Individualized Treatment Allocations with Distributional Welfare*

Yifan Cui
Department of Statistics
Zhejiang University
yifanc095@gmail.com

Sukjin Han
School of Economics
University of Bristol
sukjin.han@gmail.com

November 18, 2023

Abstract

In this paper, we explore optimal treatment allocation policies that target distributional welfare. Most literature on treatment choice has considered utilitarian welfare based on the conditional average treatment effect (ATE). While average welfare is intuitive, it may yield undesirable allocations especially when individuals are heterogeneous (e.g., with outliers)—the very reason individualized treatments were introduced in the first place. This observation motivates us to propose an optimal policy that allocates the treatment based on the conditional *quantile of individual treatment effects* (QoTE). Depending on the choice of the quantile probability, this criterion can accommodate a policymaker who is either prudent or negligent. The challenge of identifying the QoTE lies in its requirement for knowledge of the joint distribution of the counterfactual outcomes, which is generally hard to recover even with experimental data. Therefore, we introduce minimax optimal policies that are robust to model uncertainty. We then

*For helpful discussions, the authors are grateful to participants at the Advances in Econometrics Conference 2023 and the seminar at Brown University.

propose a range of identifying assumptions under which we can point or partially identify the QoTE. We establish the asymptotic bound on the regret of implementing the proposed policies. We consider both stochastic and deterministic rules. In simulations and two empirical applications, we compare optimal decisions based on the QoTE with decisions based on other criteria.

JEL Numbers: C14, C31, C54.

Keywords: Treatment choice, treatment regime, treatment rule, policy learning, distributional treatment effects, quantile treatment effects, partial identification.

1 Introduction

1.1 Policy Learning with Distributional Welfare

Individuals are heterogeneous, so are their responses to treatments or programs. When designing policies (e.g., rules of allocating treatments or programs), it is important to reflect the heterogeneity of individual treatment effects. A policymaker (PM), or equivalently an analyst, would devise a policy to achieve a specific objective (e.g., welfare). Depending on how the PM aggregates individual gains, her objective can be viewed as either utilitarian or non-utilitarian. A utilitarian PM would consider welfare that takes the sum or average of individual gains to ensure the greatest benefits for the greatest number, whereas a non-utilitarian (e.g., prioritarian, maximin) PM would prioritize specific groups of individuals. The utilitarian objective has been the most widely-used criterion in the literature of treatment allocations and policy learning (e.g., Manski (2004); see below for a further review). However, there may be settings where the utilitarian goal is less sensible. For example, the target population may exhibit skewed heterogeneity (e.g., outliers). As another example, the PM may want to target a vulnerable population or privileged individuals, or a certain share of benefited individuals.[1] The purpose of this paper is to explore objectives of a (non-utilitarian) PM who is concerned with certain aspects of the distribution (e.g., tails) of treatment effects

[1]The possibility of non-utilitarian welfare is also briefly mentioned in Manski (2004).

or who has political incentives and thus makes decisions influenced by vote shares.

In this paper, we develop a policy learning framework that concerns distributional welfare. A policy is defined as a mapping from individuals' observed characteristics to either a deterministic or stochastic decision of treatment allocation. Intuitively, the knowledge of individual treatment effects conditional on characteristics plays a crucial role in learning such a policy. We propose an objective function that is formulated based on the conditional quantile of individual treatment effects (QoTE). This objective function is robust to outliers of treatment effects and, more importantly, can reflect the PM's level of prudence toward the target population.

Suppose the PM employs the utilitarian welfare, which can be written as a function of the conditional average treatment effect (ATE). If the policy class is unconstrained, it is optimal for the utilitarian PM to treat each subgroup (defined by observed characteristics) whenever their ATE is positive. Suppose that this PM faces a target subgroup, say black females, whose distribution of treatment effects exhibits that a small share of individuals enjoys positive treatment effects that dominate the negative effects of the remaining majority. If the resulting ATE is positive, then the PM would treat *all* black females, harming the majority. The objective function based on the QoTE with the quantile probability $\tau = 0.5$ (i.e., the median of treatment effects) would not suffer from this sensitivity to outliers. Moreover, the PM can choose the quantile probability τ (i.e., the rank in individual treatment effects) to set a reference group. A large τ corresponds to a PM who is willing to focus on privileged individuals in each subgroup, ignoring the majority of less advantaged, thus being a *negligent* PM. A small τ corresponds to a PM who is concerned with the disadvantaged, treating each subgroup only if most benefit from the treatment, thus being a *prudent* PM. Relatedly, we show that the PM equipped with the QoTE can be interpreted as being concerned with vote shares when each individual casts a vote whenever he or she experiences a positive gain from the treatment.

An alternative objective function that can be robust to certain outliers is the one based on the conditional quantile treatment effect (QTE) which contrasts the quantiles of treated and untreated outcomes. We argue that this quantity may not be an appropriate basis for individualized treatment decisions because an individual represented by the quantile of treated outcomes is not necessarily the same individual represented by the same quantile of untreated outcomes. On the other hand, the QoTE by definition captures an individual with a specific rank in gains. Moreover, as shown later, there is no clear interpretation of prudence when the PM's criterion is based on the QTE.

Despite the desirable properties of the PM's objective function constructed from the QoTE, the challenge is that the QoTE is not generally point-identified even when the PM has access to experimental data. This is due to the fact that, in the definition of the QoTE, the joint distribution of counterfactual outcomes is involved. We overcome this challenge by proposing a minimax criterion that is robust to model ambiguity. In particular, we propose to minimize the worst-case regret calculated over the class of joint distributions of counterfactual outcomes that are compatible with the data and identifying assumptions. We then propose a range of identifying assumptions that can be imposed to tighten the identified set of the QoTE, sometimes to a singleton (i.e., point identification). These assumptions can be imposed by practitioners depending on their specific settings. For some assumptions, the identified set of the QoTE may not have a closed-form expression. In this case, an optimization algorithm can be used to compute the set. By using a Bernstein approximation, we show how the optimization problem becomes a simple linear programming.

We establish theoretical properties of the proposed minimax policy by providing asymptotic bounds on the regret of implementing the estimated policy. First, when the policy class is unconstrained, we show that the estimated policy is consistent if either the bounds on the QoTE are sign-determining or the QoTE is point-identified. Otherwise, the leading term of the regret bound has a magnitude that depends on the relative location of zero in the QoTE bounds. It is important to allow the policy class to be constrained as the PM may prefer a parsimonious policy or face institutional or budget constraints. In this case of constrained policy classes, we propose to use the machine learning (ML) technique of the outcome-weighting framework with a surrogate loss (Zhao et al. (2012)). We then show that the ML-estimated policy is consistent and characterize the rate in terms of approximation and estimation errors. Through numerical exercises, we show how the treatment allocations can differ across welfare criteria especially when the QoTE is partially identified and when one is preferred over the others. We find that the correct classification rate tends to be high when the welfare criterion of the estimated policy matches that of the population policy.

In this paper, we consider two empirical applications. The first application concerns the allocation of a diagnostic procedure for critically ill patients using data from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. The second application examines the allocation of job training using data from the US National Job Training Partnership Act. In both applications, a common finding is that there is heterogeneity in the distributional treatment effects and thus the corresponding allocation decisions based on the QoTE. In addition, we show in the space of covariates how the allocation decisions take place. As expected, the allocation becomes more aggressive as the quantile probability τ increases. We compare this result with the decisions based on the QTE and ATE. The QTE decisions do not exhibit the change in the degree of prudence in τ. Comparing the ATE decisions with the QoTE decisions (with $\tau = 0.5$), we can inspect whether there are outliers in these data sets. In this sense, we show how the QoTE decisions can be viewed as a means of a robustness check for the ATE decisions prevalent in the literature.

1.2 Related Literature

Learning optimal treatment regimes has received considerable interest in the past few years across multiple disciplines including computer science (Dudík et al., 2011), econometrics (Manski, 2004; Hirano and Porter, 2009; Stoye, 2009; Kitagawa and Tetenov, 2018; Athey and Wager, 2021; Mbakop and Tabord-Meehan, 2021; Ida et al., 2022), and statistics (Murphy, 2003; Kosorok and Moodie, 2015; Kosorok and Laber, 2019; Tsiatis et al., 2019; Jiang et al., 2019). In statistics, existing methods for learning optimal treatment regimes are mostly through either Q-learning (Watkins and Dayan, 1992; Qian and Murphy, 2011) or A-learning

(Murphy, 2003; Robins, 2004; Shi et al., 2018). Alternative approaches have emerged from a classification perspective (Zhao et al., 2012; Zhang et al., 2012; Rubin and van der Laan, 2012), which has proven more robust to model misspecification in some settings.

Recently, there is a growing literature on learning optimal treatment allocations that aims to relax the unconfoundedness assumption. Within this literature, a strand of work considers cases where the welfare and optimal treatment regime is point-identified, that is, the treatment decision is free from ambiguity given the observed data. Cui and Tchetgen Tchetgen (2021b); Qiu et al. (2021) consider instrumental variable (IV) approaches under a point identification and Han (2021); Cui and Tchetgen Tchetgen (2021a) consider IV methods under a sign identification. Kallus et al. (2021); Qi et al. (2023a); Shen and Cui (2023) consider optimal policy learning under the proximal causal inference framework. Another strand of work considers robust policy learning under ambiguity. Kallus and Zhou (2021) propose to learn an optimal policy in the presence of partially identified treatment effects under a sensitivity model. Pu and Zhang (2021) consider a minimax regret policy for IV models under partial identification. Cui (2021) and D'Adamo (2021) consider a variety of decision rules in general settings where treatment effects are partially identified. Stoye (2012); Yata (2021) develop finite-sample minimax regret rules under partial identification of welfare. Moreover, Han (2023) proposes optimal dynamic treatment regimes through a partial welfare ordering when the sequential randomization assumption is violated. Policy learning under ambiguity is not limited to confounded settings. There are other settings of robust decisions under ambiguity, for example, when the treatment positivity assumption is violated (Ben-Michael et al., 2021), when data sets are aggregated in meta analyses (Ishihara and Kitagawa, 2021) and when the target population is shifted from the experiment population (Adjaho and Christensen, 2022). The present paper contributes to this literature of model ambiguity by considering a distributional welfare that is partially identified.

There is also work focused on policy learning based on distributional properties under point identification. Leqi and Kennedy (2021) consider the QTE as a criterion and Qi et al. (2023b) consider maximizing the average outcomes that are below a certain quantile. Wang et al. (2018); Linn et al. (2017) consider maximizing the quantile of global welfare, which can be viewed as a special case of Kitagawa and Tetenov (2021). The latter study considers estimating the optimal treatment allocation based on individual characteristics when the objective is to maximize an equality-minded rank-dependent welfare function, which essentially puts higher weights on individuals with lower-ranked outcomes. Our work complements this line of literature by introducing a different type of distributional welfare using the distribution of treatment effects and proposing decision-making under ambiguity. Further comparisons to this line of work are made in Section 2. Finally, Manski and Tetenov (2023); Kitagawa et al. (2023) consider a distribution or nonlinear function of regret and establish admissible treatment rules within that framework. Although our welfare has distributional aspects, when showing the theoretical guarantee of the estimated rules, we use the standard the notion of the (mean) regret.

1.3 Organization of the Paper

The paper is organized as follows. The next section formally introduces our welfare criterion and compare it with criteria previously considered in the literature. Then the minimax framework is proposed. Section 3 lists a menu of identifying assumptions that can be used to narrow the bounds on the QoTE. Section 4 presents the theoretical properties of the estimated policies for constrained and unconstrained policy classes. Section 5 briefly discusses how to systematically calculate the bounds on the QoTE using linear programming. Section 6 contains numerical exercises and Section 7 presents the two empirical applications.

2 Treatment Rules and Distributional Welfare

Let $Y \in \mathcal{Y}$ be the outcome, $X \in \mathcal{X}$ be covariates, and $D \in \{0, 1\}$ be binary treatment in respective supports. Let Y_d be the potential outcome that is consistent with the observed outcome, that is, $Y = DY_1 + (1-D)Y_0$. We define a treatment allocation rule, or equivalently a *policy*, as $\delta : \mathcal{X} \to \mathcal{A} \subseteq [0, 1]$ where $\mathcal{A}$ is the action space. A deterministic rule corresponds to $\mathcal{A} = \{0,1\}$ and a stochastic rule corresponds to $\mathcal{A} = [0,1]$. Unless noted otherwise, we allow both in our general framework. Let $\delta \in \mathcal{D}$ where $\mathcal{D}$ is the (potentially constrained) space of δ. For the allocation problem, a policymaker (PM) would set an objective function that she maximizes to find the optimal allocation rule.

To motivate the objective function we propose, we first review the most common objective function considered in the literature: the average welfare.[2] The optimal policy under this welfare criterion can be defined as

$$\delta^*_{ATE} \in \arg\max_{\delta \in \mathcal{D}} E[\delta(X)Y_1 + (1-\delta(X))Y_0].$$

With deterministic rules in particular, the welfare can be written as $E[\delta(X)Y_1 + (1 - \delta(X))Y_0] = E[Y_{\delta(X)}]$. See Section A in the Appendix that shows how $E[\delta(X)Y_1+(1-\delta(X))Y_0]$ (and other welfare criteria appearing below) is compatible with stochastic rules. Because

$$E[\delta(X)Y_1 + (1-\delta(X))Y_0] = E[Y_0 + \delta(X)(Y_1 - Y_0)] = E[Y_0] + E[\delta(X)E[Y_1 - Y_0|X]],$$

δ^*_{ATE} also satisfies

$$\delta^*_{ATE} \in \arg\max_{\delta \in \mathcal{D}} E[\delta(X)E[Y_1 - Y_0|X]], \tag{2.1}$$

where the objective function corresponds to the *welfare gain*. Therefore, subject to the constraints, δ^*_{ATE} maximizes the average of conditional average treatment effects (ATEs) either chosen (in the case of deterministic policies) or weighted (in the case of stochastic policies) by δ, thus the notation "δ^*_{ATE}." For example, when $\mathcal{D}$ is not constrained, $\delta^*_{ATE}(x) = 1\{E[Y_1 - Y_0|X = x] \geq 0\}$. In general, the formulation (2.1) reveals an important fact: the conditional treatment effect is the important basis for the policy choice. This makes sense because the treatment should be allocated to those who would benefit the most from it. This idea becomes important in introducing our distributional welfare later.

[2]Welfare is sometimes called a value function in the literature.

