Policy Learning with New Treatments

Samuel D. Higbee

Department of Economics, University of Chicago

samuelhigbee@uchicago.edu

July 2025

I study the problem of a decision maker choosing a policy which allocates treatment to a heterogeneous population on the basis of experimental data that includes only a subset of possible treatment values. The effects of new treatments are partially identified by shape restrictions on treatment response. Policies are compared according to the minimax regret criterion, and I show that the empirical analog of the population decision problem has a tractable linear- and integer-programming formulation. I prove the maximum regret of the estimated policy converges to the lowest possible maximum regret at a rate which is the maximum of $N^{-1/2}$ and the rate at which conditional average treatment effects are estimated in the experimental data. In an application to designing targeted subsidies for electrical grid connections in rural Kenya, I find that nearly the entire population should be given a treatment not implemented in the experiment, reducing maximum regret by over 60% compared to the policy that restricts to the treatments implemented in the experiment.

I am grateful to Max Tabord-Meehan, Alex Torgovitsky, Stéphane Bonhomme, Guillaume Pouliot, Stefan Wager, Chen Qiu, and Davide Viviano for helpful feedback. I would also like to thank seminar participants at the University of Chicago, Washington University in St. Louis, and the North American Winter Meetings of the Econometric Society 2024 for helpful comments.

1 Introduction

Heterogeneous treatment effects are often estimated with a decision problem in mind— should a particular individual be treated? This question has fostered much research in econometrics, statistics, and machine learning. However, relatively less attention has been given to another important margin of the decision— should the treatment itself be adjusted? Whether the treatment is a medical treatment, subsidy, job training, or audit probability, decision makers can usually entertain changing the treatment value that was observed in the data. Even experiments with multivalued treatments may not implement an exhaustive list of treatment values. This is especially true in the social sciences, where testing multiple interventions can be costly, and in the medical sciences, where specific treatment doses are often tested in clinical trials. In this paper I propose a method for allocating treatment to a population when the treatment values themselves can be adjusted to values never before seen in the data. I show how combining the data on existing treatments with economically motivated shape restrictions can be used to design policies that outperform those possible when only previously implemented treatments are considered.

I first formulate a decision problem in which the decision maker observes experimental data on some treatment values and seeks to construct a mapping, or policy, from the space of covariates to the space of treatments in order to maximize some objective function. I assume all experimentation is done before the policy is constructed. This setting, which is common in econometrics, is often referred to as treatment choice or offline policy learning. Examples include Athey and Wager (2021) Bhattacharya and Dupas (2012), Kitagawa and Tetenov (2018) and other examples mentioned in the literature review thereof, Liu (2024), Mbakop and Tabord-Meehan (2021), Qian and Murphy (2011), Sasaki and Ura (2024), Zhang, Tsiatis, et al. (2012), and Zhao et al. (2012). A distinctive feature of this paper as opposed to most policy learning problems is that the set of treatments that the decision maker can consider may be a strict superset of the support of the treatment random variable observed in the data. This extends policy learning to practically relevant situations in which constraints in the design and implementation of experiments or simply differences in the objectives of the experimenter versus decision maker result in only a few treatment values being piloted in the experiment, while the decision maker may want to consider many more.

Despite the lack of data on the impacts of these never-before-implemented treatments, I show how to bound the response to new treatments using simple, economically interpretable restrictions on the shape of treatment response. For example, a financial incentive may be assumed to have a positive effect, exhibit diminishing returns, or satisfy smoothness conditions. Such shape restrictions are often exploited to partially identify treatment effects (e.g. Manski 2009, Mogstad, Santos, and Torgovitsky 2018). The empirical analysis of the present paper demonstrates that such bounds can be adequately informative for choosing whether and

how to implement new treatment values. Based on these bounds, I construct a population decision problem to choose which treatment to assign to each covariate value. I use the minimax regret criterion to evaluate treatment choice under partial identification following Manski (2007).

As in Manski (2004), Kitagawa and Tetenov (2018) and the subsequent literature on empirical welfare maximization methods, I propose a decision rule based on solving the empirical analog of the decision problem as a surrogate for the infeasible population objective. The resulting empirical minimax regret estimator is constructed by minimizing maximum regret across an estimate of the partially identified set of treatment response functions. In this way, the resulting policy is robust to model ambiguity induced by introducing new treatments. Despite involving nested, non-closed form optimization problems which characterize the identified set for treatment response, I show how the optimal policy can be computed using the same linear and integer programming tools common in the policy learning literature. The estimator is thus computationally feasible and can be implemented by widely available software.

I show that the proposed decision rule possesses desirable regret properties. The maximum regret obtained under the estimated policy converges to the smallest possible maximum regret that the decision maker could have achieved in the absence of sampling uncertainty– that is, if the population identified set were observed– uniformly across a set of data distributions. The rate at which the regret of the estimated policy converges to its optimum depends on the estimation rate of the response to the treatments which were observed in the data, and hence is an asymptotic rather than finite-sample convergence guarantee. In the case of discrete covariates, or more generally parametric rates of convergence for estimated treatment effects, the rate of convergence of maximum regret is $N^{-1/2}$. Otherwise, maximum regret converges at the nonparametric rate.

I apply the method to data from Lee, Miguel, and Wolfram (2020c), in which households in rural Kenya were offered one of four prices in 0, 15, 25, or 35 thousand shillings to connect to the electrical grid. I consider a decision maker able to offer prices in increments of 2.5 thousand shillings based on household size and income. This represents a much richer set of fifteen possible treatments, allowing for finer targeting of personalized prices to optimize the cost-effectiveness of the subsidy program. To bound the takeup at these new prices, I assume demand is downward sloping and convex. The estimated minimax regret optimal policy assigns prices that were not implemented in the experiment to nearly the entire population. Moreover, the maximum regret of the estimated policy is over 60% lower than the maximum regret of the best policy that only implements the prices implemented in the experiment, illustrating that constraining the decision maker to treatments that appear in the experimental data can result in suboptimal decisions.

1.1 Related literature

This paper contributes to a growing literature on statistical treatment rules in econometrics beginning with Manski (2004) and Kitagawa and Tetenov (2018), which introduced the now-common empirical welfare maximization framework. I follow a similar strategy of constructing an empirical analog of the population objective, but seek to minimize the worst-case regret that can occur within the identified set of treatment response.

Forecasting the effects of treatments or policies never before observed in the data is a fundamental goal of econometrics, especially when applied as a guide for public policy (see Heckman and Vytlacil (2007) and Manski (2021) for a deep discussion, including a historical overview). Nonetheless, the recent literature on policy learning and treatment choice has generally not considered the introduction of new treatments with partially identified effects.

Previous literature has treatment choice under various forms of partial identification. Manski (2006) and Manski (2010) consider minimax regret treatment choice when the decision maker only observes data under the status quo policy and uses monotonicity restrictions to partially identify the effects of counterfactual treatment intensities on welfare. The analysis is in the population, and issues of statistical estimation are not considered. Manski (2007) considers minimax regret treatment choice when some outcome data is missing not necessarily at random, leading to partial identification of treatment effects, and proposes an empirical analog. In contrast to this paper and much of the statistical policy learning literature, Manski (2006), Manski (2010), and Manski (2007) do not consider restricted policy classes, yielding a decision problem that is separable in covariates. These restricted policy classes are also important for the convergence properties of the estimated policy.

The paper most closely related to this one is Manski (2025), which studies the question of how to allocate new dosage levels of a treatment given experimental evidence on a subset of possible dosage levels. Like this paper, Manski (2025) uses shape restrictions to bound the response to new treatments, and uses the minimax regret criterion to choose a decision rule. Unlike this paper, Manski (2025) assumes population-level quantities are known and hence does not consider statistical properties of estimated decision rules, nor does it consider targeting new treatments on the basis of covariates using complexity-constrained policy classes. The two papers also use different utility functions– the present paper using a linear-in-outcome utility function, while Manski (2025) assuming four discrete outcomes associated with different utility levels. Finally, Manski (2025) considers fractional treatment assignment in addition to the deterministic treatment assignment considered in this paper.

Ben-Michael et al. (2022) develops a method for learning policies from data gathered under a deterministic

policy for which strict overlap fails; Zhang, Ben-Michael, and Imai (2024) tailors this framework to the case where the deterministic policy is a regression discontinuity design. The introduction of new treatments is similar to the deterministic policy setting considered here in that it is also a case where strict overlap fails. Ben-Michael et al. (2022) and Zhang, Ben-Michael, and Imai (2024) use a maximin gain welfare criterion, where the objective is to learn a policy that is guaranteed to weakly improve on the status quo policy. Khan, Saveski, and Ugander (2024) studies robust policy evaluation using Lipschitz constraints when strict overlap fails. Ben-Michael et al. (2022), Zhang, Ben-Michael, and Imai (2024), and Khan, Saveski, and Ugander (2024) focus on partial identification through restrictions on response as a function of covariates, while this paper focuses on shape restrictions on the response across treatment values.

Unobserved confounding can be a source of partial identification in policy choice with observational data. Kallus and Zhou (2021) studies this setting, and use bounds on the distance between the true propensity weights and the observed (biased) weights to partially identify the effect of policies. Their criterion is maximum regret relative to a baseline policy such as the status quo, and like the present paper they show that the maximum regret of the estimated policy converges to the lowest possible maximum regret. Pu and Zhang (2021) uses an instrumental variable to partially identify treatment effects when unobserved confounding precludes point identification, and proposes a classification-based approach with a surrogate loss to learn the optimal policy under a maximin welfare criterion.

While unobserved confounding threatens the internal validity of the estimated policy on the experimental population, other papers consider threats to external validity, where the experimental population is different from the target population. Adjaho and Christensen (2023) studies this setting, using Wasserstein neighborhoods to construct the identified set, and derive closed form expressions worst case welfare within these neighborhoods. Lei, Sahoo, and Wager (2023) studies policy learning when the experimental population may self-select into the experiment on the basis of unobserved characteristics. They consider maximin, maximin gain, and minimax regret policies. They solve for a closed form when the policy class is unconstrained, and propose an estimation method which is not a plug-in method.

D'Adamo (2023) studies policy learning with a binary treatment where the identified set is rectangular, meaning it is constructed by taking the product of pointwise bounds on each treatment effect. In contrast, shape restrictions on the response across treatment values generally yield nonrectangular identified sets. This leads to difficulties when estimating the optimal policy in my setting because the bounds I identify do not in general admit a closed form. However, the extra information provided by these shape restrictions can lead to lower maximum regret than one would obtain using pointwise bounds. This is illustrated in the empirical example of Section 5. D'Adamo (2023) also provides a doubly robust estimator that can improve the convergence rate of the estimated policy under a margin condition.

Stoye (2012) gives exact finite-sample results for minimax regret treatment choice in Binomial and Gaussian experiments, also with unrestricted policy classes. Yata (2025) gives exact finite-sample minimax regret results in more general Gaussian settings with binary policies. These papers do not consider treatment choice with multiple treatments. Another difference is that the present paper only delivers asymptotic performance guarantees, but does not require distributional assumptions.

Many of the previously mentioned works are concerned with binary treatments, while I am concerned with multivalued treatments. Zhou, Athey, and Wager (2023) and Kallus and Zhou (2018) consider policy learning with multivalued treatments and continuous treatments, respectively, but in point-identified settings where all possible treatment values are implemented in the experiment.

Athey and Wager (2021) extends policy learning to observational studies where exogeneity of treatment only holds after conditioning on high-dimensional covariates. In contrast, I am motivated by settings in which decision makers have data from a pilot experiment which tested a few treatment values. When this is the case, estimating the effects of policies involving new treatments only requires conditioning on the set of covariates used in the treatment rule, which is typically low-dimensional due to exogenous constraints on the policy class (Kitagawa and Tetenov 2018). Athey and Wager (2021) also considers infinitesimal, local changes to treatment values; however, I consider new treatments that are sufficiently far from the support of the data as to make local approximations or parametric extrapolations unreliable, necessitating a partial identification approach.

An alternative to the plug-in approach used in this paper and common in policy learning is to average across the parameter space according to some distribution. Christensen, Moon, and Schorfheide (2023) study optimal decisions in a discrete set under partial identification where Bayes rules and the bootstrap distribution are used to average over the space of identified parameters, while a minimax approach is taken over the partially identified parameters. An important finding is that plug-in-rules may be dominated in the limit experiment. See Hirano and Porter (2009) and Hirano and Porter (2020) for further discussion of asymptotic optimality of statistical treatment rules.

The rest of the article is organized as follows: Section 2 describes the decision problem in the population and shows how to incorporate information from shape restrictions. Section 3 describes the empirical minimax regret problem and the algorithm for estimating the optimal policy. Section 4 describes the convergence guarantees. Section 5 applies the method to study personalized subsidies to connect to the electrical grid in rural Kenya.

2 Population decision problem

2.1 General framework

A decision maker has access to experimental data and must choose a rule assigning individuals to treatments based on their observable covariates. The experimental data is described by random variables (D, X, Y) taking values in $\mathcal{D}_0 \times \mathcal{X} \times \mathcal{Y}$ where D is a treatment taking $|\mathcal{D}_0| = J_0$ values in the data, X are observed covariates, and Y is a univariate outcome of interest. D is assumed to be randomly assigned, perhaps conditionally on X.

Although the random variable D only takes values in $\mathcal{D}_0$, the decision maker can consider assigning individuals to any treatment value $d \in \mathcal{D}$ where $\mathcal{D}$ is potentially larger than $\mathcal{D}_0$. Hence, I assume the existence of potential outcomes $Y(d)$ for all $d \in \mathcal{D}$. The set $\mathcal{D}$ has cardinality $|\mathcal{D}| = J < \infty$, and its elements are denoted by d_j for $j \in \{1, \ldots, J\}$. The observed outcome Y is generated as $Y = Y(D)$. Let P denote the distribution of $(D, X, (Y(d))_{d \in \mathcal{D}})$.

The decision maker seeks a policy $\pi : \mathcal{X} \mapsto \mathcal{D}$ which assigns individuals to treatment status based on their observable covariates. The policy is chosen from some set Π which is taken as given. The treatment assigned to an individual with covariate values X is $\pi(X)$ and the realized outcome is $Y(\pi(X))$.

The decision maker has some utility function $u(d, x, y)$ which may depend on the treatment assigned, covariates, and the realized outcome of interest. I assume the decision maker is utilitarian and ultimately cares about the expected utility derived from the data realized from the policy, resulting in the following problem that the decision maker would like to solve

$$\begin{aligned} &\max_{\pi \in \Pi} \mathbb{E}_P \Big[u(\pi(X), X, Y(\pi(X))) \Big] \\ &= \max_{\pi \in \Pi} \mathbb{E}_P \Big[v_P(\pi(X), X) \Big] \end{aligned}$$

where $v_P(d, x) := \mathbb{E}_P[u(d, X, Y(d)) \mid X = x]$ is the conditional mean utility.

Two sources of ignorance on the decision maker's part make this problem infeasible to solve in practice. The first is that only a sample is observed, so the population probability distribution is unknown. The second is that even if the population distribution of the data (D, X, Y) were known, the effects of some treatments are not identified because they are never observed. In particular, the function v depends on the distribution of potential outcomes $Y(d)$ for values of d not in $\mathcal{D}_0$. Since data on these potential outcomes are not observed in the sample, the decision maker's objective is not point identified. To deal with partial identification, I will solve a proxy problem which is robust to partial identification in that it achieves uniformly low regret

across the identified set for v. Since only sample data is available, I solve the empirical or plug-in version of this problem.

Following Manski (2004) and much of the econometric literature on treatment choice, policies will be evaluated based on their expected regret. For a chosen policy π, the regret of π is the difference in expected utility obtained from implementing the first-best policy versus π. The first-best policy maps each covariate value x to $\arg\max_d v_P(d,x)$. For any chosen policy π, the expected regret of implementing π versus implementing the first-best policy is

$$R_P(\pi) := \mathbb{E}_P\left[\max_d v_P(d,X) - v_P(\pi(X),X)\right].$$

Since v_P is not identified, regret is not identified either. However, letting $\mathcal{V}_P$ be the identified set for v_P determined by the experimental data (which will be characterized shortly), the maximum expected regret that can occur if the decision maker implements policy π is given by

$$\overline{R}_P(\pi) := \max_{v\in\mathcal{V}_P} \mathbb{E}_P\left[\max_{d\in\mathcal{D}} v(d,X) - v(\pi(X),X)\right]$$

I use the minimax regret criterion to guide the choice of policy. This means π is chosen to minimize the largest regret that can occur within the identified set— that is, $\overline{R}_P(\pi)$. Therefore, the decision maker chooses π to minimize the worst-case expected regret as follows

$$\pi_P^* \in \arg\min_{\pi\in\Pi} \overline{R}_P(\pi) = \arg\min_{\pi\in\Pi} \max_{v\in\mathcal{V}_P} \mathbb{E}_P\left[\max_{d\in\mathcal{D}} v(d,X) - v(\pi(X),X)\right] \tag{1}$$

This ensures that the chosen policy minimizes regret uniformly across the identified set. If the minimizer is not unique, the decision maker is indifferent among them. Since v is unknown, the minimum maximum regret is generally larger than zero.

