Policy Learning with Compliance Guarantee

Thomas Chan, Vadim Marmer, and Kyungchul Song
Vancouver School of Economics, University of British Columbia

Abstract. We study optimal policy learning where a policymaker (PM) uses data from a source population to design treatment assignments for a target population under a budget constraint. Because of the budget constraint, the PM needs to consider both treatment effects and individuals' incentives for treatment participation to minimize wasted resources. The main challenge is that treatment participation incentives may differ between the two populations. We develop a maximin approach that maximizes the minimum of the PM's expected objective across all possible incentive configurations. We show that this optimal policy learning problem can be reformulated using stochastic dominance constraints, where the optimal assignment prioritizes individuals most likely to comply with the treatment.

Key words: Policy learning; Almost stochastic dominance; Compliance guarantee

JEL Classification: C14, C21, C44, D81

Date: December 18, 2025.
We are grateful to Gordon Anderson, Sukjin Han, Tom Parker, and the participants in the Canadian Econometrics Study Group Meeting, the Stochastic Dominance and Quantile-based Methods in Financial Econometrics Workshop in Honour of Oliver Linton and Yoon-Jae Whang, and the CUHK-Shenzhen Econometrics Workshop for valuable comments and suggestions. We gratefully acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC) under grants 435-2021-0189, 435-2025-0755 (Marmer), and 435-2025-0722 (Song).

1. Introduction

Policy learning refers to designing an optimal policy for a target population using data from a source population. As such, it is inherently a *transfer problem*, which involves transferring features learned about one population to another. For example, suppose that the policymaker (PM) has access to experimental data from a source population. From the experimental data, the PM can estimate the treatment effect for the source population and then seeks an optimal treatment assignment for a target population. The relevance of the source population data to the treatment assignment problem in a different population depends on how much of the experimental setting is "transferable" to the target population. An extreme assumption often made in the literature is that both populations are the same. However, such an assumption is often implausible in practice. There are many examples in the randomized experiments literature in which the statistical significance of causal effects disappears when programs are expanded to wider populations (Duflo, 2004; Allcott, 2015; Muralidharan and Niehaus, 2017; Wang and Yang, 2021). One can plausibly argue that the source and target populations may differ, and hence not all information about the experimental setting in the source population is relevant to the PM.

Departing from the single-population assumption and acknowledging the transfer issue opens up a wide range of possibilities. In this paper, we assume only partial transferability between the target and source populations. Specifically, we assume that the conditional distributions of potential outcomes and total costs given covariates are transferable between the source and target populations, but program participation incentives are not. In other words, the source and target populations may exhibit different take-up behaviors. This could arise, for example, because the experiments are administered by NGO workers who have high stakes in the success of the program and put extra effort into raising the take-up rate, whereas the program is administered by civil servants in the government with fewer stakes in the program's success. In this case, take-up behavior between the two populations can differ substantially.

We assume that the PM faces a budget constraint. Since the PM's costs depend on the take-up rate of subjects, the optimal decision should take into account both the take-up rate and the contribution of treatment to the social objective function. When the take-up rate cannot be transferred from the source population to the target population and is therefore unknown, the PM also faces uncertain compliance behavior. To address this issue, we adopt a maximin approach where the PM aims to find a policy that maximizes the minimum social objective across a range of participation incentives.

First, we specify that individual decisions are made through a generalized Roy model where each subject decides to participate in the program based on net expected payoff differences.

However, the PM does not know the individual payoff functions, except that they belong to a certain class of utility functions. The size of the class is regulated through a bound on maximum variation in marginal utilities. It turns out that the optimal policy design problem in such a setting can be reformulated under the Almost First-Order Stochastic Dominance (AFOSD) constraint of Leshno and Levy (2002). From this reformulation, we obtain an explicit form of an optimal policy. This policy is intuitive: it focuses on those expected to comply fully and prioritizes those with the highest returns to treatment, defined as the ratio of the conditional treatment effect to the conditional total cost given the covariates used to define the treatment assignment function.

Relative to a policy that ignores the compliance issue, our optimal policy can lead to a substantial improvement in the social objective. A policy without a compliance guarantee may allocate treatment to individuals who are likely to refuse it, which leads to underutilization of the budget. The situation is even worse when unused treatment cannot be reassigned once refused. In contrast, our optimal policy focuses on those who are likely to comply and hence avoids underuse or waste of resources. To distinguish between underuse and waste, our model allows for two types of costs: one incurred when the treatment is offered and another incurred when the treatment is actually taken up. The former cost arises regardless of whether the treatment is accepted, whereas the latter cost is incurred only when the treatment is accepted. When the former cost is non-negligible, a policy without a compliance guarantee can lead to significant waste of resources.

Note that in our framework, the PM does not optimize an aggregation of individuals' utilities. First, it may not be obvious for the PM to construct a social utility function from subjective utilities in a way that is transparent and socially agreeable. Second, the PM's decision problem has a normative nature, whereas individuals' participation decisions are positive. Hence, the socially agreeable objective of the PM does not need to conform to an aggregation of individual utilities. Compliance issues may arise because the PM and individuals have different objectives. For example, the PM may want to increase the probability of employment, whereas individuals may want to maximize expected income. In this case, individuals may refuse treatment if they do not expect it to substantially increase their income, even if the treatment is expected to increase their probability of employment. Our framework accommodates such discrepancies in objectives between the PM and individuals.

The issue of optimal policy design has received significant attention in the literature. Manski (2004) proposed an optimal policy that maximizes empirical welfare and suggested using maximal regret as its performance measure. Kitagawa and Tetenov (2018) established finite-sample upper and lower bounds for the maximal regret of the empirical welfare maximizer. Athey and Wager (2021) considered a setting where source population data are generated observationally and proposed an optimal policy motivated by semiparametric efficiency. Recent work by Chernozhukov et al. (2025) proposed an upper confidence bound approach in policy learning, which explicitly incorporates the estimation error as part of the policy objective function. Another related paper is Higbee (2023), who studies policy learning when the decision-maker considers new treatments not present in the experimental data, using shape restrictions to handle partial identification. While Higbee (2023) focuses on extrapolating treatment effects to unobserved intervention levels, we focus on the robustness of policy design to unobserved participation incentives, utilizing stochastic dominance to address the ambiguity in compliance behavior.

There is a line of research that explicitly incorporates budget constraints. Bhattacharya and Dupas (2012) considered welfare-maximizing treatment assignment and derived an optimal solution under budget constraints. More recently, Sun et al. (2025) considered a setting where the PM does not know costs exactly. Sun (2024) studied empirical welfare maximization under budget constraints. The policy learning problem under a budget constraint is mathematically similar to that under fairness considerations (Viviano and Bradic, 2024).

Our main departure from most existing policy learning literature is that we consider an *ex ante* setting where the target population has not yet implemented the policy. Consequently, we do not observe outcomes from the policy in the target population. However, there is a source population that has experimented with the policy, and the outcome data from the policy are available. Thus, the policy learning problem is summarized as one in which the PM searches for an optimal policy for the target population using data generated from the source population. Related literature studies the external validity of randomized experiments (Hotz et al., 2005; Gechter, 2024; Gechter and Meager, 2022). However, that literature focuses on identification and estimation of the treatment effect in the target population, whereas we focus on the optimal policy-design problem.

Our framework also relates to a recent paper by Firpo et al. (2023), who propose a criterion of loss aversion-sensitive dominance to rank policy interventions when individuals exhibit loss aversion. While their work focuses on welfare ranking under specific preference structures, we employ stochastic dominance to address the ambiguity regarding treatment compliance in a transfer learning setting.

The rest of thepaper is organized as follows. Section 2 presents the policy learning problem, making explicit the transferability condition. Toward the optimal policy solution, we first provide a reformulation of the policy design problem using AFOSD constraints, and then present an optimal policy at the population level. Section 2 also discusses an extension of the problem to soft budget constraints. Section 3 proposes an estimated optimal policy using the combined sample from the target and source populations and establishes its consistency. Section 4 provides numerical illustrations of our results.

2. Optimal Treatment Assignments with Compliance Guarantee

2.1. The Policy Objectives

In this section, we introduce the basic policy learning setup. In the target population, each individual is endowed with the potential outcomes $Y(1)$ and $Y(0)$, where $Y(1)$ indicates the potential outcome of the treated state and $Y(0)$ that of the control state. To formally model the incentives, we assume that each individual is endowed with the utility $u(y,v)$, when the potential outcome of the treatment and the other payoff states are realized to be y and v, respectively. Each individual observes the payoff state V as a random vector, but does not observe the potential outcome of the treatment at the time of deciding whether to participate in the treatment program. Throughout the paper, we assume that the payoff state V taking values in a space $\mathcal{V}$ includes a covariate vector X so that

$$V := (X, V_{-X}),$$

where V_{-X} denotes the payoff states other than X. We assume that the PM observes X but does not observe V_{-X}. Thus, the random utility of an agent with the outcome y is given by

$$u(y,V).$$

When offered treatment, an individual decides whether to accept it or not by comparing the expected utilities between the treatment state and the control state. Thus, the acceptance decision can be written as

$$a(V;u) := \mathbf{1}\{\mathbf{E}[u(Y(1),V) \mid V] \geq \mathbf{E}[u(Y(0),V) \mid V]\},$$

where the conditional expectation given V reflects that the individual observes V, knows their own utility function u, and has rational expectations.

A policy by the PM is a map $g : \mathcal{X} \to [0,1]$, which assigns to each covariate group $X = x$ a probability of being in the treatment group. Here, $\mathcal{X}$ denotes the set of values the covariate X takes. In considering the outcome, we assume that the PM adopts a *social objective function*, $q(y,x) : \mathcal{Y} \times \mathcal{X} \to \mathbf{R}$, so that for each realized outcome, the PM deems the outcome y more desirable for group x if $q(y,x)$ is higher. Thus, under the policy g, the PM focuses on the following aggregate quantity, which we refer to as the *expected social outcome*:

$$S(g;u) := \mathbf{E}\Big[q(Y(0),X)\Big(1 - g(X)a(V;u)\Big) + q(Y(1),X)g(X)a(V;u)\Big]. \tag{2.1}$$

The expected social outcome not only depends on the policy g but also on the take-up decision a by individuals. Potential discrepancies in the objectives between the PM and individuals are reflected in the possible differences between q and u.

As for the choice of q, we provide the following three examples. Let $w(x)$ be a given weight function that depends only on the covariates.

(Weighted Average): $q(y,x) := q_W(y,x) := y \cdot w(x)$.
(Sufficientarianism): $q(y,x) := q_S(y,x) := \min\{y, y^*\} \cdot w(x)$.
(Qualification): $q(y,x) := q_L(y,x) := \mathbf{1}\{y > y^*\} \cdot w(x)$.

The choice $q(y,x) = q_W(y,x)$ says that the PM considers the treatment a success if the weighted average of the outcomes for those treated is high. In contrast, by choosing $q(y,x) = q_S(y,x)$, the PM takes as a measure of success the average outcome below a threshold y^*. This measure reflects sufficientarianism which suggests that the PM should not be concerned about the outcome ordering among those groups with high enough outcomes (see Alcantud et al., 2022; Bossert et al., 2023).[1] The qualification objective provides an alternative criterion, defining success as the proportion of individuals whose outcomes exceed a fixed benchmark. These latter two criteria illustrate cases where the PM's evaluative standard may diverge from the incentives of the individuals affected by the policy.

The policy is subject to a budget constraint. We consider two kinds of costs, C_O and C_T, where the random variable $C_O \geq 0$ represents the cost incurred by sending out an offer of treatment and $C_T \geq 0$ the cost incurred by treating the individual. The budget constraint is written as

$$\mathbf{E}[C_O g(X)] + \mathbf{E}[C_T g(X) a(V;u)] \leq B, \tag{2.2}$$

where B denotes the budget allowed for the program. Thus, the main goal of the PM in the target population is to maximize the expected outcome in (2.1) subject to the budget constraint (2.2).

Since the PM does not know the participation decisions $a(V;u)$, we consider a maximin approach, where the PM maximizes the expected outcome under the least favorable participation incentives, which minimize the expected outcome. Let $\mathcal{U}$ denote the class of permitted utility functions. The PM considers an optimal policy that maximizes (over policies $g \in \mathcal{G}$) the minimum (over utilities $u \in \mathcal{U}$) expected social outcome subject to the budget constraint:

$$\sup_{g \in \mathcal{G}} \inf_{u \in \mathcal{U}} S(g;u) \text{ subject to (2.2).} \tag{2.3}$$

[1] We are grateful to Gordon Anderson for introducing the notion of sufficientarianism and the relevant literature to us.

Here, we consider $\mathcal{G}$ to be the collection of (measurable) maps from $\mathcal{X}$ to $\{0,1\}$. The problem appears complex, because u is involved both in the objective function and the budget constraint. Below, we show how we can reformulate the problem into a more tractable one.

2.2. Transfer from the Source Population

We assume that the source population and the target population are different, and hence not all the information on the experimental setting in the source population is relevant to the PM. Here, we clarify the available data for the PM regarding the source population and the transferable aspects of the experimental setting.

We define the conditional CDF of the potential outcomes and the Conditional Average Total Cost (CATC) as follows: for $d = 0,1$,

$$F_d(y \mid x) := P\{Y(d) \le y \mid X = x\} \text{ and } c(x) := \mathbf{E}[C_O + C_T \mid X = x], \quad x \in \mathcal{X}.$$

We assume that $(F_0(\cdot \mid x), F_1(\cdot \mid x), c(\cdot))$ is identified from the source population and transferable to the target population.

Assumption 2.1 (Identification and Transferability)**.** For each $x \in \mathcal{X}$, $(F_1(\cdot \mid x), F_0(\cdot \mid x), c(x))$ is identified in the source population and identical between the source and target populations.

It is well known that the identification of the conditional CDFs, F_d, can be obtained under the unconfoundedness condition: $(Y(1), Y(0)) \perp\!\!\!\perp D \mid X$. This condition excludes partial compliance in the experiment in the source population. We will later discuss an extension of our framework to the case with partial compliance. The assumption also requires a transferability condition that requires $(F_1(\cdot \mid x), F_0(\cdot \mid x), c(x))$ to remain the same as we move from the source to the target population. However, we assume that all other aspects of the source population are not transferable to the target population. For example, the distribution of the covariates can be different between the two populations, and hence the average treatment effect (ATE) is not transferable between the populations.

A major challenge for the PM in this setting is that the incentives for participating in the treatment program differ across the two populations. The difference arises naturally because participation decisions can involve various social and cultural factors, which may be distinct across different populations. Due to the difference in incentives for treatment participation, we cannot use, for example, the propensity scores estimated from the source population to predict individuals' participation behavior in the target population.[2]

[2]However, the propensity score from the source population can still be useful for identifying the conditional average treatment effect in the units of the PM's objective function.

2.3. Stochastic-Dominance Characterization of Policy Learning

Given the budget constraint setting, a focus on the worst possible scenario can lead to a highly conservative decision if the collection of permitted data-generating processes and the class of utility functions $\mathcal{U}$ under consideration are overly large. For example, the maximin solution can be trivial, such as treating no one, if even the slightest harm is expected for a group that is extremely sensitive to any possible harm from the treatment. To alleviate this issue of conservativeness, we introduce additional, mild conditions on the probabilities and utilities so that the optimal solution is reasonable, admits an explicit form, and is practically implementable using the data from the source population.

2.3.1. **Compliance Guarantee.** First, we place restrictions on the data generating process by introducing the following independence condition.

Assumption 2.2 (Utilities and Costs)**.** For the utilities and costs in the target population, we assume that the following conditions are satisfied:

(i) $(Y(1), Y(0)) \perp\!\!\!\perp V_{-X} \mid X$.
(ii) $(C_O, C_T) \perp\!\!\!\perp V_{-X} \mid X$.

According to Assumption 2.2(i), the potential outcomes are independent of the unobserved component of the payoff function conditional on the covariates X. This is reasonable when the effectiveness of the treatment outcomes is less related to the idiosyncratic component of individual characteristics and is transferable to any individual with the same covariate group. An extreme example would be the effect of medical treatment, which is scientifically verified so that it is deemed transferable to a wide population of the same observable category. In this case, the difference in outcome for the same policy across different populations is likely to be from the different participation incentives. Condition (ii) says that the offer cost and treatment cost are not related to the unobserved idiosyncratic payoff component, given the observed covariates. These costs are most likely to arise from the cost of administration and implementation of the treatment on the PM's side. Hence, this condition seems plausible in practice. Note that the individual cost of participating in the program is subsumed in the payoff u (hence, u is a net payoff), and not included in C_O, C_T.