Although it is the most common form of welfare, the average welfare is obviously sensitive to outliers. For example, a small share of individuals with $X = x$ and substantially large $Y_1 - Y_0$ can make $E[Y_1 - Y_0|X = x]$ positive, suggesting to treat *all* individuals with $X = x$ even though the majority suffers from receiving the treatment. This can be especially problematic when the distribution of $Y_1 - Y_0|X = x$ is skewed and heavy-tailed. This motivates us to alternatively consider the quantile of individual treatment effects $Y_1 - Y_0$ (QoTE) as the basis for a welfare criterion (analogous to (2.1)) and a corresponding optimal policy. Let $Q_\tau(Y) \equiv \inf\{y : F_Y(y) \geq \tau\}$ be the τ-quantile of Y and $Q_\tau(Y|X) \equiv \inf\{y : F_{Y|X}(y) \geq \tau\}$ be the τ-quantile of Y conditional on X. We consider

$$\delta^* \in \arg\max_{\delta \in \mathcal{D}} E[\delta(X) Q_\tau(Y_1 - Y_0|X)], \tag{2.2}$$

where $Q_\tau(Y_1 - Y_0|X)$ is the τ-quantile of $Y_1 - Y_0$ given X. That is, δ^* maximizes the average of conditional QoTEs chosen (in the case of deterministic policies) or weighted (in the case of stochastic policies) by δ. With no constraint in $\mathcal{D}$, $\delta^*(x) = 1\{Q_\tau(Y_1 - Y_0|X = x) \geq 0\}$. The QoTE is less sensitive to outliers than the ATE, so for example (2.2) with $\tau = 0.5$ may be preferred to (2.1). This aspect makes the allocation decision within the $X = x$ group not driven by treatment effects of a small share of individuals. In this sense, this aspect of robustness can be viewed as the "within-group fairness" (Leqi and Kennedy (2021)).[3] In general, τ (i.e., the rank in individual treatment effects) represents individuals in that specific quantile as a *reference group* chosen by the PM. For example, by choosing low τ, the PM allocates the treatment only if most individuals benefit from it because $Q_{\tau'}(Y_1 - Y_0|X) \geq Q_\tau(Y_1 - Y_0|X)$ for any $\tau' > \tau$. In other words, she ensures that disadvantaged individuals with poor treatment effects are not harmed from receiving the allocation. In this sense, low τ corresponds to a *prudent PM*. On the other hand, by choosing high τ, the PM focuses on benefiting solely the top-ranked individuals even though the majority would suffer from it. In this sense, high τ corresponds to a *negligent PM*. Therefore, the choice of τ reflects the

[3]In fact, we show below that the notion of within-group fairness fits better in our framework than that of Leqi and Kennedy (2021)'s framework.

level of prudence of the policy that the PM commits to.

The proposed optimal policy has another interesting interpretation that relates to the PM's incentive. Suppose individuals who benefit from the treatment would vote for it. Also suppose $\tau = 0.5$ and $\mathcal{D}$ is unconstrained. Then $\delta^*(X) = 1\{Q_{0.5}(Y_1 - Y_0|X) \geq 0\}$ can be viewed as a policy that obeys *majority vote*. To see this, note the following:

$$\begin{aligned} Q_{0.5}(Y_1 - Y_0|X) \geq 0 \Leftrightarrow \quad & F_{Y_1 - Y_0|X}(0) \leq 1/2 \\ \Leftrightarrow \quad & P[Y_1 \geq Y_0|X] \geq 1/2 \\ \Leftrightarrow \quad & P[Y_1 \geq Y_0|X] \geq P[Y_1 < Y_0|X] \end{aligned}$$

Therefore, the distributional welfare criterion (2.2) is consistent with a PM who has political incentive and whose decision is influenced by vote shares. This interpretation can be generalized by considering $Q_{0.5-\alpha/2}(Y_1 - Y_0|X) \geq 0$ for $0 \leq \alpha \leq 1$, which is equivalent to $P[Y_1 \geq Y_0|X] \geq P[Y_1 < Y_0|X] + \alpha$ where α can be viewed as the vote share margin.

Related to the proposed welfare criterion, previous studies have considered alternative criteria that are robust to outliers. Focusing on a deterministic policy (i.e., $\mathcal{A} = \{0, 1\}$), Wang et al. (2018) consider the marginal quantile of $Y_{\delta(X)}$ as their criterion, while Leqi and Kennedy (2021) focus on the average of conditional quantile $Y_{\delta(X)}$. First, Wang et al. (2018) explore the optimal policy under $Q_\tau(Y_{\delta(X)})$, which can be viewed as a sensible quantity robust to outliers. Note that the randomness in $Y_{\delta(X)}$ arises from both Y_d and X. Because of that, the optimal policy under $Q_\tau(Y_{\delta(X)})$ does not have a closed form solution, which make the interpretation of the optimal policy somewhat elusive. Moreover, Leqi and Kennedy (2021) demonstrate that the policy under this welfare criterion lacks "across-group fairness," in that the allocation decision for one group (defined by $X = x$) can be influenced by the treatment effects of other groups (defined by other $X = x'$). This issue stems from the difficulty in associating the objective function $Q_\tau(Y_{\delta(X)})$ with a clear notion of treatment effects or gains, unlike the other criteria discussed in this section.

To overcome this issue, Leqi and Kennedy (2021) consider the optimal policy under

$E[Q_\tau(Y_{\delta(X)}|X)]$, which achieves across-group fairness as X is fixed in the calculation of quantile. As shown in their paper,

$$
\begin{aligned}
E[Q_\tau(Y_{\delta(X)}|X)] &= E[\delta(X)Q_\tau(Y_1|X) + (1-\delta(X))Q_\tau(Y_0|X)] \\
&= E[Q_\tau(Y_0|X)] + E[\delta(X)\{Q_\tau(Y_1|X) - Q_\tau(Y_0|X)\}]
\end{aligned}
$$

and therefore the optimal policy also satisfies

$$
\delta^*_{QTE} \in E[\delta(X)\{Q_\tau(Y_1|X) - Q_\tau(Y_0|X)\}].
$$

That is, δ^*_{QTE} maximizes the average of conditional QTEs chosen by δ. However, allocating the treatment based on the QTE may be questionable because the individual at the τ-quantile of Y_1 may not be the same individual as the one at the τ-quantile of Y_0. This aspect is also reflected in the fact that generally $Q_\tau(Y_1|X) - Q_\tau(Y_0|X) \neq Q_\tau(Y_1 - Y_0|X)$ unlike in the expectation operator (i.e., the ATE). Since the QTE is introduced in Doksum (1974) and Lehmann (1975), its limitation as a causal parameter has been acknowledged in the literature but the problem seems more pronounced in the context of treatment allocation. Moreover, this aspect implies that there is no clear notion of a negligent or prudent PM associated with the level of τ.

Despite the desirable properties of our proposed objective function, the main challenge of using (2.2) as the welfare criterion is that the QoTE is generally not point-identified even under unconfoundedness. Therefore, we propose optimal policies that are robust to this ambiguity. One may consider maximizing the worst-case gain:

$$
\delta^*_{mmw} \in \arg\max_{\delta\in\mathcal{D}} \min_{F_{Y_1,Y_0|X}\in\mathcal{F}} E[\delta(X)Q_\tau(Y_1 - Y_0|X)], \tag{2.3}
$$

where $F_{Y_1,Y_0|X}$ is the joint distribution of (Y_1, Y_0) conditional on X and $\mathcal{F} \equiv \mathcal{F}(P)$ is the identified set of $F_{Y_1,Y_0|X}$ given the data P. However, this criterion is known to be overly pessimistic (Savage (1951)). Therefore, one may instead consider minimizing the worst-case regret:

$$\delta^*_{mmr} \in \arg\min_{\delta \in \mathcal{D}} \max_{F_{Y_1,Y_0|X} \in \mathcal{F}} E[\{\delta^\dagger(X) - \delta(X)\} Q_\tau(Y_1 - Y_0|X)], \tag{2.4}$$

where $\delta^\dagger \in \arg\max_{\delta:\mathcal{X}\to\mathcal{A}} E[\delta(X) Q_\tau(Y_1 - Y_0|X)]$ is the first-best rule. Note that $\delta^\dagger(X) = 1\{Q_\tau(Y_1 - Y_0|X) \geq 0\}$ as no restriction is imposed on the class of δ. The minimax regret criterion is free from priors and thus avoids the feature of maximin mentioned above. Therefore, our primary focus is the minimax policy.

For each x, define the identified interval for $Q_\tau(Y_1 - Y_0|X = x)$ as

$$[Q^L_\tau(x), Q^U_\tau(x)] = \{Q_\tau(Y_1 - Y_0|X = x) : F_{Y_1,Y_0|X} \in \mathcal{F}\}.$$

Using these lower and upper bounds, we can derive closed-form expressions for the inner optimization in (2.3) and (2.4). To this end, we impose a very weak assumption on the identified interval.

Assumption RC. *The identified set $\mathcal{Q}(P)$ of $Q_\tau(Y_1 - Y_0|X = \cdot)$ is rectangular, that is,*

$$\mathcal{Q}(P) = \{Q_\tau(Y_1 - Y_0|X = \cdot) : Q_\tau(Y_1 - Y_0|X = x) \in [Q^L_\tau(x), Q^U_\tau(x)]\}.$$

This assumption holds for the identified sets we derive in this paper. It will be violated if one imposes certain shape restrictions on $Q_\tau(Y_1 - Y_0|X = \cdot)$ such as monotonicity. We do not consider shape restrictions in this paper as allowing for unrestricted heterogeneity across X is important in the context of optimal allocations. Essentially, this assumption allows us to interchange the maximum or minimum over $\mathcal{F}$ with the expectation over X (Kasy (2016); D'Adamo (2021)).[4] Under Assumption RC, we can easily show that δ^*_{mmr}

[4] To illustrate this, consider a simple case of binary $X \in \{0,1\}$ and let $Q_\tau(x) \equiv Q_\tau(Y_1 - Y_0|X = x)$ and $p_x \equiv \Pr[X = x]$. Then Assumption RC imposes that $\{(Q_\tau(0), Q_\tau(1)) : Q_\tau(x) \in [Q^L_\tau(x), Q^U_\tau(x)], x \in \{0,1\}\}$

equivalently satisfies

$$\delta^*_{mmr} \in \arg\max_{\delta \in \mathcal{D}} E[\delta(X)\bar{Q}_\tau(X)], \tag{2.5}$$

where

$$\begin{aligned}\bar{Q}_\tau(x) = Q^U_\tau(x)1\{Q^L_\tau(x) \geq 0\} + Q^L_\tau(x)1\{Q^U_\tau(x) \leq 0\} \\ + \left(Q^U_\tau(x) + Q^L_\tau(x)\right) 1\{Q^L_\tau(x) < 0 < Q^U_\tau(x)\}.\end{aligned}$$

Also, we can show $\delta^*_{mmw} \in \arg\max_{\delta \in \mathcal{D}} E[\delta(X)Q^L_\tau(X)]$. In general, finding the optimal δ for (2.5) does not yield a closed-form expression when the policy class $\mathcal{D}$ is constrained. Additionally, solving $\max_{\delta \in \mathcal{D}} E[\delta(X)\bar{Q}_\tau(X)]$ proves to be a challenging task as $\bar{Q}_\tau(\cdot)$ incorporates an indicator function. Nonetheless, allowing the policy class to be constrained is important because the PM may prefer a more parsimonious rule (e.g., a linear rule) or be limited by certain institutional constraints. Following Zhao et al. (2012), we consider a convex and continuous relaxation of (2.5) by utilizing the hinge loss function $\phi(t) = \max(1 - t, 0)$ and introducing a regularization term. This is done in Section 4.2 below. The consistency of hinge loss is shown even when the class of δ is restricted (Kitagawa et al. (2021)).

3 Possible Identifying Assumptions

We now provide a menu of identifying assumptions that researchers may want to consider imposing to shrink $\mathcal{F}$ (i.e., the identified set for the joint distribution of (Y_1, Y_0) conditional on X). This would consequently tighten $[Q^L_\tau(x), Q^U_\tau(x)]$ (i.e., the bounds on the conditional QoTE), and sometimes reduce it to a singleton, achieving point identification. First, there

is rectangular, which implies that, for example,

$$\begin{aligned}\min_{F_{Y_1,Y_0|X}} E[\delta(X)Q_\tau(X)] &= \min_{F_{Y_1,Y_0|X}} [p_1\delta(1)Q_\tau(1) + p_0\delta(0)Q_\tau(0)] \\ &= p_1\delta(1) \min_{F_{Y_1,Y_0|X}} Q_\tau(1) + p_0\delta(0) \min_{F_{Y_1,Y_0|X}} Q_\tau(0) = E[\delta(X) \min_{F_{Y_1,Y_0|X}} Q_\tau(X)].\end{aligned}$$

are ways to identify the marginal distribution of Y_d. The most obvious approach is to impose conditional independence.

Assumption CI (Conditional Independence). *For $d \in \{0, 1\}$, $Y_d \perp D|X$.*

An clear example where this assumption holds is when data from randomized experiments are available. In general, one can argue that the treatment is exogenous after adequately controlling for covariates. Alternatives to Assumption CI, such as panel quantile regression models (Chernozhukov et al. (2013)), can be used to identify $Q_\tau(Y_d|X)$.

The identification of the marginal distribution of Y_d yields bounds on the QoTE, $Q_\tau(Y_1 - Y_0|X = x)$. The best-known sharp bounds on the QoTE are derived by Makarov (1982) and Williamson and Downs (1990) without imposing further restrictions on the data-generating mechanism. Henceforth, we will refer to these bounds as the *Makarov bounds*. We describe them here by trivially extending Lemma 2.3 in Fan and Park (2010) to incorporate covariates.

Proposition 3.1. *For $0 \le \tau \le 1$, $Q^L_\tau(x) \le Q_\tau(Y_1 - Y_0|X = x) \le Q^U_\tau(x)$ where*

$$Q^L_\tau(x) = \begin{cases} \inf_{u\in[\tau,1]}[Q_u(Y_1|X = x) - Q_{u-q}(Y_0|X = x)] & \text{if } q \neq 0 \\ Q_0(Y_1|X = x) - Q_1(Y_0|X = x) & \text{if } q = 0, \end{cases}$$

$$Q^U_\tau(x) = \begin{cases} \sup_{u\in[0,\tau]}[Q_u(Y_1|X = x) - Q_{1+u-q}(Y_0|X = x)] & \text{if } q \neq 1 \\ Q_1(Y_1|X = x) - Q_0(Y_0|X = x) & \text{if } q = 1. \end{cases}$$

Note that the Makarov bounds are *not* achieved at the Fréchet-Hoeffding bounds for the joint distribution of (Y_1, Y_0) (Fan and Park (2010, Lemma 2.1)). It is known that the Makarov bounds tend to be uninformative, which may result in the subsequent treatment allocation decisions being similarly uninformative. We now consider a range of identifying assumptions that can be used to yield tighter bounds, leading to more informative decisions.