The minimax regret criterion is not the only method for comparing statistical decisions with partially identified effects. In the context of treatment choice, Manski (2011) compares the minimax regret criterion with the maximin welfare and subjective expected welfare criteria, two common alternatives. Under the maximin welfare criterion, the decision maker seeks to maximize the minimum possible level of the outcome that could be attained as opposed to the minimum gap between the attained and first-best level of the outcome. The method for construction and estimation of the optimal policy that follows can be applied when using the maximin welfare criterion as well. Indeed, it can be obtained as a simplification of what follows by replacing $\max_{d\in\mathcal{D}} v(d,X)$ with 0 in (1). However, the resulting estimator will of course have different behavior and regret properties.

In some settings the maximin criterion can be quite conservative (see the discussion of Wald (1949) found in Savage (1951)). Indeed, unless a new treatment $d \in \mathcal{D} \setminus \mathcal{D}_0$ can be guaranteed to outperform the original set of treatments in every possible state of the world $v \in \mathcal{V}_P$, the maximin welfare criterion will not implement new treatments. This is because under the maximin welfare criterion the decision is driven entirely by hedging against the least favorable state of the world. In contrast, the minimax regret criterion considers the suboptimality gap in all possible states of the world. The decision maker measures the performance of the policy in each state of the world according to the benchmark of optimality in that state of the world. I follow Manski (2007) in applying the minimax regret criterion to treatment choice. This represents a particular choice of loss function and in turn delivers a point estimate of an optimal policy.

When the probability of each state of the world $v \in \mathcal{V}_P$ can be described by a probability distribution, the Bayesian approach to decision-making can be applied. This consists of setting a prior on states of the world $v \in \mathcal{V}_P$, using the data to form a posterior, and selecting a treatment policy which maximizes posterior expected welfare. One potential weakness of this approach in the context of introducing new treatments is that the lack of identification means that even in large samples, the influence of the prior on the posterior will be substantial. Yet another possible approach to estimate the effects of treatments that lie outside the support of the data could be to extrapolate using a parametric model, thus circumventing entirely the need for partial identification. However, when the new treatments are sufficiently far from the support of the data, a parametric point-identified model substantially understates the degree of model uncertainty. This is illustrated in Section 5 where a policy based on parametric extrapolation leads to substantially higher maximum regret than the estimated minimax regret policy.

2.2 Imposing shape restrictions

I now describe how a tractable characterization of the minimax regret problem (1) can be obtained using shape restrictions on the treatment response. This requires that the utility function is linear in the outcome of interest. That is, there exist known functions b and c such that

$$u(d, x, y) = b(d, x)y - c(d, x).$$

While it is often possible to avoid the assumption of linear utility by simply redefining Y as utility, in some applications (such as in Section 5) it may be more natural to impose shape restrictions in terms of the original outcome variable, which may relate to a structural economic quantity such as a demand curve. In Section 5, y will be a purchase indicator, $b(d, x)$ will be value of connections net of the cost of the subsidy,

and $c(d,x)$ represents the cost of offering the subsidy regardless of takeup, which I take to be 0.[1]

Note that the assumption of linearity implies that

$$v_P(d,x) = b(d,x)m_P(d,x) - c(d,x)$$

where $m_P(d,x) := \mathbb{E}_P[Y(d) \mid X = x]$ is the conditional mean response function. Moreover, any two probability distributions which induce the same conditional mean response function will induce the same expected utility function. I therefore will also use the notation $v_m(d,x) := b(d,x)m(d,x) - c(d,x)$ where conditional mean utilities are indexed by conditional mean response functions rather than probability distributions. Since b and c are known functions, in order to characterize maximum regret it is sufficient to characterize the identified set for m_P.

The decision maker has experimental data on the effectiveness of some treatments. This means that $m_P(d,\cdot)$ is identified for every $d \in \mathcal{D}_0$. For this information on the effects of treatments in $\mathcal{D}_0$ to be informative about the effects of treatments in $\mathcal{D} \setminus \mathcal{D}_0$, some structure must be known about the mean conditional response function m_P. For example, the decision maker may know that demand is downward sloping, that a particular intervention features decreasing returns to scale, or that the treatment response exhibits some smoothness properties. By combining knowledge of $m_P(d,\cdot)$ for $d \in \mathcal{D}_0$ with such shape restrictions, the effects of new treatments may be partially identified.

Let the set of shape-restricted mean conditional response functions be denoted by $\mathcal{S}$. The sharp identified set for m_P is

$$\mathcal{M}_P := \mathcal{S} \cap \{m : m(d,X) = m_P(d,X),\ P - \text{a.s.},\ \forall d \in \mathcal{D}_0\}$$

which represents the set of functions which obey the shape restrictions and match identified population means. I assume that $\mathcal{S}$ restricts the shape of $m(d,x)$ in d for any given x, leaving the behavior of m across x unrestricted.

Assumption 2.1: *There exist sets $\mathcal{S}_x \subset \mathbb{R}^J$ for each $x \in \mathcal{X}$ such that*

$$\mathcal{S} = \{m : m(\cdot,X) \in \mathcal{S}_X\ P - a.s.\}.$$

Under Assumption 2.1, a hypothetical conditional mean response function m is in $\mathcal{S}$ if and only if $m(\cdot,X)$ satisfies some shape restrictions almost surely in X. That is, $\mathcal{S}$ encapsulates assumptions about the shape

[1] If $b(d,x)$ and $c(d,x)$ represent preferences of a population, they may have to be estimated from the data. While this paper focuses on uncertainty about the response of y to new treatments, Appendix B discusses how to extend the methods to the case where $b(d,x)$ and $c(d,x)$ are estimated.

of m_P across d for fixed x, leaving the behavior of m_P across x unrestricted (Manski 1997, Manski 2006). This means that an individual at a particular covariate value is assumed to have an expected treatment response that is decreasing, convex, smooth, etc. The sets S_x may also stipulate that $m(\cdot, x)$ belongs to some parametric family, such as polynomials.

Under Assumption 2.1, the maximization over v (equivalently maximization over $m \in \mathcal{M}_P$) in (1) is solved by considering each value of x in isolation and finding the $m(\cdot, x)$ which maximizes regret. This allows the maximum to be interchanged with the expectation in the minimax regret problem (1)

$$\begin{aligned}
&\min_{\pi \in \Pi} \max_{m \in \mathcal{M}_P} \mathbb{E}_P \left[\max_{d \in \mathcal{D}} v_m(d, X) - v_m(\pi(X), X) \right] \\
&\quad = \min_{\pi \in \Pi} \mathbb{E}_P \left[\max_{m \in \mathcal{M}_P} \left(\max_{d \in \mathcal{D}} v_m(d, X) - v_m(\pi(X), X) \right) \right] \\
&\quad = \min_{\pi \in \Pi} \mathbb{E}_P \left[\sum_{j=1}^{J} \pi_j(X) \Gamma_{j,P}(X) \right] \qquad (2)
\end{aligned}$$

where $\pi_j(X) = \mathbb{1}[\pi(X) = d_j]$ and

$$\Gamma_{j,P}(X) = \max_{m \in \mathcal{M}_P} \left(\max_{d \in \mathcal{D}} v_m(d, X) - v_m(d_j, X) \right)$$

which can be interpreted as the contribution to maximum expected regret of assigning a person with covariate values X to treatment d_j.

The optimization problem (2) defines the policy which is optimal in terms of its population minimax regret. The maximum regret of any policy depends on the strength of the assumptions encoded in $\mathcal{S}$, and their implications for the size of the identified set $\mathcal{M}_P$. Larger identified sets will lead to higher maximum regret, since it expands the set from which the worst-case response m can be chosen. The size of the identified set also depends on the relationship between the new and existing treatments. If a new treatment d_j lies between two existing treatments, restrictions on m such as monotonicity can provide informative bounds on $m(d_j, x)$. When $d_j > d$ for all $d \in \mathcal{D}_0$, monotonicity can leave the identified set unbounded in the absence of additional assumptions.

The benefit of imposing shape restrictions only on the behavior of m_P across d is that the representation (2) is an optimization problem over a population expected loss defined by $\Gamma_{j,P}(X)$. This problem possesses a form similar to decision problems presented in Athey and Wager (2021), D'Adamo (2023) and others, with the key distinction that the covariate-level loss $\Gamma_{j,P}(X)$ is itself the solution to an optimization problem which generally will not have a closed-form solution. Nonetheless, the minimax regret problem (2) can be cast in terms of the empirical welfare maximization framework of Kitagawa and Tetenov (2018). In the

following section, I discuss how to set up the empirical analog of the nested optimization problem (2) and provide a computationally attractive algorithm for solving it.

3 Estimation

The optimization problem (2) is infeasible for the decision maker because in practice only a sample $\{(D_i, X_i, Y_i)\}_{i=1}^N$ is observed. Instead, I propose solving the empirical analog of (2) to obtain an estimate of the population optimal policy. Insofar as the constraints of this problem are constructed from consistent estimators, the optimal policy will inherit similar properties.

In this section I describe the empirical analog of (2) and provide a solution procedure. It consists of first estimating the effects of the treatments which were implemented in the experimental data, then constructing estimates of $\Gamma_{j,P}(X_i)$ for every observation i and treatment j, and finally plugging these estimates into the empirical analog of (2) where the sample mean is used instead of the population expectation. I show how these estimates of $\Gamma_{j,P}(X_i)$ can be computed using linear programming, resulting in a mixed integer-linear programming formulation for (2) for many policy classes Π.

First, I estimate the mean conditional response function for every $d \in \mathcal{D}_0$, denoted $\hat{m}_0(d, x)$. Except for high level conditions on the accuracy of the estimate detailed in Section 4, I remain agnostic about how the estimate is constructed. The estimate $\hat{m}_0$ is used to construct an estimate of the identified set for $m_P(\cdot, X_i)$ for each i as a function of d in all of $\mathcal{D}$, which represents covariate-level bounds on the effects of new treatments. The empirical analog of $\mathcal{M}_P$ is the set of functions which obey the shape restrictions and match estimated sample means, and is denoted by $\hat{\mathcal{M}} := \mathcal{S} \cap \{m : m(d, X_i) = \hat{m}_0(d, X_i)\ \forall i,\ \forall d \in \mathcal{D}_0\}$. I assume it is nonempty. As discussed in Section 4, estimates which violate the shape restrictions and hence yield an empty $\hat{\mathcal{M}}$ can be projected onto the set of functions which satisfy the shape restrictions. Since these are assumed to hold in the population, imposing such shape restrictions on estimators typically improves performance in finite samples (Chetverikov, Santos, and Shaikh 2018).

This is then used to construct estimates $\hat{\Gamma}_j(X_i)$ of the covariate-level loss $\Gamma_{j,P}(X_i)$, for every observation i and treatment j. That is,

$$\hat{\Gamma}_j(X_i) = \max_{m \in \hat{\mathcal{M}}} \left(\max_{d \in \mathcal{D}} v_m(d, X_i) - v_m(d_j, X_i) \right) \tag{3}$$

These estimates are then used in the program

$$\hat{\pi} := \arg\min_{\pi \in \Pi} \sum_{i=1}^{n} \sum_{j=1}^{J} \pi_{ij} \hat{\Gamma}_j(X_i) \tag{4}$$

where $\pi_{ij} = \mathbb{1}[\pi(X_i) = d_j]$.

Having defined the estimator for the minimax regret optimal policy (4), I turn to computationally convenient methods for estimating $\hat{\Gamma}_j(X_i)$ and thereby the policy $\hat{\pi}$. This is achieved by expressing $\hat{\Gamma}_j(X_i)$ through linear programs and considering policy classes Π which can be expressed using linear and integer constraints. In doing so, I impose some additional structure on the set of shape restricted functions $\mathcal{S}$ specified in Assumption 2.1. Specifically, I assume the a priori knowledge on shape restrictions can be summarized through ℓ linear inequalities on the treatment response vector for almost every x.

Assumption 3.1: *There exists a matrix $S \in \mathbb{R}^{\ell \times J}$ and a vector $r \in \mathbb{R}^{\ell}$ such that the conditional mean response function $m(d,x)$ is in the set $\mathcal{S}$ if and only if $Sm(\cdot, X) \leq r$ $P-$almost surely, where $m(\cdot, x) := (m(d_1, x), \ldots, m(d_J, x))'$.*

This assumption strengthens Assumption 2.1 by requiring that the shape restrictions are linear in the treatment response vector. Such linear restrictions can accommodate a wide range of shape restrictions that may be used in practice. For example, restrictions on the first, second, or higher differences of the mean conditional response can be expressed this way, allowing for m_P to be constrained to be decreasing, Lipschitz, convex, or obey higher order smoothness conditions (Mogstad, Santos, and Torgovitsky 2018).[2] Upper and lower bounds on m_P can also be expressed through such constraints. Appendix D describes in detail how the restrictions of decreasing demand and diminishing responsiveness to the subsidy are applied to the empirical example in Section 5.

The following examples convey the practical use of the assumption.

Example 3.2: Suppose $\mathcal{D} = \{1, 2, 3, 4\}$ and $m_P(d, x)$ is assumed to be increasing and concave in d. Then $m \in \mathcal{S}$ if and only if $Sm(\cdot, X) \leq r$ P-almost surely where

$$S = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \end{bmatrix} \qquad r = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Here the first three rows of S ensure that m is increasing, and the second two rows of S ensure concavity. $\lozenge$

Example 3.3: Suppose $m_P(\cdot, x)$ is assumed to be a polynomial of degree L where $J_0 - 1 \leq L \leq J - 1$, in addition to shape restrictions such as boundedness or monotonicity. These restrictions can be imposed by

[2] The analysis can be extended to allow S and r to depend on x. I have focused on the case where the shape restrictions are the same for all x for simplicity.

using Bernstein polynomials. That is,

$$\mathcal{S}_x = \left\{ m \in \mathbb{R}^J : m_j = B(d_j, L)'\beta,\ T\beta \leq r\ \beta \in \mathbb{R}^{L+1} \right\}$$

where $B(d_j, L)$ is an $L+1$ vector of Bernstein polynomials of degree L in d_j, and β is a vector of coefficients. The matrix T encodes shape restrictions on m through constraints on the coefficients β.