Under Assumption 2.2, we show that the policy learning problem (2.3) can be reformulated as one under a compliance guarantee. We define the conditional average offer cost and the conditional average treatment cost:

$$c_O(x) := \mathbf{E}[C_O \mid X = x] \text{ and } c_T(x) := \mathbf{E}[C_T \mid X = x].$$

Recall that $a(V;u)$ denotes the individual's participation (or acceptance) rule, which depends on the expected payoff from the treatment. Define the probability of participation given $X=x$ as:

$$\pi_u(x) := \mathbf{E}[a(V;u) \mid X=x],$$

and the Conditional Average Treatment Effect (CATE) in the units of the social objective function $q(Y(d),X)$:

$$\tau_q(x) := \mathbf{E}[q(Y(1),X) - q(Y(0),X) \mid X=x]. \tag{2.4}$$

Lastly, it is convenient to introduce the following notation: for a map $f : \mathcal{X} \to \mathbf{R}$,

$$\mu(g;f) := \mathbf{E}[g(X)f(X)].$$

Conditional independence in Assumption 2.2 allows us to rewrite the PM's objective as:

$$W(g,u;\tau_q,\mathbf{c}) := \begin{cases} \mu(g;\tau_q\pi_u), & \text{if } \mu(g;c_O) + \mu(g;c_T\pi_u) \le B \\ -\infty, & \text{otherwise,} \end{cases} \tag{2.5}$$

where $\mathbf{c} := (c_O, c_T)$. The policy learning problem thus becomes:

$$\sup_{g\in\mathcal{G}} \inf_{u\in\mathcal{U}} W(g,u;\tau_q,\mathbf{c}). \tag{2.6}$$

A policy g^* is optimal if it satisfies the following:

$$\inf_{u\in\mathcal{U}} W(g^*,u;\tau_q,\mathbf{c}) = \sup_{g\in\mathcal{G}} \inf_{u\in\mathcal{U}} W(g,u;\tau_q,\mathbf{c}). \tag{2.7}$$

Our budget constraint in (2.5) suggests that we focus only on the "least favorable" subjects who are at the margin in a certain sense. We say that a policy is ***harmless*** if the distribution of the outcome after treatment stochastically dominates the distribution of the outcome before treatment. The least favorable subjects participate in the program if and only if the program is harmless for them. Thus, a program that guarantees full compliance is one that is harmless.

2.3.2. **Almost First-Order Stochastic Dominance.** The solution to (2.6) can be conservative depending on the class of utility functions that are allowed in the target population. Define $\overline{\mathcal{H}}$ to be the set of all non-decreasing, differentiable functions. For each $\epsilon \in (0,1]$, let[3]

$$\mathcal{H}_\epsilon := \left\{ h \in \overline{\mathcal{H}} : \sup_{t\in[-M,M]} h'(t) \le \left(\frac{2-\epsilon}{\epsilon}\right) \inf_{t\in[-M,M]} h'(t) \right\}.$$

[3]Without imposing restrictions on the space of utility functions, the maximin approach that we propose later becomes trivial, because the worst scenario is achieved by considering strictly decreasing utility functions for an individual $X=x$ with $\tau_q(x) > 0$. Then, no one has an incentive to participate and the optimal treatment assignment is to deny treatment to everybody.

Then, when $\epsilon = 0$, we take $\mathcal{H}_0$ to be the set of non-decreasing functions on $[-M, M]$. We can think of the class $\mathcal{H}_\epsilon$ as a set in which the random utility $u(\cdot, V)$ realizes. The class $\mathcal{H}_\epsilon$ entails a restriction on how the marginal random utility varies from a low outcome to high outcome. The restriction gets stronger when ϵ is larger, so that the classes $\mathcal{H}_\epsilon$, $\epsilon \in [0, 1]$, are nested: if $\epsilon_1 < \epsilon_2$, then $\mathcal{H}_{\epsilon_2} \subseteq \mathcal{H}_{\epsilon_1}$.

The connection between $\mathcal{H}_\epsilon$ and stochastic dominance constraints is established by Leshno and Levy (2002) and plays a crucial role in our setting. Let $F_0(\cdot \mid x)$ and $F_1(\cdot \mid x)$ denote the conditional CDF of control and treated outcomes, respectively, given covariate $x \in \mathcal{X}$. Define the extent of the violation of first-order stochastic dominance (FOSD) at x as:

$$\delta(x) := \begin{cases} \dfrac{\int \left[\Delta(y \mid x)\right]^+ dy}{\int |\Delta(y \mid x)| dy}, & \text{if denominator is positive,} \\ 0, & \text{otherwise,} \end{cases} \tag{2.8}$$

where

$$\Delta(y \mid x) := F_1(y \mid x) - F_0(y \mid x).$$

According to Leshno and Levy (2002), we have the equivalence:

$$\delta(x) \leq \epsilon \ \text{ if and only if } \mathbf{E}[h(Y(1)) \mid X = x] \geq \mathbf{E}[h(Y(0)) \mid X = x] \text{ for all } h \in \mathcal{H}_\epsilon. \tag{2.9}$$

For $\epsilon \in [0, 1]$, we define

$$\mathcal{X}_c := \left\{x \in \mathcal{X} : \delta(x) \leq \epsilon\right\}.$$

The set $\mathcal{X}_c$ consists of the covariate groups such that the degree of the FOSD violation is bounded by ϵ.

We introduce the following assumption on $\mathcal{U}$, the utility functions in the target population. Let $\tilde{\mathcal{U}}$ be the class of real-valued measurable functions on $\mathcal{Y} \times \mathcal{V}$.

Assumption 2.3. For the target population, there is a constant $\epsilon \in [0, 1]$ known to the PM such that

$$\mathcal{U} := \{u \in \tilde{\mathcal{U}} : u(\cdot, v) \in \mathcal{H}_\epsilon \text{ for all } v \in \mathcal{V}\}.$$

This assumption says that $\mathcal{U}$ is flexible enough that each individual in the target population has a random utility that can take the value of any element in $\mathcal{H}_\epsilon$. We assume that the PM knows this constant ϵ. Under Assumption 2.3, the PM knows that the individuals in $\mathcal{X}_c$ will comply with the treatment if offered. For individuals with $x \notin \mathcal{X}_c$, the PM cannot determine their compliance behavior.

Assumption 2.3 has the following implication for the budget constraint in (2.5).

Lemma 2.1. *Under Assumption 2.3, one of the following two statements holds:*

(i) $\inf_{u\in\mathcal{U}} W(g,u;\tau_q,\mathbf{c}) = -\infty$ *and* $\inf_{u\in\mathcal{U}} W(g,u;\tau_q,(c_O+c_T,0)) = -\infty$*; or*

(ii) $\inf_{u\in\mathcal{U}} W(g,u;\tau_q,\mathbf{c}) = \inf_{u\in\mathcal{U}} W(g,u;\tau_q,(c_O+c_T,0))$*, where* $\mathbf{c} := (c_O, c_T)$*.*

According to Lemma 2.1, the acceptance rule a is eliminated from the budget constraint. This is intuitive: in this framework, budget violations outweigh all benefits, so a policy must remain feasible even under full compliance.[4] With full compliance, the constraint involves the full cost $c_O + c_T$, and there is no distinction between c_O and c_T. We therefore define

$$c := c_O + c_T,$$

where $c(x)$ can be interpreted as the group level marginal treatment cost when the group x complies with the treatment.

The following lemma describes the worst-case (over permitted utilities) PM's objective given a policy g that satisfies the full compliance budget constraint. The lemma plays a crucial role in characterizing the optimal treatment assignment.

Lemma 2.2. *Suppose that Assumptions 2.1-2.3 hold, and that a policy g satisfies the full compliance budget constraint $\mu(g;c) \le B$. Then,*

$$\inf_{u\in\mathcal{U}} W(g,u;\tau_q,\mathbf{c}) = \mathbf{E}\Big[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}\Big] + \mathbf{E}\Big[g(X)\tau_q(X)\mathbf{1}\{X\notin\mathcal{X}_c, \tau_q(X)<0\}\Big].$$

The intuition behind the lemma is as follows. We can partition individuals into three groups: (i) those with $x\in\mathcal{X}_c$, who will comply with treatment; (ii) those with $x\notin\mathcal{X}_c$ and $\tau_q(x)\ge 0$, who will not comply in the worst-case scenario for the PM; and (iii) those with $x\notin\mathcal{X}_c$ and $\tau_q(x)<0$, who will comply in the worst-case scenario (because their utility worsens under treatment). Under the minimax approach, the PM must therefore consider only the $\tau_q(x)$ contributions from the first and third groups when designing the optimal policy.

2.4. Optimal Policy

Our main result is to provide an explicit characterization of the optimal solution g^* in (2.7). We now describe a solution to (2.7). Define

$$\rho(x) := \frac{\tau_q(x)}{c(x)}.$$

Thus $\rho(x)$ is the CATE (in the units of the PM's objective function q) relative to cost. We can view it as the inverse of the cost-effectiveness ratio (CER) used in health-care evaluations.[5] We

[4] In Section 2.5 below, we present an extension that permits limited budget violations.

[5] There is debate about the use of ICER (incremental cost-effectiveness ratio) in health-care policies.

call $\rho(x)$ the ***returns-to-treatment (RTT)***. The result below provides an explicit form of the optimal policy in (2.7). Define

$$\beta(k) := \mathbf{E}[\mathbf{1}\{X \in \mathcal{X}_c, \rho(X) > k\}c(X)].$$

The quantity $\beta(k)$ gives the total expense from treating all the compliant individuals with RTTs above k.

Theorem 2.1. *Under Assumptions 2.1-2.3, an optimal policy g^* to (2.7) is given by the following. Define the treatment threshold for RTTs as*

$$k^* := \inf\Big\{k \geq 0 : \beta(k) \leq B\Big\}.$$

(i) If $k^* > 0$, then the optimal policy is given by

$$g^*(x) = \begin{cases} 1 & \text{if } x \in \mathcal{X}_c \text{ and } \rho(x) > k^*, \\ r(x) & \text{if } x \in \mathcal{X}_c \text{ and } \rho(x) := k^*, \\ 0 & \text{otherwise,} \end{cases}$$

where $r(x)$ is a randomization probability such that

$$\mathbf{E}[\mathbf{1}\{X \in \mathcal{X}_c, \rho(X) = k^*\}c(X)r(X)] = B - \beta(k^*).$$

(ii) If $k^* = 0$, then the optimal policy is given by

$$g^*(x) = \begin{cases} 1 & \text{if } x \in \mathcal{X}_c \text{ and } \tau_q(x) > 0, \\ 0 & \text{if } x \in \mathcal{X}_c \text{ and } \tau_q(x) \leq 0, \\ r(x) & \text{if } x \notin \mathcal{X}_c \text{ and } \tau_q(x) > 0, \end{cases}$$

where $r(x)$ is a randomization probability such that

$$\mathbf{E}\Big[\mathbf{1}\big\{X \notin \mathcal{X}_c, \tau_q(X) > 0\big\}c(X)r(X)\Big] = B - \beta(0).$$

The threshold k^* represents the minimum threshold k on RTT for which the total cost does not exceed the budget B when we treat groups $x \in \mathcal{X}_c$ with RTTs above this threshold. Thus, when $k^* > 0$, our optimal policy is to only treat the groups willing to comply with the treatment (that is, $x \in \mathcal{X}_c$) with RTTs $\rho(x) \geq k^*$. It is possible that the boundary group $\{x \in \mathcal{X}_c : \rho(x) = k^*\}$ has a non-zero probability. In such cases, the individuals in the boundary group receive treatment with a probability $r(x)$, where the randomization probability $r(x)$ is chosen to satisfy the remaining budget after treating the groups with RTTs strictly above the threshold. The policy ensures that only the individuals willing to comply with the treatment are treated and, among those individuals, those with the highest RTTs are prioritized.

When $k^* = 0$, the budget is large enough to treat all groups in $\mathcal{X}_c$ with positive treatment effects $\tau_q(x)$; that is, all individuals with a positive treatment effect that will comply. The remaining budget can be used to treat some individuals who have positive RTTs $\rho(x)$ but may not comply ($x \notin \mathcal{X}_c$). Our rule proposes randomizing treatment allocations among those individuals. However, since in the worst-case scenario the groups $x \notin \mathcal{X}_c$ will not comply, any rule that respects the budget constraint yields the same worst-case outcome. For example, instead of randomizing, the PM may choose to prioritize groups $x \notin \mathcal{X}_c$ with the highest RTTs.

2.5. Soft Budget Constraints

We have assumed a hard budget constraint, where no violation of the budget constraint is allowed. However, in reality, there might be some flexibility in budget violation. We can modify the procedure depending on the degree of violation that the PM is willing to tolerate.

Let $\overline{\lambda} > 0$ be an upper bound on the reduction in the expected social outcome per unit of budget violation. We define the Lagrangian-based version of the PM's objective as

$$W_{\mathcal{L}}(g) := \inf_{u \in \mathcal{U}} \inf_{\lambda \in [0,\overline{\lambda}]} \Big(\mu(g; \tau_q \pi_u) + \lambda(B - \mu(g; c_O + \pi_u c_T)) \Big).$$

Lemma 2.3. *Under Assumptions 2.3,*

$$\sup_g W_{\mathcal{L}}(g) = \sup_g \tilde{W}_{\mathcal{L}}(g),$$

where

$$\begin{aligned} \tilde{W}_{\mathcal{L}}(g) := {} & \mathbf{E}\big[g(X) \big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\} \big) \tau_q(X) \big] \\ & + \inf_{\lambda \in [0,\overline{\lambda}]} \Big\{ \lambda \Big(B - \mathbf{E}\big[g(X) \big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\} \big) c(X) \big] \Big) \Big\}, \end{aligned} \tag{2.10}$$

and $c(\cdot) = c_O(\cdot) + c_T(\cdot)$.

Unlike Lemma 2.2, we do not have an equality such as $W_{\mathcal{L}}(g) = \tilde{W}_{\mathcal{L}}(g)$ that holds for all g. This is primarily because under $W_{\mathcal{L}}(g)$, every dollar that g overspends is punished by $\overline{\lambda}$, regardless of whom that dollar is spent on. On the other hand, such overspending is not punished under $\tilde{W}_{\mathcal{L}}$ if that dollar is spent on x such that $x \notin \mathcal{X}_c$ and $\tau_q(x) > 0$.

Nonetheless, Lemma 2.3 provides an easy way to verify whether a policy maximizes $W_{\mathcal{L}}$. Because $\tilde{W}_{\mathcal{L}}$ does not involve the infimum over u, its maximization is much simpler. Let such a solution be denoted by $\tilde{g}^*_{\mathcal{L}}$. If a policy $g^*_{\mathcal{L}}$ satisfies $W_{\mathcal{L}}(g^*_{\mathcal{L}}) = \tilde{W}_{\mathcal{L}}(\tilde{g}^*_{\mathcal{L}})$, then $g^*_{\mathcal{L}}$ must maximize $W_{\mathcal{L}}$.

Theorem 2.2. *A maximizer of $W_{\mathcal{L}}$ is given by*

$$g^*_{\mathcal{L}}(x) = \begin{cases} g^*(x), & \text{if } k^* \le \overline{\lambda}, \\ \mathbf{1}\{x \in \mathcal{X}_c, \rho(x) > \overline{\lambda}\}, & \text{otherwise}, \end{cases} \tag{2.11}$$

where g^* is as given in Theorem 2.1.

If, in addition to Assumption 2.3, the cost function $c(X)$ is bounded between $\underline{c}$ and $\overline{c}$ with $0 < \underline{c} \le \overline{c}$ almost surely, we can simplify this as

$$g^*_{\mathcal{L}}(x) = \mathbf{1}\{x \in \mathcal{X}_c, \rho(x) > \min(k^*, \overline{\lambda})\},$$

where k^* is as given in Theorem 2.1.

3. Empirical Implementation

3.1. Estimation of Optimal Policy

The optimal solution described in the previous section depends on unknown objects F_0, F_1, τ_q, δ, and c. In this section, we describe a practical implementation that uses a plug-in estimator of the optimal policy, which replaces the unknown objects with their estimators. We assume that the econometrician can construct the estimators $\hat{F}_0$, $\hat{F}_1$, $\hat{\tau}_q$, $\hat{\delta}$, and $\hat{c}$ based on i.i.d. data from the source population, $(X_i, Y_i, D_i)_{i=1}^{n_S}$, where n_S denotes the size of the random sample from the source population. Further, define $\hat{\rho} := \hat{\tau}_q / \hat{c}$. For the target population, we assume that the econometrician observes a random sample of the covariates $(X_i)_{i=1}^{n_T}$, where n_T denotes the size of the sample from the target population.

Recall that we measure FOSD violations using $\delta(x)$ defined in (2.8). Its estimator is given by

$$\hat{\delta}(x) := \frac{\displaystyle\int \left[\hat{\Delta}(y \mid x)\right]^+ dy}{\displaystyle\int \left|\hat{\Delta}(y \mid x)\right| dy},$$

where $\hat{\Delta}(y \mid x)$ is the estimator of $\Delta(y \mid x)$:

$$\hat{\Delta}(y \mid x) := \hat{F}_1(y \mid x) - \hat{F}_0(y \mid x).$$

To describe the estimator of the optimal policy g^*, we define[6]

$$\mathcal{X}_\varepsilon := \{x \in \mathcal{X} : \delta(x) \le \varepsilon\},$$

[6]Note that $\mathcal{X}_\epsilon = \mathcal{X}_c$.

and its estimator

$$\hat{\mathcal{X}}_{\varepsilon} := \left\{x \in \mathcal{X} : \hat{\delta}(x) \le \varepsilon\right\}.$$

For $\xi, \kappa > 0$ and $k \in \mathbf{R}$, we define the following estimator of the expected cost of treating the individuals in the target population with $x \in \hat{\mathcal{X}}_{\epsilon+\kappa}$ and $\hat{\rho}(X_i) > k$:

$$\hat{\beta}_{\epsilon+\kappa}(k) := \frac{1}{n_T} \sum_{i=1}^{n_T} \mathbf{1}\left\{X_i \in \hat{\mathcal{X}}_{\epsilon+\kappa}, \hat{\rho}(X_i) > k\right\} \hat{c}(X_i).$$

Thus, $\hat{\beta}_{\epsilon+\kappa}(k)$ is a sample version of $\beta(k)$. We construct the estimated threshold $\hat{k}_{\epsilon+\kappa,\xi}$ as

$$\hat{k}_{\epsilon+\kappa,\xi} := \inf\left\{k \ge \xi : \hat{\beta}_{\epsilon+\kappa}(k) \le B\right\}. \tag{3.1}$$

Lastly, the estimator of the policy g^* is given by

$$\hat{g}_{\epsilon+\kappa,\xi}(x) := \mathbf{1}\left\{x \in \hat{\mathcal{X}}_{\epsilon+\kappa}, \hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}\right\}.$$

We derive the properties of the estimated policy under the following assumptions. We first introduce the assumption on the distributional properties of the observed random variables and potential outcomes.