Assumption SI (Stochastic Increasing). *For $x \in \mathcal{X}$, $P[Y_1 \le y_1|Y_0 = \cdot, X = x]$ and $P[Y_0 \le y_0|Y_1 = \cdot, X = x]$ are non-increasing.*

Assumption SI imposes positive dependence between Y_1 and Y_0 (Joe (2014)). This assumption makes sense when individuals with high Y_1 (e.g., potential health with the medical treatment) tend to have high Y_0 (e.g., potential health without the medical treatment) and vice versa. Due to its plausibility in many settings, we consider this assumption as our leading one in later analyses.[5] While maintaining Assumption CI, Assumption SI is helpful to obtain more informative bounds on the conditional QoTE. For example, Frandsen and Lefgren (2021) derive bounds on the (unconditional) distribution of treatment effects under SI. Instead of assuming positive dependence between Y_1 and Y_0, one may want to impose stochastic dominance of Y_d between treatment and control groups or stochastic dominance between Y_1 and Y_0 for each subgroup:

Assumption SD (Stochastic Dominance). *For $x \in \mathcal{X}$, either (i) $P[Y_d \leq y|D = 1, X = x] \leq P[Y_d \leq y|D = 0, X = x]$; or (ii) $P[Y_1 \leq y|D = d, X = x] \leq P[Y_0 \leq y|D = d, X = x]$.*

Either under Assumption CI or the existence of instrumental variables (IVs), Assumption SD(i) or SD(ii) can be used to narrow the bounds on the distribution of treatment effects (Blundell et al. (2007), Lee (2021)) and thus on the QoTE.

Next, we present assumptions that help point-identify the conditional QoTE.

Assumption CI2 (Joint Conditional Independence). *$(Y_1, Y_0) \perp D|X$.*

This assumption is stronger than Assumption CI.

Assumption DC (Deconvolution). *$Y_1 - Y_0 \perp Y_0|X$.*

Heckman and Smith (1995) show how Assumption DC can be useful to point identify $F_{Y_1,Y_0|X}$ when combined with Assumption CI2. To see this, let $\Delta \equiv Y_1 - Y_0$ for convenience. First note that $Y = Y_0 + \Delta D$. Under CI2, $F_{Y_0}(y_0|X)$ and $F_{Y_1}(y_1|X) = F_{Y_1}(y_0 + \Delta|X)$ are identified by $F_Y(y|X, D = 0)$ and $F_Y(y|X, D = 1)$, respectively. Then under DC, the densities satisfy $f_{Y_1}(y_1|X) = f_\Delta(\Delta|X) * f_{Y_0}(y_0|X)$, where "$*$" denotes convolution. Then

[5] One can conversely impose stochastic decreasingness although it may be easier to find contexts in which stochastic increasingness is more plausible.

the characteristic functions satisfy $E[e^{itY_1}|X] = E[e^{it\Delta}|X]E[e^{itY_0}|X]$ or $E[e^{it\Delta}|X] = \frac{E[e^{itY_1}|X]}{E[e^{itY_0}|X]}$ where the right-hand-side terms are known. Therefore, by the inversion theorem, we can recover $f_\Delta(\Delta|X)$. Note that we can also recover the full joint distribution of (Y_1, Y_0) given X. Interested readers can refer to Section 2.5.5 of Abbring and Heckman (2007), which shows that this assumption relates to a normal random coefficient model. The next set of assumptions explicitly posits that the treatment selection is determined by the net gain from the treatment.

Assumption RY (Roy Model)**.** *$D = 1\{Y_1 \geq Y_0\}$ and $X = (X_0, X_1, X_c)$ where (i) $Y_1 = g_1(X_1, X_c) + U_1$ and $Y_0 = g_0(X_0, X_c) + U_0$, (ii) $(U_0, U_1) \perp (X_0, X_1, X_c)$, (iii) (U_0, U_1) are absolutely continuous with Supp$(U_0, U_1) = \mathbb{R}^2$, (iv) for each X_c and $d \in \{0, 1\}$, $g_d(X_d, X_c) : \mathbb{R}^{k_d} \to \mathbb{R}$ for all X_{1-d}, Supp$(g_d(X_d, X_c)|X_c, X_{1-d}) = \mathbb{R}$ for all X_c, X_{1-d}, and Supp$(X_d|X_{1-d}, X_c) =$ Supp$(X_d) = \mathbb{R}$ for all X_c, X_{1-d}, and (v) for $d \in \{0, 1\}$, U_d has zero median.*

Under Assumption RY, g_0, g_1, and F_{U_0,U_1} are point identified (Heckman and Smith (1995, Theorem A-1)); see Heckman and Honore (1990) for Gaussian case.

Assumption RY2 (Extended Roy Model)**.** *$D = 1\{Y_1 \geq h(Y_0, X, Z)\}$ where (i) $(Y_0, Y_1) \perp Z|X$, (ii) Supp$(Y_0, Y_1|X) = \mathbb{R}^2$, (iii) $h(y_0, x, \cdot)$ and $h(\cdot, x, z)$ are strictly increasing for any (y_0, x, z), and (iv) $h(y_0, x, \cdot)$ is differentiable.*

Under Assumption RY2, Lee and Park (2023) show that $F_{Y_1,Y_0|X}(y_1, y_0|x)$ is point identified for $(y_1, y_0) \in \mathcal{H}(x)$ where $\mathcal{H}(x) \equiv \{(y_1, y_0) \in \mathbb{R}^2 : y_1 = h(y_0, x, z)$ for some $z \in \text{Supp}(Z|X = x)\}$. Its implication for our purpose is that $Q_\tau(Y_1 - Y_0|X = x)$ is identified if and only if $\{(y_1, y_0) \in \mathbb{R}^2 : y_1 - y_0 = Q_\tau(Y_1 - Y_0|X = x)\} \subseteq \mathcal{H}(x)$. The next assumption is a special instance of Assumption SI.

Assumption RI (Rank Invariance)**.** *For $d \in \{0, 1\}$, $Y_d = m_d(X, U_d)$ where $m_d(x, \cdot)$ is strictly increasing and $U_d|X = x$ is absolutely continuous and satisfies $U_1|_{X=x} = U_0|_{X=x}$ for given $x \in \mathcal{X}$.*

Heckman et al. (1997) and Chernozhukov and Hansen (2005) show the identifying power of Assumption RI. This assumption essentially restricts heterogeneity by holding the ranks in Y_1 and Y_0 the same. This implies that, under this assumption, the QTE can be interpreted as the difference between Y_1 and Y_0 for the same individual. Yet, the QTE is *not* identical to the OoTE even under this assumption. Moreover, Assumption RI implies Assumption SI because, suppressing X, $\Pr[Y_1 \leq y_1 | Y_0 = y_0] = \Pr[m_1(U) \leq y_1 | m_0(U) = y_0] = \Pr[U \leq m_1^{-1}(y_1) | U = m_0^{-1}(y_0)]$ and thus the probability is 1 when $y_0 \leq m_0(m_1^{-1}(y_1))$ and 0 otherwise. Under Assumption RI, $F_1(\cdot|x)$ and $F_0(\cdot|x)$ are strictly monotone for all x, and $\phi_0(\cdot|x) = F_0^{-1}(F_1(\cdot|x)|x)$ maps the outcome of a treated individual with covariates $X = x$ (which is identified under CI) into their untreated outcome (which is unobserved). Similarly, $\phi_1(\cdot|x) = F_1^{-1}(F_0(\cdot|x)|x)$ maps the outcome of an untreated individual with covariates $X = x$ into their treated outcome.[6] Then, $F_{\Delta|X}(\delta)$ is point identified by $F_{\Delta|X}(\delta) = \Pr[D(Y - \phi_0(Y|X)) + (1-D)(\phi_1(Y|X) - Y) \leq \delta | X]$. More generally, Heckman et al. (1997) consider Markov kernels M and $\tilde{M}$ so that $F_{Y_1|X}(y_1) = \int M(y_1, y_0|X) dF_{Y_0|X}(y_0)$ and $F_{Y_0|X}(y_0) = \int \tilde{M}(y_1, y_0|X) dF_{Y_1|X}(y_1)$. Also see Vuong and Xu (2017) for the case of endogenous treatment with IVs. Abbring and Heckman (2007, Section 2.5.3) also consider perfect *negative* dependence.[7]

Assumption SY (Symmetric Distribution)**.** *The distribution of* $Y_1 - Y_0|X$ *is symmetric.*

Under this assumption, $Q_{0.5}(Y_1 - Y_0|X) = E[Y_1 - Y_0|X]$, which is point-identified under Assumption CI. Other possible assumptions for point identification can be found in Abbring and Heckman (2007).

4 Theoretical Properties of Estimated Policy

Let $\Delta \equiv Y_1 - Y_0$ for notational simplicity. Focusing on the optimal policy δ^*_{mmr} based on the minimax criterion, we provide theoretical guarantees for the estimated policy. The

[6] These mappings are called counterfactual mappings (Vuong and Xu (2017)).

[7] Related to the previous footnote, we can also consider perfect negative dependence by $U_1|_{X=x} = -U_0|_{X=x}$.

policy can be readily estimated once the bounds $[Q^L_\tau(X), Q^U_\tau(X)]$ on $Q_\tau(\Delta|X)$ are estimated using parametric or nonparametric methods with the sample of (Y, D, X). We denote such estimators as $\hat{Q}^L_\tau(X)$ and $\hat{Q}^U_\tau(X)$ and corresponding estimated stochastic and deterministic policies as $\hat{\delta}^{stoch}$ and $\hat{\delta}^{determ}$, respectively. The theory includes the case of point identification as a special case in which $Q_\tau(\Delta|X) = Q^L_\tau(X) = Q^U_\tau(X)$.

Recall that our objective function is

$$V(\delta) \equiv E[\delta(X) Q_\tau(\Delta|X)]$$

and the regret of this "classification" is

$$R(\delta) \equiv V(\delta^\dagger) - V(\delta) = E[|Q_\tau(\Delta|X)| 1\{\delta(X) \neq sign(Q_\tau(\Delta|X))\}],$$

where $\delta^\dagger(X) = 1\{Q_\tau(\Delta|X) \geq 0\}$ and $sign(q) = 1$ when $q \geq 0$ and $sign(q) = 0$ when $q < 0$. Note that $R(\delta)$ is generally not point-identified and thus we define maximum regret as

$$\bar{R}(\delta) \equiv \max_{Q_\tau(\Delta|\cdot) \in [Q^L_\tau(\cdot), Q^U_\tau(\cdot)]} E[|Q_\tau(\Delta|X)| 1\{\delta(X) \neq sign(Q_\tau(\Delta|X))\}].$$

The maximum regret can be expressed in different ways, which are useful in the analysis below.

Lemma 4.1. *Suppose Assumption RC hold. For a stochastic or deterministic rule δ, the maximum regret can be expressed as*

$$\bar{R}(\delta) = E\left[\max\left\{[1 - \delta(X)] \max(Q^U_\tau(X), 0), \delta(X) \max(-Q^L_\tau(X), 0)\right\}\right] \tag{4.1}$$

$$= -E[\delta(X) \bar{Q}_\tau(X)] + E\left[Q^U_\tau(X) \left(1\{Q^L_\tau(X) \geq 0\} + 1\{Q^U_\tau(X) > 0\}\right)\right] \tag{4.2}$$

where

$$\bar{Q}_\tau(X) = Q^U_\tau(X)1\{Q^L_\tau(X) \geq 0\} + Q^L_\tau(X)1\{Q^U_\tau(X) \leq 0\} \\ + (Q^U_\tau(X) + Q^L_\tau(X))1\{Q^L_\tau(X) < 0 < Q^U_\tau(X)\}.$$

For a deterministic rule δ*,*

$$\bar{R}(\delta) = E[|\bar{Q}_\tau(X)|1\{\delta(X) \neq sign(\bar{Q}_\tau(X))\}] \\ + E\Big[\min(Q^U_\tau(X), -Q^L_\tau(X))1\{Q^L_\tau(X) < 0 < Q^U_\tau(X)\}\Big]. \tag{4.3}$$

Note that (4.2) is used in expressing (2.5). Below, (4.1) is used in Section 4.1 and (4.3) in Section 4.2.

Proof. Fix x and let $\bar{R}(\delta; x) \equiv \max_{Q_\tau(\Delta|x) \in [Q^L_\tau(x), Q^U_\tau(x)]} |Q_\tau(\Delta|x)|1\{\delta(x) \neq sign(Q_\tau(\Delta|x))\}$. If $0 \leq Q^L_\tau(x) < Q^U_\tau(x)$,

$$\bar{R}(\delta; x) = |Q^U_\tau(x)|1\{\delta(x) \neq 1\} = Q^U_\tau(x)1\{\delta(x) \neq 1\}$$

and if $0 \geq Q^U_\tau(x) > Q^L_\tau(x)$,

$$\bar{R}(\delta; x) = |Q^L_\tau(x)|1\{\delta(x) \neq 0\} = -Q^L_\tau(x)1\{\delta(x) \neq 0\}.$$

Finally, if $Q^L_\tau(x) < 0 < Q^U_\tau(x)$,

$$\bar{R}(\delta; x) = |Q^U_\tau(x)|1\{\delta(x) \neq 1\} + |Q^L_\tau(x)|1\{\delta(x) \neq 0\} \\ = Q^U_\tau(x)1\{\delta(x) \neq 1\} - Q^L_\tau(x)1\{\delta(x) \neq 0\}.$$

Therefore, by the law of iterated expectation and Assumption RC,

$$\bar{R}(\delta) = E\Big[Q_\tau^U(X)[1-\delta(X)]1\{Q_\tau^L(X) \geq 0\} - Q_\tau^L(X)\delta(X)1\{Q_\tau^U(X) \leq 0\}$$
$$+ Q_\tau^U(X)[1-\delta(X)]1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\} - Q_\tau^L(X)\delta(X)1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\Big]. \tag{4.4}$$

From (4.4), it is straightforward to show (4.1) and (4.2). To show (4.3), note that, if $Q_\tau^L(x) < 0 < Q_\tau^U(x)$,

$$\begin{aligned}
&Q_\tau^U(x)1\{\delta(x) \neq 1\} - Q_\tau^L(x)1\{\delta(x) \neq 0\} \\
&= |Q_\tau^U(x) + Q_\tau^L(x)|1\{\delta(x) \neq sign(Q_\tau^U(x) + Q_\tau^L(x))\} + \min(Q_\tau^U(x), -Q_\tau^L(x)).
\end{aligned}$$

This can be shown by inspecting each case of $\delta(x) = 1$ and $\delta(x) = 0$. If $0 \leq Q_\tau^L(x) < Q_\tau^U(x)$, it is obvious that $Q_\tau^U(x)1\{\delta(x) \neq 1\} = Q_\tau^U(x)1\{\delta(x) \neq sign(Q_\tau^U(x))\}$ and similarly for the case of $0 \geq Q_\tau^U(x) > Q_\tau^L(x)$. □

Now, we provide asymptotic bounds on these regrets evaluated at the estimated stochastic and deterministic policies when $\mathcal{D}$ is unconstrained and constrained.