Let B be the $J \times (L+1)$ matrix of Bernstein polynomials of degree L. Since B has linearly independent columns, it has a left inverse $B^\dagger$ and $m \in S_x$ if and only if $Sm \leq r$ where $S = TB^\dagger$. $\lozenge$

To ensure that $\hat{\mathcal{M}}$ consists of functions which match sample analogs of identified means on $\mathcal{D}_0$, I introduce the $J_0 \times J$ matrix F where $F_{kj} = 1$ if d_j is the kth element of $\mathcal{D}_0$ and $F_{kj} = 0$ otherwise. Then the identified set $\mathcal{M}_P$ is the set of all m such that $Sm(\cdot, X) \leq r$ and $Fm(\cdot, X) = m_{0,P}(\cdot, X)$ almost surely, where $m_{0,P}(\cdot, X) = (m_P(d, X))_{d \in \mathcal{D}_0}$. That is,

$$\mathcal{M}_P = \{m : Sm(\cdot, X) \leq r,\ Fm(\cdot, X) = m_{0,P}(\cdot, X),\ P - \text{a.s.}\}$$

which allows the empirical analog $\hat{\mathcal{M}}$ to be expressed as

$$\hat{\mathcal{M}} = \{m : Sm(\cdot, X_i) \leq r\ \forall i, Fm(\cdot, X_i) = \hat{m}_0(\cdot, X_i)\ \forall i\}$$

where $\hat{m}_0(\cdot, x)' = (\hat{m}_0(d, x))_{d \in \mathcal{D}_0}$. By expressing $\hat{\mathcal{M}}$ this way it is possible to express $\hat{\Gamma}_j(X_i)$ as the maximum of J linear programs. Define the estimate

$$\hat{\Gamma}_{jk}(X_i) := \max_{m \in \hat{\mathcal{M}}} v_m(d_k, X_i) - v_m(d_j, X_i)$$

which measures the contribution to expected regret of assigning an individual d_j instead of assigning them d_k, conditional on X_i. Recall that $b(d, x)$ and $c(d, x)$ parametrize the linear utility function. For each observation i and treatments j and k, construct $b_{jk}(X_i)$ as a vector in $\mathbb{R}^J$ with $b(d_k, X_i)$ in the k^{th} entry and $b(d_j, X_i)$ in the j^{th} entry, and zeros everywhere else. Construct $c_{jk}(X_i)$ likewise. Then

$$\begin{aligned} \hat{\Gamma}_{jk}(X_i) = \max_{m \in \mathbb{R}^J} \quad & b_{jk}(X_i)'m - c_{jk}(X_i) \\ s.t. \quad & Sm \leq r \\ & Fm = \hat{m}_0(\cdot, X_i). \end{aligned} \tag{5}$$

After defining $\hat{\Gamma}_j(X_i) = \max_k \hat{\Gamma}_{jk}(X_i)$, these estimates can be used in the program (4).

Despite the linear programming representation of $\hat{\Gamma}_{jk}(X_i)$, computing $\hat{\Gamma}_j(X_i)$ for all i and j appears to require NJ^2 linear programs total (one for each i, j, and k combination). However, a dual formulation detailed in Appendix D demonstrates that $\hat{\Gamma}_j(X_i)$ can be computed with a single linear program. Moreover, the linear programs for all observations can be stacked together and solved simultaneously. This can be done prior to or in conjunction with the policy optimization over π. Additionally, in the case of discrete covariates, it is only necessary to construct $\hat{\Gamma}_j(X_i)$ for unique values of X_i, which may be substantially smaller than the sample size N.

Having computed the covariate-level loss estimates $\hat{\Gamma}_j(X_i)$ that appears in (4), optimization of the policy π over the set Π can be performed according to established methods in policy learning. In many cases, Π can be represented by linear and integer constraints. Examples include linear eligibility scores, decision trees, and treatment sets with piecewise linear boundaries (Kitagawa and Tetenov 2018, Mbakop and Tabord-Meehan 2021, Zhou, Athey, and Wager 2023). When this is the case, the problem (4) is a mixed integer-linear program for which highly optimized solvers are readily available. In Section 5, I use a class of linear eligibility score policies, which is described using linear and integer constraints in Appendix D.

Since optimization of π using mixed integer-linear programming is standard practice in policy learning problems, the only additional computational burden resulting from considering new treatments is that of solving the linear programs corresponding to $\hat{\Gamma}_j(X_i)$ as described above. For the example in Section 5 I found the computation time for $\hat{\Gamma}_j(X_i)$ to be at most a similar order of magnitude as that of the estimation of $\hat{\pi}$ and sometimes much shorter, depending on the complexity of the policy class. Constructing $\hat{\Gamma}_j(X_i)$ can often benefit from parallelization so that the overall computational burden is not much larger than the point identified case.

4 Regret convergence

In this section I investigate theoretical guarantees on the performance of the estimated policy $\hat{\pi}$. Following Manski (2004), I evaluate the performance of policies in terms of their statistical regret. In particular, I show that the regret of the estimated policy $\hat{\pi}$ converges to the lowest possible maximum regret the decision maker could achieve if the population identified set under distribution P were observed, uniformly across P. Specifically, the regret guarantees will be of the form

$$\sup_P \left(\mathbb{E}_P[\overline{R}_P(\hat{\pi})] - \overline{R}_P(\pi_P^*) \right) \leq \mathcal{O}(N^{-1/2} \vee \rho_N^{-1})$$

where P ranges across an appropriate set defined below, implying that

$$\sup_P \mathbb{E}_P[\overline{R}_P(\hat{\pi})] \leq \sup_P \overline{R}_P(\pi_P^*) + \mathcal{O}(N^{-1/2} \vee \rho_N^{-1})$$

for an appropriate sequence $\rho_N \to \infty$. Since the estimated policy $\hat{\pi}$ is constructed using consistent estimates of the partially identified set, these guarantees will generally be asymptotic in nature. Above, the expectation only averages across realizations of the estimator $\hat{\pi}$ because $\overline{R}_P(\cdot)$ is defined using the population probability measure P.

The interpretation of this bound is that in large samples the performance of the estimated policy $\hat{\pi}$ as measured by its maximum regret across distributions P (and the identified set under P) approaches the performance of the population optimal policy π_P^*. This bound on the difference between the maximum regret of $\hat{\pi}$ and the best-in-class policy π_P^* is similar to the bounds often obtained in the empirical welfare maximization or empirical risk minimization literature in the point-identified case (e.g. Kitagawa and Tetenov (2018)) after replacing (unidentified) welfare with maximum regret. Since $0 \leq \overline{R}_P(\pi_P^*) \leq \overline{R}_P(\hat{\pi})$, this means that averaging across realizations of the estimate $\hat{\pi}$, the worst-case expected regret of $\hat{\pi}$ is growing arbitrarily close to $\overline{R}_P(\pi_P^*)$, the lowest possible maximum regret the decision maker could achieve in the absence of sampling uncertainty. In general no policy π can achieve zero maximum regret across the entire identified set, resulting in $\overline{R}_P(\pi_P^*) \geq 0$ typically holding with strict inequality for the population optimal π_P^*.

I now discuss assumptions sufficient for such guarantees. The main assumptions on the joint distribution of the data are random assignment of treatment and boundedness of the components of utility.

Assumption 4.1: *$\{(Y_i, D_i, X_i)\}_{i=1}^n$ are i.i.d. copies of (Y, D, X), generated by P which satisfies*

1. *$D \perp Y(d) \mid X$ for all $d \in \mathcal{D}$*
2. *There exists $C < \infty$ such that $m_P(d, X)$, $b(d, X)$, and $c(d, X)$ are all bounded in absolute value by C almost surely, for each $d \in \mathcal{D}$*

Assumption 4.1.1 reflects the standard exogeneity condition that holds in the randomized experiment settings I use as a motivating example. It may also hold in observational studies, in which case it may be a strong assumption. In many randomized experiments the stronger condition $D \perp (Y(d), X)$ is satisfied. When this is true, it is sufficient to estimate $m_P(d, x) = \mathbb{E}_P[Y(d) \mid \tilde{X} = x]$, where $\tilde{X}$ is a subset of covariates X which directly enter the policy. $\tilde{X}$ may be of a much lower dimension than X since policies are often restricted to be relatively simple (Kitagawa and Tetenov 2018). In Section 5, there are two covariates which enter the policy. Henceforth, I do not distinguish between the covariates required for Assumption 4.1.1 and the covariates used for the policy. Assumption 4.1.2 restricts decision maker preferences by requiring that the

mean response function is bounded, as well as the parameters of the linear utility function. In the example of Section 5 the outcome Y is bounded while $b(d,x)$ is constant in x and $c(d,x)=0$, satisfying this condition trivially.

The estimate $\hat{m}$ also must be sufficiently accurate in the following sense

Assumption 4.2: *For some sequence $\rho_N \to \infty$ and some class of distributions $\mathcal{P}$, the estimate $\hat{m}$ satisfies*

1. $\limsup_{N\to\infty} \sup_{P\in\mathcal{P}} \rho_N \mathbb{E}_P \left[\frac{1}{N} \sum_{i=1}^{N} \|\hat{m}_0(\cdot, X_i) - m_{0,P}(\cdot, X_i)\| \right] < \infty$
2. $\hat{\mathcal{M}} = \{m : Sm(\cdot, X) \leq r,\ Fm(\cdot, X) = \hat{m}_0(\cdot, X)\}$ *is nonempty, almost surely, for all $P \in \mathcal{P}$.*

One common setting in which Assumption 4.2.1 holds with $\rho_N = N^{1/2}$ is when the covariates are discrete and sample averages may be used. Alternatively, $m_P(d,\cdot)$ may be assumed to belong to a parametric family, for each value of $d \in \mathcal{D}_0$. Since $m_P(d,\cdot)$ is identified holding $d \in \mathcal{D}_0$ fixed, parametric assumptions on the relationship between covariates and the outcome of interest conditional on treatment values observed in the data may be weaker assumptions than the kinds of parametric assumptions that would allow one to extrapolate to new treatments, in the sense that the former are testable. Kitagawa and Tetenov (2018) provides more general conditions under which Assumption 4.2.1 is satisfied when $\hat{m}$ is constructed via local polynomial regression.

Assumption 4.2.2 is not a restrictive assumption. Unless m_P is on the boundary of $\mathcal{S}$, the estimated set $\hat{\mathcal{M}}$ will typically be nonempty with high probability as N grows even if 4.2.2 is not assumed. In finite samples, an estimator that yields an empty $\hat{\mathcal{M}}$ can be projected onto the set of all $\hat{m}$ such that $\hat{\mathcal{M}}$ is nonempty. Since $\hat{m}_0(\cdot, X)$ is a vector in $\mathbb{R}^{J_0}$ and $\mathcal{S}$ is described by linear inequalities, this is a convex minimum norm problem that can be solved by quadratic programming.

Finally, the choice set Π is assumed to satisfy a standard condition on its complexity.

Assumption 4.3: *For each $d \in \mathcal{D}$, the class of sets $\{x : \pi(x) = d,\ \pi \in \Pi\}$ is a VC-class of sets with VC dimension at most $V < \infty$.*

For a formal definition of the VC dimension, see Van Der Vaart and Wellner (1996). The assumption of finite VC dimension limits the complexity of the class Π; specifically, Assumption 4.3 ensures that Π cannot be so flexible as to assign any arbitrary subset of a collection of $V+1$ points in $\mathcal{X}$ to treatment d. This assumption is commonly invoked in offline policy learning settings as a way to express the constraints faced by decision makers (Kitagawa and Tetenov 2018); this may be for the sake of interpretation, fairness, ease of implementation, political constraints, etc. The types of rules discussed in Section 3 which can be expressed using linear and integer constraints, like linear eligibility scores and decision trees, satisfy this assumption under bounds on the number of inputs to the eligibility score or the depth of the decision tree.

The assumption of VC dimension also plays an important role in the convergence of the regret of the optimal policy by ensuring the policy does not overfit the sample data. This assumption can be relaxed by instead using a holdout validation sample which regularizes estimation of the policy (Mbakop and Tabord-Meehan 2021).

Under these assumptions, the following regret bound is obtained:

Theorem 4.4: *Let $\mathcal{P}_C$ be a set of distributions for which (1) Assumptions 4.1 holds with constant C and (2) Assumption 4.2 holds. Under Assumptions 3.1 and 4.3,*

$$\sup_{P \in \mathcal{P}_C} \left(\mathbb{E}_P[\overline{R}_P(\hat{\pi})] - \overline{R}_P(\pi_P^*) \right) \leq \mathcal{O}(N^{-1/2} \vee \rho_N^{-1})$$

is satisfied. As a result,

$$\sup_{P \in \mathcal{P}_C} \mathbb{E}_P[\overline{R}_P(\hat{\pi})] \leq \sup_{P \in \mathcal{P}_C} \overline{R}_P(\pi_P^*) + \mathcal{O}(N^{-1/2} \vee \rho_N^{-1}) \tag{6}$$

The rate of convergence of the maximum regret is the slower of two rates: $N^{-1/2}$, and the estimation rate of $\hat{m}_0$ in Assumption 4.2. This first rate is driven by the convergence of an empirical process uniformly over the policy class, which is $N^{-1/2}$ under Assumption 4.3 (Van Der Vaart and Wellner 1996). The second rate reflects that the regret of the estimated policy depends on the behavior of the linear program (5), the constraints of which depend on identified moments of the data and must be estimated. In turn, the value of the linear program can be shown to converge to its population counterpart at the same rate as the constraints (see Hoffman (1952) and Rockafellar and Wets (2009); related results in econometrics include Fang et al. (2023) and Freyberger and Horowitz (2015)). Because Assumption 4.2 is only a condition on the rate of convergence of this estimator, the bound of Theorem 4.4 is a rate result. If non-asymptotic bounds on the estimator $\hat{m}_0$ are available, (for example if covariates are discrete and outcomes are bounded), then the proof of Lemma 4.6 can be used to obtain non-asymptotic regret bounds. This is developed in more detail in Appendix B.

I give a heuristic sketch of the proof and defer the details to Appendix A. I first define the quantities

$$\tilde{R}_{N,P}(\pi) := \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \pi_{ij} \Gamma_{j,P}(X_i)$$
$$\overline{R}_N(\pi) := \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} \pi_{ij} \hat{\Gamma}_j(X_i)$$

$\tilde{R}_{N,P}(\pi)$ measures the in-sample or empirical maximum regret of policy π, supposing the true m_P and hence

the true $\Gamma_{j,P}$ were known. $\overline{R}_N(\pi)$ is the objective function of the empirical minimax regret problem (4). The difference between the maximum regret of the estimated policy and that of the minimax regret optimal policy can then be decomposed in terms of these quantities as follows:

$$\begin{aligned} 0 \leq \overline{R}_P(\hat{\pi}) - \overline{R}_P(\pi^*) = \overline{R}_P(\hat{\pi}) - \tilde{R}_{N,P}(\hat{\pi}) \\ + \tilde{R}_{N,P}(\hat{\pi}) - \overline{R}_N(\hat{\pi}) \\ + \overline{R}_N(\hat{\pi}) - \overline{R}_N(\pi^*) \\ + \overline{R}_N(\pi^*) - \tilde{R}_{N,P}(\pi^*) \\ + \tilde{R}_{N,P}(\pi^*) - \overline{R}_P(\pi^*) \end{aligned} \tag{7}$$

The first and last lines of (7) each concern the difference between a sample mean and the population expectation, holding the policy and distribution fixed and assuming $\Gamma_{j,P}(X)$ is known. They are each bounded by

$$\sup_{\pi \in \Pi} \left| \overline{R}_P(\pi) - \tilde{R}_{N,P}(\pi) \right|$$

The second and fourth lines of (7) concern the difference between sample means of the true quantities $\Gamma_{j,P}(X_i)$ and their estimated counterparts, holding the policy and distribution fixed. They are each bounded by

$$\sup_{\pi \in \Pi} \left| \tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi) \right|$$

The third line of (7) concerns the difference between the in-sample performances of $\hat{\pi}$ and π^*. This is always negative because $\hat{\pi}$ is optimal for the empirical minimax regret problem (4). Hence, the decomposition (7) yields

$$\overline{R}_P(\hat{\pi}) - \overline{R}_P(\pi^*) \leq 2 \sup_{\pi \in \Pi} \left| \overline{R}_P(\pi) - \tilde{R}_{N,P}(\pi) \right| \tag{8}$$

$$+ 2 \sup_{\pi \in \Pi} \left| \tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi) \right| \tag{9}$$

Term (8) is the sup-Π norm of a centered empirical process. Its expectation can be shown to converge uniformly at $N^{-1/2}$ rate using techniques in empirical process theory.

Lemma 4.5: *Under assumptions 3.1, 4.1, and 4.3,*

$$\sup_{P \in \mathcal{P}_C} \mathbb{E}_P \left[\sup_{\pi \in \Pi} \left| \overline{R}_P(\pi) - \tilde{R}_{N,P}(\pi) \right| \right] \leq K \sqrt{\frac{V}{N}}$$

for some constant K depending only on C and J.

The constant K hides a dependence on the number of treatments J. This dependence represents a cost to introducing arbitrarily large sets of new treatments. Just as Assumption 4.3 restricts the complexity of the sets of covariate values assigned to each treatment, the assumption of a fixed J represents an exogenous constraint on the overall complexity of the policy.

Term (9) concerns the difference between the value of the linear program defining $\hat{\Gamma}_{jk}(X_i)$, in which the constraints are estimated, versus the linear program defining $\Gamma_{jk,P}(X_i)$, in which the true value of the constraint vector is used. When the estimated constraints converge at $N^{-1/2}$ rate, the value of the linear program can be shown to exhibit similar convergence uniformly across Π. More generally, the value of the linear program converges at the same rate as the estimated constraints. This is because the feasible set of a linear program is Lipschitz in its constraints with respect to the Hausdorff metric.