Assumption 3.1 (Sampling)**.**

(i) The econometrician observes i.i.d. data $(X_i, Y_i, D_i)_{i=1}^{n_S}$ and $(X_i)_{i=1}^{n_T}$ from the source and target populations, respectively.
(ii) The support of the distributions of $Y(0)$ and $Y(1)$ is in $[-M, M]$ for some $M > 0$.

The bounded support condition in Assumption 3.1(ii) is imposed for convenience and can be relaxed. Similarly, we assume that treatment costs are bounded.

Assumption 3.2 (Bounds for Costs)**.** For some $\bar{c}, \underline{c} > 0$, $P(\underline{c} \le c(X) \le \bar{c}) = 1$.

To implement the optimal policy in practice, one needs to estimate $\delta(x)$, which measures the extent of FOSD violations. Because $\delta(x)$ depends on the distance between the distributions F_0 and F_1 through its denominator, the finite sample properties of any estimator of $\delta(x)$ may deteriorate when this distance is small. Recall that when $k^* > 0$, under the optimal policy, treatment will be assigned only to groups with sufficiently high RTT.

Assumption 3.3 below imposes a restriction on the PM's objective function q to ensure that large values of the treatment effect $\tau_q(x)$ imply that the distance between F_0 and F_1 distributions is bounded away from zero.

Assumption 3.3. There exists a function $\ell : (0, \infty) \to (0, \infty)$ such that for all $x \in \mathcal{X}$ and $\xi > 0$, $\tau_q(x) \ge \xi$ implies that $\int_{-M}^{M} |\Delta(y \mid x)| dy \ge \ell(\xi)$.

Assumption 3.3 is satisfied, for example, when $\mathcal{X}$ is finite, that is, the covariates have a discrete distribution with a finite support. When the covariates are continuously distributed, Assumption 3.3 holds if, for each x, $q(y,x)$ is non-decreasing in y, and the induced Lebesgue-Stieltjes measure is absolutely continuous with respect to the Lebesgue measure, with a density that is bounded uniformly in x. This includes objective functions based on sufficientarianism, but not those defined by the qualification criterion; for the latter, we verify Assumption 3.3 directly in the following lemma.

Lemma 3.1 (Qualification Objective)**.** *Suppose that the set $\{x \in \mathcal{X} : \tau_q(x) \geq \xi\}$ is compact for each $\xi > 0$, and the conditional CDF difference $\Delta(y|\cdot)$ is continuous for each y. Then,*

(i) Assumption 3.3 is satisfied for $q(y,x) = q_L(y,x) = \mathbf{1}\{y > y^*\}w(x)$.

(ii) Furthermore, if there exists $\eta > 0$ such that $F_1(\cdot|x)$ is right differentiable at y^* and differentiable at each $y \in (y^*, y^* + \eta]$ for each x with derivatives bounded uniformly over x by $D > 0$, then we can take

$$\ell(\xi) = \xi^2/(2D),$$

for small enough $\xi > 0$.

We impose the following condition to ensure strong identification of k^* when the budget constraint is binding.

Assumption 3.4 (Identification of k^*)**.** Suppose $k^* > 0$. There exists a function $s : (0,\infty) \to (0,\infty)$ such that for all $\eta > 0$,

$$\beta(k^* + \eta) + s(\eta) \leq \beta(k^*) \leq \beta(k^* - \eta) - s(\eta).$$

In the assumption above, the behavior of the function $s(\cdot)$ near zero measures the strength of the identification of k^*. If $s(\eta)$ is close to zero for small η's, the function $\beta(\cdot)$ is nearly flat around k^*. As a result, it would be difficult to learn the RTT threshold k^* in the optimal policy from the data. This is reflected in Theorem 3.1 below, where $s(\cdot)$ enters the expression for the tail probability of the estimation error for k^*.

We use the following objects to describe the properties of the estimated threshold $\hat{k}_{\epsilon+\kappa,\xi}$. For $\xi \geq 0$, define

$$\varphi_\rho(\xi) := \sup_{t \in \mathbf{R}} P\{t \leq \rho(X) \leq t + \xi\} \text{ and} \tag{3.2}$$
$$\varphi_\delta(\xi) := \sup_{t \geq 0} P\{t \leq \delta(X) \leq t + \xi\}.$$

Let

$$\hat{\nu} := \sup_{x \in \mathcal{X}} \int |\hat{\Delta}(y \mid x) - \Delta(y \mid x)| dy.$$

Lastly, for $\xi, \kappa \geq 0$, we define

$$p_n(\kappa, \xi) := 1 - 2P\left\{\frac{2M\,\hat{\nu}}{\ell(\xi)}\left(\frac{1}{\ell(\xi) - \hat{\nu}} + 3\right) > \kappa, \xi > \hat{\nu}\right\}. \tag{3.3}$$

The following theorem provides a finite sample bound for the tail probability of the estimation error of $\hat{k}_{\epsilon+\kappa,\xi}$.

Theorem 3.1. *Suppose that Assumptions 3.1-3.4 hold. Then, there exists a universal constant $C > 0$ such that for each $\eta, \xi, \kappa > 0$,*

$$P\{|\hat{k}_{\epsilon+\kappa,\xi} - k^*| > \eta\} \leq CP\left\{\sup_{x\in\mathcal{X}}|\hat{c}(x) - c(x)| > \frac{B - \beta(\eta)}{2} \wedge s(\eta)\right\} + C\tilde{p}_n\left(\kappa, \frac{B - \beta(\eta)}{2} \wedge s(\eta), \xi\right),$$

where

$$\tilde{p}_n(\kappa, b, \xi) := \alpha\left(n_T^{1/2}\left(\frac{b}{4\bar{c}} - \varphi_\rho(\xi/2)\right)\right) + (1 - p_n(\kappa, \xi)) + \alpha\left(n_T^{1/2}\left(\frac{b}{16\bar{c}} - \varphi_\delta(2\kappa)\right)\right) + P\left\{\sup_{x\in\mathcal{X}}|\rho(x) - \hat{\rho}(x)| > \frac{\xi}{2}\right\}, \tag{3.4}$$

and $\alpha(t) = (t \vee 1) \cdot \exp(-2t^2)$, $t \in \mathbf{R}$.

The theorem establishes a comprehensive finite sample bound on the estimation error of $\hat{k}_{\epsilon+\kappa,\xi}$ that yields different convergence rates depending on the underlying distribution and the choice of constants. For example, the rate of convergence becomes faster if the identification of k^* is strong (like when $s(\eta)$ is continuous and strictly increasing in η) or the estimation errors of $\hat{c}$ and $\hat{\rho}$ converge to zero faster in probability. It also depends on the anti-concentration behavior of $\rho(X)$ and $\delta(X)$ (through $\varphi_\rho(\xi)$ and $\varphi_\delta(\xi)$), especially when ξ is chosen to decrease to zero as $n_T \to \infty$.

Next, we present a result that shows the behavior of the estimated policy relative to the optimal policy. The behavior is measured in terms of the probability that a covariate group is mis-assigned into a treatment group or a non-treatment group.

Theorem 3.2. *Suppose that the conditions of Theorem 3.1 hold. Then, for each $\eta_1, \eta_2 > 0$, and on the event that $|\hat{k}_{\epsilon+\kappa,\xi} - k^*| \leq \eta_1$ and $\sup_{x\in\mathcal{X}}|\hat{\rho}(x) - \rho(x)| \leq \eta_2$,*

$$\int \mathbf{1}\left\{x : \hat{g}_{\epsilon+\kappa,\xi}(x) \neq g^*(x)\right\} dP_X(x) \leq \varphi_\delta(2\kappa) + \varphi_\rho(\max\{\xi - k^*, 0\}) + 2\,\varphi_\rho(\eta_1 + \eta_2),$$

with probability $p_n(\kappa, \xi)$.

The theorem says that the probability of the covariate groups on which the estimated policy and the true optimal policy differ goes to zero as $n \to \infty$.

3.2. Choice of Tuning Parameters

The procedure described in Section 3.1 can be viewed as constructing a set C_{n_S}, defined as

$$C_n := \left\{x \in \hat{\mathcal{X}}_{\varepsilon_{n_S}} : \hat{\rho}(x) > \xi_{n_S}\right\},$$

where recall that n_S denotes the sample size for source data, and the sequence $\{\varepsilon_{n_S}\}$ is such that $\varepsilon_{n_S} \geq \epsilon$ for all n_S. The goal of the construction is to satisfy the following two properties: as $n_S \to \infty$,

$$P\left(\left\{x \in \mathcal{X}_c : \rho(x) > \xi_{n_S}/2\right\} \subseteq C_{n_S}\right) \to 1, \tag{3.5}$$

and

$$P\left(X \in \left\{x \notin \mathcal{X}_c : \rho(x) > \xi_{n_S}/2\right\} \cap C_{n_S}\right) \to 0. \tag{3.6}$$

In words, C_{n_S} includes valid covariates with large enough RTT, and excludes invalid covariates appropriately. Given these properties, an empirical rule can be obtained by restricting treatment offers to covariates in C_{n_S}, with the RTT cutoff estimated accordingly.

The tuning parameters ξ_{n_S} and ε_{n_S} play crucial roles. For each covariate x, when the conditional CDFs are not identical across treatment and control,

$$\delta(x) = \frac{\displaystyle\int \left[\Delta(y \mid x)\right]^{+} dy}{\displaystyle\int |\Delta(y \mid x)| dy},$$

and it is possible for both the denominator here and $\delta(x)$ itself to be arbitrarily small. This means $\hat{\delta}(x)$ may fail to be uniformly consistent across x despite uniform consistency of the conditional CDF estimates. By enforcing that covariates need to satisfy $\hat{\rho}(x) \geq \xi_{n_S}$ to be offered treatment, we impose a positive lower bound on the denominator. If this lower bound goes to 0 at a slow enough rate, $\hat{\delta}(x)$ will be uniformly consistent across x, and can detect FOSD violations of size $\varepsilon_{n_S} - \epsilon$ given that this does not go to zero too quickly, and leads to (3.5) and (3.6).

Furthermore, if budget B is such that

$$\mathbf{E}[B - \mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > 0\}] > 0 \text{ but } \mathbf{E}[B - \mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) \geq 0\}] < 0,$$

and there is a point mass at $\rho(X) = 0$, then

$$\int |\mathbf{1}\{x \in \mathcal{X}_c : \rho(x) > 0\} - \mathbf{1}\{x \in \mathcal{X}_c : \hat{\rho}(x) > 0\}| dP_X(x)$$

may not converge to 0 in probability, as we could be offering treatment to covariates with zero conditional average treatment effect, and budget violation may not go to 0 in probability. This can be dealt with by again requiring that $\hat{\rho}(x) > \xi_{n_S}$ for x to be considered for a treatment offer: this is the same as only offering treatment to covariate values x that display strong evidence of positive effects, since ξ_{n_S} will converge at a slower rate than $\hat{\rho}$.

Based on Theorem 3.1, a feasible choice of tuning parameters is of the form

$$\varepsilon_{n_S} := \epsilon + c_1\left(\frac{\log n_S}{n_S}\right)^{\alpha/2} \quad \text{and} \quad \xi_{n_S} := c_2\left(\frac{\log n_S}{n_S}\right)^{\alpha/4},$$

for some positive constants c_1 and c_2. If $\rho(X)$ has no point mass in a neighborhood around 0, we can set $\xi_{n_S} = 0$ and use a faster rate for $\varepsilon_{n_S} \to \epsilon$ such as

$$\varepsilon_{n_S} := \epsilon + c_1\left(\frac{\log n_S}{n_S}\right)^{\frac{4\alpha}{5}}.$$

While the empirical rule may mistakenly offer treatments to covariates which violate the AFOSD constraint due to the denominator in $\delta(x)$, which is $\|F_1(\cdot|x) - F_0(\cdot|x)\|_{L_1}$, being too close to 0, such x must have $\rho(x)$ close to 0, and overall that proportion goes to 0 in probability.

Similarly, if the budget constraint is known to be binding, we can also set ξ_{n_S} to be any arbitrary non-negative constant smaller than k^*, including 0, and set a faster rate for $\varepsilon_{n_S} \to \epsilon$. Even though C_n may contain covariates $x \notin \mathcal{X}_c$ due to the denominator in $\delta(x)$ being close to 0, such x would never be considered for a treatment offer anyway due to the budget constraint.

3.3. Confidence-Band Approach

The key goal of constructing C_n is to ensure that the conditions (3.5) and (3.6) are satisfied. An alternative approach instead defines

$$C_n := \left\{x \in \mathcal{X} : L(x) \le \epsilon, \hat{\rho}(x) > \xi_{n_S}\right\},$$

where $L(x)$ is a lower bound of a uniform confidence band for $\delta(x)$. Two common methods of constructing $L(x)$ are the unstudentized and studentized bootstraps. In the unstudentized bootstrap case, let $\hat{\varrho}_t$ be the bootstrap estimator of the t-th quantile of the distribution of

$$\sup_{x:\hat{\rho}(x)>\xi_{n_S}} \left|\hat{\delta}(x) - \delta(x)\right|.$$

One can choose the quantile order t close to one, and set $L(x) := \hat{\delta}(x) - \hat{\varrho}_t$. For the studentized bootstrap method, first one estimates pointwise standard errors $\hat{\sigma}(x)$ via the bootstrap. Next, let $\hat{\varsigma}_t$ denote the bootstrap estimator of the t-th quantile of the distribution of

$$\sup_{x:\hat{\rho}(x)>\xi_{n_S}} \left| \frac{\hat{\delta}(x)-\delta(x)}{\hat{\sigma}(x)} \right|.$$

Again, one can choose a large quantile order t, and set $L(x) := \hat{\delta}(x) - \hat{\varsigma}_t \hat{\sigma}(x)$.

When computing $\hat{\varrho}_t$ and $\hat{\varsigma}_t$, we only consider the groups $\{x : \hat{\rho}(x) > \xi_{n_S}\}$. This excludes covariate values x for which the denominator of $\delta(x)$ is close to zero and the estimator $\hat{\delta}(x)$ may be ill-behaved.

4. Numerical Illustration

We illustrate how the optimal policy works by comparing its performance to that of a naive policy. The naive policy offers treatment to those with the highest RTT. It mirrors the approach of Sun et al. (2025), who assume full uptake once treatment is offered. On the other hand, the optimal policy considers the individual's take-up decision, based on the AFOSD (called the AFOSD-based policy here). This rule first checks whether the treatment outcome for each potential recipient satisfies the AFOSD criterion up to a tolerance level ϵ. Then, eligible agents are ranked by RTT. This policy takes into account the possibility that some individuals may reject the offer of treatment.

We consider different variance configurations for the conditional distribution of the potential outcomes given the covariates. It is not easy to induce highly risk-averse individuals to participate in the treatment, when the treatment is highly effective on average, yet can carry some negative net utilities with a positive probability. Thus, risk-averse people have different incentives to participate in the treatment, if the treatment outcomes differ in their variances, yielding different compliance behaviors.