4.1 Regret Bounds with Unconstrained Policy Class

For this part, we assume that we are equipped with the consistent estimators for $Q_\tau^L(X)$ and $Q_\tau^U(X)$.

Assumption EST. *$Q_\tau(\Delta|X)$ is bounded almost surely and*

$$\begin{aligned}
\hat{Q}_\tau^L(X) - Q_\tau^L(X) &= o_p(1), \\
\hat{Q}_\tau^U(X) - Q_\tau^U(X) &= o_p(1).
\end{aligned}$$

When $Q_\tau^L(X)$ and $Q_\tau^U(X)$ are known functions of $F_{Y_1|X}$ and $F_{Y_0|X}$, Assumption EST is implied from the consistency of $\hat{F}_{Y_1|X}$ and $\hat{F}_{Y_0|X}$ by the continuous mapping theorem; see

Section 5 for the case of bounds that are computationally derived. Let δ^{stoch} and δ^{determ} are the optimal policies that minimize $\bar{R}(\delta)$ when δ is stochastic and deterministic policies, respectively. Let $\hat{\delta}^{stoch}$ and $\hat{\delta}^{determ}$ are the estimates of δ^{stoch} and δ^{determ}, respectively.

Theorem 4.1. *Suppose Assumption EST holds and $|Y| \leq M$ for some constant M. The regret of $\hat{\delta}^{stoch}$ is bounded by*

$$R(\hat{\delta}^{stoch}) \leq E\left[\frac{Q_\tau^L(X)Q_\tau^U(X)}{Q_\tau^L(X) - Q_\tau^U(X)} 1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\right] + o_p(1),$$

where the ratio is defined to be 0 whenever its denominator is 0. The regret regret of $\hat{\delta}^{determ}$ is bounded by

$$R(\hat{\delta}^{determ}) \leq E\left[\min(\max(Q_\tau^U(X), 0), \max(-Q_\tau^L(X), 0))\right] + o_p(1).$$

The leading term in each asymptotic regret bound collapses to zero when either (i) the bounds on $Q_\tau(\Delta|X)$ exclude zero almost surely or (ii) $Q_\tau(\Delta|X)$ is point-identified. These are the situations in which we can identify the sign of $Q_\tau(\Delta|X)$. Recalling $\delta^\dagger(X) = 1\{Q_\tau(\Delta|X) \geq 0\}$, this is enough to achieve consistency $R \to 0$ as the second term in each regret bound is the sampling error. In general, the leading term becomes larger as the endpoints $[Q_\tau^L(X), Q_\tau^U(X)]$ are farther away from zero, which is intuitive. An immediate corollary of Theorem 4.1 establishes the bound for the regret averaged over the sample of estimated policies. Let $\mathbb{E}_n$ denote the expectation over the sample of (Y, D, X).

Corollary 4.1. *Suppose Assumption EST holds. Then,*

$$\mathbb{E}_n\left[R(\hat{\delta}^{stoch})\right] \leq E\left[\frac{Q_\tau^L(X)Q_\tau^U(X)}{Q_\tau^L(X) - Q_\tau^U(X)} 1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\right] + o_p(1),$$

where the ratio is defined to be 0 whenever its denominator is 0, and

$$\mathbb{E}_n\left[R(\hat{\delta}^{determ})\right] \leq E\left[\min(\max(Q_\tau^U(X), 0), \max(-Q_\tau^L(X), 0))\right] + o_p(1).$$

Proof. [Proof of Theorem 4.1]Given the expression (4.1), a simple calculation yields

$$\delta^{stoch}(x) = \begin{cases} 1 & \text{if } Q_\tau^L(x) > 0, \\ 0 & \text{if } Q_\tau^U(x) < 0, \\ \frac{Q_\tau^U(x)}{Q_\tau^U(x) - Q_\tau^L(x)} & \text{if } Q_\tau^L(x) < 0 < Q_\tau^U(x). \end{cases}$$

Therefore, the maximum risk of δ^{stoch} is

$$\bar{R}(\delta^{stoch}) = E\left[\frac{Q_\tau^L(X)Q_\tau^U(X)}{Q_\tau^L(X) - Q_\tau^U(X)} 1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\right].$$

Without loss of generality, we suppose $X \in [0,1]^p$. As shown below, $R(\delta^{stoch}) \leq \bar{R}(\delta^{stoch})$. Then

$$R(\hat{\delta}^{stoch}) \leq E\left[\frac{Q_\tau^L(X)Q_\tau^U(X)}{Q_\tau^L(X) - Q_\tau^U(X)} 1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\right] + o_p(1),$$

because $V(\hat{\delta}^{stoch}) - V(\delta^{stoch}) = o_p(1)$, which can be shown as follows:

$$V(\hat{\delta}^{stoch}) - V(\delta^{stoch}) = E[(\hat{A}(X) - A(X))Q(\Delta|X)] = o_p(1)O(1) = o_p(1)$$

and $E[\hat{A}(X) - A(X)] = E[\hat{\delta}^{stoch}(X) - \delta^{stoch}(X)] = o_p(1)$ by the definition of $A(X)$ in Section A and the definition of $\hat{A}(X)$ with the estimated Bernoulli probability $\hat{\delta}^{stoch}(X)$.

Again, given the expression (4.1), a simple calculation yields

$$\delta^{determ}(x) = \begin{cases} 1 & \text{if } Q_\tau^L(x) > 0, \\ 0 & \text{if } Q_\tau^U(x) < 0, \\ 1 & \text{if } Q_\tau^L(x) < 0 < Q_\tau^U(x) \text{ and } |Q_\tau^L(x)| < |Q_\tau^U(x)|, \\ 0 & \text{if } Q_\tau^L(x) < 0 < Q_\tau^U(x) \text{ and } |Q_\tau^L(x)| > |Q_\tau^U(x)|. \end{cases}$$

Therefore, the maximum risk of δ^{determ} is

$$\bar{R}(\delta^{determ}) = E[\min(\max(Q^U_\tau(X), 0), \max(-Q^L_\tau(X), 0))].$$

Again, as shown below, $R(\delta^{stoch}) \leq \bar{R}(\delta^{stoch})$. Then

$$\bar{R}(\hat{\delta}^{determ}) \leq E[\min(\max(Q^U_\tau(X), 0), \max(-Q^L_\tau(X), 0))] + o_p(1),$$

because $V(\hat{\delta}^{determ}) - V(\delta^{determ}) = o_p(1)$, which can be shown as follows:

$$\begin{aligned}
&V(\hat{\delta}^{determ}) - V(\delta^{determ}) \\
=&E[1(\hat{\delta}^{determ}(X) = \delta^{determ}(X)) \times 0 + 1(\hat{\delta}^{determ}(X) = 1, \delta^{determ}(X) = 0) \times Q(\Delta|X) \\
&- 1(\hat{\delta}^{determ}(X) = 0, \delta^{determ}(X) = 1) \times Q(\Delta|X)] \\
=&0 + o_p(1)O(1) = o_p(1),
\end{aligned}$$

and $E[1(\hat{\delta}^{determ}(X) \neq \delta^{determ}(X))] = P[\hat{\delta}^{determ}(X) \neq \delta^{determ}(X)] = o_p(1)$ by the definition of $\hat{\delta}^{determ}$ and δ^{determ}.

Finally, we are remained to show that $R(\delta) \leq \bar{R}(\delta)$ for any δ. To see this inequality, note that

$$\begin{aligned}
V(\delta^\dagger) - V(\delta) &= E[\delta^\dagger(X)Q_\tau(\Delta|X)] - E[\delta(X)Q_\tau(\Delta|X)] \\
&= E[|Q_\tau(\Delta|X)|1\{\delta(X) \neq sign(Q_\tau(\Delta|X))\}],
\end{aligned}$$

and recall $\bar{R}(\delta) \equiv \max_{Q_\tau(\Delta|\cdot)\in[Q^L_\tau(\cdot), Q^U_\tau(\cdot)]} E[|Q_\tau(\Delta|X)|1\{\delta(X) \neq sign(Q_\tau(\Delta|X))\}]$. □

4.2 Regret Bounds with Constrained Policy Class

As mentioned, allowing for a constrained policy class is crucial for practical and institutional reasons. Our proposed method readily extends to a scenario in which the policy class $\mathcal{D}$ is constrained. Define the estimator of $\bar{Q}_\tau(\cdot)$ as

$$\widehat{\bar{Q}}_\tau(X) \equiv \hat{Q}_\tau^U(X)1\{\hat{Q}_\tau^U(X) \geq 0\} + \hat{Q}_\tau^L(X)1\{\hat{Q}_\tau^L(X) \leq 0\}$$

by noting that $\bar{Q}_\tau(x)$ also satisfies $\bar{Q}_\tau(x) = Q_\tau^U(x)1\{Q_\tau^U(x) \geq 0\} + Q_\tau^L(x)1\{Q_\tau^L(x) \leq 0\}$. We assume that the consistent estimators $\hat{Q}_\tau^L(X)$ and $\hat{Q}_\tau^U(X)$ are consistent with the specified rate of convergence.

Assumption EST2. *$Q_\tau(\Delta|X)$ is bounded almost surely and*

$$\begin{aligned}
\hat{Q}_\tau^L(X) - Q_\tau^L(X) &= O_p(n^{-\alpha}), \\
\hat{Q}_\tau^U(X) - Q_\tau^U(X) &= O_p(n^{-\alpha})
\end{aligned}$$

for some $\alpha > 0$.

To overcome the computational problem of obtaining δ_{mmr}^*, we adopt the outcome weighted learning framework (Zhao et al. (2012)). We are interested in finding a decision function $f : \mathcal{X} \to \mathbb{R}$ such that $\delta(x) = 1\{f(x) \geq 0\}$. Note that by (4.3), we have

$$\begin{aligned}
\bar{R}(1\{f(\cdot) \geq 0\}) = {} & E[|\bar{Q}_\tau(X)|1\{sign(f(X)) \neq sign(\bar{Q}_\tau(X))\}] \\
& + E\Big[\min(Q_\tau^U(X), -Q_\tau^L(X))1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\Big].
\end{aligned}$$

Accordingly, we define the surrogate regret as

$$\begin{aligned}
\bar{R}^S(f) = {} & E[|\bar{Q}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))f(X)\}] \\
& + E\Big[\min(Q_\tau^U(X), -Q_\tau^L(X))1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\Big].
\end{aligned}$$

Motivated by this expression, let $\hat{f}$ be the ML estimator of f from the following problem:

$$\hat{f} = \arg\min_{f \in \mathcal{H}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \left| \widehat{\bar{Q}}_\tau(X_i) \right| \phi\{sign(\widehat{\bar{Q}}_\tau(X_i))f(X_i)\} + \lambda_n ||f||^2 \right\}, \tag{4.5}$$

where $\phi(t) = \max\{1-t, 0\}$ is the hinge loss, λ_n is the regularization parameter, and $||\cdot||$ is the norm in a function space. We focus on the reproducing kernel Hilbert space (RKHS) $\mathcal{H}_k$ associated with Gaussian radial basis function kernels $k(x,z) = \exp(-\sigma_n^2||x-z||^2)$. By Theorem 2.1 of Steinwart and Scovel (2007), the complexity of $\mathcal{H}_k$ in terms of the covering number satisfies

$$\sup_{P_n} \log N\{B_{\mathcal{H}_k}, \epsilon, L_2(P_n)\} \leq c_n \epsilon^{-p},$$

where P_n is the distribution of (Y, D, X), $c_n = c_{p,\delta,d}\sigma_n^{(1-p/2)(1+\delta)d}$, $B_{\mathcal{H}_k}$ is the closed unit ball of $\mathcal{H}_k$, $p \in (0, 2]$, $\delta > 0$, and $c_{p,\delta,d}$ is a constant. Define the approximation error function as

$$a(\lambda_n) = \inf_{f\in\mathcal{H}_k} E[|\bar{Q}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))f(X)\}+\lambda_n||f||^2]-\inf_f E[|\bar{Q}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))f(X)\},$$

where the second infimum is over the unrestricted space of f. Note that $a(\lambda_n)$ goes to zero if the RKHS is rich enough. The following theorem establishes the asymptotic bound on $\bar{R}(f) \equiv \bar{R}(1\{f(x) \geq 0\})$. The asymptotic bound on the true regret can be similarly obtained.

Theorem 4.2. *Under Assumption EST2, and suppose that $\lambda_n = o(1)$ and $\lambda_n n^{\min(2\alpha,1)} \to \infty$, we have that, with probability larger than $1 - \exp(-2\eta)$,*

$$\bar{R}(\hat{f}) \leq \inf_f \bar{R}(f) + a(\lambda_n) + O_p(n^{-\alpha}\lambda_n^{-1/2}) + M_p c_n^{\frac{2}{p+2}} n^{-\frac{2}{p+2}} \left(\lambda_n^{-\frac{2}{p+2}} + \lambda_n^{-1/2}\right) + \frac{K\eta}{n\lambda_n}(1 + 2\lambda_n^{1/2}),$$

where M_p and K are constants.

The first term of the regret bound is $E\Big[\min(Q_\tau^U(X), -Q_\tau^L(X))1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\Big]$ because f is not restricted and the following f

$$f(x) = \begin{cases} 1 & \text{if } Q_\tau^L(X) > 0 \\ 0 & \text{if } Q_\tau^U(X) < 0 \\ sign(Q_\tau^L(X) + Q_\tau^U(X)) & \text{if } Q_\tau^L(X) < 0 < Q_\tau^U(X) \end{cases}$$

satisfies $\inf_f \bar{R}(f) = \bar{R}(f) = E\Big[\min(Q_\tau^U(X), -Q_\tau^L(X))1\{Q_\tau^L(X) < 0 < Q_\tau^U(X)\}\Big]$. Note that this term coincides with the leading term (i.e., $E\left[\min(\max(Q_\tau^U(X), 0), \max(-Q_\tau^L(X), 0))\right]$) derived in Theorem 4.1 for the deterministic rule. The second term is the approximation error due to using the RKHS. The third term is the estimation error in estimating the bounds. The rest of the terms are statistical errors in estimating the policy.

