{\text{minimize}} \quad \mathbb{E}\left\{\left(Y^1 - Y^0 - \beta^\top \boldsymbol{b}(W)\right)^2\right\} + \sum_{j=1}^m \lambda_j \left[\beta^\top \mathbb{E}\left\{\text{uf}_j(Z)\boldsymbol{b}(W)\right\}\right]^2 \tag{10}$$

Since our constraint set consists of only linear constraints strong duality holds, and there exists a dual solution $\lambda \geq 0$ to (P) such that any solution $\widetilde{\beta}$ in (P) minimizes

$$\mathbb{E}\left\{\left(Y^1 - Y^0 - \beta^\top \boldsymbol{b}(W)\right)^2\right\} + \sum_{j=1}^m \lambda_j \left(\left[\beta^\top \mathbb{E}\left\{\text{uf}_j(Z)\boldsymbol{b}(W)\right\}\right]^2 - \delta_j\right).$$

Hence, $\widetilde{\beta}$ is a solution in (10) as well.

Now, consider the following parametrized program

$$\underset{\beta \in \mathcal{B}}{\text{minimize}} \quad f(\beta, \xi) := \mathbb{E}\left\{\left(Y^1 - Y^0 - \beta^\top \boldsymbol{b}(W)\right)^2\right\} + \sum_{j=1}^m \left(\beta^\top \xi_j\right)^2, \tag{P(ξ)}$$

with $\xi = (\xi_1, ..., \xi_m)$, $\xi_j \in R^k$, and let $\beta(\xi)$ denote an optimal solution of P(ξ). (9) and (10) correspond to (P(ξ)) with the parameter $\xi_j = 0$ and $\xi_j = \sqrt{\lambda_j}\mathbb{E}\left\{\text{uf}_j(Z)\boldsymbol{b}(W)\right\}$, $\forall j$, respectively.

By positive-definiteness of $\mathbb{E}[\boldsymbol{b}(W)\boldsymbol{b}(W)^\top]$, $\mathsf{P}(0)$ is strongly convex everywhere. Consequently the quadratic growth condition holds at $\beta(0)$ such that

$$c\|\beta(0)-\beta'\|_2^2 \le f(\beta',0) - f\left(\beta(0),0\right)$$

for some constant $c>0$ and all $\beta' \in \mathbb{R}^k$. Hence,

$$\begin{aligned} c\|\beta(0)-\beta(\xi)\|_2^2 &\le f\left(\beta(\xi),0\right) - f\left(\beta(0),0\right) \\ &= f\left(\beta(\xi),0\right) - f\left(\beta(\xi),\xi\right) + f\left(\beta(\xi),\xi\right) - f\left(\beta(0),\xi\right) + f\left(\beta(0),\xi\right) - f\left(\beta(0),0\right) \\ &\le \left\{\|\beta(\xi)\|_2^2 + \|\beta(0)\|_2^2\right\}\|\xi\|_2^2, \end{aligned}$$

where the last inequality follows by $f\left(\beta(\xi),\xi\right) - f\left(\beta(0),\xi\right) \le 0$.

From the above argument, by the Cauchy–Schwarz and Jensen's inequality we have that

$$\begin{aligned} \|\widetilde{\beta}-\beta^*\|_2 &\lesssim \left\|\mathbb{E}\left\{\left(\sum_j \sqrt{\lambda_j}\mathrm{uf}_j(Z)\right)\boldsymbol{b}(W)\right\}\right\|_2 \\ &\le \left\|\sum_j \sqrt{\lambda_j}\mathrm{uf}_j(Z)\right\|_{2,\mathbb{P}} \sqrt{\sum_{i=1}^k \mathbb{E}\{\boldsymbol{b}(W)_i{}^2\}} \\ &\le \left\|\sum_j \sqrt{\lambda_j}\mathrm{uf}_j(Z)\right\|_{2,\mathbb{P}} \|\boldsymbol{b}(W)\|_{2,\mathbb{P}}. \end{aligned}$$

Further, under the conditions (A1)-(A3), (A5), by Theorem 3.2, it follows that

$$\|\beta^*-\widehat{\beta}\|_2 = O_{\mathbb{P}}\left(n^{-\frac{1}{2}} + \max_a \|\widehat{\pi}_a - \pi_a\|_{2,\mathbb{P}} \max \|\widehat{\tau}-\tau\|_{2,\mathbb{P}}\right).$$

Since $0<\alpha<\infty$, we obtain the desired result by putting the pieces together due to Minkowski's inequality.

The part (iii) follows by the exact same logic as used for the part (i).

The part (ii) immediately follows by the fact that

$$\begin{aligned} \mathbb{P}\left\{\widehat{g}(W) \ne g^*(W)\right\} &= \mathbb{P}\left\{|\widehat{g}(W)-g^*(W)\|_{2,\mathbb{P}}\right\} \\ &\le \mathbb{P}\left[\mathbb{1}\left\{|\tau(W)\|_{2,\mathbb{P}} \le \sum \left|\tau(W) - \widehat{\beta}^\top \boldsymbol{b}(W)\right\|_{2,\mathbb{P}}\right\}\right] \\ &\lesssim \left\|\tau(W) - \widehat{\beta}^\top \boldsymbol{b}(W)\right\|_{\infty,\mathbb{P}}^{\alpha}. \end{aligned}$$

□