4.1. Data Generating Process

First, we generate covariates and control outcomes as

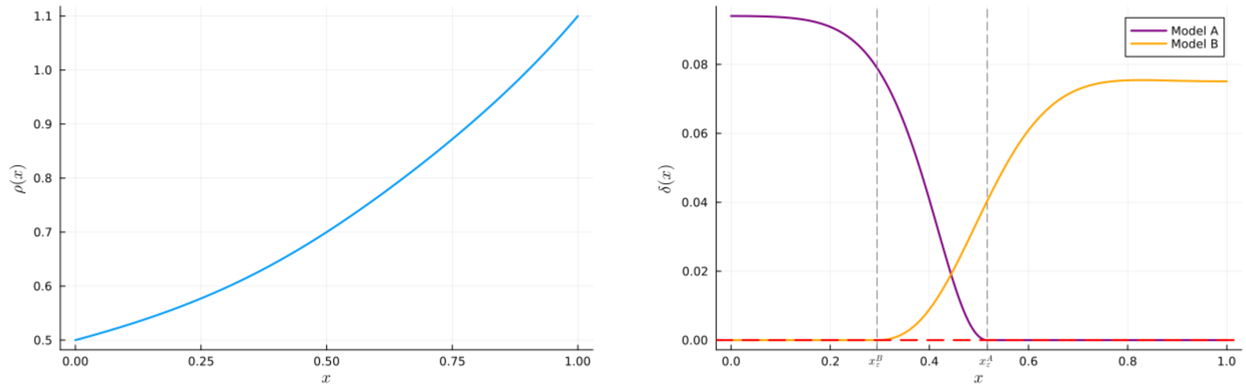
$$X_i \sim \text{Uniform}(0,1) \quad \text{and} \quad Y_i(0) \sim \text{Uniform}(1.2, 1.4)$$

which are independent. Conditional on $X_i = x$, treated outcome $Y_i(1)$ has the mean:

$$m_1(x) := 1.8 + 0.05 \frac{x^2}{x^2 + (1-x)^2},$$

but for their variances, we consider two specifications:

$$\textbf{Model A: } \sigma^2_{1,A}(x) := \left[0.7 \cos^2\left(\frac{\pi}{2} \frac{x^2}{x^2 + (1-x)^2} \right) \right]^2, \text{ and}$$



(A) Returns-to-treatment (RTT) by covariate

(B) FOSD violation by covariate

FIGURE 1. Returns-to-treatment and stochastic-dominance violation

Note: The left figure shows the returns-to-treatment (RTT) against covariate values, whereas the right figure plots FOSD violation against covariates. In our design, RTT increases with x in both models A and B. However, due to the different behavior of the conditional variance of $Y(1)$ given $X = x$ between models, the degree of FOSD violation increases with x in Model B, whereas it decreases with x in Model A.

$$\textbf{Model B: } \sigma^2_{1,B}(x) := \left[0.7\cos^2\left(\frac{x^2}{x^2+(1-x)^2}-1\right)\right]^2.$$

In Model A, the conditional variance of the treated outcome declines with x while CATE increases in x. In Model B, the CATE and the conditional variance of the treated outcome move in the same direction: In both cases, conditional on $X_i = x$, the treated outcomes in both models are normally distributed, but truncated $\pm 2\sigma_1(x)$ around their respective means.

Next, conditional on X_i, treatment costs C_i are drawn from Uniform$(0.5-0.5x, 1.5-0.5x)$, so that $\mathbf{E}[C_i \mid X_i = x] = 1-0.5x$.

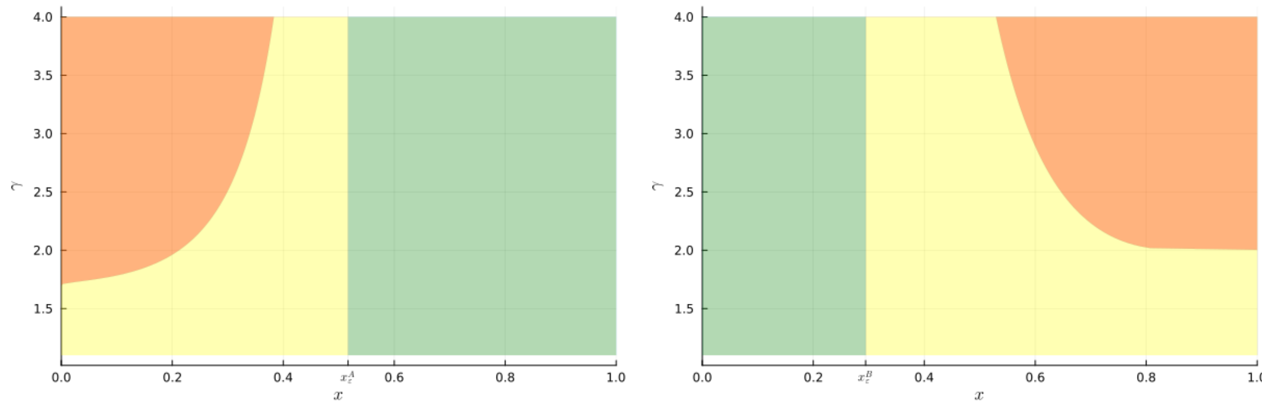
Agents evaluate outcomes using expected utility from a CRRA utility function

$$h(y;\gamma_i) := \frac{(y-0.44)^{1-\gamma_i}-1}{1-\gamma_i}$$

where each $\gamma_i \sim$ Uniform$(1.1, 4)$ is drawn independently of all other variables in both models. Consequently, an agent with (x,γ) will accept a treatment offer if and only if

$$\mathbf{E}[h(Y_i(1),\gamma_i)|X_i = x, \gamma_i = \gamma] \geq \mathbf{E}[h(Y_i(0),\gamma_i)|X_i = x, \gamma_i = \gamma].$$

Both Models A and B share the same conditional means for treatment effect and cost. As x increases, the conditional average treatment effect rises, while expected cost falls, so the RTT is increasing in x, see Figure 1(A). However, the models differ in how the conditional variance behaves, resulting in opposite patterns of stochastic dominance violation. This leads to a threshold x^A_ϵ such that $\delta(x) \leq \epsilon$ if and only if $x \geq x^A_\epsilon$. By contrast, CATE and stochastic



(A) Covariate-utility decision under Model A

(B) Covariate-utility decision under Model B

FIGURE 2. FOSD Violation and Treatment Acceptance

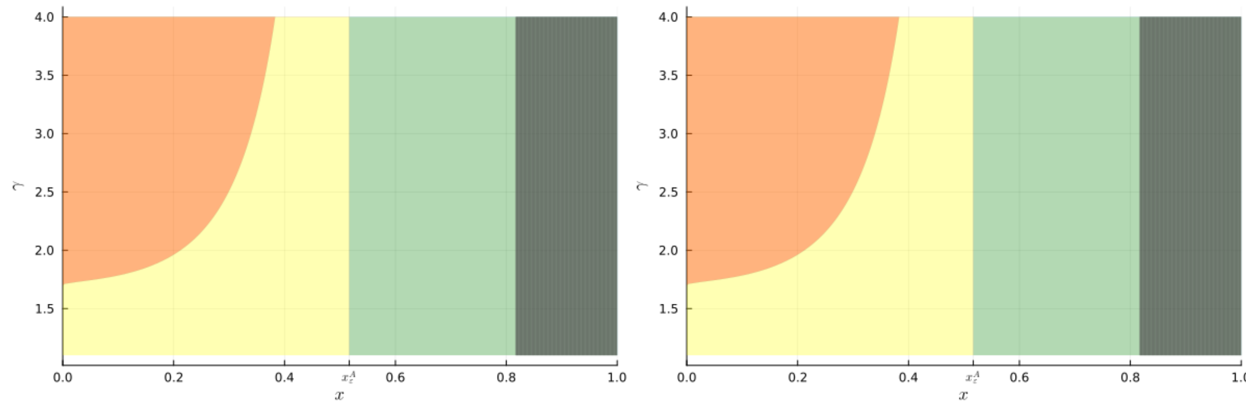
Note: In each plot, the green region indicates covariates x for which $\delta(x) \leq \epsilon$. Agents in the yellow region have $\delta(x) > \epsilon$ but are not risk averse enough to reject treatment. This is in contrast to agents in the red region who have $\delta(x) > \epsilon$ and are highly risk averse, so that they would reject treatment if offered.

dominance do not align in Model B, as conditional variance of $Y_i(1)$ increases with x, and $\delta(x) \leq \epsilon$ if and only if $x \leq x_\epsilon^B$ for some threshold x_ϵ^B, see Figure 1(B).

For the calculations below, we assume $\epsilon = 0$. Based on the characterization of δ in Section 2, $\mathcal{H}_\epsilon$ contains every $h(y;\gamma)$ for γ in the support of γ_i. As a result, any individual with x such that $\delta(x) \leq \epsilon$ will accept treatment. On the other hand, when $\delta(x) > \epsilon$, whether an individual accepts treatment depends on his utility function. Since higher γ implies a greater risk aversion and because the conditional variance of $Y(1)$ grows in x, a portion of the population, characterized by high x and γ, will in fact reject treatment. The rest, despite having $\delta(x) > \epsilon$, still accepts treatment if offered. Figure 2 illustrates these three regions in the (x, γ) plane for each model, where: (i) the green region contains those with $\delta(x) \leq \epsilon$; (ii) the yellow region contains those with $\delta(x) > \epsilon$ but would still accept treatment; and (iii) the red region contains those with $\delta(x) > \epsilon$ and would reject treatment.

4.2. Comparing Population Solutions

We fix budget at $B = 0.1$, a level too small to offer treatment to everyone even though $\rho(x) > 0$ for all x. Under the naive rule, we simply set a cutoff k_{naive} so that offering treatments to all agents with $\rho(x) > k_{\text{naive}}$ exactly exhausts the budget. By contrast, our AFOSD rule requires two simultaneous conditions for an offer at covariate value x: i) a sufficiently high $\rho(x) > k_{\text{AFOSD}}$, and ii) a small enough FOSD violation, i.e. $\delta(x) \leq \epsilon$. The extent to which these two rules agree depend on the joint distribution of potential outcomes.



(A) Agents offered treatment under the naive rule (grey region)

(B) Agents offered treatment under the AFOSD rule (grey region)

FIGURE 3. Naive and AFOSD policy under Model A

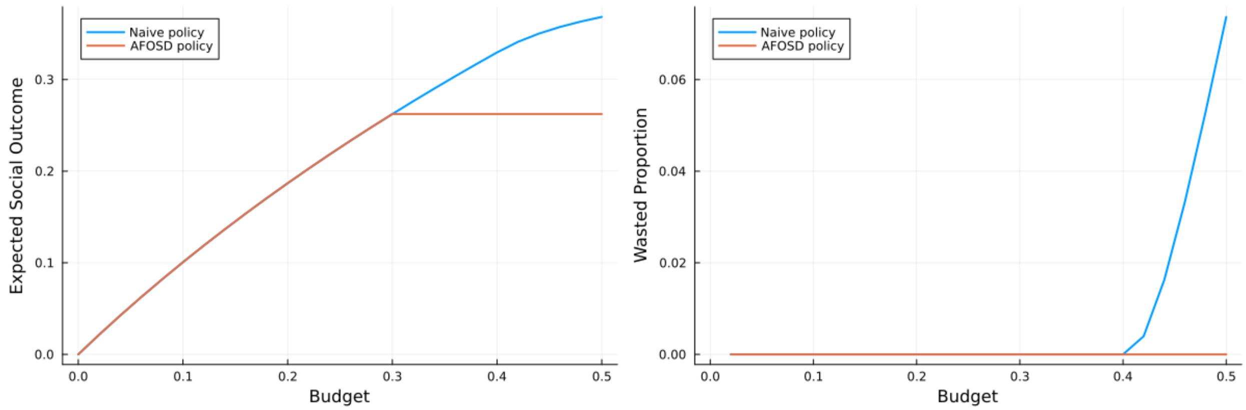
Note: In each plot, the green region indicates covariates x for which $\delta(x) \leq \epsilon$. Agents in the yellow region have $\delta(x) > \epsilon$ but are not risk averse enough to reject treatment. This is in contrast to agents in the red region who have $\delta(x) > \epsilon$ and are highly risk averse, so that they would reject treatment if offered.

In Model A, high RTT is associated with low FOSD violation, so for low budgets, both rules target the same subpopulation with high x. As shown in Figure 3, at a budget of 0.1, the naive and AFOSD rules align perfectly, offering treatment to exactly the same set. Because treatment offers are based solely on x, both rules result in vertical slices in the (x, γ) plane.

As the budget increases, the policies diverge. The naive rule ignores stochastic dominance violations and continues expanding offers until all x are included, regardless of compliance. The AFOSD rule, however, stops expanding once all covariates satisfying $\delta(x) \leq \epsilon$ (i.e. $x \geq x_\epsilon^A$) are included; it excludes the remaining x values that violate the AFOSD constraint. This difference is visible in Figure 4, which shows the reduction in the expected social outcome and the resource wastage as functions of the budget. At higher budgets, AFOSD treats a strict subset of those offered by the naive rule, resulting in lower expected social outcome but zero wastage, since every dollar goes to an agent who accepts.

Model B presents a stark contrast. Here, higher RTT is correlated with greater stochastic dominance violation. The naive policy starts at the maximum $x = 1$ and expand leftwards until the budget is depleted as before, whereas the AFOSD decision rule starts expanding from $x = x_\epsilon^B$ instead of $x = 1$, thereby reserving offers for those covariate values where risk-averse agents are guaranteed to accept.

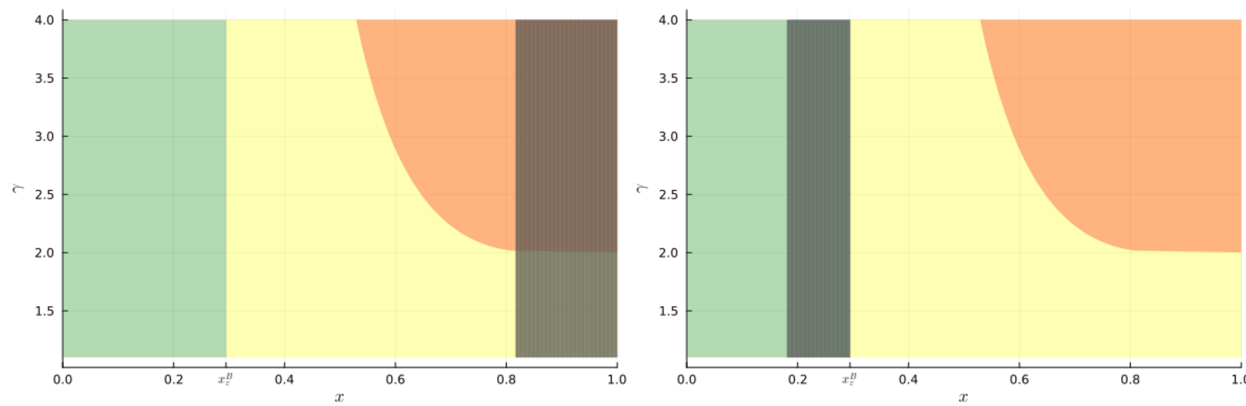
Figure 5 highlights a key weakness of the naive policy: by targeting the highest RTT values without regard to compliance in Model B, it ends up offering treatment to many agents who will refuse. For example, at $x = 1.0$, roughly 69 percent of individuals possess a γ large enough



(A) Expected social outcome under naive and AFOSD treatment rules as budget varies

(B) Proportion of budget wasted on rejected offers under naive and AFOSD treatment rule as budget varies

FIGURE 4. Expected social outcome and resource wastage against budget under Model A



(A) Agents offered treatment under the naive rule (grey region)

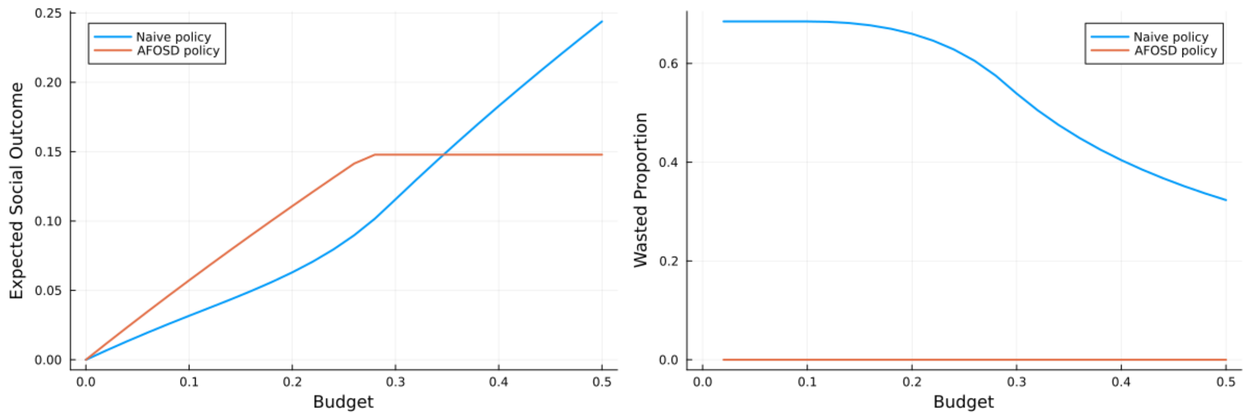
(B) Agents offered treatment under the AFOSD rule (grey region)

FIGURE 5. Naive and AFOSD policy under Model B

Note: In each plot, the green region indicates covariates x for which $\delta(x) \leq \epsilon$. Agents in the yellow region have $\delta(x) > \epsilon$ but are not risk averse enough to reject treatment. This is in contrast to agents in the red region who have $\delta(x) > \epsilon$ and are highly risk averse, so that they would reject treatment if offered.

to reject treatment. Although the naive rule maximizes allocation based on RTT, this translates poorly into social outcome gains once refusal is accounted for.

In contrast, the AFOSD rule confines offers to the region where $\delta(x) \leq \epsilon$. While this constraint may exclude some high RTT individuals, every treatment offered under the AFOSD rule is taken up, ensuring that each allocation fully contributes to the expected social outcome.



(A) Expected social outcome under naive and AFOSD treatment rules as budget varies

(B) Proportion of budget wasted due to rejection under naive and AFOSD treatment rule as budget varies

FIGURE 6. Expected social outcome and resource wastage against budget under Model B

Consequently, the AFOSD rule yields more reliable social outcome improvements by aligning allocation decisions with agents' true willingness to accept treatment.