Proof. [Proof of Theorem 4.2] By Theorem 3.2 of Zhao et al. (2012), we have that

$$\bar{R}(\hat{f}) - \inf_f \bar{R}(f) \leq \bar{R}^S(\hat{f}) - \inf_f \bar{R}^S(f).$$

We essentially need to bound the right-hand side. Let

$$\tilde{f} = \arg\min_{f \in \mathcal{H}_k} E[|\bar{Q}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))f(X)\} + \lambda_n ||f||^2].$$

Note that

$$
\begin{aligned}
&\bar{R}^S(\hat{f}) - \inf_f \bar{R}^S(f) \\
=&\bar{R}^S(\hat{f}) - \bar{R}^S(\tilde{f}) + \bar{R}^S(\tilde{f}) - \inf_{f \in \mathcal{H}_k}[\bar{R}^S(f) + \lambda_n ||f||^2] + \inf_{f \in \mathcal{H}_k}[\bar{R}^S(f) + \lambda_n ||f||^2] - \inf_f \bar{R}^S(f) \\
\leq& \inf_{f \in \mathcal{H}_k}[\bar{R}^S(f) + \lambda_n ||f||^2] - \inf_f \bar{R}^S(f) \\
&- \frac{1}{n}\sum_{i=1}^n [|\widehat{\bar{Q}}_\tau(X_i)|\phi\{sign(\hat{Q}_\tau(X_i))\hat{f}(X_i)\} + \lambda_n ||\hat{f}||^2 - |\widehat{\bar{Q}}_\tau(X_i)|\phi\{sign(\widehat{\bar{Q}}_\tau(X_i))\tilde{f}(X_i)\} - \lambda_n ||\tilde{f}||^2] \\
&+ E[|\widehat{\bar{Q}}_\tau(X)|\phi\{sign(\hat{Q}_\tau(X))\hat{f}(X)\} + \lambda_n ||\hat{f}||^2 - |\widehat{\bar{Q}}_\tau(X)|\phi\{sign(\widehat{\bar{Q}}_\tau(X))\tilde{f}(X)\} - \lambda_n ||\tilde{f}||^2] \\
&+ E[|\widehat{\bar{Q}}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))\hat{f}(X)\}] - E[|\widehat{\bar{Q}}_\tau(X)|\phi\{sign(\widehat{\bar{Q}}_\tau(X))\hat{f}(X)\}] \\
&+ E[|\widehat{\bar{Q}}_\tau(X)|\phi\{sign(\hat{Q}_\tau(X))\tilde{f}(X)\}] - E[|\bar{Q}_\tau(X)|\phi\{sign(\bar{Q}_\tau(X))\tilde{f}(X)\}].
\end{aligned}
$$

Following the proof of Theorem 1 of Zhao et al. (2015), we have that

$$
\begin{aligned}
&\bar{R}^S(\hat{f}) - \inf_f \bar{R}^S(f) \\
\leq & a(\lambda_n) + M_p c_n^{\frac{2}{p+2}} (n\lambda_n)^{-\frac{2}{p+2}} + M_p \lambda_n^{-1/2} c_n^{\frac{2}{p+2}} n^{-\frac{2}{p+2}} + K\eta \frac{1}{n\lambda_n} + 2K\eta \frac{1}{n\lambda_n^{1/2}} + O_p(n^{-\alpha}\lambda_n^{-1/2})
\end{aligned}
$$

with probability larger than $1 - 2\exp(-\eta)$. □

5 Calculating Bounds

When $Q_\tau(\Delta|X)$ is partially identified, we need a practical way of calculating its bounds

$$[Q_\tau^L(x), Q_\tau^U(x)] = \{Q_\tau(\Delta|X = x) : F_{Y_1,Y_0|X} \in \mathcal{F}\}.$$

Unlike the Makarov bounds, the closed-form expression of the bounds is not always available especially under Assumption SI. Therefore, it is fruitful to have a systematic procedure of calculating the bounds. To this end, let $C(u_1, u_2|X)$ be the copula for $(U_1, U_2) \equiv (F_{Y_1}(Y_1), F_{Y_0}(Y_0))$ conditional on X. By Sklar's Theorem, $C(u_1, u_2|X) = F_{Y_1,Y_0|X}(Q_{u_1}(Y_1|X), Q_{u_2}(Y_0|X))$. Then, it satisfies

$$P[Y_1 - Y_0 \leq t|X] = \int 1\{Q_{u_1}(Y_1|X) - Q_{u_2}(Y_0|X) \leq t\} dC(u_1, u_2|X).$$

Therefore, we can calculate the lower and upper bounds on the distribution of $\Delta|X$ by

$$F^L_{\Delta|X}(t) = \inf_{C(\cdot,\cdot|X)\in\mathcal{C}} \int 1\{Q_{u_1}(Y_1|X) - Q_{u_2}(Y_0|X) \leq t\} dC(u, v|X), \tag{5.1}$$

$$F^U_{\Delta|X}(t) = \sup_{C(\cdot,\cdot|X)\in\mathcal{C}} \int 1\{Q_{u_1}(Y_0|X) - Q_{u_2}(Y_0|X) \leq t\} dC(u, v|X), \tag{5.2}$$

where $\mathcal{C}$ is the class of copulas $C(\cdot, \cdot|X = x)$ restricted by identifying assumptions. Note that (5.1) and (5.2) can be viewed as the (constrained version of the) Monge-Kantorovich problem of finding the optimal coupling of marginal distributions in the optimal transport theory

(Villani (2009)). Then, for τ-quantile Q_τ of Δ, we can obtain its lower and upper bounds as $Q^L_\tau(X) = F^{U,-1}_{\Delta|X}(\tau)$ and $Q^U_\tau(X) = F^{L,-1}_{\Delta|X}(\tau)$, where the inverse denotes the generalized inverse. In practice, (5.1)–(5.2) are infinite dimensional programs, and thus infeasible. To transform them into a linear program, we propose to approximate $C(u, v|x)$ using the Bernstein copula $C_B(u, v|x)$ (Sancetta and Satchell (2004)).

Definition 5.1 (Bernstein Copula). *For $j \in \{1, 2\}$, let $P^{m_j}_{v_j}(u_j) \equiv \begin{pmatrix} m_j \\ v_j \end{pmatrix} u_j^{v_j}(1-u_j)^{m_j-v_j}$. Then, $C_B : [0, 1]^2 \to [0, 1]$ is a conditional Bernstein copula for any $m_j \geq 1$ and $x \in \mathcal{X}$ if*

$$C_B(u_1, u_2|x) = \sum_{v_1=0}^{m_1}\sum_{v_2=0}^{m_2} \beta\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, x\right) P^{m_1}_{v_1}(u_1)P^{m_2}_{v_2}(u_2)$$

satisfies the usual properties of the copula function.

Then we can compute a feasible version of (5.1)–(5.2) as

$$\min_{\beta\in\mathcal{B}} \sum_{v_1=0}^{m_1}\sum_{v_2=0}^{m_2} \beta\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, X\right) \int_0^1\int_0^1 1\{Q_{u_1}(Y_1|X) - Q_{u_2}(Y_0|X) \leq t\}dP^{m_1}_{v_1}(u_1)dP^{m_2}_{v_2}(u_2), \tag{5.3}$$

$$\max_{\beta\in\mathcal{B}} \sum_{v_1=0}^{m_1}\sum_{v_2=0}^{m_2} \beta\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, X\right) \int_0^1\int_0^1 1\{Q_{u_1}(Y_1|X) - Q_{u_2}(Y_0|X) \leq t\}dP^{m_1}_{v_1}(u_1)dP^{m_2}_{v_2}(u_2), \tag{5.4}$$

where $\mathcal{B}$ is the restricted set of $\beta(\cdot)$ to impose identifying assumptions and guarantee that C_B is a proper copula. We omitted the latter restrictions for succinctness; see Theorem 2 in Sancetta and Satchell (2004) for details. To impose Assumption SI, for example, it is necessary to ensure that $C_B(u_1|u_2, x) = \partial C_B(u_1, u_2, x)/\partial u_2$ and $C_B(u_2|u_1, x) = \partial C_B(u_1, u_2, x)/\partial u_1$ are non-increasing. Then, by the desirable property of Bernstein, this corresponds to $\beta\left(\frac{v_1}{m_1}, \frac{v_2}{m_2}, X\right)$ being weakly increasing in v_1 and v_2. The use of Bernstein approximation for the systematic calculation of bounds on treatment effects also appears in Han (2023) and Han and Yang (2023) in different contexts. As an alternative to the

Bernstein approximation, one can discretize the space of $(U_1, U_2) \in [0,1]^2$. This approach is considered in Frandsen and Lefgren (2021); see also Blundell et al. (2007). Finally, in practice, the inputs $Q_{u_1}(Y_1|X)$ and $Q_{u_2}(Y_0|X)$ of the linear program can be estimated using standard nonparametric or parametric methods. When they are estimated consistently, we can show that Assumption EST holds for the estimated outputs, $\hat{Q}^L_\tau(X)$ and $\hat{Q}^U_\tau(X)$, of the linear program:

Lemma 5.1. *Suppose that, for $d \in \{0,1\}$, $F_{Y_d|X}(y|X)$ and $Q_\tau(Y_d|X)$ are absolutely continuous in $y \in \mathcal{Y}$ and $\tau \in (0,1)$, respectively, and $\hat{Q}_\tau(Y_d|X)$ is a consistent estimator of $Q_\tau(Y_d|X)$ for any $\tau \in (0,1)$, almost surely. Then, $|\hat{Q}^L_\tau(X) - Q^L_\tau(X)| = o_p(1)$ and $|\hat{Q}^U_\tau(X) - Q^U_\tau(X)| = o_p(1)$.*

Proof. [Proof of Lemma 5.1] In terms of notation, let $Q_\tau(Y_d|X) = F^{-1}_{d|X}(\tau)$. For any $\epsilon > 0$, as n goes to infinity, $P[|\{\hat{F}^{-1}_{1|X}(v) - \hat{F}^{-1}_{0|X}(u)\} - \{F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u)\}| \geq \epsilon] \to 0$. Therefore, on a set with probability converging to 1, we have for $F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \notin (t-\epsilon, t+\epsilon)$,

$$\left| \int\int 1\{\hat{F}^{-1}_{1|X}(v) - \hat{F}^{-1}_{0|X}(u) \leq t\} c(u,v) du dv - \int\int 1\{F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \leq t\} c(u,v) du dv \right| = 0,$$

where $c(u,v)$ is the copula density (which is bounded), because $1\{\hat{F}^{-1}_{1|X}(v) - \hat{F}^{-1}_{0|X}(u) \leq t\} = 1\{F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \leq t\}$. For $F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \in (t-\epsilon, t+\epsilon)$,

$$\left| \int\int 1\{\hat{F}^{-1}_{1|X}(v) - \hat{F}^{-1}_{0|X}(u) \leq t\} c(u,v) du dv - \int\int 1\{F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \leq t\} c(u,v) du dv \right| \leq O_p(\epsilon),$$

because $\int\int_{(u,v): F^{-1}_{1|X}(v) - F^{-1}_{0|X}(u) \in (t-\epsilon, t+\epsilon)} c(u,v) du dv = O_p(\epsilon)$. Hence, for the infeasible optimal value $\tilde{F}^L_{\Delta|X}(t)$ of the linear program using $\hat{F}^{-1}_{1|X}(v)$ and $\hat{F}^{-1}_{0|X}(u)$, we have

$$|\tilde{F}^L_{\Delta|X}(t) - F^L_{\Delta|X}(t)| = o_p(1).$$

For the feasible optimal value $\hat{F}^L_\Delta(t)$ using the discretization approach, we can show that

$$\hat{F}^L_\Delta(t) = \min_{c(\cdot,\cdot)} \sum_{j=1}^{k}\sum_{i=1}^{k} 1\{\hat{F}^{-1}_{Y_1}(r(i)) - \hat{F}^{-1}_{Y_0}(r(j)) \leq t\} c(i,j) \rightarrow \tilde{F}^L_\Delta(t),$$

as $k = k(n)$ goes to infinity. Therefore, $|\hat{F}^L_{\Delta|X}(t) - F^L_{\Delta|X}(t)| = o_p(1)$. We can similarly prove the claim for the upper bound $\hat{F}^U_{\Delta|X}(t)$ and bounds that are obtained using the Bernstein approximation. □

6 Numerical Illustrations

The question we want to answer via numerical exercises is: how the performance of treatment allocations differ across welfare criteria, especially when the QoTE is partially identified. To facilitate illustration, we focus on in the case of unconstrained $\mathcal{D}$ and no X.

We consider the following data-generating processes (DGPs). We draw either (Y_1, Y_0) or $(\log Y_1, \log Y_0)$ from $N(\mu, \Sigma)$, where $\mu = (\mu_1, \mu_0)'$ and $\Sigma = \left(\begin{smallmatrix} \sigma_1 & \rho_{10}\sqrt{\sigma_0\sigma_1} \\ \rho_{10}\sqrt{\sigma_0\sigma_1} & \sigma_0 \end{smallmatrix}\right)$, and D independently from $Bernoulli(0.5)$. Then, the observed outcome is generated by $Y = DY_1 + (1-D)Y_0$. Note that, under the bivariate normal distribution, $Y_1|Y_0 \sim N(\mu_1 + \rho_{10}\sigma_1 Z_0, (1 - \rho^2\sigma_1))$ where $Z_0 = \frac{Y_0 - \mu_0}{\sigma_0}$. Therefore, Y_1 and Y_0 are stochastically increasing when $\rho_{10} \geq 0$, satisfying Assumption SI. In fact, this is also true when Y_1 and Y_0 are bivariate log-normal; they are stochastically increasing when $\rho_{10} \geq 0$. When $\mathcal{D}$ is unrestricted, the true optimal policies based on the QoTE, QTE and ATE can be written as follows:

- $\delta^* = 1\{Q_\tau(Y_1 - Y_0) > 0\}$ where $Q_\tau(Y_1 - Y_0) = \mu_1 - \mu_0 + \Phi^{-1}(\tau)\sqrt{\sigma_1^2 + \sigma_0^2 - 2\rho_{10}\sigma_1\sigma_0}$
- $\delta^*_{QTE} = 1\{Q_\tau(Y_1) - Q_\tau(Y_0) > 0\}$ where $Q_\tau(Y_1) - Q_\tau(Y_0) = \mu_1 - \mu_0 + \Phi^{-1}(\tau)(\sigma_1 - \sigma_0)$
- $\delta^*_{ATE} = 1\{E[Y_1 - Y_0] > 0\}$ where $E[Y_1 - Y_0] = \mu_1 - \mu_0$

Note that these policies are first-best regardless of whether we consider a deterministic or stochastic policy. Unlike δ^*_{QTE} and δ^*_{ATE}, the proposed δ^* involves model uncertainty. Therefore, given (2.5) for deterministic δ^*, we have

$$\delta^*_{mmr} = \begin{cases} 1\{|Q^U_\tau| \geq |Q^L_\tau|\} & \text{if } Q^L_\tau < 0 < Q^U_\tau \\ 1 & \text{if } Q^L_\tau \geq 0 \\ 0 & \text{if } Q^U_\tau \leq 0 \end{cases}$$

ignoring the case of point identification. When we implement a stochastic rule, then we have

$$\delta^*_{mmr} \sim \begin{cases} Bernoulli\left(\frac{Q_U}{Q_U - Q_L}\right) & \text{if } Q^L_\tau < 0 < Q^U_\tau \\ 1 & \text{if } Q^L_\tau \geq 0 \\ 0 & \text{if } Q^U_\tau \leq 0 \end{cases}$$

ignoring the case of point identification.