Lemma 4.6: *Under assumptions 3.1, 4.1, and 4.2,*

$$\sup_{P \in \mathcal{P}_C} \mathbb{E}_P \left[\sup_{\pi \in \Pi} \left| \tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi) \right| \right] \leq \mathcal{O}(\rho_N^{-1})$$

Taking the expectation of the bound given by (8) and (9) and combining this with Lemmas 4.5 and 4.6 yields the bound of Theorem 4.4.

5 Application to rural electrification

Investment in energy infrastructure is an important focus of development aid and there is a large body of research in development economics devoted to its study (reviews include Lee, Miguel, and Wolfram 2020b, Peters and Sievert (2016), and Van De Walle et al. (2015)). Lee, Miguel, and Wolfram (2020c) examines the relationship between the price of connections to the electrical grid and takeup in rural Kenya. This particular setting provides a compelling use case for the procedure outlined in this paper. There are only four prices observed in the data, leading to substantial model ambiguity in the form of partial identification of the demand curve outside these four prices. Further, the treatments are subsidies valued at hundreds of US dollars, making subsequent experimentation with new treatments expensive. In this section, I take experimental data collected to study the economics of rural electrification (Lee, Miguel, and Wolfram 2020a) and illustrate how the method outlined in the present paper can be used to design cost-effective targeted subsidy policies to maximize household takeup.

Prices of d-thousand Kenyan shillings for $d \in \mathcal{D}_0 = \{0, 15, 25, 35\}$ are randomly offered to households, who have an eight-month period in which to decide whether to purchase the connection at the offered price.

After the period is over, households continue to have the option to connect at the full price of 35 thousand shillings. Here D is price, Y is a takeup indicator, and X is a two-dimensional random vector containing household size and income.

Given this experimental data, I consider a decision maker able to offer subsidies to households. However, the decision maker has no reason to restrict themselves to the four prices that appear in the data. In my baseline analysis, I examine an expanded treatment set of $\mathcal{D} = \{0, 2.5, 5, \ldots, 35\}$ thousand shillings. The sensitivity of results to coarser and finer treatment sets is reported in Appendix E. I assume the decision maker values each connection at α-thousand Kenyan shillings and must pay the value of the subsidy if the recipient purchases a connection. There is no fixed cost for offering the subsidy. This means $u(d, x, y) = (\alpha - (35 - d))y$ so that $b(d, x) = \alpha - (35 - d)$ and $c(d, x) = 0$. As a baseline specification, I take α to be the full market price of 35 thousand shillings and explore policies under other valuations in Appendix E.

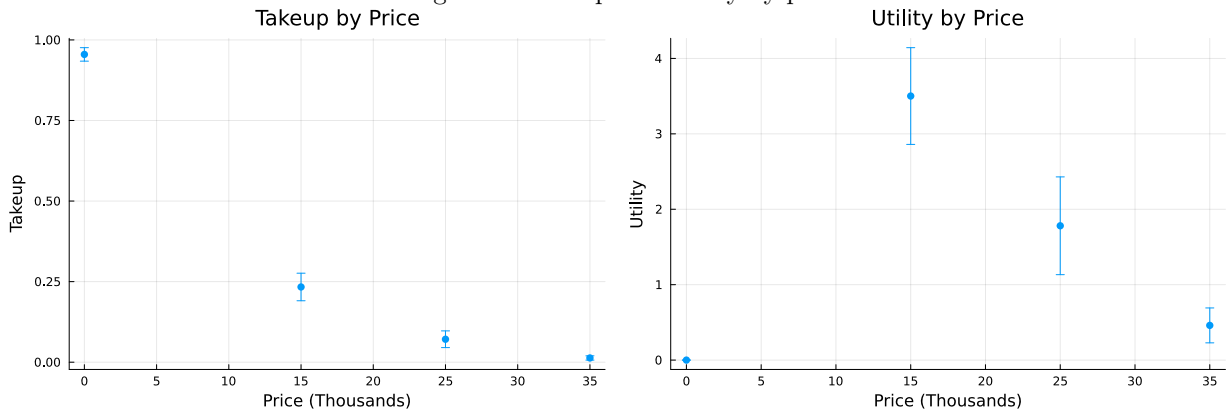
5.1 Optimal policy without covariates

Before estimating the optimal policy mapping covariate values to prices, I illustrate the method for the simple case of no covariates. Ignoring covariates for the time being makes the process easier to visualize, and transparently demonstrates how combining the experimental data, shape restrictions, and the minimax regret criterion drives the choice of whether and how to implement new treatments. In the next subsection, where I consider policies which target prices on the basis of covariates, the worst-case regret is computed for each covariate value similarly to the no-covariate case of this subsection.

I first estimate the average takeup at each price. Using these first stage estimates, I construct bounds for the effects of each new treatment and explain the difference between these pointwise bounds on outcomes and the estimated identified set $\hat{\mathcal{M}}$. Then I consider a fixed policy which assigns a single price to the entire population and find the regret-maximizing demand curve. I find the minimax regret optimal policy by finding the policy for which the maximum regret is as small as possible.

For each d in the experimental data, I plot the mean takeup and utility in Figure 1. Mean takeup $m_{0,P}(d) = \mathbb{E}_P[Y(d)]$ is identified from the experimental data for $d \in \mathcal{D}_0$, and estimated mean takeup $\hat{m}_0(d)$ is simply the sample mean at each price. Expected utility for experimental subsidy values is given by $v_{m_P}(d) = (\alpha - (35 - d))m_P(d)$ and is estimated for $d \in \mathcal{D}_0$ by plugging in $\hat{m}_0(d)$. The price $d = 0$ represents a fully subsidized connection, which is clearly undesirable from the decision maker's perspective because the decision maker will receive 0 utility, which is the minimum possible, regardless of whether the household connects. Amongst the treatment values that appear in the data, $d = 15$ achieves the highest utility on average. While not shown here, this is largely true of estimated mean utility conditional on X as well.

Figure 1: Takeup and utility by price



Mean estimated takeup and utility at each price included in the experiment of Lee, Miguel, and Wolfram (2020c). 95% confidence intervals shown in bars.

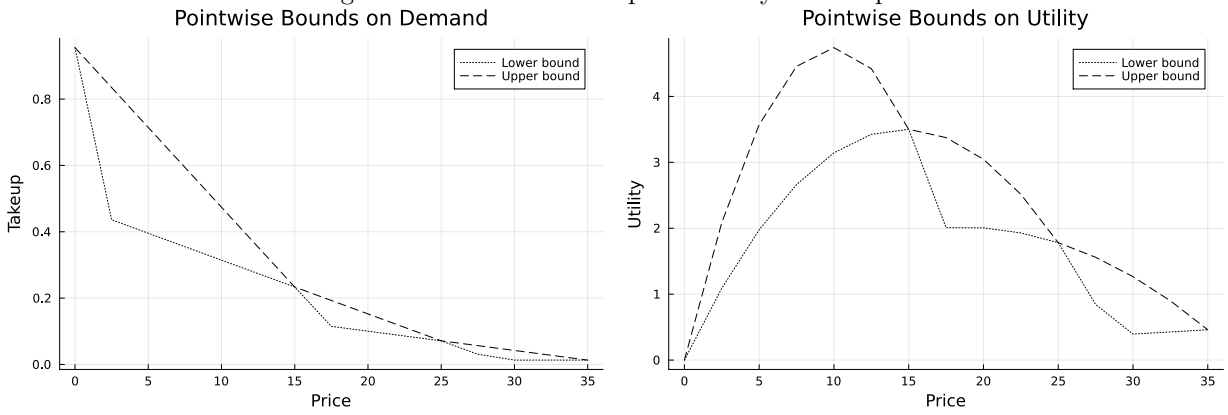
Indeed, setting $\mathcal{D} = \mathcal{D}_0$ and solving the empirical welfare maximization problem as in Kitagawa and Tetenov (2018) with a linear eligibility score as the policy class assigns all individuals to a price of 15.

A key question from the decision maker's perspective is whether prices not in the support of D in the data could yield higher utility, and how data from the experiment can provide information on the magnitude of such gains. To answer this, I impose shape restrictions which imply bounds on takeup at new prices. The shape restrictions I study here are that demand is downward sloping and the price subsidy exhibits diminishing returns. Takeup is also bounded between zero and one. Downward sloping demand is expected to be satisfied in all but a few exceptional markets, and represents one of the weaker assumptions a researcher may impose. Diminishing sensitivity to treatment may be more context specific, and can be motivated by a simple binary choice model where the density of valuations is decreasing on the support of treatments. Another setting where such a restriction may be applied is the analysis of production functions (Manski 1997). The shape restrictions I impose, which can be expressed as linear inequalities involving the J-dimensional vector m as shown in Appendix D, define the constraint $Sm \leq r$ in the linear program (5).

To explore the potential effects of new treatments informally in the simple case of no covariates, in Figure 2 I plot pointwise upper and lower bounds on takeup at each possible price. These are obtained by calculating $\min_{m \in \hat{\mathcal{M}}} m(d)$ and $\max_{m \in \hat{\mathcal{M}}} m(d)$ for each d. Note that the lower bound is not convex. This is an illustration of the non-rectangularity induced by the shape restrictions– there is no $m \in \hat{\mathcal{M}}$ that simultaneously minimizes takeup for all prices d. More generally, not every curve that lies within the pointwise bounds of Figure 2 satisfies the shape restrictions. This can be expressed formally as

$$\hat{\mathcal{M}} \subset \left\{ \tilde{m} \in \mathbb{R}^J : \min_{m \in \hat{\mathcal{M}}} m \leq \tilde{m}_j \leq \max_{m \in \hat{\mathcal{M}}} m, \forall j \right\}$$

Figure 2: Bounds on takeup and utility at each price



The maximal and minimal possible expected takeup at each price, and the corresponding bounds on expected utility generated by these bounds on takeup.

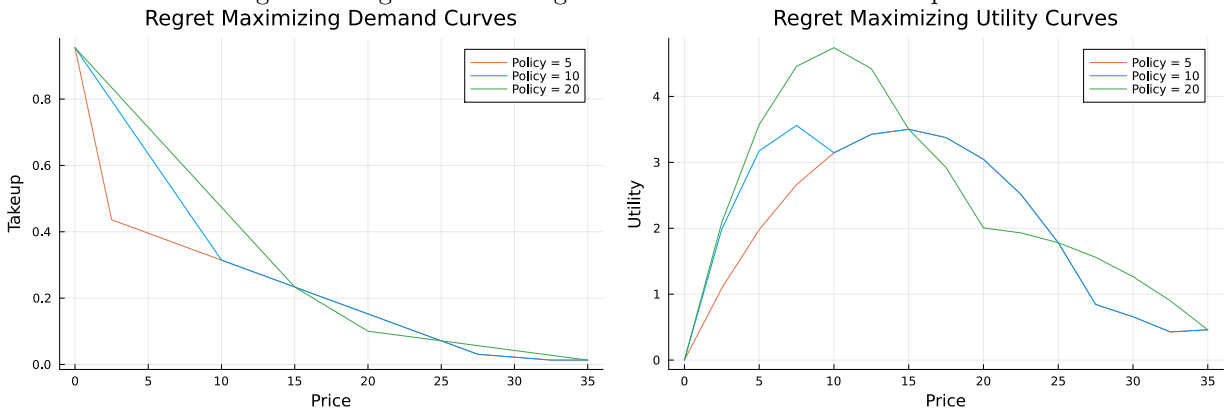
with strict containment. This difference is key for the informativeness of the linear program (5) because regret is defined by comparing the outcomes under the chosen policy to those of the first-best policy under the same demand curve m. If the chosen policy achieves low utility for one demand curve and the first-best policy achieves high utility only for a different demand curve, this does not contribute to high regret.

Along with the bounds on takeup in Figure 2, I also plot the bounds on expected utility v_m generated by the bounds on takeup. These curves illustrate a range of possible outcomes that may result from implementing new treatments. The upper bounds on utility illustrate the potential for much better outcomes as a result of implementing new treatments, especially in the range of 7.5 to 12.5. The lower bounds imply the possibility of worse outcomes as well. A maximin welfare approach to this problem would not assign a price in $\mathcal{D} \setminus \mathcal{D}_0$ to anyone for whom that price was not guaranteed to outperform the prices in $\mathcal{D}$. This ends up assigning a price of 15 to the entire sample, which seems excessively conservative in this example[3]. On the other hand, the minimax regret approach considers losses relative to the ex-post optimal decision in each state of the world represented by $m \in \hat{\mathcal{M}}$.

Given the bounds in Figure 2, one could imagine naively constructing $\hat{\Gamma}_{jk}$ by comparing the worst possible $v_m(d_j)$ to the best possible $v_m(d_k)$. For example, taking $d_k = 7.5$ and $d_j = 10$ would result in an estimate of about $4.5 - 3.2 = 1.3$. However, recall that these bounds on v_m were constructed from the bounds on m. Observing the bounds on m, it can be seen that the demand curve m which achieves maximal takeup at $d = 7.5$ and minimal takeup at $d = 10$ is not convex, and thus the regret estimate obtained by comparing the pointwise bounds is unnecessarily pessimistic. Likewise, taking $d_k = 12.5$ and $d_j = 10$ and comparing the pointwise bounds would yield an estimate of about $4.4 - 3.2 = 1.2$, but a demand curve which achieves

[3]It is not generally true that the maximin welfare policy restricts to the original set of treatments. See Appendix E.1 for an example in which the maximin welfare policy assigns a new treatment to the population.

Figure 3: Regret-maximizing demand curves for alternative policies



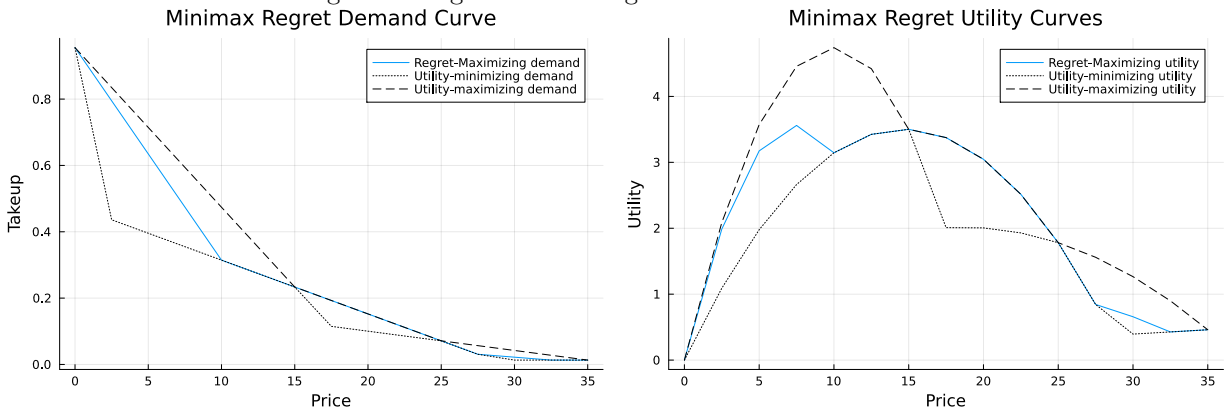
The demand curves that maximize regret for each of three policies– assigning $d = 5$ to everyone, assigning $d = 10$ to everyone, and assigning $d = 20$ to everyone. Also plotted are the utility curves generated by these regret-maximizing demand curves. Given a policy, the regret-maximizing demand curve is chosen to achieve low utility at the chosen price but high utility elsewhere.

these bounds is not decreasing. Formally, $\max_m [v_m(d_k) - v_m(d_j)] \leq \max_m v_m(d_k) - \min_m v_m(d_j)$. Thus, it is necessary to construct the regret estimates by finding a demand curve m which maximizes regret while satisfying the shape restrictions. This illustrates that the linear program (5) defining $\hat{\Gamma}_{jk}$, while requiring more computations than pointwise bounds for each $d_j \in \mathcal{D}$, carries additional useful information.