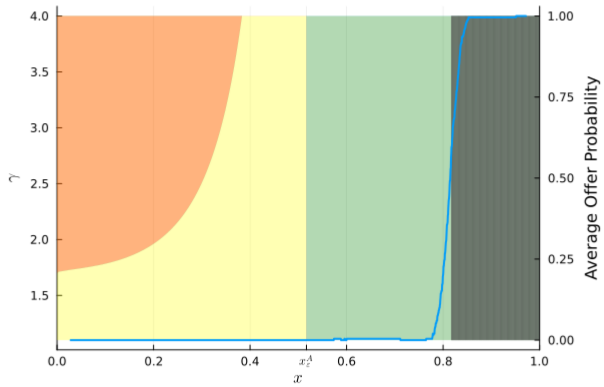
Figure 6(A) shows that, in Model B, targeting only individuals with a compliance guarantee is advantageous at low to medium budget levels. As the budget increases, the number of agents offered treatment under the AFOSD rule rises, leveling off at a budget of approximately 0.28, once all agents who satisfy the AFOSD criterion are included. Up to this threshold, the compliance guarantee property of AFOSD results in a higher expected social outcome than the naive rule. Beyond this point, however, only the naive rule continues to expand the offer set. Due to this difference in targeting, the naive rule ultimately overtakes AFOSD at high budget levels, capturing additional high RTT agents who may not always accept treatment.

This pattern is mirrored in Figure 6(B), which breaks down wastage. Under the naive policy, a nontrivial share of spending goes to agents who reject treatment, and is especially high for low budget levels, whereas AFOSD incurs zero wasted expenditure.

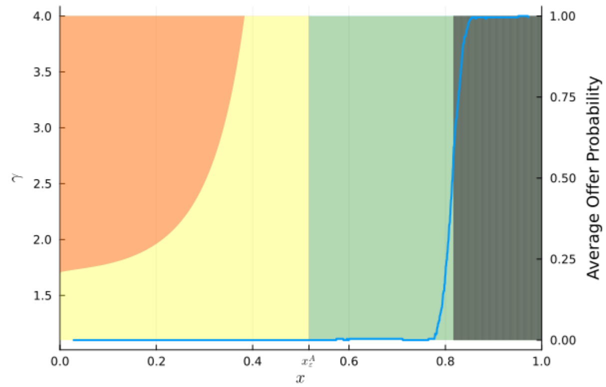
4.3. Estimated Policies

In this section, we evaluate the finite sample performance of the estimated policies. For simplicity, we assume the covariate distributions are the same in the source and target populations. The estimation procedure is as follows. For the naive policy, let

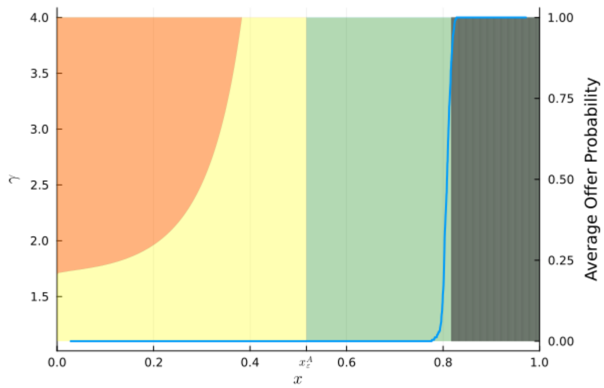
$$\hat{\rho}(x) := \frac{\hat{\tau}_q(x)}{\hat{c}(x)},$$



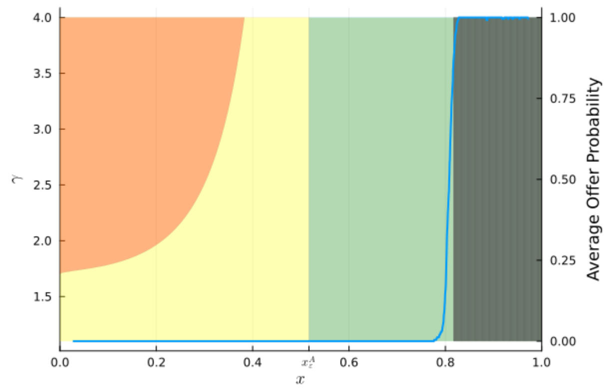
(A) Naive, $n = 500$



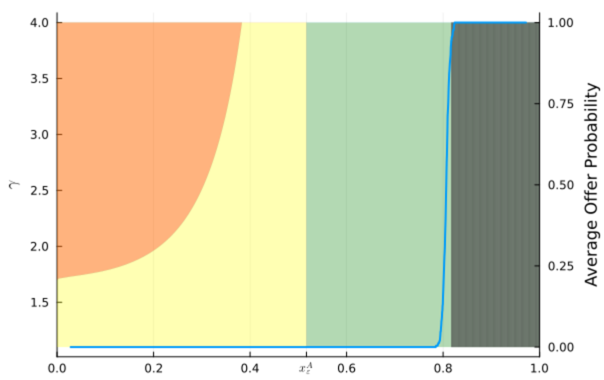
(B) AFOSD, $n = 500$



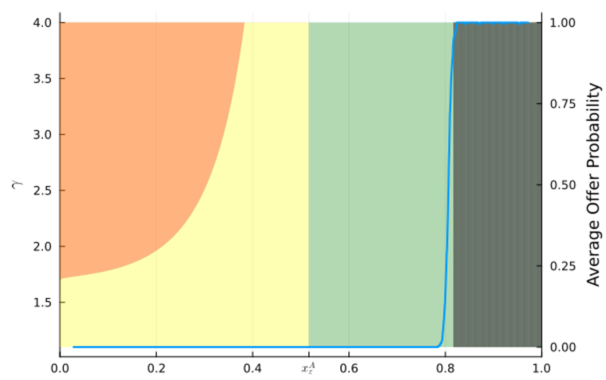
(C) Naive, $n = 2000$



(D) AFOSD, $n = 2000$



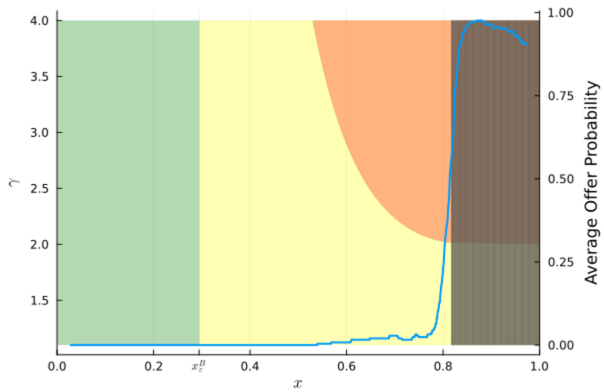
(E) Naive, $n = 4000$



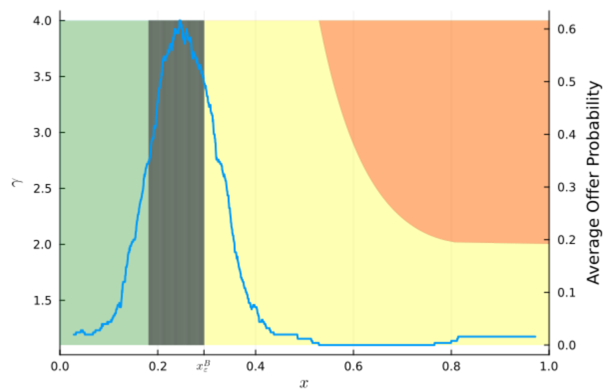
(F) AFOSD, $n = 4000$

FIGURE 7. Comparison of estimated policies under Model A for different sample sizes

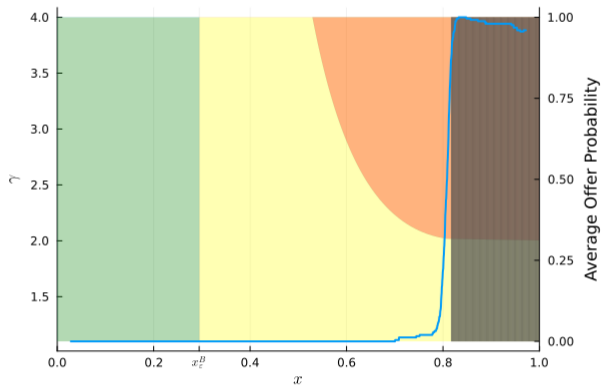
Note: The blue line shows the average (over Monte Carlo replications) offer probability under the estimated policies. The grey shaded area indicates the individuals offered treatment under the corresponding population policies. Risk-aversion γ values are on the left vertical axis, and the average offer probabilities are on the right vertical axis.



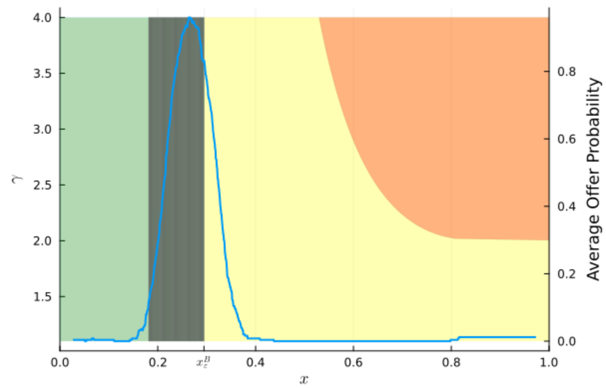
(A) Naive, $n = 500$



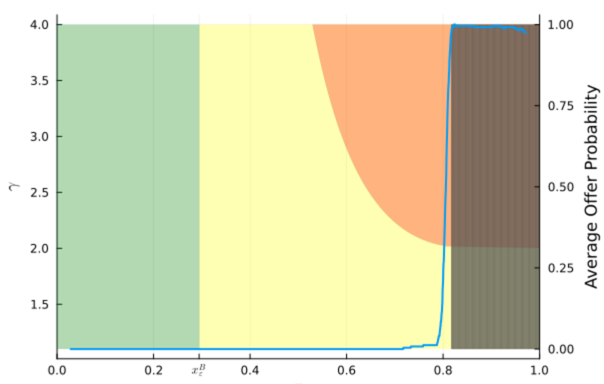
(B) AFOSD, $n = 500$



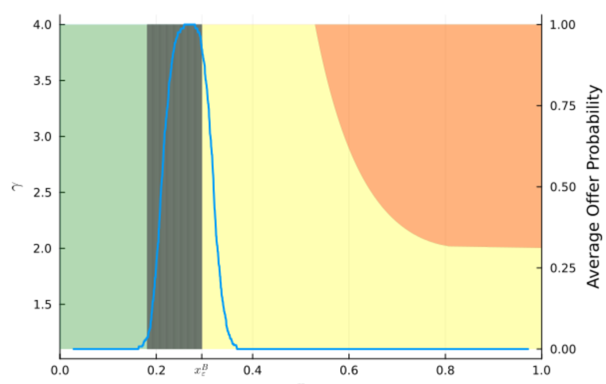
(C) Naive, $n = 2000$



(D) AFOSD, $n = 2000$



(E) Naive, $n = 4000$



(F) AFOSD, $n = 4000$

FIGURE 8. Comparison of estimated policies under Model B for different sample sizes

Note: The blue line shows the average (over Monte Carlo replications) offer probability under the estimated policies. The grey shaded area indicates the individuals offered treatment under the corresponding population policies. Risk-aversion γ values are on the left vertical axis, and the average offer probabilities are on the right vertical axis.

where $\hat{\tau}_{q_q}, \hat{c}$ are estimates for CATE and conditional mean cost using the Nadaraya-Watson estimator. The naive cutoff is then obtained by

$$\hat{k}_{\mathsf{naive}} := \sup_k \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\hat{\rho}(X_i) \geq k) \hat{c}(X_i) - B \geq 0 \right\},$$

which leads to the estimated naive policy

$$\hat{g}_{\mathsf{naive}}(x) := \mathbf{1}\big(\hat{\rho}(x) > \hat{k}_{\mathsf{naive}}\big).$$

We use $B = 0.1$.

For the AFOSD policy, let $\hat{\rho}(x)$ be defined as above. We also estimate the FOSD violation using

$$\hat{\delta}(x) := \frac{\frac{1}{M} \sum_{j=1}^{M} \left[\hat{F}_1(y_j|x) - \hat{F}_0(y_j|x) \right]^+}{\frac{1}{M} \sum_{j=1}^{M} \left| \hat{F}_1(y_j|x) - \hat{F}_0(y_j|x) \right|},$$

where $\{y_1, \ldots, y_M\}$ is a fine grid of y values in the support of Y. With this, we obtain lower bound $L(x)$ of a uniform confidence band for $\delta(x)$ using studentized bootstrap as outlined in Section 3.3, solve for the threshold

$$\hat{k}_{\mathrm{AFOSD}} := \sup_k \left\{ \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(\hat{\rho}(X_i) \geq k) \mathbf{1}(L(X_i) \leq \epsilon) \hat{c}(X_i) - B \geq 0 \right\},$$

and compute the policy estimate as

$$\hat{g}_{\mathrm{AFOSD}}(x) := \mathbf{1}\big(\hat{\rho}(x) \geq \hat{k}_{\mathrm{AFOSD}}\big) \mathbf{1}\big(\hat{L}(x) \leq \epsilon\big).$$

For each sample size $n \in \{500, 2000, 4000\}$, we proceed as follows:

(1) Draw a dataset of size n and then estimate both the naive and AFOSD rules. This is repeated 250 times. In each of these 250 replications, the AFOSD lower-bound threshold is constructed using the studentized bootstrap (method 3) with $q_n := 1 - 0.5n^{-\frac{1}{4}}$.

(2) For each replication $r = 1, \ldots, 250$, and each covariate value x in our grid, we record whether the naive policy $\hat{g}^{(r)}_{\mathrm{naive}}(x)$ and the AFOSD policy $\hat{g}^{(r)}_{\mathrm{AFOSD}}$ offers treatment.

(3) For each point on a fine grid of x values, compute the relative frequency with which each rule would offer treatment:

$$\bar{g}_{\mathrm{naive}}(x) := \frac{1}{250} \sum_{r=1}^{250} \mathbf{1}\big\{\hat{g}^{(r)}_{\mathrm{naive}}(x) = 1\big\},$$

$$\bar{g}_{\mathrm{AFOSD}}(x) := \frac{1}{250} \sum_{r=1}^{250} \mathbf{1}\big\{\hat{g}^{(r)}_{\mathrm{AFOSD}}(x) = 1\big\},$$

The curves $\bar{g}_{\text{naive}}(x)$ and $\bar{g}_{\text{AFOSD}}(x)$ represent the "average learned" policies as functions of x. We can overlay $\bar{g}_{\text{naive}}(x)$ and $\bar{g}_{\text{AFOSD}}(x)$ on the true population policy boundaries to assess their convergence.

Figure 7 shows the average learned policy for each rule in Model A. At a budget of 0.1, the naive and AFOSD policies coincide in both theory and practice. At this budget level, since $\delta(x) \leq \epsilon$ for all x in the shaded area, and the lower confidence band lies below δ with high probability, the estimated policies depend essentially on only $\hat{\rho}$ which tracks ρ reasonably well.

In the case of Model B, in the left panels of Figure 8, which correspond to the naive policy estimator, the learned policy closely tracks the true population rule even at $n = 500$. This consistency reflects the fact that ranking by the estimated $\hat{\rho}(x)$ mirrors the ranking by the true $\rho(x)$, so the cutoff procedure reliably selects the same covariate region regardless of sample size.

By contrast, for the estimated AFOSD policy to work well in Model B, it has to be able to eliminate points that violate the stochastic dominance constraints, and this is challenging. The right panels in Figure 8, which corresponds to the AFOSD policy estimates, reveals noticeable discrepancies for smaller n. Although the RTT ranking remains accurate, the one-sided confidence lower bound $\hat{L}(x)$ significantly understates the true FOSD violation $\delta(x)$, causing the estimator to offer treatment at $x > x_\epsilon$ when it should not. As n grows, however, $\hat{L}(x)$ converges more tightly to $\delta(x)$, and the estimated policy increasingly refrains from offering treatment beyond the true threshold. This shift is evident in the declining "overshoot" of the average offer probability and the gradual alignment of its peak with x_ϵ.

Appendix A. Proofs

Proof of Lemma 2.1: Suppose $\mu(g; c_O + c_T) > B$. Consider a utility function $u(y, V) = 1$ for all y. Since subjects with such u are indifferent between outcomes, $u \in \mathcal{H}_\epsilon$ and $\pi_u(X) = 1$ with probability one because in this case. Therefore,

$$\mu(g; c_O) + \mu(g; c_T \pi_u) = \mu(g; c_O + c_T) > B.$$

Budget violation means that the value of the PM's objective under this u satisfies $W(g, u; \tau_q, \mathbf{c}) = -\infty$ and $W(g, u; \tau_q, (c_O + c_T, 0)) = -\infty$.