In simulation, the bounds Q^L_τ and Q^U_τ are calculated under either no assumption (i.e., Makarov bounds) or Assumption SI. Under the latter, we use discretization to calculate the bounds. For the population policies δ^*_{mmr}, δ^*_{QTE} and δ^*_{ATE}, we estimate their sample counterparts $\hat{\delta}^*$, $\hat{\delta}^*_{QTE}$ and $\hat{\delta}^*_{ATE}$ by estimating Q^U_τ, Q^L_τ, $Q_\tau(Y_d)$, and $E[Y_d]$ $(d = 0, 1)$. Since D is exogenous in our simulated data, $Q_\tau(Y_d) = Q_\tau(Y|D = d)$ and $E[Y_d] = E[Y|D = d]$. For each estimate $\hat{\delta}_j$ of δ^*_j $(j \in \{\emptyset, QTE, ATE\})$, the misclassification error is $\mathbb{E}_n[1\{\hat{\delta}_j \neq \delta^*_j\}]$ and regret is defined as $\mathbb{E}_n[|T_j| \cdot 1\{\hat{\delta}_j \neq \delta^*_j\}]$ where $T = Q_\tau(Y_1 - Y_0)$, $T_{QTE} = Q_\tau(Y_1) - Q_\tau(Y_0)$, and $T_{ATE} = E[Y_1 - Y_0]$ are the corresponding treatment effects (or equivalently the welfare gains). We focus on $\tau = 0.25$.

Tables 1–2 present the simulated correct classification rates of the estimated policies relative to the (true) population policies. We set $n = 1000$ for Table 1 and $n = 50$ for Table 2. To calculate each classification rate, we replicate each experiment 200 times. We consider both the correct specification of Assumption SI and misspecification. We also vary the parameter values in the normal and log-normal distributions. We treat each DGP as a subgroup of population (as if it corresponds to a particular value of X if covariates were to be introduced). Subgroups 1–4 and 7 follow the normal distribution and SI, and Subgroups 5–6 are where SI is violated. Under bivariate normality and SI, if $0 < \tau < 0.5$, $Q_\tau(Y_1 - Y_0) < Q_\tau(Y_1) - Q_\tau(Y_0)$ and $Q_\tau(Y_1 - Y_0) < E(Y_1) - E(Y_0)$. The purpose of Subgroup 8 is to break this mechanical relationship. Subgroup 8 follows the log-normal distribution and SI.

Estimated Policy / Optimal Policy	$\hat{\delta}^{SI,stoch}$	$\hat{\delta}^{stoch}$	$\hat{\delta}^{SI,determ}$	$\hat{\delta}^{determ}$	$\hat{\delta}_{QTE}$	$\hat{\delta}_{ATE}$
	Subgroup 1					
δ^*	100%	100%	100%	100%	0%	100%
δ^*_{QTE}	0%	0%	0%	0%	100%	0%
δ^*_{ATE}	100%	100%	100%	100%	0%	100%
	Subgroup 2					
δ^*	100%	99%	100%	100%	0%	0%
δ^*_{QTE}	0%	1%	0%	0%	100%	100%
δ^*_{ATE}	0%	1%	0%	0%	100%	100%
	Subgroup 3					
δ^*	89%	59%	100%	95%	100%	100%
δ^*_{QTE}	89%	59%	100%	95%	100%	100%
δ^*_{ATE}	89%	59%	100%	95%	100%	100%
	Subgroup 4					
δ^*	21%	43%	0%	20%	0%	0%
δ^*_{QTE}	79%	57%	100%	80%	100%	100%
δ^*_{ATE}	79%	57%	100%	80%	100%	100%
	Subgroup 5					
δ^*	92%	74%	100%	100%	100%	0%
δ^*_{QTE}	92%	74%	100%	100%	100%	0%
δ^*_{ATE}	8%	26%	0%	0%	0%	100%
	Subgroup 6					
δ^*	34%	48%	6.5%	86%	0%	0%
δ^*_{QTE}	66%	52%	93.5%	14%	100%	100%
δ^*_{ATE}	66%	52%	93.5%	14%	100%	100%
	Subgroup 7					
δ^*	76.5%	53%	100%	31.5%	100%	100%
δ^*_{QTE}	76.5%	53%	100%	31.5%	100%	100%
δ^*_{ATE}	76.5%	53%	100%	31.5%	100%	100%
	Subgroup 8 (log normal)					
δ^*	100%	60%	100%	97%	100%	70.5%
δ^*_{QTE}	100%	60%	100%	97%	100%	70.5%
δ^*_{ATE}	0%	40%	0%	3%	0%	29.5%

Table 1: Correct Classification Rate ($n = 1000$)

optimal decision \ estimated decision	$\hat{\delta}^{SI,stoch}$	$\hat{\delta}^{stoch}$	$\hat{\delta}^{SI,determ}$	$\hat{\delta}^{determ}$	$\hat{\delta}_{QTE}$	$\hat{\delta}_{ATE}$
	Subgroup 1					
δ^*	100%	100%	100%	100%	33.5%	93%
δ^*_{QTE}	0%	0%	0%	0%	66.5%	7%
δ^*_{ATE}	100%	100%	100%	100%	33.5%	93%
	Subgroup 2					
δ^*	92%	90%	90.5%	94%	1%	17%
δ^*_{QTE}	8%	10%	9.5%	6%	99%	83%
δ^*_{ATE}	8%	10%	9.5%	6%	99%	83%
	Subgroup 3					
δ^*	79%	51%	84%	62%	99.5%	100%
δ^*_{QTE}	79%	51%	84%	62%	99.5%	100%
δ^*_{ATE}	79%	51%	84%	62%	99.5%	100%
	Subgroup 4					
δ^*	26%	49.5%	14.5%	44%	1.5%	0.5%
δ^*_{QTE}	74%	50.5%	85.5%	56%	98.5%	99.5%
δ^*_{ATE}	74%	50.5%	85.5%	56%	98.5%	99.5%
	Subgroup 5					
δ^*	82%	81%	83%	93.5%	66.5%	4%
δ^*_{QTE}	82%	81%	83%	93.5%	66.5%	4%
δ^*_{ATE}	18%	19%	17%	6.5%	33.5%	96%
	Subgroup 6					
δ^*	43%	61.5%	40%	61%	24%	0%
δ^*_{QTE}	57%	38.5%	60%	39%	76%	100%
δ^*_{ATE}	57%	38.5%	60%	39%	76%	100%
	Subgroup 7					
δ^*	71.5%	48%	80%	49.5%	96%	99.5%
δ^*_{QTE}	71.5%	48%	80%	49.5%	96%	99.5%
δ^*_{ATE}	71.5%	48%	80%	49.5%	96%	99.5%
	Subgroup 8 (log normal)					
δ^*	95.5%	68%	95.5%	73%	100%	76.5%
δ^*_{QTE}	95.5%	68%	95.5%	73%	100%	76.5%
δ^*_{ATE}	4.5%	32%	4.5%	28%	0%	23.5%

Table 2: Correct Classification Rate ($n = 50$)

Here are the summary of the features in the DGP and corresponding results in Tables 1–2. Recall that $\tau = 0.25$.

- Overall, the correct classification rate tends to be high when the welfare criterion of the estimated policy matches that of the population policy.

- Subgroup 1: Both intervals under SI and no assumption exclude 0 and lie relatively far from it; therefore, both $\hat{\delta}$ and $\hat{\delta}^{SI}$ perform well; $\delta^*_{QTE} \neq \delta^* = \delta^*_{ATE}$.

- Subgroup 2: Both intervals under SI and no assumption exclude 0; $\hat{\delta}^{determ}$ does not perform worse than $\hat{\delta}^{determ,SI}$ for δ^* because $Q_\tau^{L,SI} - Q_\tau^L > Q_\tau^U - Q_\tau^{U,SI}$; $\delta^* \neq \delta^*_{QTE} = \delta^*_{ATE}$.

- Subgroup 3: Both intervals under SI and no assumption cover 0 (and the same holds for Subgroups 4–7); $\hat{\delta}^{SI}$ performs better than $\hat{\delta}$; $Q_\tau^{L,SI} - Q_\tau^L > Q_\tau^U - Q_\tau^{U,SI}$ and, under the bivariate normal distribution and SI, $Q_\tau(Y_1 - Y_0) < Q_\tau(Y_1) - Q_\tau(Y_0)$ and $Q_\tau(Y_1 - Y_0) < E(Y_1) - E(Y_0)$ always hold, and thus both $\hat{\delta}_{QTE}$ and $\hat{\delta}_{ATE}$ perform well; $\delta^* = \delta^*_{QTE} = \delta^*_{ATE} = 1$.

- Subgroup 4: Both $\hat{\delta}^{SI}$ and $\hat{\delta}$ perform poorly for δ^* because the bound on $Q_\tau(Y_1 - Y_0)$ covers zero, and the difference between the upper bound and zero is larger than the difference between the lower bound and zero; $\delta^* \neq \delta^*_{QTE} = \delta^*_{ATE}$.

- Subgroup 5: SI is false but $\hat{\delta}^{SI}$ does not perform so poorly because $Q_\tau(Y_1 - Y_0)$ is still covered by a relatively long interval; for the same reason, $\hat{\delta}$ does not perform significantly better; $\delta^* = \delta^*_{QTE} \neq \delta^*_{ATE}$.

- Subgroup 6: SI is false and $\hat{\delta}^{SI}$ performs poorly because $Q_\tau(Y_1 - Y_0)$ is *not* covered by a relatively long interval; $\hat{\delta}$ does not perform well because $Q_\tau^{L,SI} - Q_\tau^L < Q_\tau^U - Q_\tau^{U,SI}$; $\delta^* \neq \delta^*_{QTE} = \delta^*_{ATE}$.

- Subgroup 7: $\hat{\delta}^{SI}$ makes a correct decision while $\hat{\delta}$ performs worse; meanwhile, $\hat{\delta}_{QTE}$ and $\hat{\delta}_{ATE}$ perform well.

- Subgroup 8: The interval under SI excludes 0 while the interval under no assumption covers 0; therefore, $\hat{\delta}^{SI}$ performs better than $\hat{\delta}$; under this log-normal setting and SI, $Q_\tau(Y_1 - Y_0) < E(Y_1) - E(Y_0)$ may be violated, which occurs in the current subgroup and thus $\hat{\delta}^{SI}$ performs better than $\hat{\delta}_{ATE}$.

7 Empirical Applications

7.1 Application I: Allocation of Right Heart Catheterization

We consider the right heart catheterization (RHC) dataset from the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments (SUPPORT) (Hirano and Imbens (2001)). The treatment D in question is the RHC (1 if received and 0 otherwise), a diagnostic procedure for critically ill patients. The outcome Y is the number of days from admission to death within 30 days (t3d30), whose value ranges from 2 to 30. In contrast to the belief of practitioners that the RHC is beneficial, studies like Connors et al. (1996) found that patient survival is lower with the RHC than without. Therefore, a relevant policy question in this critical situation is to find patients for whom allocating (or avoiding) the RHC is life-saving. In the dataset, 5735 patients are divided into a treatment group (2184 patients) and a control group (3551 patients). We consider the following covariates as X: age, sex, coma in primary disease 9-level category (cat1_coma), coma in secondary disease 6-level category, (cat2_coma), do not resuscitate (DNR) status on day 1 (i.e., DNR when heart stops) (dnr1), estimated probability of surviving 2 months (surv2md1), and APACHE III score ignoring coma (i.e., ICU mortality score) (aps1).

To estimate the counterfactual distributions $F_{Y_1|X}$ and $F_{Y_0|X}$ of the outcome (t3d30) for different groups defined by the covariates, we conduct a kernel regression in the treatment and control groups separately with bandwidth under Scott's rule of thumb.[8] Then we calculate the upper and lower bounds of the QoTE under SI and no assumption and make the decisions

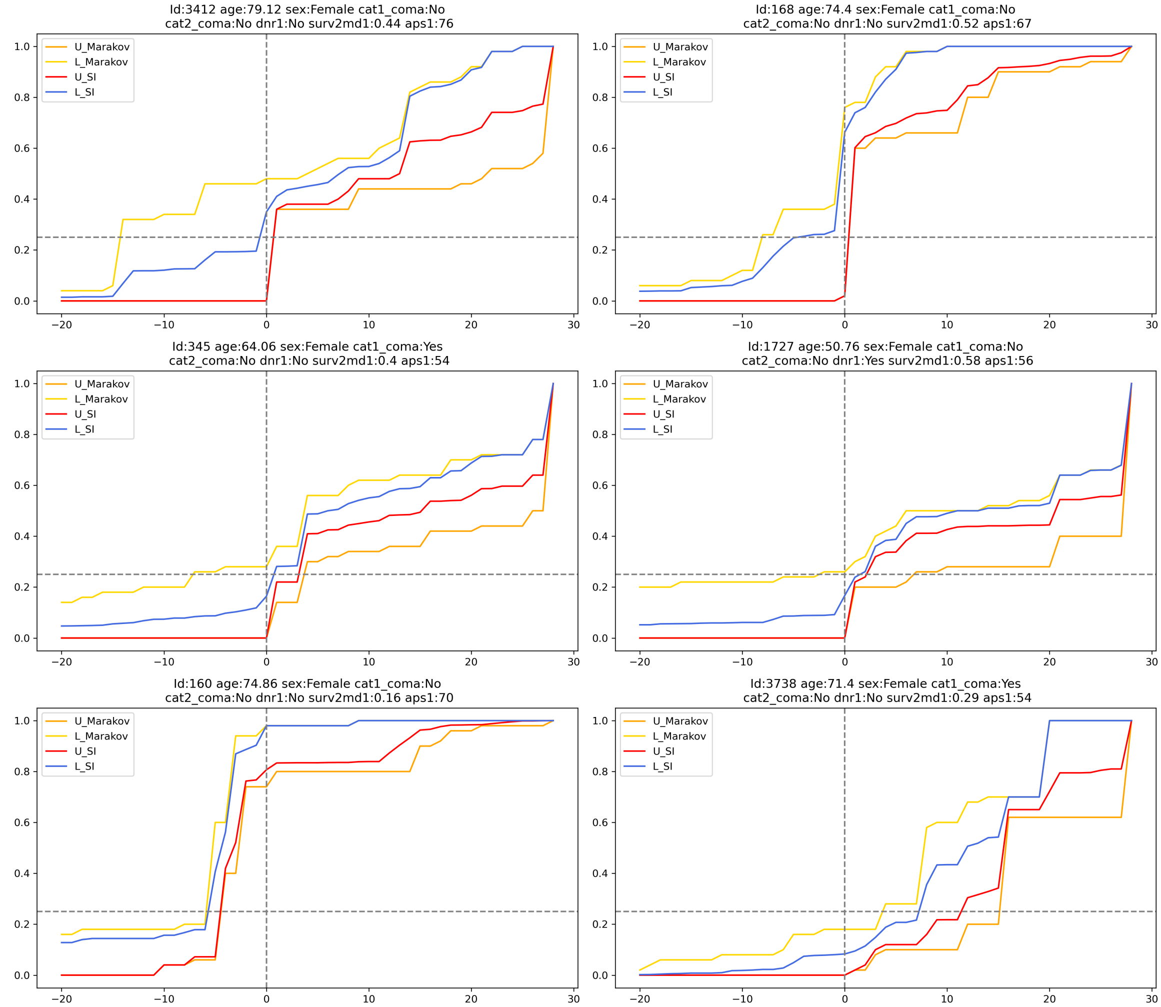
[8]To simplify this process, we run the regression $P[Y < y_j|X = x] = E[1\{Y < y_j\}|X = x]$ on a series of $y_j = F_Y^{-1}(\frac{2j-1}{2k})$, where $k = 1000$ and $j = 1, ..., k$.