To understand how maximal regret is computed for each policy, I plot regret-maximizing demand curves for each of three different policies in Figure 3. In this case, a policy is a single value of d that will be assigned to the entire population. The regret-maximizing demand curve is the vector m which solves (3) with no covariates. To compute it, I solve (5) for each k and find the m corresponding to the optimal k. Supposing the decision maker assigns a price of $d = 5$ to the entire population, the regret-maximizing demand curve is chosen to yield low expected utility when $d = 5$ but high utility for some other price, thus incurring high regret in the sense that the chosen policy of $d = 5$ was ex-post a poor policy compared to, say, a price of $d = 15$. The same process is enacted for the policies which assign $d = 10$ to the entire population and $d = 20$ to the entire population. Under the policy $d = 20$, regret is very high because the difference between expected utility at $d = 20$ and the optimal expected utility under the regret-maximizing demand curve is very large. Comparatively, the maximum regret incurred under the policy $d = 10$ is small. Importantly, the regret-maximizing demand curves which generate these worst-case utility curves obey the shape restrictions, as can be seen in the left-hand pane of Figure 3.

Finally, I compute the optimal policy which does not target based on covariates. The solution to the empirical minimax regret problem without covariates is given by the policy which assigns the price $d = 10$ to the population. This means that across all demand curves $m \in \hat{\mathcal{M}}$, $v_m(10)$ is uniformly as close as possible to $\max_d v_m(d)$. To visualize this, in Figure 4 I overlay the regret-maximizing demand curve for the policy

Figure 4: Regret-maximizing demand curves for $d = 10$



The demand curve that achieves maximum regret under the minimax regret-optimal price of $d = 10$ and the resulting welfare curve. The curves lie between the pointwise bounds examined in Figure 2.

$d = 10$ on top of the bounds on takeup and welfare plotted in Figure 2. The m which maximizes regret is the one which maximizes utility at $d = 7.5$ but performs somewhat worse when $d = 10$. This difference between the best possible outcome and the outcome realized under the chosen policy is the regret that nature seeks to maximize through the choice of m and the decision maker seeks to minimize through the choice of π. Observe that maximum regret, given by $\max_m[v_m(7.5) - v_m(10)]$, is much smaller than a naive comparison of the bounds. Hence, an adversarially chosen demand curve in $\hat{\mathcal{M}}$ can make a price of $d = 10$ perform only mildly suboptimally.

5.2 Optimal policy with covariates

Having illustrated the method for estimating $\hat{m}_0$, constructing $\hat{\mathcal{M}}$, and constructing $\hat{\pi}$ in the simple case of no covariates, I now solve for the optimal policy when the decision maker can target subsidies based on household size and income. I construct the estimate $\hat{m}_0(d, x)$ using a Lasso-penalized logistic regression of takeup on a dictionary of Chebyshev polynomials in household size and income, for each $d \in \mathcal{D}_0$. As before, I use the shape restrictions that demand is decreasing and convex in d for every x. For some observations, the estimates $\hat{m}_0(d, X_i)$ violate these shape restrictions. When this happens, I replace the estimates with $\arg\min_m \|m(d) - \hat{m}_0(d, X_i)\|$, where the minimum is taken over all $m(d) \in \mathbb{R}^{J_0}$ that are decreasing and convex in d and bounded between 0 and 1. This ensures that $\hat{\mathcal{M}}$ is nonempty. These estimates are used to obtain $\hat{\Gamma}_j(X_i)$ for each i and j.

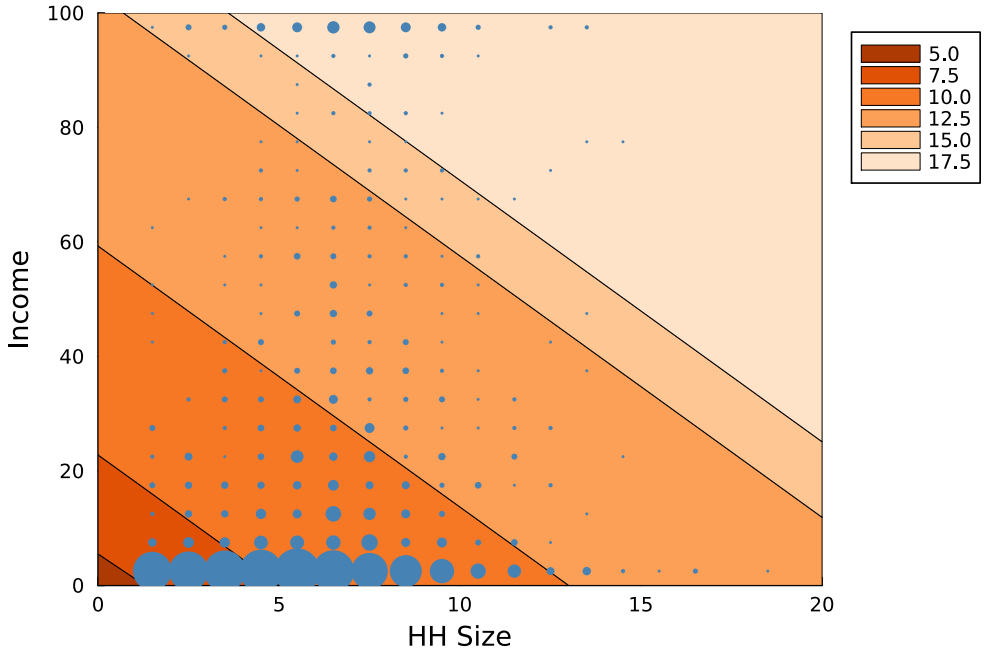
Finally, to estimate the optimal policy, I consider a policy class of linear eligibility score rules where each treatment shares the same eligibility score, but different cutoffs. The decision maker chooses a vector of covariate weights β and a vector of increasing cutoffs $\{c_j\}_{j=0}^{J-1}$. A household with covariates X_i receives treatment d_j if $c_{j-1} < X_i'\beta \leq c_j$, where $c_0 = -\infty$ and $c_J = \infty$. I impose that the eligibility score increases

with income, implying that poorer households receive lower prices. Formally,

$$\Pi = \left\{ \pi : \pi(x) = \sum_{j=1}^{J-1} (d_{j+1} - d_j) \mathbb{1}[X_i'\beta > c_j],\ c_{j-1} \leq c_j,\ \beta_1 > 0 \right\} \tag{10}$$

Appendix D discusses how this class can be formulated with linear and integer constraints, resulting in a mixed integer-linear program formulation for the empirical minimax regret problem (4).

Figure 5: Optimal policy



The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Table 1: Optimal policy

d	5.0	7.5	10.0	12.5	15.0	17.5
% Treated	6.6%	27.4%	52.2%	8.3%	1.3%	4.2%
Cutoff ($\beta = [0.266, 1.221]$)	1.48	6.1	15.81	27.57	31.09	42.23

Percent of population assigned to each treatment and eligibility score cutoff for each treatment for which a nonzero share of the population was assigned. Households were assigned to treatment j if their score was below cutoff j and above cutoff $j-1$.

The optimal allocation is illustrated in Figure 5, and exact estimates of the optimal policy along with the fraction of the population assigned to each treatment are presented in Table 1. The optimal allocation assigns poor, small households the lowest prices as they have the lowest willingness to pay. Almost the entire

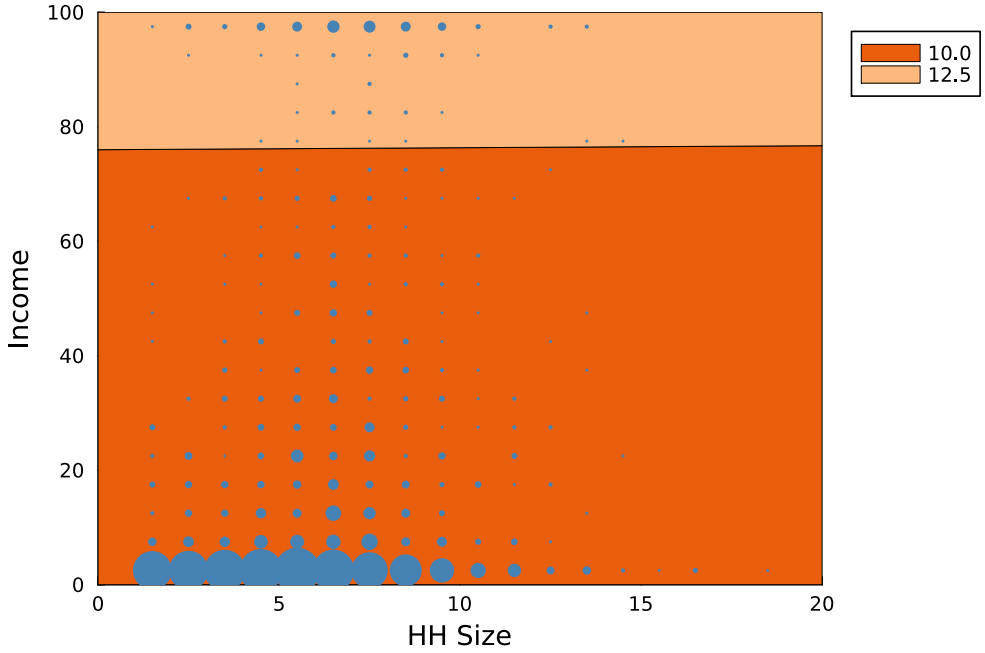
population is assigned a price not observed in the experimental data, with only 1.3% of the population being assigned to the price $d = 15$ which was optimal amongst the prices that were used in the experiment. Most of the population is assigned to a price of $d = 7.5$ or $d = 10$.

5.3 Comparison with other policies

I compare the optimal policy in Figure 5 with two other policies that a decision maker might use in the absence of the method proposed in this paper. I evaluate the performance of these policies in terms of their estimated maximum regret, and compare this to the estimated maximum regret of the policy proposed in this paper. These heuristic policies, which are not designed to control maximum regret, have the potential to perform substantially worse than the minimax regret optimal policy.

For the first benchmark, I estimate the optimal policy using an ad-hoc parametric interpolation that a decision maker might use to forecast the effects of new treatments. I estimate takeup using OLS with linear and quadratic terms in household size and income, motivated by the near-quadratic response to price observed in Figure 1. This means that the identified set $\hat{\mathcal{M}}$ is a singleton, making the maximization over $\hat{\mathcal{M}}$ trivial. These takeup estimates are then used to construct a policy which maximizes estimated utility. The resulting treatment allocation is shown in Figure 6, and the associated policy is given in Table 2.

Figure 6: Optimal policy with parametric interpolation



The estimated optimal treatment allocation under the parametric interpolation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

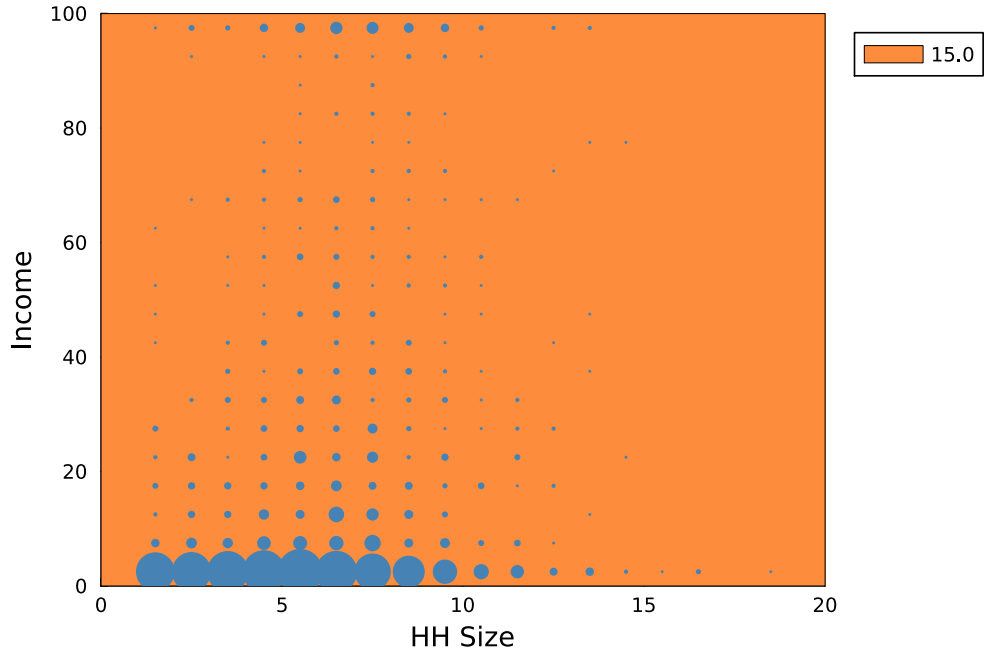
Table 2: Optimal policy with parametric interpolation

d	10.0	12.5
% Treated	95.1%	4.9%
Cutoff ($\beta = [0.039, -0.001]$)	2.99	3.89

Percent of population assigned to each treatment and eligibility score cutoff under the parametric interpolation, for each treatment for which a nonzero share of the population was assigned. Households were assigned to treatment j if their score was below cutoff j and above cutoff $j-1$.

For the second benchmark, I estimate the optimal policy under the restriction that the policy cannot assign new treatments. This reflects what a decision maker might do if they did not have a method for evaluating and choosing policies which involve new treatments. That is, I take the regret estimates $\hat{\Gamma}_j$ from the Lasso model and use it to construct a minimax regret policy under the additional restriction that the policy cannot assign new treatments. This reflects the maximum regret that a policymaker would incur by choosing not to assign new treatments, even if it were possible to do so. The resulting treatment allocation is shown in Figure 7, and the associated policy is given in Table 3.

Figure 7: Optimal policy with no new treatments



The estimated optimal treatment allocation under the restriction that the policy cannot assign new treatments, as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

In Table 4, I compare the estimated maximum regret of the estimated policy $\hat{\pi}$ with the estimated

Table 3: Optimal policy with no new treatments

d	15.0
% Treated	100.0%
Cutoff ($\beta = [0.0, -5.263]$)	-5.26

Percent of population assigned to each treatment and eligibility score cutoff under the restriction that the policy cannot assign new treatments, for each treatment for which a nonzero share of the population was assigned. Households were assigned to treatment j if their score was below cutoff j and above cutoff $j-1$.

Table 4: Comparison of maximum regret of different policies

Policy	Maximum Regret	Percent Increase
Lasso	488.19	
Parametric	620.75	27.15 %
Restricted	1381.38	182.96 %

Estimated maximum regret of the optimal policy, the policy with parametric interpolation, and the policy with no new treatments.

maximum regret of the other two policies described above. By construction, the estimated policy $\hat{\pi}$ minimizes the estimated maximum regret.

The parametric extrapolation results in higher estimated regret than the method proposed in this paper, which takes into account non-identification of the effects of new treatments. Thus, whether the decision maker should use the parametric extrapolation is sensitive to how much they trust the parametric form. If the parametric model is used only for convenience and the actual identified set is described by $\mathcal{M}$, then the parametric model may lead to higher regret than the robust method proposed in this paper. When the decision maker uses the parametric extrapolation, they fail to consider the worst-case effects of new treatments, and hence are overconfident in the benefits of new treatments.

Restricting to the support of the experimental data greatly increases maximum regret. When the decision maker chooses not to assign new treatments, the worst-case utility function v_m will be high on the set of new treatments to make the regret of the chosen policy large (in the case without covariates, the worst-case utility curve coincides with the upper bound on utility at $d = 10$ in Figure 2). By implementing new treatments, the decision maker can ensure they don't miss out on potentially large gains from these new treatments. However, it is not guaranteed that the decision maker will choose to implement new treatments, as they must also ensure the potential downside of implementing new treatments is not so large as to make worst-case regret higher than restricting to old treatments.

6 Conclusion

Experiments may not pilot all possible treatments a decision maker may consider. The existing literature on policy learning and treatment choice does not offer much guidance for how to use data on some treatment values to design policies involving new treatment values. I use data on previously observed treatments, partial identification, and the minimax regret criterion to extend empirical welfare maximization methods to settings where new treatments may be considered. Since the effects of new treatments are partially identified, a single policy is chosen to uniformly minimize regret across the identified set. The empirical minimax regret estimator is computationally tractable and possesses favorable regret convergence properties. In the setting of targeting subsidies to connect to the electrical grid, the estimator takes information on a small set of treatments and provides informative bounds on the effects of a much richer set new treatments, resulting in policies that implement new treatments which are uniformly close to optimal in every state of the world.