Next, suppose $\mu(g; c_O + c_T) \leq B$, then for any $u \in \mathcal{H}_\epsilon$,

$$\mu(g; c_O) + \mu(g; c_T \pi_u) \leq \mu(g; c_O + c_T) \leq B,$$

so that

$$W\big(g, u; \tau_q, \mathbf{c}\big) = \mu\big(g; \tau_q \pi_u\big) = W\big(g, u; \tau_q, (c_O + c_T, 0)\big).$$

The last equality follows because $W\big(g,u;\tau_q,\mathbf{c}\big)$ does not depend on $\mathbf{c}$ as long as $\mathbf{c}$ satisfies the budget constraint. In both cases, we have

$$\inf_{u\in\mathcal{U}} W\big(g,u;\tau_q,\mathbf{c}\big) = \inf_{u\in\mathcal{U}} W\big(g,u;\tau_q,(c_O+c_T,0)\big).$$

■

Proof of Lemma 2.2: First, note that for each $x\in\mathcal{X}_c$ and $v=(x,v_{-x})$ with some $v_{-x}\in\mathcal{V}_{-X}$,

$$\begin{aligned}
&\mathbf{E}[u(Y(1),V)\mid V=v]-\mathbf{E}[u(Y(0),V)\mid V=v]\\
&=\int u(y,v)dP_{Y(1)|V}(y\mid v)-\int u(y,v)dP_{Y(0)|V}(y\mid v)\\
&\geq \inf_{v'\in\mathcal{V}}\left\{\int u(y,v')dP_{Y(1)|V}(y\mid v)-\int u(y,v')dP_{Y(0)|V}(y\mid v)\right\}\\
&=\inf_{v'\in\mathcal{V}}\left\{\int u(y,v')dF_1(y\mid x)-\int u(y,v')dF_0(y\mid x)\right\}\\
&\geq 0,
\end{aligned}$$

where $P_{Y(d)|V}$ denotes the regular conditional distribution of $Y(d)$ given V, the equality in the line before the last holds by Assumption 2.2(i), and the inequality in the last line holds by Assumption 2.3. Therefore,

$$\pi_u(x)=1,\ \text{if } x\in\mathcal{X}_c. \tag{A.1}$$

For g such that $\mu(g;c)\leq B$, we can write

$$\begin{aligned}
W(g,u;\tau_q,\mathbf{c}) &= \mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}\big]+\mathbf{E}\big[g(X)\tau_q(X)\pi_u(X)\mathbf{1}\{X\notin\mathcal{X}_c\}\big]\\
&\geq \mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}\big]+\inf_{u\in\mathcal{U}}\mathbf{E}\big[g(X)\tau_q(X)\pi_u(X)\mathbf{1}\{X\notin\mathcal{X}_c,\tau_q(X)<0\}\big]\\
&\quad+\inf_{u\in\mathcal{U}}\mathbf{E}\big[g(X)\tau_q(X)\pi_u(X)\mathbf{1}\{X\notin\mathcal{X}_c,\tau_q(X)>0\}\big]\\
&\geq \mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}\big]+\inf_{u\in\mathcal{U}}\mathbf{E}\big[g(X)\tau_q(X)\pi_u(X)\mathbf{1}\{X\notin\mathcal{X}_c,\tau_q(X)<0\}\big]\\
&\geq \mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}\big]+\mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X\notin\mathcal{X}_c,\tau_q(X)<0\}\big],
\end{aligned}$$

where the last inequality follows by choosing $u(y,(x,v_{-x}))$ such that $\pi_u(x)=1$ for $x\notin\mathcal{X}_c$ with $\tau_q(x)<0$. This lower bound is attained by such a choice of u, yielding the desired result. ■

Proof of Theorem 2.1: We first prove the result in part (i), where $k^*>0$. In this case, the budget constraint is insufficient to treat everyone with $x\in\mathcal{X}_c$. By Lemma 2.2, we can set $g^*(x)=0$ for all $x\notin\mathcal{X}_c$. The PM's problem becomes maximizing $\mathbf{E}[g(X)\tau_q(X)\mathbf{1}\{X\in\mathcal{X}_c\}]$

subject to $\mathbf{E}[g(X)c(X)\mathbf{1}\{X \in \mathcal{X}_c\}] \leq B$. The Lagrangian for this problem is

$$\begin{aligned}\mathcal{L}(g,\lambda) &= \mathbf{E}\big[g(X)\tau_q(X)\mathbf{1}\{X \in \mathcal{X}_c\}\big] - \lambda(\mathbf{E}[g(X)c(X)\mathbf{1}\{X \in \mathcal{X}_c\}] - B)\\ &= \mathbf{E}\big[g(X)\big(\tau_q(X) - \lambda c(X)\big)\mathbf{1}\{X \in \mathcal{X}_c\}\big] + \lambda B, \;\; \lambda \geq 0.\end{aligned}$$

The optimal policy g^* maximizes the Lagrangian pointwise, so that for each $x \in \mathcal{X}_c$, we can set $g^*(x) = 1$ if $\tau_q(x) - \lambda c(x) \geq 0$, and $g^*(x) = 0$ if $\tau_q(x) - \lambda c(x) < 0$. Lastly, setting $k^* := \lambda$, $g^*(x) = r(x)$ for $x \in \mathcal{X}_c$ such that $\tau_q(x) - \lambda c(c) = 0$, where r is chosen to satisfy the budget constraint, and $g^*(x) = 0$ for $x \notin \mathcal{X}_c$, we obtain the desired result.

For part (ii), we have $k^* = 0$. In this case, the budget is sufficient to treat everyone with $x \in \mathcal{X}_c$ and $\tau_q(x) \geq 0$, and setting $g^*(x) = 1$ for such individuals maximizes the PM's objective while satisfying the budget constraint. We can again set $g^*(x) = 0$ for all $x \notin \mathcal{X}_c$ with $\tau_q(x) < 0$. For $x \notin \mathcal{X}_c$ with $\tau_q(x) > 0$, any treatment assignment rule satisfying the remaining budget constraint results in the same worst-case expected social outcome by Lemma 2.2. ■

Proof of Lemma 2.3: For any $u \in \mathcal{U}$, $\lambda \in [0, \bar{\lambda}]$, and measurable function $g : \mathcal{X} \to [0,1]$, we have

$$\begin{aligned}&\mathbf{E}\big[g(X)\pi_u(X)\tau_q(X)\big] + \lambda(B - \mathbf{E}[g(X)c_O(X) + g(X)\pi_u(X)c_T(X)])\\
&= \mathbf{E}\big[g(X)\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\}\tau_q(X)\big] + \mathbf{E}\big[g(X)\pi_u(X)\mathbf{1}\{\tau_q(X) \leq 0\}\tau_q(X)\big]\\
&\quad + \mathbf{E}\big[g(X)\pi_u(X)\mathbf{1}\{X \notin \mathcal{X}_c, \tau_q(X) > 0\}\tau_q(X)\big]\\
&\quad + \lambda\Big(B - \mathbf{E}[g(X)c_O(X)] - \mathbf{E}[g(X)\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\}c_T(X)]\\
&\qquad\qquad - \mathbf{E}[g(X)\pi_u(X)\mathbf{1}\{\tau_q(X) \leq 0\}c_T(X)] - \mathbf{E}[g(X)\pi_u(X)\mathbf{1}\{X \notin \mathcal{X}_c, \tau_q(X) > 0\}c_T(X)]\Big)\\
&\geq \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\}\big)\tau_q(X)\big]\\
&\qquad + \mathbf{E}\big[g(X)\pi_u(X)\mathbf{1}\{X \notin \mathcal{X}_c, \tau_q(X) > 0\}\tau_q(X)\big]\\
&\qquad + \lambda\Big(B - \mathbf{E}\big[g(X)c_O(X)\big] - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\}\big)c_T(X)\big]\\
&\qquad\qquad - \mathbf{E}\big[g(X)\pi_u(X)\mathbf{1}\{X \notin \mathcal{X}_c, \tau_q(X) > 0\}c_T(X)\big]\Big).\\
&= \mathbf{E}\big[g(X)\pi_{\bar{u}}(X)\tau_q(X)\big] + \lambda\big(B - \mathbf{E}\big[g(X)c_O(X) + g(X)\pi_{\bar{u}}(X)c_T(X)\big]\big)\\
&\geq \inf_u \inf_{\lambda \in [0,\bar{\lambda}]} \mathbf{E}\big[g(X)\pi_u(X)\tau_q(X)\big] + \lambda\big(B - \mathbf{E}\big[g(X)c_O(X) + g(X)\pi_u(X)c_T(X)\big]\big),\end{aligned}$$

where $\bar{u}$ in the second last line is defined as

$$\bar{u}(\cdot, v) := \begin{cases} 1, & \text{if } v = (x, v_{-x}) \text{ with } x \in \mathcal{X}_c \text{ or } \tau_q(x) \leq 0,\\ u(\cdot, v), & \text{otherwise,}\end{cases}$$

and the last inequality follows since $\overline{u}$ is in $\mathcal{U}$. By taking infimum over u and λ, we have

(A.2)

$$W_{\mathcal{L}}(g) = \inf_{u} \inf_{\lambda\in[0,\overline{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)\tau_q(X)\big] + \lambda\Big(B - \mathbf{E}\big[g(X)c_O(X)\big] - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)c_T(X)\big]\Big) + \mathbf{E}\Big[g(X)\pi_u(X)\mathbf{1}\{X \notin \mathcal{X}_c, \tau_q(X) > 0\}\big(\tau_q(X) - \lambda c(X)\big)\Big]\Big\}.$$

Now, for any g and $\lambda \ge 0$, since u can be chosen such that $\pi_u(x) = 0$ for any $x \notin \mathcal{X}_c$ with $\tau_q(x) > 0$, we have

$$W_{\mathcal{L}}(g) \le \inf_{\lambda\in[0,\overline{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)\tau_q(X)\big] + \lambda\Big(B - \mathbf{E}\big[g(X)c_O(X)\big] - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)c_T(X)\big]\Big)\Big\}.$$

This means

(A.3)

$$\begin{aligned}\sup_g W_{\mathcal{L}}(g) &\le \sup_g \inf_{\lambda\in[0,\overline{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)\tau_q(X)\big] \\ &\quad + \lambda\Big(B - \mathbf{E}\big[g(X)c_O(X)\big] - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)c_T(X)\big]\Big)\Big\} \\ &\le \sup_g \inf_{\lambda\in[0,\overline{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)\tau_q(X)\big] \\ &\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)(c_O(X) + c_T(X))\big]\Big)\Big\} \\ &= \sup_g \tilde{W}_{\mathcal{L}}(g).\end{aligned}$$

On the other hand, since PM can always choose $g(x) = 0$ for any $x \notin \mathcal{X}_c$ with $\tau_q(x) > 0$, (A.2) leads to

(A.4)

$$\begin{aligned}\sup_g W_{\mathcal{L}}(g) &\ge \sup_{\substack{g: g(x)=0 \text{ if} \\ x\notin\mathcal{X}_c,\ \tau_q(x)>0}} W_{\mathcal{L}}(g) \\ &= \sup_{\substack{g: g(x)=0 \text{ if} \\ x\notin\mathcal{X}_c,\ \tau_q(x)>0}} \inf_{\lambda\in[0,\overline{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c,\ \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)\tau_q(X)\big] \\ &\quad + \lambda\Big(B - \mathbf{E}[g(X)c_O(X)] - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X \in \mathcal{X}_c\} + \mathbf{1}\{\tau_q(X) \le 0\}\big)c_T(X)\big]\Big)\Big\}\end{aligned}$$

$$
\begin{aligned}
&= \sup_{\substack{g:g(x)=0 \text{ if} \\ x\notin\mathcal{X}_c,\ \tau_q(x)>0}} \inf_{\lambda\in[0,\bar{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big)\Big\} \\
&\ge \sup_{g} \inf_{\lambda\in[0,\bar{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big)\Big\} \\
&= \sup_{g} \tilde{W}_{\mathcal{L}}(g).
\end{aligned}
$$

The last inequality here can be justified as follows. For any g, define the function J as

$$
Jg(x) := g(x)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big).
$$

Then for any $\lambda \ge 0$,

$$
\begin{aligned}
&\mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big) \\
&= \mathbf{E}\big[Jg(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[Jg(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big).
\end{aligned}
$$

It then follows that

$$
\begin{aligned}
&\sup_{g} \inf_{\lambda\in[0,\bar{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big)\Big\} \\
&= \sup_{g} \inf_{\lambda\in[0,\bar{\lambda}]} \Big\{ \mathbf{E}\big[Jg(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[Jg(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big)\Big\} \\
&\le \sup_{\substack{g:g(x)=0 \text{ if} \\ x\notin\mathcal{X}_c,\ \tau_q(x)>0}} \inf_{\lambda\in[0,\bar{\lambda}]} \Big\{ \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)\tau_q(X)\big] \\
&\quad + \lambda\Big(B - \mathbf{E}\big[g(X)\big(\mathbf{1}\{X\in\mathcal{X}_c,\ \tau_q(X)>0\}+\mathbf{1}\{\tau_q(X)\le 0\}\big)(c_O(X)+c_T(X))\big]\Big)\Big\},
\end{aligned}
$$

since for any g, the function Jg satisfies $g(x) = 0$ if $x \notin \mathcal{X}_c$ with $\tau_q(x) < 0$. Lemma 2.3 then follows from (A.3) and (A.4). ■

Proof of Theorem 2.2: For the rest of this proof, let k^* and r be as defined in Theorem 2.1. We first show that $g^*_{\mathcal{L}}$ is in fact a maximizer of $\tilde{W}_{\mathcal{L}}$, by considering cases.

Case 1 (Non-binding constraint): Suppose the budget constraint is not binding, i.e.,

$$\mathbf{E}[\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > 0\}c(X)] \leq B.$$

Then $k^* = 0$, and $g^*_{\mathcal{L}} = g^*$. In this case, we see that g^* maximizes the first term in (2.10) by considering sign of τ_q, while the last term (which is non-positive due to infimum over u) is zero, so that $g^*_{\mathcal{L}}$ maximizes $\tilde{W}_{\mathcal{L}}$ in this case.

Case 2 (Binding constraint): Suppose the budget constraint is binding, so that $k^* > 0$. We want to show

$$\tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}) \geq \tilde{W}_{\mathcal{L}}(\tilde{g}),$$

for any measurable $\tilde{g}$. We do this by splitting into the subcases depending on values of k^* relative to $\overline{\lambda}$, as well as spending for policy $\tilde{g}$:

Case 2a: Suppose $k^* \leq \overline{\lambda}$. Then $g^*_{\mathcal{L}} = g^*$, so that

$$\mathbf{E}(g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] \leq \mathbf{E}[g^*_{\mathcal{L}}(X)c(X)] = B$$

in which the last equality is due to a binding constraint, and thus

$$\begin{aligned}\tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}) &= \mathbf{E}[\mathbf{1}\{\rho(X) > k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}\tau_q(X)] + r\mathbf{E}[\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) = k^*\}\tau_q(X)] \\ &= \inf_{u \in \mathcal{U}} W(g^*, u; \tau_q, \mathbf{c}),\end{aligned}$$

where the last equality follows from Lemma 2.2.

Case 2a(i): Consider the case where $\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] \leq B$. Define

$$\bar{g}(x) := \tilde{g}(x)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\}).$$

Since $\bar{g}$ satisfies

$$\mathbf{E}[\bar{g}(X)c(X)] = \mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] \leq B,$$

we have

$$\begin{aligned}\tilde{W}_{\mathcal{L}}(\tilde{g}) = \tilde{W}_{\mathcal{L}}(\bar{g}) &= \mathbf{E}[\bar{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\tau_q(X)] + \mathbf{E}[\bar{g}(X)\tau_q(X)\mathbf{1}\{\tau_q(X) < 0\}] \\ &= \inf_{u \in \mathcal{U}} W(\bar{g}, u; \tau_q, \mathbf{c}) \\ &\leq \inf_{u \in \mathcal{U}} W(g^*, u; \tau_q, \mathbf{c}) = \tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}),\end{aligned}$$

where the third equality follows from Lemma 2.2, while the inequality follows from Theorem 2.1.

Case 2a(ii): If $\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] > B$, then

$$
\begin{aligned}
&\tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}) - \tilde{W}_{\mathcal{L}}(\tilde{g}) \\
&= \mathbf{E}[(1-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > k^*\}\tau_q(X)] + \mathbf{E}[(r-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) = k^*\}\tau_q(X)] \\
&\quad - \mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{0 < \rho(X) < k^*\}\tau_q(X)] - \mathbf{E}[\tilde{g}(X)\mathbf{1}\{\tau_q(X) \leq 0\}\tau_q(X)] \\
&\quad - \overline{\lambda}(B - \mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)]) \\
&\geq k^*\mathbf{E}[(1-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > k^*\}c(X)] + k^*\mathbf{E}[(r-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) = k^*\}c(X)] \\
&\quad - k^*\mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{0 < \rho(X) < k^*\}c(X)] - k^*\mathbf{E}[\tilde{g}(X)\mathbf{1}\{\tau_q(X) \leq 0\}c(X)] \\
&\quad - \overline{\lambda}(B - \mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)]) \\
&= (k^* - \overline{\lambda})\big(B - \mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)]\big) \geq 0,
\end{aligned}
$$

where the first inequality uses the fact that $\tilde{g}(x) \in [0,1]$ and $c(x) \geq 0$, and the last inequality is because both bracketed terms on the LHS are non-positive.