Patient ID	(Q_τ^L, Q_τ^U)	$\hat{\delta}$	$(Q_\tau^{L,SI}, Q_\tau^{U,SI})$	$\hat{\delta}^{SI}$
3412	$(-14.27, 0.69)$	0	$(-0.65, 0.69)$	1
168	$(-8.07, 0.40)$	0	$(-4.63, 0.40)$	0
345	$(-7.17, 3.69)$	0	$(0.73, 3.16)$	1
1727	$(-2.50, 6.75)$	1	$(1.50, 2.13)$	1
160	$(-5.88, -4.44)$	0	$(-5.69, -4.49)$	0
3738	$(3.7, 15.12)$	1	$(7.24, 11.37)$	1

Table 3: Bounds on the QoTE ($\tau = 0.25$) and Estimated Policies

by using the proposed criterion based on the QoTE. As seen in the simulation results in Section 6, the SI and no-assumption bounds will not always give the same decisions, and the information provided by the bounds differs from person to person.

In Table 3 and Figure 1, we present six cases to show the SI and no-assumption bounds of the QoTE. We only focus on deterministic policies and $\tau = 0.25$. These results illustrate how the actual implementation of our proposed policies would look like for each individual. It is shown that there is much heterogeneity in terms of the QoTEs and thus the corresponding optimal decisions.



In the figure, the vertical line indicates zero and the horizontal line indicates $\tau = 0.25$.

Figure 1: Bounds on the QoTE of Six Representative Patients

7.2 Application II: Allocation of Job Training

The dataset is collected from the National Job Training Partnership Act (JTPA) Study (Bloom et al. (1997)). We use a subset that includes 9,223 adults; 6,133 of them received job training, while the remaining 3,090 did not. The treatment D in question is the job training. In this experiment, we use the 30-month earnings after the job training program as the measure of outcome Y and the sex, years of education, high school diploma, and previous earnings in \$10K before the program as covariates X. Based on the data, the kernel regression has been conducted in the treatment and control groups separately to obtain the

$\hat{F}_{Y_1|X}$ and $\hat{F}_{Y_0|X}$. From the estimated conditional distributions, we obtain the upper and lower bounds under SI and no assumption for each individual.

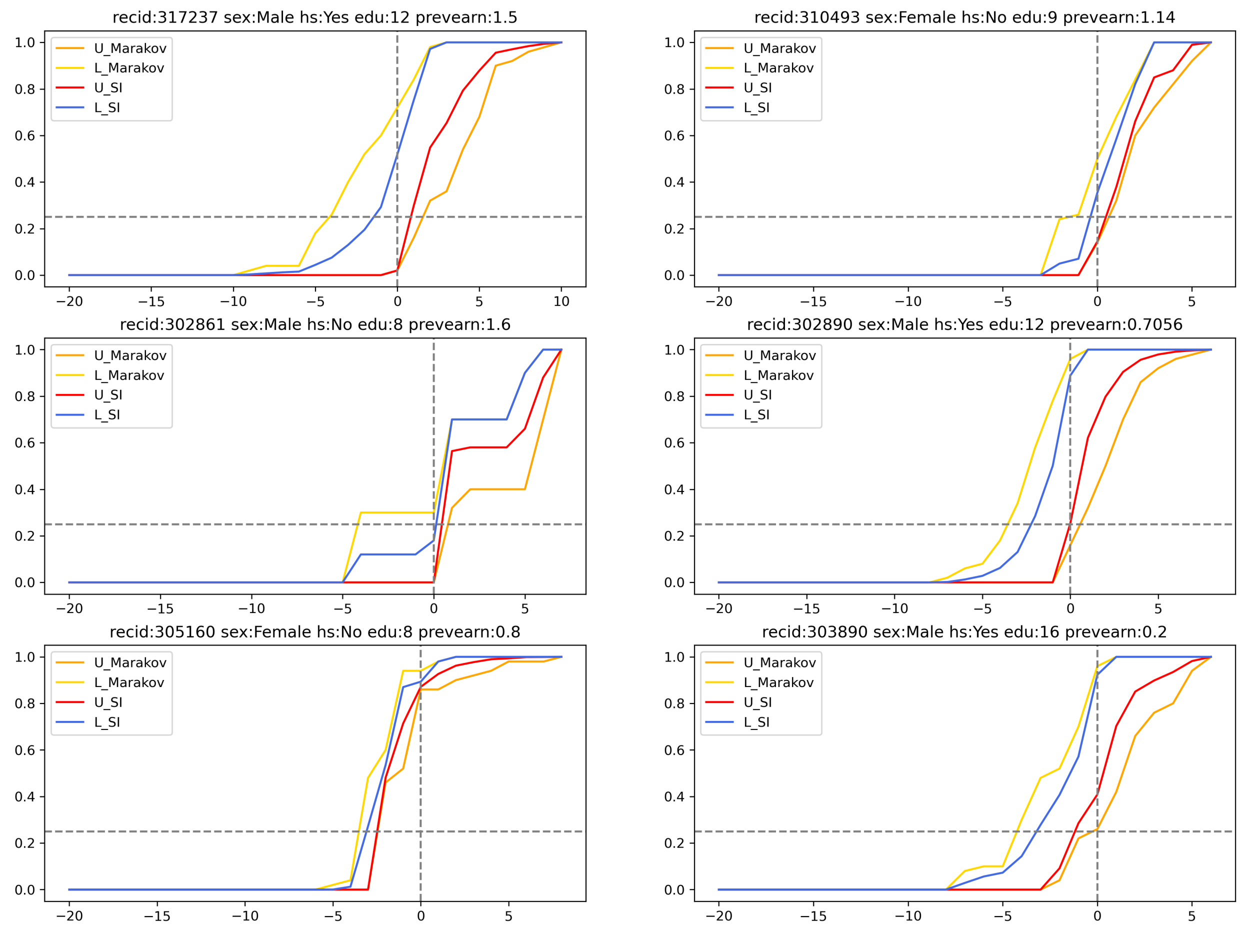
In Table 4 and Figure 2, we present six cases and their covariates to show bounds on the QoTE (i.e., the effect of job training on earnings) under SI and no assumption. Similar to the first application, we find heterogeneity in the treatment effects and thus the optimal decisions, but less so than the first application.

Next, in Figure 3, we present the decisions of allocating the job training to the female group without a high school diploma in the space of education and previous earnings (i.e., the two important covariates for the allocation decision). We use the 0.25-quantile, median, and 0.75-quantile QoTE bounds to represent prudent, majority-minded, and negligent PMs, respectively. As expected, the 0.75-quantile bounds suggest the treatment option more often than the bounds with the other quantile probabilities. Given that the 0.25-quantile bounds suggest the most prudent decisions, the treatment option suggested in this figure can be viewed as a compelling recommendation.

For comparison, in Figure 4, we present the allocation decisions based on the 0.25-quantile, median, and 0.75-quantile QTE and the ATE. Interestingly, there is no obvious tendency in decisions when the quantile probability increases from 0.25 to 0.75, which reflects the limitation of using the QTE as the basis for decisions (e.g., the quantile probability does not capture the level of prudence). The decisions based on the ATE show how they can be viewed as the most common approaches in the literature. They look very similar to the decisions based on the median QoTE bounds, which suggests that the issue of outliers is not serious in this application. In this sense, the policy based on the median QoTE bounds can be viewed as a robustness check for the policy based on the ATE (e.g., Kitagawa and Tetenov (2018)).

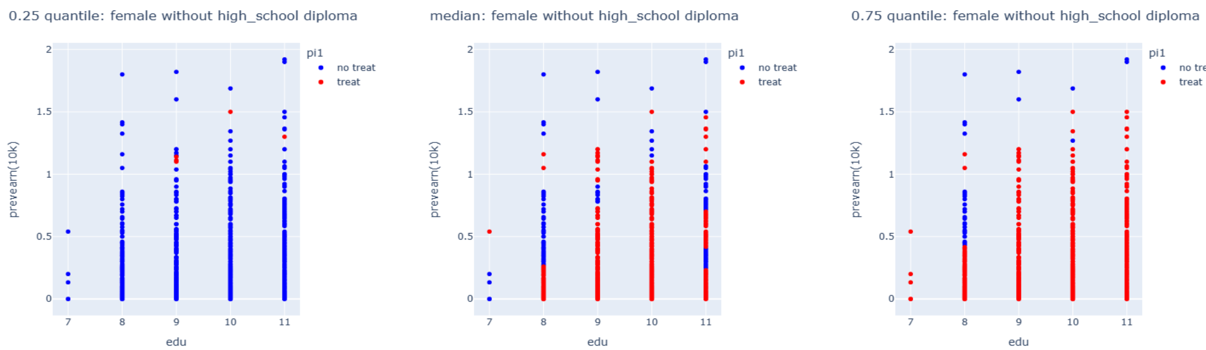
Worker ID	(Q_τ^L, Q_τ^U)	$\hat{\delta}$	$(Q_\tau^{L,SI}, Q_\tau^{U,SI})$	$\hat{\delta}^{SI}$
317237	$(-14.27, 0.69)$	0	$(-1.44, 0.83)$	0
310493	$(-8.07, 0.40)$	0	$(-0.37, 0.45)$	1
302861	$(-7.17, 3.69)$	0	$(0.13, 0.44)$	1
302890	$(-2.50, 6.75)$	1	$(-2.23, -0.02)$	0
305160	$(-5.88, -4.44)$	0	$(-3.09, -2.48)$	0
303890	$(3.7, 15.12)$	1	$(-3.21, -1.19)$	0

Table 4: Bounds on the QoTE ($\tau = 0.25$) and Estimated Policies



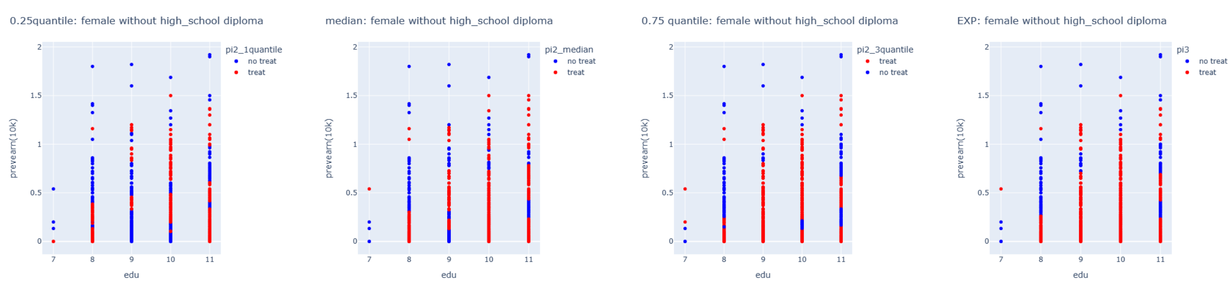
In the figure, the vertical line indicates zero and the horizontal line indicates $\tau = 0.25$.

Figure 2: Bounds on the QoTE of Six Representative Workers



(a) QoTE with $\tau = 0.25$ (b) QoTE with $\tau = 0.5$ (c) QoTE with $\tau = 0.75$

Figure 3: Treatment Decisions for Female Workers Without High School Diploma



(a) QTE with $\tau = 0.25$ (b) QTE with $\tau = 0.5$ (c) QTE with $\tau = 0.75$ (d) ATE

Figure 4: Treatment Decisions Based on the QTEs and the ATE

A Welfare Criteria with Stochastic Rules

We present a more rigorous formulation of the welfare criteria in Section 2 when a stochastic rule is considered. Let $A(x)$ is a r.v. representing the stochastic rule drawn from Bernoulli with parameter $\delta(x) \equiv \Pr[A(x) = 1|X = x]$. Then, by assuming $A(X) \perp Y_d|X$ for any d (and using it in the third equality below), we have

$$
\begin{aligned}
E[A(X)Y_1 + (1 - A(X))Y_0] &= E[Y_0] + E[A(X)(Y_1 - Y_0)] \\
&= E[Y_0] + E[A(X)E[Y_1 - Y_0|A(X), X]] \\
&= E[Y_0] + E[A(X)E[Y_1 - Y_0|X]] \\
&= E[Y_0] + E[E[A(X)E[Y_1 - Y_0|X]|X]] \\
&= E[Y_0] + E[E[Y_1 - Y_0|X]E[A(X)|X]] \\
&= E[Y_0] + E[E[Y_1 - Y_0|X]\delta(X)].
\end{aligned}
$$

Note that the last expression can be written as $E[\delta(X)Y_1 + (1 - \delta(X))Y_0]$.