References

Adjaho, Christopher and Timothy Christensen (July 2, 2023). *Externally Valid Policy Choice*. DOI: 10.48550/arXiv.2205.05561. arXiv: 2205.05561 [econ]. URL: http://arxiv.org/abs/2205.05561 (visited on 04/03/2025). Pre-published.

Athey, Susan and Stefan Wager (2021). "Policy Learning With Observational Data". In: *Econometrica* 89.1, pp. 133–161. ISSN: 0012-9682. DOI: 10.3982/ECTA15732. URL: https://www.econometricsociety.org/doi/10.3982/ECTA15732 (visited on 07/24/2024).

Ben-Michael, Eli, D. James Greiner, Kosuke Imai, and Zhichao Jiang (Feb. 15, 2022). *Safe Policy Learning through Extrapolation: Application to Pre-trial Risk Assessment*. DOI: 10.48550/arXiv.2109.11679. arXiv: 2109.11679 [stat]. URL: http://arxiv.org/abs/2109.11679 (visited on 03/27/2025). Pre-published.

Bhattacharya, Debopam and Pascaline Dupas (Mar. 2012). "Inferring Welfare Maximizing Treatment Assignment under Budget Constraints". In: *Journal of Econometrics* 167.1, pp. 168–196. ISSN: 03044076. DOI: 10.1016/j.jeconom.2011.11.007. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407611002697 (visited on 10/17/2024).

Chetverikov, Denis, Andres Santos, and Azeem M. Shaikh (Aug. 2, 2018). "The Econometrics of Shape Restrictions". In: *Annual Review of Economics* 10.1, pp. 31–63. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-080217-053417. URL: https://www.annualreviews.org/doi/10.1146/annurev-economics-080217-053417 (visited on 04/03/2025).

Christensen, Timothy, Hyungsik Roger Moon, and Frank Schorfheide (May 14, 2023). *Optimal Decision Rules When Payoffs Are Partially Identified*. arXiv: 2204.11748 [econ, stat]. URL: http://arxiv.org/abs/2204.11748 (visited on 07/24/2024). Pre-published.

D'Adamo, Riccardo (Jan. 3, 2023). "Orthogonal Policy Learning Under Ambiguity". In:

Fang, Zheng, Andres Santos, Azeem M. Shaikh, and Alexander Torgovitsky (2023). "Inference for Large-Scale Linear Systems With Known Coefficients". In: *Econometrica* 91.1, pp. 299–327. ISSN: 0012-9682. DOI: 10.3982/ECTA18979. URL: https://www.econometricsociety.org/doi/10.3982/ECTA18979 (visited on 04/03/2025).

Freyberger, Joachim and Joel L. Horowitz (Nov. 2015). "Identification and Shape Restrictions in Nonparametric Instrumental Variables Estimation". In: *Journal of Econometrics* 189.1, pp. 41–53. ISSN: 03044076. DOI: 10.1016/j.jeconom.2015.06.020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407615001918 (visited on 04/03/2025).

Heckman, James J. and Edward J. Vytlacil (2007). "Chapter 70 Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation". In: *Handbook of Econometrics*. Vol. 6. Elsevier, pp. 4779–4874. ISBN: 978-0-444-53200-8. DOI: 10.1016/S1573-4412(07)06070-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S1573441207060709 (visited on 10/08/2024).

Hirano, Keisuke and Jack R. Porter (2009). "Asymptotics for Statistical Treatment Rules". In: *Econometrica* 77.5, pp. 1683–1701. ISSN: 1468-0262. DOI: 10.3982/ECTA6630. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA6630 (visited on 02/26/2024).

Hirano, Keisuke and Jack R. Porter (2020). "Asymptotic Analysis of Statistical Decision Rules in Econometrics". In: *Handbook of Econometrics*. Vol. 7. Elsevier, pp. 283–354. ISBN: 978-0-444-63649-2. DOI: 10.1016/bs.hoe.2020.09.001. URL: https://linkinghub.elsevier.com/retrieve/pii/S1573441220300040 (visited on 01/29/2024).

Hoffman, Alan J (Oct. 1952). "On Approximate Solutions of Systems of Linear Inequalities". In: *Journal of Research of the National Bureau of Standards* 49.4.

Kallus, Nathan and Angela Zhou (2018). "Policy Evaluation and Optimization with Continuous Treatments". In:

Kallus, Nathan and Angela Zhou (May 2021). "Minimax-Optimal Policy Learning Under Unobserved Confounding". In: *Management Science* 67.5, pp. 2870–2890. ISSN: 0025-1909, 1526-5501. DOI: 10.1287/mnsc.2020.3699. URL: https://pubsonline.informs.org/doi/10.1287/mnsc.2020.3699 (visited on 03/27/2025).

Khan, Samir, Martin Saveski, and Johan Ugander (Mar. 8, 2024). *Off-Policy Evaluation beyond Overlap: Partial Identification through Smoothness*. DOI: 10.48550/arXiv.2305.11812. arXiv: 2305.11812 [stat]. URL: http://arxiv.org/abs/2305.11812 (visited on 04/15/2025). Pre-published.

Kitagawa, Toru and Aleksey Tetenov (2018). "Who Should Be Treated? Empirical Welfare Maximization Methods for Treatment Choice". In: *Econometrica* 86.2, pp. 591–616. ISSN: 1468-0262. DOI: 10.3982/ECTA13288. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA13288 (visited on 02/26/2024).

Lee, Kenneth, Edward Miguel, and Catherine Wolfram (Apr. 2020a). *Data Archive for: Experimental Evidence on the Economics of Rural Electrification*. URL: https://www.journals.uchicago.edu/doi/suppl/10.1086/705417/suppl_file/2016544data.zip (visited on 07/14/2025).

Lee, Kenneth, Edward Miguel, and Catherine Wolfram (Feb. 1, 2020b). "Does Household Electrification Supercharge Economic Development?" In: *Journal of Economic Perspectives* 34.1, pp. 122–144. ISSN:

0895-3309. DOI: 10.1257/jep.34.1.122. URL: https://pubs.aeaweb.org/doi/10.1257/jep.34.1.122 (visited on 04/03/2025).

Lee, Kenneth, Edward Miguel, and Catherine Wolfram (Apr. 2020c). "Experimental Evidence on the Economics of Rural Electrification". In: *Journal of Political Economy* 128.4, pp. 1523–1565. ISSN: 0022-3808, 1537-534X. DOI: 10.1086/705417. URL: https://www.journals.uchicago.edu/doi/10.1086/705417 (visited on 04/03/2025).

Lei, Lihua, Roshni Sahoo, and Stefan Wager (Apr. 23, 2023). *Policy Learning under Biased Sample Selection*. DOI: 10.48550/arXiv.2304.11735. arXiv: 2304.11735 [econ]. URL: http://arxiv.org/abs/2304.11735 (visited on 04/15/2025). Pre-published.

Liu, Yan (Mar. 1, 2024). *Policy Learning under Endogeneity Using Instrumental Variables*. DOI: 10.48550/arXiv.2206.09883. arXiv: 2206.09883 [econ]. URL: http://arxiv.org/abs/2206.09883 (visited on 04/03/2025). Pre-published.

Manski, Charles F (1997). "Monotone Treatment Response". In: *Econometrica: Journal of the Econometric Society*, pp. 1311–1334. ISSN: 0012-9682.

Manski, Charles F (2009). *Identification for Prediction and Decision*. Harvard University Press. ISBN: 0-674-03366-3.

Manski, Charles F. (July 2004). "Statistical Treatment Rules for Heterogeneous Populations". In: *Econometrica* 72.4, pp. 1221–1246. ISSN: 0012-9682, 1468-0262. DOI: 10.1111/j.1468-0262.2004.00530.x. URL: http://www.blackwell-synergy.com/links/doi/10.1111%2Fj.1468-0262.2004.00530.x (visited on 04/03/2025).

Manski, Charles F. (Nov. 1, 2006). "Search Profiling with Partial Knowledge of Deterrence". In: *The Economic Journal* 116.515, F385–F401. ISSN: 0013-0133, 1468-0297. DOI: 10.1111/j.1468-0297.2006.01128.x. URL: https://academic.oup.com/ej/article/116/515/F385-F401/5089391 (visited on 03/27/2025).

Manski, Charles F. (July 2007). "Minimax-Regret Treatment Choice with Missing Outcome Data". In: *Journal of Econometrics* 139.1, pp. 105–115. ISSN: 03044076. DOI: 10.1016/j.jeconom.2006.06.006. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407606001047 (visited on 04/03/2025).

Manski, Charles F. (Mar. 2, 2010). "Vaccination with Partial Knowledge of External Effectiveness". In: *Proceedings of the National Academy of Sciences* 107.9, pp. 3953–3960. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0915009107. URL: https://pnas.org/doi/full/10.1073/pnas.0915009107 (visited on 03/27/2025).

Manski, Charles F. (Sept. 1, 2011). "Choosing Treatment Policies Under Ambiguity". In: *Annual Review of Economics* 3.1, pp. 25–49. ISSN: 1941-1383, 1941-1391. DOI: 10.1146/annurev-economics-061109-

080359. URL: https://www.annualreviews.org/doi/10.1146/annurev-economics-061109-080359 (visited on 04/03/2025).

Manski, Charles F. (2021). "Econometrics for Decision Making: Building Foundations Sketched by Haavelmo and Wald". In: *Econometrica* 89.6, pp. 2827–2853. ISSN: 0012-9682. DOI: 10.3982/ECTA17985. URL: https://www.econometricsociety.org/doi/10.3982/ECTA17985 (visited on 07/29/2024).

Manski, Charles F. (Jan. 2025). "Using Limited Trial Evidence to Credibly Choose Treatment Dosage When Efficacy and Adverse Effects Weakly Increase with Dose". In: *Epidemiology* 36.1, pp. 60–65. ISSN: 1044-3983, 1531-5487. DOI: 10.1097/EDE.0000000000001793. URL: https://journals.lww.com/10.1097/EDE.0000000000001793 (visited on 04/03/2025).

Mbakop, Eric and Max Tabord-Meehan (2021). "Model Selection for Treatment Choice: Penalized Welfare Maximization". In: *Econometrica* 89.2, pp. 825–848. ISSN: 0012-9682. DOI: 10.3982/ECTA16437. URL: https://www.econometricsociety.org/doi/10.3982/ECTA16437 (visited on 08/09/2024).

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2018). "Using Instrumental Variables for Inference About Policy Relevant Treatment Parameters". In: *Econometrica* 86.5, pp. 1589–1619. ISSN: 0012-9682. DOI: 10.3982/ECTA15463. URL: https://www.econometricsociety.org/doi/10.3982/ECTA15463 (visited on 04/03/2025).

Peters, Jörg and Maximiliane Sievert (July 2, 2016). "Impacts of Rural Electrification Revisited – the African Context". In: *Journal of Development Effectiveness* 8.3, pp. 327–345. ISSN: 1943-9342, 1943-9407. DOI: 10.1080/19439342.2016.1178320. URL: https://www.tandfonline.com/doi/full/10.1080/19439342.2016.1178320 (visited on 04/03/2025).

Pu, Hongming and Bo Zhang (Apr. 1, 2021). "Estimating Optimal Treatment Rules with an Instrumental Variable: A Partial Identification Learning Approach". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 83.2, pp. 318–345. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/rssb.12413. URL: https://academic.oup.com/jrsssb/article/83/2/318/7056005 (visited on 03/27/2025).

Qian, Min and Susan A. Murphy (Apr. 1, 2011). "Performance Guarantees for Individualized Treatment Rules". In: *The Annals of Statistics* 39.2. ISSN: 0090-5364. DOI: 10.1214/10-AOS864. URL: https://projecteuclid.org/journals/annals-of-statistics/volume-39/issue-2/Performance-guarantees-for-individualized-treatment-rules/10.1214/10-AOS864.full (visited on 01/31/2025).

Rockafellar, R Tyrrell and Roger J-B Wets (2009). *Variational Analysis*. Vol. 317. Springer Science & Business Media. ISBN: 3-642-02431-9.

Sasaki, Yuya and Takuya Ura (Sept. 16, 2024). "Welfare Analysis via Marginal Treatment Effects". In: *Econometric Theory*, pp. 1–24. ISSN: 0266-4666, 1469-4360. DOI: 10.1017/S0266466624000227. URL:

https://www.cambridge.org/core/product/identifier/S0266466624000227/type/journal_article (visited on 04/03/2025).

Savage, Leonard J (1951). "The Theory of Statistical Decision". In: *Journal of the American Statistical association* 46.253, pp. 55–67. ISSN: 0162-1459.

Stoye, Jörg (Jan. 2012). "Minimax Regret Treatment Choice with Covariates or with Limited Validity of Experiments". In: *Journal of Econometrics* 166.1, pp. 138–156. ISSN: 03044076. DOI: 10.1016/j.jeconom.2011.06.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0304407611001254 (visited on 03/27/2025).

Van De Walle, Dominique, Martin Ravallion, Vibhuti Mendiratta, and Gayatri Koolwal (Oct. 14, 2015). "Long-Term Gains from Electrification in Rural India". In: *The World Bank Economic Review*, lhv057. ISSN: 0258-6770, 1564-698X. DOI: 10.1093/wber/lhv057. URL: https://academic.oup.com/wber/article-lookup/doi/10.1093/wber/lhv057 (visited on 04/03/2025).

Van Der Vaart, Aad W and Jon A Wellner (1996). *Weak Convergence*. Springer. ISBN: 1-4757-2547-7.

Vershynin, Roman (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. Vol. 47. Cambridge university press. ISBN: 1-108-24454-8.

Wald, Abraham (1949). "Statistical Decision Functions". In: *The Annals of Mathematical Statistics*, pp. 165–205. ISSN: 0003-4851.

Yata, Kohei (Mar. 2, 2025). *Optimal Decision Rules Under Partial Identification*. DOI: 10.48550/arXiv.2111.04926. arXiv: 2111.04926 [econ]. URL: http://arxiv.org/abs/2111.04926 (visited on 03/27/2025). Pre-published.

Zhang, Baqun, Anastasios A. Tsiatis, Marie Davidian, Min Zhang, and Eric Laber (Oct. 2012). "Estimating Optimal Treatment Regimes from a Classification Perspective". In: *Stat* 1.1, pp. 103–114. ISSN: 2049-1573, 2049-1573. DOI: 10.1002/sta.411. URL: https://onlinelibrary.wiley.com/doi/10.1002/sta.411 (visited on 01/31/2025).

Zhang, Yi, Eli Ben-Michael, and Kosuke Imai (Sept. 4, 2024). *Safe Policy Learning under Regression Discontinuity Designs with Multiple Cutoffs*. DOI: 10.48550/arXiv.2208.13323. arXiv: 2208.13323 [stat]. URL: http://arxiv.org/abs/2208.13323 (visited on 03/27/2025). Pre-published.

Zhao, Yingqi, Donglin Zeng, A. John Rush, and Michael R. Kosorok (Sept. 2012). "Estimating Individualized Treatment Rules Using Outcome Weighted Learning". In: *Journal of the American Statistical Association* 107.499, pp. 1106–1118. ISSN: 0162-1459, 1537-274X. DOI: 10.1080/01621459.2012.695674. URL: https://www.tandfonline.com/doi/full/10.1080/01621459.2012.695674 (visited on 01/31/2025).

Zhou, Zhengyuan, Susan Athey, and Stefan Wager (Jan. 2023). "Offline Multi-Action Policy Learning: Generalization and Optimization". In: *Operations Research* 71.1, pp. 148–183. ISSN: 0030-364X, 1526-

5463. DOI: 10.1287/opre.2022.2271. URL: https://pubsonline.informs.org/doi/10.1287/opre.2022.2271 (visited on 04/03/2025).

A Proof of Theorem 4.4

A.1 Intermediate results

I first introduce some notation and state existing results that I will use. Given a class of functions $\mathcal{F}$, the Rademacher complexity of $\mathcal{F}$ is defined as

$$\mathcal{R}_N(\mathcal{F}) := \mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_i \epsilon_i f(X_i)\right|\right]$$

where ϵ_i are i.i.d. Rademacher random variables.

We say $\mathcal{T}$ is a δ-cover of a metric space $(\mathcal{F}, h)$ if for every $f \in \mathcal{F}$, there is some $f_i \in \mathcal{T}$ such that $h(f_i, f) \leq \delta$. The cardinality of the smallest δ-cover of $\mathcal{F}$ is called the *δ-covering number* of $\mathcal{F}$ and is denoted $N(\delta, \mathcal{F}, h)$.