Case 2b: On the other hand, if $k^* > \overline{\lambda}$, we have

$$
\begin{aligned}
&\mathbf{E}[g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] \\
&\geq \mathbf{E}[g^*(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] = \mathbf{E}[g^*(X)c(X)] = B.
\end{aligned}
$$

Inequality here is because everyone who is treated with positive probability under g^* is treated with probability 1 under $g^*_{\mathcal{L}}$. Comparing g^* and $g^*_{\mathcal{L}}$ yields

$$
\begin{aligned}
\tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}) = \tilde{W}_{\mathcal{L}}(g^*) &+ (1-r)\mathbf{E}[\mathbf{1}\{\rho(X) = k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}\tau_q(X)] \\
&+ \mathbf{E}[\mathbf{1}\{\overline{\lambda} < \rho(X) < k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}\tau_q(X)] \\
&- \overline{\lambda}(1-r)\mathbf{E}[\mathbf{1}\{\rho(X) = k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}c(X)] \\
&- \overline{\lambda}\mathbf{E}[\mathbf{1}\{\overline{\lambda} < \rho(X) < k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}c(X)] \\
&\geq \tilde{W}_{\mathcal{L}}(g^*) + (k^* - \overline{\lambda})(1-r)\mathbf{E}[\mathbf{1}\{\rho(X) = k^*\}\mathbf{1}\{X \in \mathcal{X}_c\}c(X)] \geq \tilde{W}_{\mathcal{L}}(g^*).
\end{aligned}
$$

Case 2b(i) If $\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] \leq B$, then by similar arguments as Case 2a(i), we can show that

$$
\tilde{W}_{\mathcal{L}}(\tilde{g}) \leq \tilde{W}_{\mathcal{L}}(g^*) \leq \tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}).
$$

Case 2b(ii) If $\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \leq 0\})c(X)] > B$, then

$$
\begin{aligned}
&W_{\mathcal{L}}(g^*_{\mathcal{L}}) - W_{\mathcal{L}}(\tilde{g}) \\
&= \mathbf{E}[(1-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > \overline{\lambda}\}\tau_q(X)] - \mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) = \overline{\lambda}\}\tau_q(X)] \\
&\quad - \mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{0 < \rho(X) < \overline{\lambda}\}\tau_q(X)] - \mathbf{E}[\tilde{g}(X)\mathbf{1}\{\tau_q(X) \leq 0\}\tau_q(X)]
\end{aligned}
$$

$$-\bar{\lambda}\mathbf{E}[g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})c(X)]$$
$$+\bar{\lambda}\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})c(X)])$$
$$\ge \bar{\lambda}\mathbf{E}[(1-\tilde{g}(X))\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) > \bar{\lambda}\}c(X)] - \bar{\lambda}\mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{\rho(X) = \bar{\lambda}\}c(X)]$$
$$-\bar{\lambda}\mathbf{E}[\tilde{g}(X)\mathbf{1}\{X \in \mathcal{X}_c\}\mathbf{1}\{0 < \rho(X) < \bar{\lambda}\}c(X)] - \bar{\lambda}\mathbf{E}[\tilde{g}(X)\mathbf{1}\{\tau_q(X) \le 0\}c(X)]$$
$$-\bar{\lambda}\mathbf{E}[g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})c(X)]$$
$$+\bar{\lambda}\mathbf{E}[\tilde{g}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})c(X)])$$
$$\ge 0,$$

and the inequality again uses that $\tilde{g}(x) \in [0,1]$ and $c(x) > 0$ for all x.

Combining all cases, we have shown that $g^*_{\mathcal{L}}$ maximizes $\tilde{W}_{\mathcal{L}}$. Finally, since

$$\begin{aligned} W_{\mathcal{L}}(g^*_{\mathcal{L}}) &= \mathbf{E}[g^*_{\mathcal{L}}(x)\tau_q(X)] + \inf_{\lambda \in [0,\bar{\lambda}]} \lambda(B - \mathbf{E}[g^*_{\mathcal{L}}(X)c(X)]) \\ &= \mathbf{E}[g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})\tau_q(X)] \\ &\quad + \inf_{\lambda \in [0,\bar{\lambda}]} \lambda(B - \mathbf{E}[g^*_{\mathcal{L}}(X)(\mathbf{1}\{X \in \mathcal{X}_c, \tau_q(X) > 0\} + \mathbf{1}\{\tau_q(X) \le 0\})c(X)]) \\ &= \tilde{W}_{\mathcal{L}}(g^*_{\mathcal{L}}), \end{aligned}$$

where the first equality uses the fact that $\pi_u(x) = 1$ for any u if $x \in \mathcal{X}_c$, and second equality is due to the the implication $g^*_{\mathcal{L}}(x) = 1 \implies x \in \mathcal{X}_c$ and $\tau_q(x) > 0$. By Lemma 2.3, $g^*_{\mathcal{L}}$ must also be a maximizer for $W_{\mathcal{L}}$. ■

Proof of Lemma 3.1: Define

$$\mathcal{I}(x) := \int |\Delta(y|x)| dy,$$

Since $\Delta(y|x)$ is right-continuous at y^* for each x, if x satisfies $\tau_q(x) = -\Delta(y^*|x) \ge \xi$, there must exist some $\ell(x,\xi) > 0$ such that $\mathcal{I}(x) > \ell(x,\xi)$. The proof is complete if we show that

$$\ell(\xi) = \inf_{x \in \{x \in \mathcal{X} : \tau_q(x) \ge \xi\}} \mathcal{I}(x) > 0.$$

Suppose this is not true, so that $\ell(\xi) = 0$. Then by the extreme value theorem, continuity of $\mathcal{I}$ implies that $\ell(x,\xi) = 0$ for some $x \in \{x \in \mathcal{X} : \tau_q(x) \ge \xi\}$, a contradiction. ■

Lemmas A.1–A.4 below are used in the proofs of Theorem 3.1 and 3.2.

Lemma A.1. *Suppose that Assumption 3.2 and 3.3 holds. Then, for any constant $\xi > 0$, on the event that $\xi > \hat{\nu}$,*

$$\sup_{x:\tau_q(x) > \xi} \left|\hat{\delta}(x) - \delta(x)\right| \le \frac{2M\hat{\nu}}{\ell(\xi)}\left(\frac{1}{\ell(\xi) - \hat{\nu}} + 3\right).$$

Proof: We can write

$$
\begin{aligned}
&\sup_{x:\tau_q(x)>\xi}\left|\hat{\delta}(x)-\delta(x)\right| \\
&= \sup_{x:\tau_q(x)>\xi}\left|\frac{\int \hat{\Delta}(y\mid x)\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}dy}{\int\left|\hat{\Delta}(y\mid x)\right|dy}-\frac{\int \Delta(y\mid x)\mathbf{1}\left\{\Delta(y\mid x)>0\right\}dy}{\int|\Delta(y\mid x)|dy}\right| \\
&\leq \sup_{x:\tau_q(x)>\xi}\left|\frac{\int \hat{\Delta}(y\mid x)\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}dy}{\int\left|\hat{\Delta}(y\mid x)\right|dy}-\frac{\int \hat{\Delta}(y\mid x)\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}dy}{\int|\Delta(y\mid x)|dy}\right| \\
&\quad + \sup_{x:\tau_q(x)>\xi}\left|\frac{\int \hat{\Delta}(y\mid x)\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}-\Delta(y\mid x)\mathbf{1}\left\{\Delta(y\mid x)>0\right\}dy}{\int|\Delta(y\mid x)|dy}\right|.
\end{aligned}
$$

The last sum of two suprema is bounded by

$$
\begin{aligned}
&\sup_{x:\tau_q(x)>\xi}\int\left|\hat{\Delta}(y\mid x)\right|\left|\frac{1}{\int\left|\hat{\Delta}(y\mid x)\right|dy}-\frac{1}{\int|\Delta(y\mid x)|dy}\right|dy \\
&\quad + \sup_{x:\tau_q(x)>\xi}\frac{1}{\ell(\xi)}\left|\int\left[\hat{\Delta}(y\mid x)-\Delta(y\mid x)\right]\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}dy\right| \\
&\quad + \sup_{x:\tau_q(x)>\xi}\frac{1}{\ell(\xi)}\left|\int \Delta(y\mid x)\left(\mathbf{1}\left\{\hat{\Delta}(y\mid x)>0\right\}-\mathbf{1}\left\{\Delta(y\mid x)>0\right\}\right)dy\right|, \qquad \text{(A.5)}
\end{aligned}
$$

due to Assumption 3.3. By the boundedness condition of potential outcomes in Assumption 3.2, the first term in (A.5) is bounded above by

$$
\begin{aligned}
&\sup_{x:\tau_q(x)>\xi} 2M\left|\frac{1}{\int\left|\hat{\Delta}(y\mid x)\right|dy}-\frac{1}{\int|\Delta(y\mid x)|dy}\right| \\
&\leq \frac{2M}{\ell(\xi)}\frac{\sup_{x:\tau_q(x)>\xi}\left|\int|\Delta(y\mid x)|-\left|\hat{\Delta}(y\mid x)\right|dy\right|}{\inf\limits_{x:\tau_q(x)>\xi}\int\left|\hat{\Delta}(y\mid x)\right|dy} \\
&\leq \frac{2M}{\ell(\xi)}\frac{\sup_{x\in\mathcal{X}}\int\left|\hat{\Delta}(y\mid x)-\Delta(y\mid x)\right|dy}{\inf\limits_{x:\tau_q(x)>\xi}\int|\Delta(y\mid x)|dy-\sup_{x\in\mathcal{X}}\int\left|\hat{\Delta}(y\mid x)-\Delta(y\mid x)\right|dy} \leq \frac{2M}{\ell(\xi)}\frac{\hat{\nu}}{\ell(\xi)-\hat{\nu}}.
\end{aligned}
$$

The second term in (A.5) is bounded by

$$
\sup_{x:\tau_q(x)>\xi}\frac{2M}{\ell(\xi)}\int\left|\hat{\Delta}(y\mid x)-\Delta(y\mid x)\right|dy \leq \frac{2M}{\ell(\xi)}\sup_{x\in\mathcal{X}}\int\left|\hat{\Delta}(y\mid x)-\Delta(y\mid x)\right|dy=\frac{2M\hat{\nu}}{\ell(\xi)}.
$$

The last term in (A.5) is bounded by

$$\begin{aligned}
&\frac{1}{\ell(\xi)} \sup_{x:\tau_q(x)>\xi} \int |\Delta(y \mid x)| \big| \mathbf{1}(\Delta(y \mid x) > 0) - \mathbf{1}(\hat{\Delta}(y \mid x) > 0) \big| dy \\
&= \frac{1}{\ell(\xi)} \sup_{x:\tau_q(x)>\xi} \int |\Delta(y \mid x)| \mathbf{1}\big\{\Delta(y \mid x) - \hat{\Delta}(y \mid x) \geq \Delta(y \mid x) > 0\big\} dy \\
&\quad + \sup_{x:\tau_q(x)>\xi} \int |\Delta(y \mid x)| \mathbf{1}\big\{0 \geq \Delta(y \mid x) > \Delta(y \mid x) - \hat{\Delta}(y \mid x)\big\} dy \\
&\leq \frac{4M}{\ell(\xi)} \sup_{x \in \mathcal{X}} \int \big|\Delta(y \mid x) - \hat{\Delta}(y \mid x)\big| dy = \frac{4M\hat{\nu}}{\ell(\xi)}.
\end{aligned}$$

Combining the terms, we obtain the desired result. ■

Lemma A.2. *Suppose that Assumption 3.2 holds. Then, for each $t, \kappa, \xi \geq 0$, the event*

$$\big\{x \in \mathcal{X}_{t-\kappa} : \tau_q(x) > \xi\big\} \subseteq \big\{x \in \hat{\mathcal{X}}_t : \tau_q(x) > \xi\big\} \subseteq \big\{x \in \mathcal{X}_{t+\kappa} : \tau_q(x) > \xi\big\}$$

occurs with probability at least $p_n(\kappa, \xi)$.

Proof: Denote the complement of a set S in the probability space by $S^{\complement}$. Since

$$\begin{aligned}
P((A \cap E) \subseteq (B \cap E) \subseteq (D \cap E)) &= P\Big(\big((A \cap E) \cap (B \cap E)^{\complement}\big) \cup \big((B \cap E) \cap (D \cap E)^{\complement}\big) = \varnothing\Big) \\
&= P\Big(\big\{(A \cap E) \cap \big(B^{\complement} \cup E^{\complement}\big)\big\} \cup \big\{(B \cap E) \cap \big(D^{\complement} \cup E^{\complement}\big)\big\} = \varnothing\Big) \\
&= P\big(\big(A \cap E \cap B^{\complement}\big) \cup \big(B \cap E \cap D^{\complement}\big) = \varnothing\big) \\
&= 1 - P\big(\big(A \cap E \cap B^{\complement}\big) \cup \big(B \cap E \cap D^{\complement}\big) \neq \varnothing\big) \\
&\geq 1 - P\big(A \cap E \cap B^{\complement} \neq \varnothing\big) - P\big(B \cap E \cap D^{\complement} \neq \varnothing\big),
\end{aligned}$$

we can write

$$\begin{aligned}
&P\Big(\big\{x \in \mathcal{X}_{t-\kappa}, \tau_q(x) > \xi\big\} \subseteq \big\{x \in \hat{\mathcal{X}}_t, \tau_q(x) > \xi\big\} \subseteq \big\{x \in \mathcal{X}_{t+\kappa}, \tau_q(x) > \xi\big\}\Big) \\
&\geq 1 - P\Big(\big\{x \in \mathcal{X}_{t-\kappa} \cap \hat{\mathcal{X}}_t^{\complement}, \tau_q(x) > \xi\big\} \neq \varnothing\Big) - P\Big(\big\{x \in \hat{\mathcal{X}}_t \cap \mathcal{X}_{t+\kappa}^{\complement}, \tau_q(x) > \xi\big\} \neq \varnothing\Big) \\
&= 1 - P\big\{\delta(x) \leq t - \kappa, \hat{\delta}(x) > t, \text{ for some } x \text{ with } \tau_q(x) > \xi\big\} \\
&\quad - P\big\{\delta(x) > t + \kappa, \hat{\delta}(x) \leq t, \text{ for some } x \text{ with } \tau_q(x) > \xi\big\} \\
&\geq 1 - 2P\left\{\sup_{x:\tau_q(x)>\xi} \big|\hat{\delta}(x) - \delta(x)\big| > \kappa\right\},
\end{aligned}$$

where the last line uses Lemma A.1. ■

Lemma A.3. *Suppose that Assumption 3.2 holds. Then, for each* $t, \kappa, \xi \geq 0$,

$$\int \mathbf{1}\left\{x \in \mathcal{X} : \mathbf{1}\left\{\rho(x) > \xi, x \in \hat{\mathcal{X}}_{t+\kappa}\right\} \neq \mathbf{1}\{\rho(x) > \xi, x \in \mathcal{X}_t\}\right\} dP_X(x) \leq \varphi_\delta(2\kappa),$$

with probability at least $p_n(\kappa, \xi)$*, where* $\varphi_\delta(\cdot)$ *is defined in (3.2).*

Proof: The random subset in the outer indicator is equal to the union of the following two random subsets:

$$\left\{x \in \mathcal{X}_t \cap \hat{\mathcal{X}}_{t+\kappa}^{\complement} : \rho(x) > \xi\right\} \text{ and } \left\{x \in \mathcal{X}_t^{\complement} \cap \hat{\mathcal{X}}_{t+\kappa} : \rho(x) > \xi\right\}.$$

We will show that the integration of the union with respect to the distribution P_X of X is bounded by $\varphi(2\kappa)$ with probability $p_n(\kappa, \xi)$. Note that $\rho(x) > \xi$ implies $\tau_q(x) > \underline{c}\xi$ for some $\underline{c} > 0$ due to Assumption 3.2. By Lemma A.2,

$$P_X\left\{x \in \mathcal{X}_t \cap \hat{\mathcal{X}}_{t+\kappa}^{\complement} : \rho(x) > \xi\right\} = 0,$$

with probability $p_n(\kappa, \underline{c}\xi)$. As for the second random set,

$$\begin{aligned}
&P_X\left\{x \in \mathcal{X}_t^{\complement} \cap \hat{\mathcal{X}}_{t+\kappa} \text{ and } \rho(x) > \xi\right\} \\
&= P_X\left\{x \in \mathcal{X}_t^{\complement} \cap \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_{t+2\kappa}^{\complement} \text{ and } \tau_q(x) > \xi\underline{c}\right\} + P_X\left\{x \in \mathcal{X}_t^{\complement} \cap \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_{t+2\kappa} \text{ and } \tau_q(x) > \xi\underline{c}\right\} \\
&\leq P_X\left\{x \in \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_{t+2\kappa}^{\complement} \text{ and } \tau_q(x) > \xi\underline{c}\right\} + P_X\left\{x \in \mathcal{X}_t^{\complement} \cap \mathcal{X}_{t+2\kappa}\right\}.
\end{aligned}$$

The second-to-last term is equal to zero with probability $p_n(\kappa, \underline{c}\xi)$. The last term is equal to

$$P\{t \leq \delta(X) \leq t + 2\kappa\} \leq \varphi_\delta(2\kappa).$$

Thus, we obtain the desired result. ■

Define

$$\begin{aligned}
\tilde{\beta}_t(k) &:= \frac{1}{n_T}\sum_{i=1}^{n_T} \mathbf{1}\left\{X_i \in \hat{\mathcal{X}}_t, \hat{\rho}(X_i) > k\right\} c(X_i), \text{ and} \\
\bar{\beta}_t(k) &:= \frac{1}{n_T}\sum_{i=1}^{n_T} \mathbf{1}\{X_i \in \mathcal{X}_t, \rho(X_i) > k\} c(X_i).
\end{aligned}$$

For each $t \in \mathbf{R}$, recall

$$\alpha(t) := (t \vee 1) \cdot \exp(-2t^2).$$

Lemma A.4. *For each* $b, \kappa, \xi > 0$ *and* $t \geq 0$,

$$P\left\{\sup_{k \geq \xi}\left|\tilde{\beta}_{t+\kappa}(k) - \beta_t(k)\right| \leq b\right\} \geq 1 - 2\tilde{p}_n(\kappa, b, \xi).$$