Similarly, motivated from the third line above

$$
\begin{aligned}
E[A(X)Q(Y_1 - Y_0|X)] &= E[E[A(X)Q(Y_1 - Y_0|X)|X]] \\
&E[Q(Y_1 - Y_0|X)E[A(X)|X]] \\
&= E[Q(Y_1 - Y_0|X)\delta(X)].
\end{aligned}
$$

B Details of the DGPs of Subgroups in Simulation

Table 5 shows the details of the DGP for each subgroup used in the simulation of Section 6.

Subgroup	(μ_1, μ_0)	(σ_1^2, σ_0^2)	ρ_{10}	δ^*	δ^*_{QTE}	δ^*_{ATE}
1	$(2, 3)$	$(1, 9)$	0.5	0 (-2.97)	1 (0.34)	0 (-1)
2	$(4, 3)$	$(1, 25)$	0.5	0 (-2.09)	1 (3.7)	1 (1)
3	$(7, 3)$	$(9, 25)$	0.5	1 (1.1)	1 (5.35)	1 (4)
4	$(3, 1)$	$(5, 5)$	0.1	0 (-0.23)	1 (2)	1 (2)
5	$(3, 2)$	$(9, 1)$	-0.5	0 (-1.43)	0 (-0.35)	1 (1)
6	$(3, 0)$	$(25, 4)$	-0.5	0 (-1.21)	1 (0.98)	1 (3)
7	$(2, 0)$	$(8, 4)$	0.5	1 (0.30)	1 (1.44)	1 (2)
8	$(3, 0)$	$(2, 8)$	0.8	1 (1.72)	1 (7.59)	0 (0)

Subgroup	$Q_\tau^{L,SI}$	$Q_\tau^{U,SI}$	Q_τ^L	Q_τ^U	$\delta^{SI,stoch}$	δ^{stoch}	$Y_1 - Y_0$
1	-3.48	-1.84	-5.1	-1.1	100%	100%	$N(-1, 7)$
2	-2.8	-1.19	-4.50	-0.17	100%	100%	$N(1, 21)$
3	-0.74	3.91	-4.83	5.63	84%	54%	$N(4, 19)$
4	-0.68	2.54	-3.1	3.38	21%	48%	$N(2, 9)$
5	-1.48	0.16	-3.1	0.9	90%	77%	$N(1, 13)$
6	-1.24	1.92	-4.3	3.5	39%	55%	$N(3, 39)$
7	-0.86	2.17	-3.37	3.21	72%	49%	$N(2, 12 - 4\sqrt{2})$
8	1.13	7.1	-5.5	7.75	100%	58%	-

Subgroup 8 is under log-normal transformation, i.e., $(\log Y_1, \log Y_0) \sim N(\mu, \Sigma)$.
In the parentheses of δ^*, δ^*_{QTE}, and δ^*_{ATE} are the values of $Q_\tau(Y_1 - Y_0)$, $Q_\tau(Y_1) - Q_\tau(Y_0)$, and $E[Y_1 - Y_0]$, respectively.

Table 5: DGP and Population Values

References

Abbring, J. H. and J. J. Heckman (2007): "Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation," *Handbook of econometrics*, 6, 5145–5303. 3, 3, 3

Adjaho, C. and T. Christensen (2022): "Externally valid treatment choice," *arXiv preprint arXiv:2205.05561*. 1.2

Athey, S. and S. Wager (2021): "Policy learning with observational data," *Econometrica*, 89, 133–161. 1.2

Ben-Michael, E., D. J. Greiner, K. Imai, and Z. Jiang (2021): "Safe policy learning through extrapolation: Application to pre-trial risk assessment," *arXiv preprint arXiv:2109.11679*. 1.2

Bloom, H. S., L. L. Orr, S. H. Bell, G. Cave, F. Doolittle, W. Lin, and J. M. Bos (1997): "The benefits and costs of JTPA Title II-A programs: Key findings from the National Job Training Partnership Act study," *Journal of human resources*, 549–576. 7.2

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the distribution of male and female wages accounting for employment composition using bounds," *Econometrica*, 75, 323–363. 3, 5

Chernozhukov, V., I. Fernández-Val, J. Hahn, and W. Newey (2013): "Average and quantile effects in nonseparable panel models," *Econometrica*, 81, 535–580. 3

Chernozhukov, V. and C. Hansen (2005): "An IV model of quantile treatment effects," *Econometrica*, 73, 245–261. 3

Connors, A. F., T. Speroff, N. V. Dawson, C. Thomas, F. E. Harrell, D. Wagner, N. Desbiens, L. Goldman, A. W. Wu, R. M. Califf, et al. (1996): "The effectiveness of right heart catheterization in the initial care of critically III patients," *Jama*, 276, 889–897. 7.1

Cui, Y. (2021): "Individualized decision making under partial identification: three perspectives, two optimality results, and one paradox," *Harvard Data Science Review*, 3, 1–19. 1.2

Cui, Y. and E. Tchetgen Tchetgen (2021a): "On a necessary and sufficient identification condition of optimal treatment regimes with an instrumental variable," *Statistics & Probability Letters*, 178, 109180. 1.2

——— (2021b): "A semiparametric instrumental variable approach to optimal treatment regimes under endogeneity," *Journal of the American Statistical Association*, 116, 162–173. 1.2

D'Adamo, R. (2021): "Orthogonal Policy Learning Under Ambiguity," *arXiv preprint arXiv:2111.10904*. 1.2, 2

Doksum, K. (1974): "Empirical probability plots and statistical inference for nonlinear models in the two-sample case," *The annals of statistics*, 267–277. 2

Dudík, M., J. Langford, and L. Li (2011): "Doubly robust policy evaluation and learning," *arXiv preprint arXiv:1103.4601*. 1.2

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

Frandsen, B. R. and L. J. Lefgren (2021): "Partial identification of the distribution of treatment effects with an application to the Knowledge is Power Program (KIPP)," *Quantitative Economics*, 12, 143–171. 3, 5

Han, S. (2021): "Comment: Individualized treatment rules under endogeneity," *Journal of the American Statistical Association*, 116, 192–195. 1.2

——— (2023): "Optimal dynamic treatment regimes and partial welfare ordering," *Journal of the American Statistical Association*, 1–11. 1.2, 5

Han, S. and S. Yang (2023): "A Computational Approach to Identification of Treatment Effects for Policy Evaluation," *arXiv preprint arXiv:2009.13861*. 5

Heckman, J. J. and B. E. Honore (1990): "The empirical content of the Roy model," *Econometrica: Journal of the Econometric Society*, 1121–1149. 3

Heckman, J. J., J. Smith, and N. Clements (1997): "Making the most out of programme evaluations and social experiments: Accounting for heterogeneity in programme impacts," *The Review of Economic Studies*, 64, 487–535. 3

Heckman, J. J. and J. A. Smith (1995): "Assessing the case for social experiments," *Journal of economic perspectives*, 9, 85–110. 3, 3

Hirano, K. and G. W. Imbens (2001): "Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization," *Health Services and Outcomes research methodology*, 2, 259–278. 7.1

Hirano, K. and J. R. Porter (2009): "Asymptotics for statistical treatment rules," *Econometrica*, 77, 1683–1701. 1.2

Ida, T., T. Ishihara, K. Ito, D. Kido, T. Kitagawa, S. Sakaguchi, and S. Sasaki (2022): "Choosing Who Chooses: Selection-driven targeting in energy rebate programs," Tech. rep., National Bureau of Economic Research. 1.2

Ishihara, T. and T. Kitagawa (2021): "Evidence aggregation for treatment choice," *arXiv preprint arXiv:2108.06473*. 1.2

Jiang, B., R. Song, J. Li, and D. Zeng (2019): "Entropy learning for dynamic treatment regimes," *Statistica Sinica*, 29, 1633. 1.2

Joe, H. (2014): *Dependence modeling with copulas*, CRC press. 3

Kallus, N., X. Mao, and M. Uehara (2021): "Causal inference under unmeasured confounding with negative controls: A minimax learning approach," *arXiv preprint arXiv:2103.14029*. 1.2

Kallus, N. and A. Zhou (2021): "Minimax-optimal policy learning under unobserved confounding," *Management Science*, 67, 2870–2890. 1.2

Kasy, M. (2016): "Partial identification, distributional preferences, and the welfare ranking of policies," *Review of Economics and Statistics*, 98, 111–131. 2

Kitagawa, T., S. Lee, and C. Qiu (2023): "Treatment Choice, Mean Square Regret and Partial Identification," *arXiv preprint arXiv:2310.06242*. 1.2

Kitagawa, T., S. Sakaguchi, and A. Tetenov (2021): "Constrained classification and policy learning," *arXiv preprint arXiv:2106.12886*. 2

Kitagawa, T. and A. Tetenov (2018): "Who should be treated? empirical welfare maximization methods for treatment choice," *Econometrica*, 86, 591–616. 1.2, 7.2

——— (2021): "Equality-minded treatment choice," *Journal of Business & Economic Statistics*, 39, 561–574. 1.2

Kosorok, M. R. and E. B. Laber (2019): "Precision medicine," *Annual review of statistics and its application*, 6, 263–286. 1.2

Kosorok, M. R. and E. E. Moodie (2015): *Adaptive treatment strategies in practice: planning trials and analyzing data for personalized medicine*, SIAM. 1.2

Lee, J. H. and B. G. Park (2023): "Nonparametric identification and estimation of the extended Roy model," *Journal of Econometrics*, 235, 1087–1113. 3

Lee, S. (2021): "Partial identification and inference for conditional distributions of treatment effects," *arXiv preprint arXiv:2108.00723*. 3

Lehmann, E. L. (1975): "Statistical methods based on ranks," *Nonparametrics. San Francisco, CA, Holden-Day.* 2

Leqi, L. and E. H. Kennedy (2021): "Median optimal treatment regimes," *arXiv preprint arXiv:2103.01802.* 1.2, 2, 3

Linn, K. A., E. B. Laber, and L. A. Stefanski (2017): "Interactive q-learning for quantiles," *Journal of the American Statistical Association*, 112, 638–649. 1.2

Makarov, G. (1982): "Estimates for the distribution function of a sum of two random variables when the marginal distributions are fixed," *Theory of Probability & its Applications*, 26, 803–806. 3

Manski, C. F. (2004): "Statistical treatment rules for heterogeneous populations," *Econometrica*, 72, 1221–1246. 1.1, 1, 1.2

Manski, C. F. and A. Tetenov (2023): "Statistical decision theory respecting stochastic dominance," *The Japanese Economic Review*, 1–23. 1.2

Mbakop, E. and M. Tabord-Meehan (2021): "Model selection for treatment choice: Penalized welfare maximization," *Econometrica*, 89, 825–848. 1.2

Murphy, S. A. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65, 331–355. 1.2

Pu, H. and B. Zhang (2021): "Estimating optimal treatment rules with an instrumental variable: A partial identification learning approach," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83, 318–345. 1.2

Qi, Z., R. Miao, and X. Zhang (2023a): "Proximal learning for individualized treatment regimes under unmeasured confounding," *Journal of the American Statistical Association*, 1–14. 1.2

QI, Z., J.-S. PANG, AND Y. LIU (2023b): "On robustness of individualized decision rules," *Journal of the American Statistical Association*, 118, 2143–2157. 1.2

QIAN, M. AND S. A. MURPHY (2011): "Performance guarantees for individualized treatment rules," *Annals of statistics*, 39, 1180. 1.2

QIU, H., M. CARONE, E. SADIKOVA, M. PETUKHOVA, R. C. KESSLER, AND A. LUEDTKE (2021): "Optimal individualized decision rules using instrumental variable methods," *Journal of the American Statistical Association*, 116, 174–191. 1.2

ROBINS, J. M. (2004): "Optimal structural nested models for optimal sequential decisions," in *Proceedings of the Second Seattle Symposium in Biostatistics: analysis of correlated data*, Springer, 189–326. 1.2

RUBIN, D. B. AND M. J. VAN DER LAAN (2012): "Statistical issues and limitations in personalized medicine research with clinical trials," *The international journal of biostatistics*, 8, 18. 1.2

SANCETTA, A. AND S. SATCHELL (2004): "The Bernstein copula and its applications to modeling and approximations of multivariate distributions," *Econometric theory*, 20, 535–562. 5, 5

SAVAGE, L. J. (1951): "The theory of statistical decision," *Journal of the American Statistical association*, 46, 55–67. 2

SHEN, T. AND Y. CUI (2023): "Optimal treatment regimes for proximal causal learning," *NeurIPS*. 1.2

SHI, C., A. FAN, R. SONG, AND W. LU (2018): "High-dimensional A-learning for optimal dynamic treatment regimes," *Annals of statistics*, 46, 925. 1.2

STEINWART, I. AND C. SCOVEL (2007): "Fast rates for support vector machines using Gaussian kernels," *The Annals of Statistics*, 35, 575–607. 4.2

Stoye, J. (2009): "Minimax regret treatment choice with finite samples," *Journal of Econometrics*, 151, 70–81. 1.2

——— (2012): "Minimax regret treatment choice with covariates or with limited validity of experiments," *Journal of Econometrics*, 166, 138–156. 1.2

Tsiatis, A. A., M. Davidian, S. T. Holloway, and E. B. Laber (2019): *Dynamic treatment regimes: Statistical methods for precision medicine*, CRC press. 1.2

Villani, C. (2009): *Optimal transport: old and new*, vol. 338, Springer. 5

Vuong, Q. and H. Xu (2017): "Counterfactual mapping and individual treatment effects in nonseparable models with binary endogeneity," *Quantitative Economics*, 8, 589–610. 3, 6

Wang, L., Y. Zhou, R. Song, and B. Sherwood (2018): "Quantile-optimal treatment regimes," *Journal of the American Statistical Association*, 113, 1243–1254. 1.2, 2

Watkins, C. J. and P. Dayan (1992): "Q-learning," *Machine learning*, 8, 279–292. 1.2

Williamson, R. C. and T. Downs (1990): "Probabilistic arithmetic. I. Numerical methods for calculating convolutions and dependency bounds," *International journal of approximate reasoning*, 4, 89–158. 3

Yata, K. (2021): "Optimal decision rules under partial identification," *arXiv preprint arXiv:2111.04926*. 1.2

Zhang, B., A. A. Tsiatis, E. B. Laber, and M. Davidian (2012): "A robust method for estimating optimal treatment regimes," *Biometrics*, 68, 1010–1018. 1.2

Zhao, Y., D. Zeng, A. J. Rush, and M. R. Kosorok (2012): "Estimating individualized treatment rules using outcome weighted learning," *Journal of the American Statistical Association*, 107, 1106–1118. 1.1, 1.2, 2, 4.2, 4.2

Zhao, Y. Q., D. Zeng, E. B. Laber, R. Song, M. Yuan, and M. R. Kosorok (2015): "Doubly robust learning for estimating individualized treatment with censored data," *Biometrika*, 102, 151–168. 4.2