I make use of the following existing results:

Lemma A.1: *(Kitagawa and Tetenov (2018) Lemma A.1) Let $\mathcal{G}$ be a VC-class of subsets of $\mathcal{X}$ with VC dimension $V < \infty$. Let g and h be two given functions from $\mathcal{Y} \times \mathcal{D} \times \mathcal{X}$ to $\mathbb{R}$. Then*

$$\mathcal{F} = \{f : f(y, d, x) = g(y, d, x)\mathbb{1}\{x \in G\} + h(y, d, x)\mathbb{1}\{x \notin G\}, G \in \mathcal{G}\}$$

is a VC subgraph class of functions with VC dimension less than or equal to V.

Lemma A.2: *(Symmetrization) (Van Der Vaart and Wellner (1996) Lemma 2.3.1) For a class of measurable functions $\mathcal{F}$ and i.i.d random variables $X_1, \ldots, X_N$,*

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{N}\sum_{i=1}^{N} f(X_i) - \mathbb{E}[f(X_i)]\right|\right] \leq 2\mathcal{R}_N(\mathcal{F})$$

Lemma A.3: *(Dudley's entropy integral inequality) (Van Der Vaart and Wellner (1996) Corollary 2.2.8) Let $(Z_f)_{f\in\mathcal{F}}$ be a separable process with sub-Gaussian increments. Then for some constant K, we have for any f_0*

$$\mathbb{E}[\sup_{f\in\mathcal{F}}|Z_f|] \leq \mathbb{E}[|Z_{f_0}|] + K\int_0^\infty \sqrt{\log N(t, \mathcal{F}, h)}dt$$

for some constant K.

Lemma A.4: *(Van Der Vaart and Wellner (1996) Theorem 2.6.7) Suppose $\mathcal{F}$ is a VC subgraph class with VC dimension at most $V < \infty$ and suppose $\mathcal{F}$ has a measurable envelope function F. For $q \geq 1$ let P*

be a probability measure such that $\|F\|_{q,P} > 0$. *Then*

$$N(\delta\|F\|_{L_q(P)}, \mathcal{F}, L_q(P)) \leq KV(16e)^V(1/\delta)^{q(V-1)}$$

for some constant K *and* $0 < \delta < 1$.

I now state and prove a useful result which establishes the relationship between the complexity of the policy class Π and the complexity of the class of regret estimates. Define the following function classes:

$$\mathcal{F} := \{f : f(x) = \sum_{j=1}^{J} \pi_j(x)\Gamma_j(x), \pi \in \Pi\}$$

$$\Pi_j := \{\pi_j : \pi_j(x) = \mathbb{1}[\pi(x) = d_j], \pi \in \Pi\}, \quad \forall j \in \{1, \ldots, J\}$$

The assumption that $\mathcal{F}$ is separable will be maintained.

Lemma A.5: *Under Assumption* 4.1.2,

$$N(\delta, \mathcal{F}, L_2(P_N)) \leq \prod_{j=1}^{J} N(\epsilon, \Pi_j, L_2(P_N))$$

where $\epsilon = \delta/(B\sqrt{J})$ *and* B *is the bound on* $v(d, x)$ *implied by Assumption 4.1.2.*

Proof. Let $f \in \mathcal{F}$ be given. From the discussion in Section 2, f can be written as

$$f(x) = \max_k \max_{m \in \mathcal{M}_P} \left(v_m(d_k, x) - \sum_{j=1}^{J} \pi_j(x) v_m(d_j, x) \right) \tag{11}$$

for some $\pi \in \Pi$. For each treatment j, let $\mathcal{T}_j$ be an ϵ-cover of Π_j and let $\tilde{\pi}_j$ be an element of $\mathcal{T}_j$ satisfying $\mathbb{E}_N[(\pi_j(X) - \tilde{\pi}_j(X))^2]^{1/2} \leq \epsilon$. Define the approximating function $\tilde{f}$ by

$$\tilde{f}(x) = \max_k \max_{m \in \mathcal{M}_P} v_m(d_k, x) - \sum_{j=1}^{J} \tilde{\pi}_j(x) v_m(d_j, x) \tag{12}$$

Finally, let $k^*, m^*(x)$ be maximizers of (11) and let $\tilde{k}, \tilde{m}(x)$ be maximizers of (12).

For each x we have by the optimality of k^* and m^*

$$\begin{aligned} f(x) - \tilde{f}(x) &= v_{m^*}(d_{k^*}, x) - v_{\tilde{m}}(d_{\tilde{k}}, x) + \sum_{j=1}^{J} \left[\tilde{\pi}_j(x) v_{m^*}(d_j, x) - \pi_j(x) v_{\tilde{m}}(d_j, x) \right] \\ &\geq v_{\tilde{m}}(d_{\tilde{k}}, x) - v_{\tilde{m}}(d_{\tilde{k}}, x) + \sum_{j=1}^{J} \left[\tilde{\pi}_j(x) v_{\tilde{m}}(d_j, x) - \pi_j(x) v_{\tilde{m}}(d_j, x) \right] \end{aligned}$$

$$= \sum_{j=1}^{J} \left[(\tilde{\pi}_j(x) - \pi_j(x)) v_{\tilde{m}}(d_j, x) \right]$$

Likewise, optimality of $\tilde{m}$ and $\tilde{k}$ imply

$$f(x) - \tilde{f}(x) \leq \sum_{j=1}^{J} \left[(\tilde{\pi}_j(x) - \pi_j(x)) v_{m^*}(d_j, x) \right]$$

Together, we have

$$\begin{aligned}
|f(x) - \tilde{f}(x)| &\leq \max \left\{ \left| \sum_{j=1}^{J} (\tilde{\pi}_j(x) - \pi_j(x)) v_{\tilde{m}}(d_j, x) \right|, \left| \sum_{j=1}^{J} (\tilde{\pi}_j(x) - \pi_j(x)) v_{m^*}(d_j, x) \right| \right\} \\
&\leq \max \left\{ \|(\tilde{\pi}_j(x) - \pi_j(x))_{j=1}^{j}\| \|v_{\tilde{m}}(\cdot, x)\|, \|(\tilde{\pi}_j(x) - \pi_j(x))_{j=1}^{j}\| \|v_{m^*}(\cdot, x)\| \right\} \\
&\leq B \left(\sum_j (\pi_j(x) - \tilde{\pi}_j(x))^2 \right)^{1/2}
\end{aligned}$$

where the second line is by the Cauchy-Schwartz inequality. Squaring and integrating over x,

$$\begin{aligned}
\mathbb{E}_{N,P}[(f(X) - \tilde{f}(X))^2] &\leq B^2 \mathbb{E}_{N,P} \left[\sum_j (\pi_j(X) - \tilde{\pi}_j(X))^2 \right] \\
&\leq B^2 J \epsilon^2 \\
&\leq \delta^2
\end{aligned}$$

Taking the square root of both sides shows that $\|f - \tilde{f}\|_{L_2(P_N)} \leq \delta$. Consider the set of all such functions constructed this way,

$$\mathcal{T} := \left\{ \tilde{f} : \tilde{f}(x) = \max_k \max_{m \in \mathcal{M}} v_m(d_k, x) - \sum_{j=1}^{J} \tilde{\pi}_j(x) v_m(d_j, x),\ \tilde{\pi} = (\tilde{\pi}_1, \ldots, \tilde{\pi}_J), \tilde{\pi}_j \in \mathcal{T}_j \right\}$$

we see that $|\mathcal{T}| = \prod_j |\mathcal{T}_j|$, and $\mathcal{T}$ is a δ-cover of $\mathcal{F}$. □

We can now prove the main results of the paper.

A.2 Proof of Lemma 4.5

Proof. To simplify notation, for now we consider P fixed and suppress dependence on P. By the definition of $\mathcal{F}$, we have

$$\mathbb{E}\left[\sup_{\pi \in \Pi} \left| \overline{R}(\pi) - \tilde{R}_N(\pi) \right| \right] = \mathbb{E}\left[\sup_{f \in \mathcal{F}} \left| \frac{1}{N} \sum_{i=1}^{N} f(X_i) - \mathbb{E}[f(X_i)] \right| \right]$$

Hence, we can apply Lemma A.2 to obtain

$$\mathbb{E}\left[\sup_{\pi\in\Pi}\left|\overline{R}(\pi)-\tilde{R}_N(\pi)\right|\right]\leq 2\mathcal{R}_N(\mathcal{F})$$

Now, define $Z_f=\frac{1}{\sqrt{N}}\sum_{i=1}^N\epsilon_i f(X_i)$. The increments of $(Z_f)_{f\in\mathcal{F}}$ are given by $\frac{1}{\sqrt{N}}\sum_{i=1}^N\epsilon_i(f(X_i)-g(X_i))$. Conditional on $(X_i)_{i=1}^N$, we can apply Hoeffding's inequality (e.g. Van Der Vaart and Wellner (1996) Lemma 2.2.7) to establish that this is sub-Gaussian with parameter $\left(\frac{1}{N}\sum_{i=1}^N(f(X_i)-g(X_i))^2\right)^{1/2}=\|f-g\|_{L_2(P_N)}$. We can then apply Lemma A.3 conditional on $(X_i)_{i=1}^N$ to obtain

$$\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}}|Z_f|]\leq\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[|Z_{f_0}|]+\frac{K}{\sqrt{N}}\int_0^\infty\sqrt{\log N(t,\mathcal{F},L_2(P_N))}dt$$

for some $f_0\in\mathcal{F}$. Moreover, since f_0 is bounded by $2B$, $|Z_{f_0}|\leq|\sum_{i=1}^N\epsilon_i\frac{2B}{\sqrt{N}}|$. Again by Hoeffding's inequality, this implies that Z_{f_0} is sub-Gaussian with parameter $K'\left(\frac{1}{N}\sum_{i=1}^N(2B)^2\right)^{1/2}=K'2B$ which does not depend on N. Basic properties of sub-Gaussian random variables (e.g. Vershynin (2018) Proposition 2.5.2) imply that $\mathbb{E}_\epsilon[|Z_{f_0}|]\leq K''$ for some constant K''. Thus,

$$\begin{aligned}\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}}|Z_f|]&\leq\frac{1}{\sqrt{N}}\left(K''+K\int_0^\infty\sqrt{\log N(t,\mathcal{F},L_2(P_N))}dt\right)\\&=\frac{1}{\sqrt{N}}\left(K''+K\int_0^{4B}\sqrt{\log N(t,\mathcal{F},L_2(P_N))}dt\right)\end{aligned}$$

where we have used the fact that since the diameter of $\mathcal{F}$ is $2B$, the integrand is 0 for $t>4B$. By Lemma A.1, each class Π_j is a VC subgraph class of functions with VC dimension at most v. Then applying Lemma A.5 and a change of variables yields

$$\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}}|Z_f|]\leq\frac{1}{\sqrt{N}}\left(K''+K\int_0^{4/\sqrt{J}}\sqrt{\sum_j\log N(t,\Pi_j,L_2(P_N))}B\sqrt{J}dt\right)$$

Apply Lemma A.4 to each VC subgraph class of functions Π_j, which have envelope 1, and allow K to subsume other constants to obtain

$$\begin{aligned}\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}}|Z_f|]&\leq\frac{1}{\sqrt{N}}\left(K''+K\int_0^{4/\sqrt{J}}\sqrt{\log K'''V(16e)^V(1/t)^{2(V-1)}}dt\right)&(13)\\&\leq\frac{1}{\sqrt{N}}\left(K''+K\sqrt{V}\right)&(14)\\&\leq K\sqrt{\frac{V}{N}}&(15)\end{aligned}$$

Finally, since $\mathcal{R}_N(\mathcal{F}) = \mathbb{E}[\frac{1}{\sqrt{N}}\mathbb{E}_\epsilon[\sup_{f\in\mathcal{F}}|Z_f|]]$, we obtain

$$\mathbb{E}\left[\sup_{\pi\in\Pi}\left|\overline{R}(\pi) - \tilde{R}_N(\pi)\right|\right] \leq K\sqrt{\frac{V}{N}}$$

Note that the constant K does depend on B and J. □

A.3 Proof of Lemma 4.6

Proof. To simplify notation, for now we consider P fixed and suppress dependence on P. For every $1 \leq j,k \leq J$, let $\gamma_{jk} : \mathbb{R}^{J_0} \mapsto 2^{\mathbb{R}^J}$ be the identified set for covariate-level regret, viewed as a set-valued mapping from the first stage conditional mean response vector to subsets of $\mathbb{R}^J$. That is, hold x fixed and define $\gamma_{jk}(w) = \{b_{jk}(x)'m - c_{jk}(x) : Fm = w, Sm \leq r\}$. For this proof, we view Γ_{jk} as a function of the first stage conditional mean response m_0 to consider how Γ_{jk} changes with perturbations to m_0. Thus, $\Gamma_{jk}(m_0(\cdot, X_i)) = \max\{\gamma_{jk}(m_0(\cdot, X_i))\}$.

For any matrix A, let $A^\dagger : \text{null}(A)^\perp \mapsto \text{range}(A)$ denote the Moore-Penrose pseudoinverse operator, where $\text{null}(A)$ and $\text{range}(A)$ denote the null space and range of A, respectively. For any $w \in \mathbb{R}^{J_0}$,

$$\{m : Fm = w\} = \{m : m = F^\dagger w + y,\ y \in \text{null}(F)\}$$

Let $\tilde{w} \in \mathbb{R}^{J_0}$ be given. For any J, let $\mathcal{B}_J$ be the unit ball in $\mathbb{R}^J$, $\mathcal{B}_J := \{w \in \mathbb{R}^J : \|w\| \leq 1\}$. Since $\tilde{w} \in w + \|w - \tilde{w}\|\mathcal{B}_{J_0}$,

$$\{m : m = F^\dagger \tilde{w} + y,\ y \in \text{null}(F)\} \subseteq \{m : m = F^\dagger z + y,\ y \in \text{null}(F),\ z \in w + \|w - \tilde{w}\|\mathcal{B}_{J_0}\}.$$

Let $\|A\|$ denote the operator norm of a matrix A. Since $F^\dagger\|w - \tilde{w}\|\mathcal{B}_{J_0} \subseteq \|F^\dagger\|\|w - \tilde{w}\|\mathcal{B}_J$, we have

$$\begin{aligned}&\{m : m = F^\dagger z + y,\ y \in \text{null}(F),\ z \in w + \|w - \tilde{w}\|\mathcal{B}_{J_0}\}\\ &\subseteq \{m : m = F^\dagger w + y,\ y \in \text{null}(F)\} + \|F^\dagger\|\|\tilde{w} - w\|\mathcal{B}_J.\end{aligned}$$

This implies

$$\{m : Fm = \tilde{w}\} \cap \{m : Sm \leq r\} \subseteq \left(\{m : m = F^\dagger w + y, y \in \text{null}(F)\} + \|F^\dagger\|\|\tilde{w} - w\|\mathcal{B}_J\right) \cap \{m : Sm \leq r\}$$

and since $\gamma_{jk}(\tilde{w})$ consists of scalars of the form $b_{jk}(x)'m - c_{jk}(x)$ for m in this set, we have

$$\gamma_{jk}(\tilde{w}) \subseteq \gamma_{jk}(w) + \|b_{jk}(x)\|\|F^\dagger\|\|\tilde{w} - w\|\mathcal{B}_1$$

and likewise

$$\gamma_{jk}(w) \subseteq \gamma_{jk}(\tilde{w}) + \|b_{jk}(x)\|\|F^\dagger\|\|\tilde{w} - w\|\mathcal{B}_1$$

Therefore, the correspondence γ_{jk} is Lipschitz with respect to the Hausdorff distance, with Lipschitz constant $\sup_x \|b_{jk}(x)\|\|F^\dagger\| := \kappa < \infty$ (Rockafellar and Wets 2009). Importantly, this implies $|\Gamma_{jk}(\hat{m}_0(\cdot, X_i) - \Gamma_{jk}(m_0(\cdot, X_i))| \leq \kappa\|\hat{m}_0(\cdot, X_i) - m_0(\cdot, X_i)\|$. While not essential to our analysis, we note that in our case $\|F^\dagger\| = 1$.

We can now prove the main claim of the lemma.