Proof: We bound the probability of the negation by

$$P\left(\sup_{k\geq\xi}\left|\tilde{\beta}_{t+\kappa}(k)-\bar{\beta}_t(k)\right|>\frac{b}{2}\right)+P\left(\left|\sup_{k\geq\xi}\bar{\beta}_t(k)-\beta_t(k)\right|>\frac{b}{2}\right), \tag{A.6}$$

in which the first term can be further bounded by:

$$\begin{aligned}P&\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i\in\hat{\mathcal{X}}_{t+\kappa}\right\}|\mathbf{1}\{\rho(X_i)>k\}-\mathbf{1}\{\hat{\rho}(X_i)>k\}|>\frac{b}{2\bar{c}}\right)\\&+P\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\{\rho(X_i)>k\}\left|\mathbf{1}\left\{X_i\in\hat{\mathcal{X}}_{t+\kappa}\right\}-\mathbf{1}\{X_i\in\mathcal{X}_t\}\right|>\frac{b}{2\bar{c}}\right).\end{aligned} \tag{A.7}$$

For the leading term on the right-hand side, we have

$$\begin{aligned}&P\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i\in\hat{\mathcal{X}}_{t+\kappa}\right\}|\mathbf{1}\{\rho(X_i)>k\}-\mathbf{1}\{\hat{\rho}(X_i)>k\}|>\frac{b}{2\bar{c}}\right)\\\leq&P\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}|\mathbf{1}\{\rho(X_i)>k\geq\hat{\rho}(X_i)\}+\mathbf{1}\{\hat{\rho}(X_i)>k\geq\rho(X_i)\}|>\frac{b}{2\bar{c}}\right).\end{aligned}$$

We bound the last probability by

$$\begin{aligned}&P\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{\rho(X_i)\in\left[k,k+\frac{\xi}{2}\right]\right\}>\frac{b}{4\bar{c}}\right)\\&\quad+P\left(\sup_{k\geq\xi}\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{\rho(X_i)\in\left[k-\frac{\xi}{2},k\right]\right\}>\frac{b}{4\bar{c}}\right)+P\left(\sup_{x\in\mathcal{X}}|\rho(x)-\hat{\rho}(x)|>\frac{\xi}{2}\right)\\&\leq P\left(\sup_{k\geq\xi}\left(\mathbb{E}_{n_T}-\mathbf{E}\right)\mathbf{1}\left\{\rho(X_i)\in\left[k,k+\frac{\xi}{2}\right]\right\}>\frac{b}{4\bar{c}}-\varphi_\rho(\xi/2)\right)\\&\quad+P\left(\sup_{k\geq\xi}\left(\mathbb{E}_{n_T}-\mathbf{E}\right)\mathbf{1}\left\{\rho(X_i)\in\left[k-\frac{\xi}{2},k\right]\right\}>\frac{b}{4\bar{c}}-\varphi_\rho(\xi/2)\right)\\&\quad+P\left(\sup_{x\in\mathcal{X}}|\rho(x)-\hat{\rho}(x)|>\frac{\xi}{2}\right),\end{aligned} \tag{A.8}$$

where $\mathbb{E}_{n_T}$ denotes the expectation with respect to the empirical distribution of the covariates in the sample from the target population. By Theorem 2.14.28 of Van Der Vaart and Wellner (2023), there are universal constants $C_1', C_2', C_3', C'' > 0$ such that

$$\begin{aligned}&P\left(\sup_{k\geq\xi}\left(\mathbb{E}_{n_T}-\mathbf{E}\right)\mathbf{1}\left\{\rho(X_i)\in\left[k,k+\frac{\xi}{2}\right]\right\}>\frac{b}{4\bar{c}}-\varphi_\rho(\xi/2)\right)\\&\leq C_1'\cdot\left(n_T^{C_2'/2}\left(\frac{b}{4\bar{c}}-\varphi_\rho(\xi/2)\right)^{C_2'}\vee C_2'\right)\cdot\exp\left(-2n_T\left(\frac{b}{4\bar{c}}-\varphi_\rho(\xi/2)\right)^2\right)\end{aligned}$$

$$\leq C'' \cdot \alpha\left(n_T^{1/2}\left(\frac{b}{4\bar{c}} - \varphi_\rho(\xi/2)\right)\right),$$

for each $b, \kappa > 0$ and $t \geq 0$. A similar observation applies to the probability in (A.8).

We turn to the last probability in (A.7). It is bounded by

$$P\left(\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i \in \hat{\mathcal{X}}_{t+\kappa}^{\complement} \cap \mathcal{X}_t\right\} > \frac{b}{4\bar{c}}\right) + P\left(\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i \in \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_t^{\complement} \cap \mathcal{X}_{t+2\kappa}^{\complement}\right\} > \frac{b}{8\bar{c}}\right)$$
$$+ P\left(\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i \in \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_t^{\complement} \cap \mathcal{X}_{t+2\kappa}\right\} > \frac{b}{8\bar{c}}\right).$$

By Lemma A.2, the sets $\left\{x \in \hat{\mathcal{X}}_{t+\kappa}^{\complement} \cap \mathcal{X}_t\right\}$ and $\left\{x \in \hat{\mathcal{X}}_{t+\kappa} \cap \mathcal{X}_{t+2\kappa}^{\complement}\right\}$ are empty with probability $p_n(\kappa, \xi)$. Hence, the sum of the first two probabilities are bounded by $2(1 - p_n(\kappa, \xi))$. The last probability is bounded by

$$P\left(\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\left\{X_i \in \mathcal{X}_t^{\complement} \cap \mathcal{X}_{t+2\kappa}\right\} > \frac{b}{8\bar{c}}\right)$$
$$\leq P\left(\sup_{t \in \mathbf{R}}\left(\mathbb{E}_{n_T} - \mathbf{E}\right)\mathbf{1}\left\{X_i \in \mathcal{X}_t^{\complement} \cap \mathcal{X}_{t+2\kappa}\right\} > \frac{b}{8\bar{c}} - \varphi_\delta(2\kappa)\right)$$
$$\leq C''' \cdot \alpha\left(n_T^{1/2}\left(\frac{b}{8\bar{c}} - \varphi_\delta(2\kappa)\right)\right),$$

for some universal constant $C''' > 0$.

Finally, for the second probability in (A.6), since each term in the sum of $\bar{\beta}_t(k)$ non-negative and bounded above by $\bar{c}$, for some universal constant C_4', it satisfies

$$P\left(\bar{\beta}_t(k) - \beta_t(k) > \frac{b}{2}\right) = P\left(\frac{n_T^{1/2}}{\bar{c}}(\bar{\beta}_t(k) - \beta_t(k)) > \frac{b n_T^{1/2}}{2\bar{c}}\right) \leq C_4' \alpha\left(\frac{b n_T^{1/2}}{2\bar{c}}\right).$$

Putting it together, the sum of the two probabilities in (A.6), up to some multiple of universal constants, is bounded by

$$\alpha\left(n_T^{1/2}\left(\frac{b}{4\bar{c}} - \varphi_\rho(\xi/2)\right)\right) + P\left(\sup_{x \in \mathcal{X}}|\rho(x) - \hat{\rho}(x)| > \frac{\xi}{2}\right) + (1 - p_n(\kappa, \xi))$$
$$+ \alpha\left(n_T^{1/2}\left(\frac{b}{8\bar{c}} - \varphi_\delta(2\kappa)\right)\right) + \alpha\left(\frac{b n_T^{1/2}}{2\bar{c}}\right).$$

The desired result follows because the last term is bounded by the first term. ■

We now present the proofs of Theorems 3.1 and 3.2.

Proof of Theorem 3.1: First, suppose that the budget constraint is not binding and, therefore, $k^* = 0$. Then, for any $\eta > 0$, we have

$$r_\eta := B - \beta(\eta) > 0,$$

and for $\eta > \xi$,

$$\text{(A.9)} \qquad \begin{aligned} P\big(\hat{k}_{\epsilon+\kappa,\xi} > \eta\big) &= P\big(\hat{\beta}_{\epsilon+\kappa}(\eta) > B\big) = P\big(\hat{\beta}_{\epsilon+\kappa}(\eta) - \beta(\eta) > r_\eta\big) \\ &\leq P\bigg(\sup_{x\in\mathcal{X}} |\hat{c}(x) - c(x)| > \frac{r_\eta}{2}\bigg) + R_n(r_\eta/2), \end{aligned}$$

where

$$R_n(r) := P\left(\sup_{k\geq\xi}\left|\frac{1}{n_T}\sum_{i=1}^{n_T}\mathbf{1}\big\{X_i \in \hat{\mathcal{X}}_{\epsilon+\kappa}, \hat{\rho}(X_i) > k\big\}c(X_i) - \mathbf{E}[\mathbf{1}\{X_i \in \mathcal{X}_c, \rho(X_i) > k\}c(X_i)]\right| > r\right).$$

By Lemma A.4, for each $r \geq 0$, this satisfies

$$\text{(A.10)} \qquad R_n(r) \leq 2\tilde{p}_n(\kappa, r, \xi).$$

Second, suppose that the budget constraint is binding, so that $k^* > 0$. We have

$$P\big\{|\hat{k}_{\epsilon+\kappa,\xi} - k^*| > \eta\big\} = P\big\{\hat{k}_{\epsilon+\kappa,\xi} > k^* + \eta\big\} + P\big\{\hat{k}_{\epsilon+\kappa,\xi} < k^* - \eta\big\}.$$

By Assumption 3.4,

$$\begin{aligned} P\big\{\hat{k}_{\epsilon+\kappa,\xi} > k^* + \eta\big\} &\leq P\big\{\hat{\beta}_{\epsilon+\kappa}(k^* + \eta) \geq \beta(k^*)\big\} \\ &\leq P\big\{\hat{\beta}_{\epsilon+\kappa}(k^* + \eta) \geq \beta(k^* + \eta) + s(\eta)\big\}, \end{aligned}$$

and

$$\begin{aligned} P\big\{\hat{k}_{\epsilon+\kappa,\xi} < k^* - \eta\big\} &\leq P\big\{\beta(\hat{k}_{\epsilon+\kappa,\xi}) - s(\eta) > \beta(k^*)\big\} \\ &\leq P\big\{\beta(\hat{k}_{\epsilon+\kappa,\xi}) - \hat{\beta}_{\epsilon+\kappa}(\hat{k}_{\epsilon+\kappa,\xi}) > s(\eta)\big\}. \end{aligned}$$

Therefore,

$$\begin{aligned} P\big\{|\hat{k}_{\epsilon+\kappa,\xi} - k^*| > \eta\big\} &\leq 2P\left\{\sup_{k\geq\xi}\left|\hat{\beta}_{\epsilon+\kappa}(k) - \beta(k)\right| > s(\eta)\right\} \\ &\leq 2P\left\{\sup_{x\in\mathcal{X}}|\hat{c}(x) - c(x)| > s(\eta)\right\} + 2R_n(s(\eta)), \end{aligned}$$

from (A.9). Combining the two cases, we obtain that

$$P\big\{|\hat{k}_{\epsilon+\kappa,\xi} - k^*| > \eta\big\} \leq 2P\left\{\sup_{x\in\mathcal{X}}|\hat{c}(x) - c(x)| > \frac{r_\eta}{2} \wedge s(\eta)\right\} + 2R_n\Big(\frac{r_\eta}{2} \wedge s(\eta)\Big).$$

Hence, we obtain the desired result by (A.10). ■

Proof of Theorem 3.2: We can write

$$\begin{aligned}
&\left|\mathbf{1}\left\{x \in \hat{\mathcal{X}}_{\epsilon+\kappa}, \hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}\right\} - \mathbf{1}\{x \in \mathcal{X}_c, \rho(x) > k^*\}\right| \\
&\le \left|\mathbf{1}\{\rho(x) > k^*\} - \mathbf{1}\left\{\hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}\right\}\right| + \mathbf{1}\{\rho(x) > k^*\}\left|\mathbf{1}(x \in \mathcal{X}_c) - \mathbf{1}\left(x \in \hat{\mathcal{X}}_{\epsilon+\kappa}\right)\right| \\
&\le \left|\mathbf{1}\{\rho(x) > k^*\} - \mathbf{1}\left\{\hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}\right\}\right| + \mathbf{1}\{\rho(x) > \xi\}\left|\mathbf{1}(x \in \mathcal{X}_c) - \mathbf{1}\left(x \in \hat{\mathcal{X}}_{\epsilon+\kappa}\right)\right| \\
&\quad + \mathbf{1}\{k^* < \rho(x) \le \xi\}.
\end{aligned}$$

For the second term, Lemma A.3 states that

$$P_X\left\{x : \rho(x) > \xi, \mathbf{1}(x \in \mathcal{X}_c) \neq \mathbf{1}\left(x \in \hat{\mathcal{X}}_{\epsilon+\kappa}\right)\right\} \le \varphi_\delta(2\kappa),$$

with probability $p_n(\kappa, \xi)$. For the third term, note that

$$P\{k^* < \rho(X_i) \le \xi\} \le \varphi_\rho(\max\{\xi - k^*, 0\}).$$

We tackle the first term. We focus on the event that $\left|\hat{k}_{\epsilon+\kappa,\xi} - k^*\right| \le \eta_1$ and $\sup_{x\in\mathcal{X}}|\hat{\rho}(x) - \rho(x)| \le \eta_2$. Then, on this event,

$$\begin{aligned}
&P_{X_i}\left(\left\{x : \mathbf{1}\{\rho(x) > k^*\} \neq \mathbf{1}\left\{\hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}\right\}\right\}\right) \\
&\le P_{X_i}\left(\left\{x : \hat{\rho}(x) > \hat{k}_{\epsilon+\kappa,\xi}, \rho(x) \le k^*\right\}\right) + P_{X_i}\left(\left\{x : \hat{\rho}(x) \le \hat{k}_{\epsilon+\kappa,\xi}, \rho(x) > k^*\right\}\right) \\
&\le P_{X_i}(\{x : 0 < k^* - \eta_1 - \eta_2 < \rho(x) \le k^*\}) + P_{X_i}(\{k^* < \rho(x) \le k^* + \eta_1 + \eta_2\}) \\
&\le 2\,\varphi_\rho(\eta_1 + \eta_2),
\end{aligned}$$

where the last inequality holds by the conditions of the theorem. ∎

References

ALCANTUD, J. C. R., M. MARIOTTI, AND R. VENEZIANI (2022): "Sufficientarianism," *Theoretical Economics*, 17, 1529–1557.

ALLCOTT, H. (2015): "Site Selection Bias in Program Evaluation," *Quarterly Journal of Economics*, 130, 1117–1165.

ATHEY, S. AND S. WAGER (2021): "Policy Learning with Observational Data," *Econometrica*, 89, 133–161.

BHATTACHARYA, D. AND P. DUPAS (2012): "Inferring Welfare Maximizing Treatment Assignment Under Budget Constraints," *Journal of Econometrics*, 167, 168–196.

BOSSERT, W., S. CATO, AND K. KAMAGA (2023): "Thresholds, Critical levels, and Generalized Sufficientarian Principles," *Economic Theory*, 75, 1099–1139.

CHERNOZHUKOV, V., S. LEE, A. ROSEN, AND L. SUN (2025): "Policy Learning with Confidence," *arXiv:2502.10653*.

DUFLO, E. (2004): "Scaling Up and Evaluation," *Annual World Bank Conference on Development Economics*, 341–369.

FIRPO, S., A. F. GALVAO, M. KOBUS, T. PARKER, AND P. ROSA-DIAS (2023): "Loss aversion and the welfare ranking of policy interventions," *Journal of Econometrics*, 105643.

GECHTER, M. (2024): "Generalizing the Results from Social Experiments: Theory and Evidence from India," *Journal of Business and Economic Statistics*, 42, 801–811.

GECHTER, M. AND R. MEAGER (2022): "Combining Experimental and Observational Studies in Meta-Analysis: A Debiasing Approach," *Working Paper*.

HIGBEE, S. (2023): "Policy learning with new treatments," arXiv:2210.04703.

HOTZ, J. V., G. W. IMBENS, AND J. H. MORTIMER (2005): "Predicting the Efficacy of Future Training Programs Using Past Experiences at Other Locations," *Journal of Econometrics*, 125, 241–270.

KITAGAWA, T. AND A. TETENOV (2018): "Who Should be Treated? Empirical Welfare Maximization Methods for Treatment Choice," *Econometrica*, 86, 591–616.

LESHNO, M. AND H. LEVY (2002): "Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance," *Management Science*, 48, 1074–1085.

MANSKI, C. F. (2004): "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221–1246.

MURALIDHARAN, K. AND P. NIEHAUS (2017): "Experimentation at Scale," *Journal of Economic Perspectives*, 31, 103–214.

SUN, H., E. MUNRO, G. KALASHNOV, S. DU, AND S. WAGER (2025): "Treatment Allocation under Uncertain Costs," *arXiv:2103.11066*.

SUN, L. (2024): "Empirical Welfare Maximization with Constraints," arXiv:2103.15298.

VAN DER VAART, A. W. AND J. A. WELLNER (2023): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer, second ed.

VIVIANO, D. AND J. BRADIC (2024): "Fair Policy Targeting," *Journal of the American Statistical Association*, 119, 730–743.

WANG, S. AND D. Y. YANG (2021): "Policy Experimentation in China: The Political Economy of Policy Learning," Tech. rep., National Bureau of Economic Research.