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\pi\in\Pi}\left|\tilde{R}_N(\pi) - \overline{R}_N(\pi)\right|\right] \\
&= \mathbb{E}\left[\sup_{\pi\in\Pi}\left|\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{J}\pi_{ij}(\max_k \Gamma_{jk}(\hat{m}_0(\cdot, X_i)) - \max_k \Gamma_{jk}(m_0(\cdot, X_i)))\right|\right] \\
&\leq \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\max_{j,k}\left|\Gamma_{jk}(\hat{m}_0(\cdot, X_i)) - \Gamma_{jk}(m_0(\cdot, X_i))\right|\right] \\
&\leq \mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\kappa\|\hat{m}_0(\cdot, X_i) - m_0(\cdot, X_i)\|\right]
\end{aligned}$$

Finally, we bound this uniformly in P by Assumption 4.2. We conclude that

$$\sup_{P\in\mathcal{P}_C}\mathbb{E}_P\left[\sup_{\pi\in\Pi}\left|\tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi)\right|\right] \leq \mathcal{O}(\rho_N^{-1})$$

□

A.4 Proof of Theorem 4.4

Proof. Combining the bounds in Lemmas 4.5 and 4.6 establishes Theorem 4.4. □

B Extensions of theoretical results

B.1 Finite-sample bounds

In this section we consider strengthening the asymptotic result of Theorem 4.4 to hold for finite samples. We replace Assumption 4.2 with the following assumption that the estimate $\hat{m}_0$ has finite-sample error bounds.

Assumption B.1: *There exists a constant $K > 0$ and a sequence $\rho_N \to \infty$ such that for all $N \geq 1$, the estimate $\hat{m}_0$ satisfies*

1. $\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\frac{1}{N} \sum_{i=1}^N \|\hat{m}_0(\cdot, X_i) - m_{0,P}(\cdot, X_i)\| \right] \leq K \rho_N^{-1}$
2. $\hat{\mathcal{M}} = \{m : Sm(\cdot, X) \leq r,\ Fm(\cdot, X) = \hat{m}_0(\cdot, X)\}$ *is nonempty, almost surely, for all $P \in \mathcal{P}$.*

This assumption leads immediately to an extension of Lemma 4.6.

Corollary B.2: *Under assumptions 4.1, 3.1, and B.1,*

$$\sup_{P \in \mathcal{P}_C} \mathbb{E}_P \left[\sup_{\pi \in \Pi} \left| \tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi) \right| \right] \leq K \rho_N^{-1}.$$

Proof. By the exact same argument as in the proof of Lemma 4.6, we have that for any $P \in \mathcal{P}$,

$$\mathbb{E}_P \left[\sup_{\pi \in \Pi} \left| \tilde{R}_{N,P}(\pi) - \overline{R}_N(\pi) \right| \right] \leq \mathbb{E}_P \left[\frac{1}{N} \sum_{i=1}^N \kappa \|\hat{m}_0(\cdot, X_i) - m_{0,P}(\cdot, X_i)\| \right]$$

and the conclusion follows by Assumption B.1. □

Combining the bounds of Lemma 4.5 and Corollary B.2, we obtain the following simple extension of Theorem 4.4.

Corollary B.3: *Let $\mathcal{P}_C$ be a set of distributions for which (1) Assumptions 4.1 holds with constant C and (2) Assumption B.1 holds. Under Assumptions 3.1 and 4.3,*

$$\sup_{P \in \mathcal{P}_C} \left(\mathbb{E}_P[\overline{R}_P(\hat{\pi})] - \overline{R}_P(\pi_P^*) \right) \leq K(N^{-1/2} \vee \rho_N^{-1})$$

for some constant K depending only on C and J.

We now show that the finite-sample bound in Assumption B.1 is satisfied with $\rho_N = N^{1/2}$ when covariates are discrete, and therefore Corollary B.3 is satisfied with $\rho_N = N^{1/2}$.

Proposition B.4: *Let $\mathcal{P}$ be a set of distributions for which*

1. $|\mathcal{X}| = M < \infty$ *and* Y_i *is binary*

2. $P(X_i = x, D_i = d) \geq \delta > 0$ *for all* $x \in \mathcal{X}$ *and* $d \in \mathcal{D}_0$.

Then there exists a sample size $\bar{N}$ *such that for all* $N \geq \bar{N}$,

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\frac{1}{N} \sum_{i=1}^{N} \|\hat{m}_0(\cdot, X_i) - m_{0,P}(\cdot, X_i)\| \right] \leq \left(\frac{J_0 \times |\mathcal{X}|}{\delta} \right)^{1/2} N^{-1/2}.$$

Proof. Define $Z_i = (D_i, X_i)$, and let $L = J_0 \times M$ denote the number of possible combinations of D_i and X_i. For any $P \in \mathcal{P}$, let $p_\ell = \mathbb{P}[Z_i = \ell]$. By Chernoff's inequality (Vershynin 2018 Theorem 2.3.1),

$$P\left(\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] \leq \frac{Np_\ell}{2} \right) \leq \exp\left((\ln(2) - 1) \frac{Np_\ell}{2} \right). \tag{16}$$

For any $\ell \in [L]$, let $m_\ell = \mathbb{E}[Y_i \mid Z_i = \ell]$. Conditional on $\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell]$, we have

$$\begin{aligned} \mathbb{E}\left[\left(\frac{1}{\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell]} \sum_{i=1}^{N} y_i \mathbb{1}[Z_i = \ell] - m_\ell \right)^2 \right] &= \frac{m_\ell(1 - m_\ell)}{\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell]} \\ &\leq \frac{1}{4 \sum_{i=1}^{N} \mathbb{1}[Z_i = \ell]} \end{aligned}$$

for any $\ell \in [L]$, since $m_\ell \in [0, 1]$.

Since $p_\ell \geq \delta$ by assumption, we have that conditional on $\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] > \frac{Np_\ell}{2}$,

$$\mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2 \right] \leq \frac{1}{2N\delta} \tag{17}$$

and therefore

$$\begin{aligned} \mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2 \right] \leq{}& \mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2 \mid \sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] > \frac{Np_\ell}{2} \right] \mathbb{P}\left(\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] > \frac{Np_\ell}{2} \right) \\ &+ \mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2 \mid \sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] \leq \frac{Np_\ell}{2} \right] \mathbb{P}\left(\sum_{i=1}^{N} \mathbb{1}[Z_i = \ell] \leq \frac{Np_\ell}{2} \right) \\ \leq{}& \frac{1}{2N\delta} + \exp\left((\ln(2) - 1) \frac{N\delta}{2} \right) \end{aligned}$$

where in the last lines we have used the bounds (16) and (17), and the fact that $(\hat{m}_\ell - m_\ell)^2 \leq 1$. There exists $\bar{N}$ such that for all $N \geq \bar{N}$, the second term is less than $\frac{1}{2N\delta}$ and therefore

$$\mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2 \right] \leq \frac{1}{N\delta}.$$

By Jensen's inequality, the overall error is bounded by

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\|\hat{m}_0(\cdot,X_i)-m_{0,P}(\cdot,X_i)\|\right] \leq \left(\sum_{\ell=1}^{L}\mathbb{E}\left[(\hat{m}_\ell - m_\ell)^2\right]\right)^{1/2}$$

and therefore

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\|\hat{m}_0(\cdot,X_i)-m_{0,P}(\cdot,X_i)\|\right] \leq \left(\frac{L}{N\delta}\right)^{1/2}$$

for all $N \geq \bar{N}$. Since L and δ do not depend on P, this bound is uniform over $\mathcal{P}$. □

B.2 Estimation of the parameters of the utility function

In this section we relax the assumption that $b(d,x)$ and $c(d,x)$ are known. We instead assume that estimates of $b(d,x)$ and $c(d,x)$ are available. This assumption is stronger than the assumption that $\hat{m}_0(d,x)$ converges to $m_0(d,x)$, because it implies a consistent estimate of $b(d,x)$ and $c(d,x)$ is available for new treatments as well as existing ones. This assumption may be satisfied if $b(d,x) = b(x)$ and $c(d,x) = c(x)$ are known to not depend on d. For example, $b(x)$ may represent the value of a good to a household with characteristics x, and $c(x)$ may represent the cost of delivering the good to the household. Alternatively, it is also possible that $b(d,x)$ and $c(d,x)$ depend on d in a known way, such as price minus marginal cost, $b(d,x) = d - b(x)$.

Redefine the regret score analogously to (5), using the estimated utility function.

$$\begin{aligned}\hat{\Gamma}_{jk}(X_i) = \max_{m\in\mathbb{R}^J} \quad & \hat{b}_{jk}(X_i)'m - \hat{c}_{jk}(X_i) \\ \text{s.t.} \quad & Sm \leq r \\ & Fm = \hat{m}_0(X_i)\end{aligned}$$

We assume that the estimated parameters converge to the true parameters at the same rate as the estimated mean of the treatment effect. We also assume that mean conditional response is bounded, as it is in Section 5.

Assumption B.5: *For some sequence $\rho_N \to \infty$ and some class of distributions $\mathcal{P}$,*

1. $\limsup_{N\to\infty}\sup_{P\in\mathcal{P}}\rho_N\mathbb{E}_P\left[\frac{1}{N}\sum_{i=1}^{N}\|\hat{m}_0(\cdot,X_i)-m_{0,P}(\cdot,X_i)\|\right] \leq \infty$
2. $\limsup_{N\to\infty}\sup_{P\in\mathcal{P}}\rho_N\mathbb{E}_P\left[\frac{1}{N}\sum_{i=1}^{N}\|\hat{b}(\cdot,X_i)-b(\cdot,X_i)\|\right] \leq \infty$
3. $\limsup_{N\to\infty}\sup_{P\in\mathcal{P}}\rho_N\mathbb{E}_P\left[\frac{1}{N}\sum_{i=1}^{N}\|\hat{c}(\cdot,X_i)-c(\cdot,X_i)\|\right] \leq \infty$

4. $\hat{\mathcal{M}} = \{m : Sm(\cdot, X) \leq r,\ Fm(\cdot, X) = \hat{m}_0(\cdot, X)\}$ *is nonempty, almost surely, for all* $P \in \mathcal{P}$.

5. *There exists a bounded set* $\overline{\mathcal{M}} \subset \mathbb{R}^J$ *such that* $\hat{\mathcal{M}} \subset \overline{\mathcal{M}}$ *almost surely, for all* $P \in \mathcal{P}$.

We show that under this assumption, the conclusions of Lemma 4.6 and Theorem 4.4 hold.

Proposition B.6: *Under Assumptions 3.1, 4.3, and B.5,*

$$\sup_{P \in \mathcal{P}} \mathbb{E}_P \left[\overline{R}_P(\hat{\pi}) - \overline{R}_P(\pi_P^*) \right] \leq \mathcal{O}(N^{-1/2} \vee \rho_N^{-1})$$

Proof. We write the difference between the estimated and true regret scores as

$$\begin{aligned}
\left| \hat{\Gamma}_{jk}(X_i) - \Gamma_{jk}(X_i) \right| &= \left| \max_{m \in \hat{\mathcal{M}}} \hat{b}_{jk}(X_i)'m - \hat{c}_{jk}(X_i) - \left(\max_{m \in \mathcal{M}} b_{jk}(X_i)'m - c_{jk}(X_i) \right) \right| \\
&\leq \left| \max_{m \in \hat{\mathcal{M}}} \hat{b}_{jk}(X_i)'m - \max_{m \in \hat{\mathcal{M}}} b_{jk}(X_i)'m \right| \\
&\quad + \left| \max_{m \in \hat{\mathcal{M}}} b_{jk}(X_i)'m - \max_{m \in \mathcal{M}} b_{jk}(X_i)'m \right| \\
&\quad + \left| \hat{c}_{jk}(X_i) - c_{jk}(X_i) \right|
\end{aligned}$$

The first term is bounded by

$$\begin{aligned}
\left| \max_{m \in \hat{\mathcal{M}}} (\hat{b}_{jk}(X_i) - b_{jk}(X_i))'m \right| &\leq \left| \max_{m \in \overline{\mathcal{M}}} (\hat{b}_{jk}(X_i) - b_{jk}(X_i))'m \right| \\
&\leq \|\hat{b}_{jk}(X_i) - b_{jk}(X_i)\| \overline{M} \\
&\leq 2\|\hat{b}(\cdot, X_i) - b(\cdot, X_i)\| \overline{M}
\end{aligned}$$

where $\overline{M} = \max_{m \in \overline{\mathcal{M}}} \|m\|$. The second term is bounded by

$$\kappa \|\hat{m}_0(X_i) - m_{0,P}(X_i)\|$$

by Lemma 4.6. The third term is bounded by

$$|c_j(X_i) - \hat{c}_j(X_i)| + |c_k(X_i) - \hat{c}_k(X_i)| \leq 2\|\hat{c}(\cdot, X_i) - c(\cdot, X_i)\|$$

and therefore all three terms are $\mathcal{O}(\rho_N^{-1})$ uniformly in $P \in \mathcal{P}$.

Together, we have that

$$\sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\sup_{\pi\in\Pi}\left|\tilde{R}_N(\pi) - \overline{R}_N(\pi)\right|\right] \leq \sup_{P\in\mathcal{P}} \mathbb{E}_P\left[\frac{1}{N}\sum_{i=1}^{N}\max_{j,k}\left|\hat{\Gamma}_{jk}(X_i) - \Gamma_{jk}(X_i)\right|\right]$$
$$\leq \mathcal{O}(\rho_N^{-1})$$

Combining this with Lemma 4.5 yields the conclusion. □

C Simulation study

This section presents the results of a simulation study that examines how the choice of estimator $\hat{m}_0$ affects the performance of the resulting estimated policy.

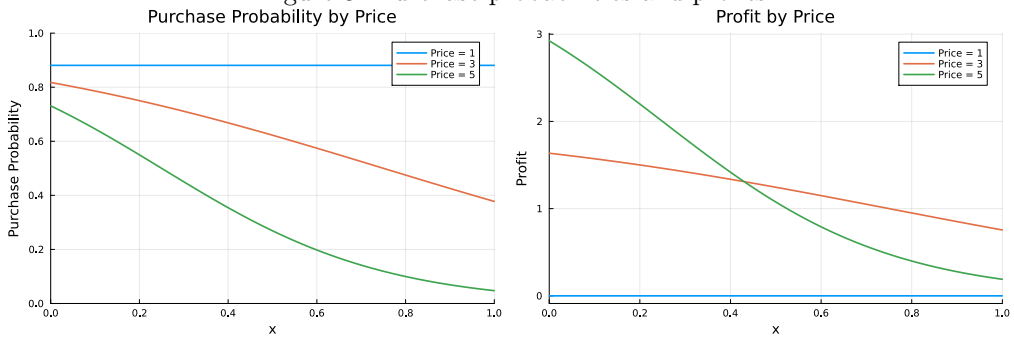
We consider a firm choosing a price to maximize profit. The firm faces logit demand given by

$$P(y_i = 1|z_i) = \frac{\exp(z_i'\beta)}{1+\exp(z_i'\beta)}$$

and constant marginal cost of 1. Here $z_i = (1, x_i, p_i, x_i p_i)$ and $x_i \sim U[0,1]$ is a continuous covariate. The price p_i is randomly assigned from $\mathcal{D}_0 = \{1,3,5\}$ independently of x_i. The set of possible prices is $\mathcal{D} = \{1,2,3,4,5\}$. The conditional mean response is assumed to be decreasing in price, holding x_i fixed.

For the simulation, we set $\beta = (2.25, 1, -0.25, -1)$. The purchase probabilities and profits are plotted in Figure 8.

Figure 8: Purchase probabilities and profits



A policy is defined by a scalar β and a vector of cutoffs $\{\kappa_j\}_{j=1}^4$ with $\kappa_{j-1} \leq \kappa_j$ for all j. Unit i is assigned to price j if $\kappa_{j-1} < \beta x_i \leq \kappa_j$ (letting $\kappa_0 = -\infty$ and $\kappa_5 = \infty$). Only the sign of β matters, since the cutoffs must be increasing.

We consider $N \in \{100, 250, 500, 750, 1000\}$ observations. We perform 200 Monte Carlo simulations for each value of N. For each simulation, we estimate two models for the response function: a logit model

in x_i separately for each price, and a Lasso model with a logit link function on a dictionary of Chebyshev polynomials in x_i, separately for each price. The regularization parameter is selected by cross-validation. The maximum regret (estimated on holdout data) is plotted in Figure 9.

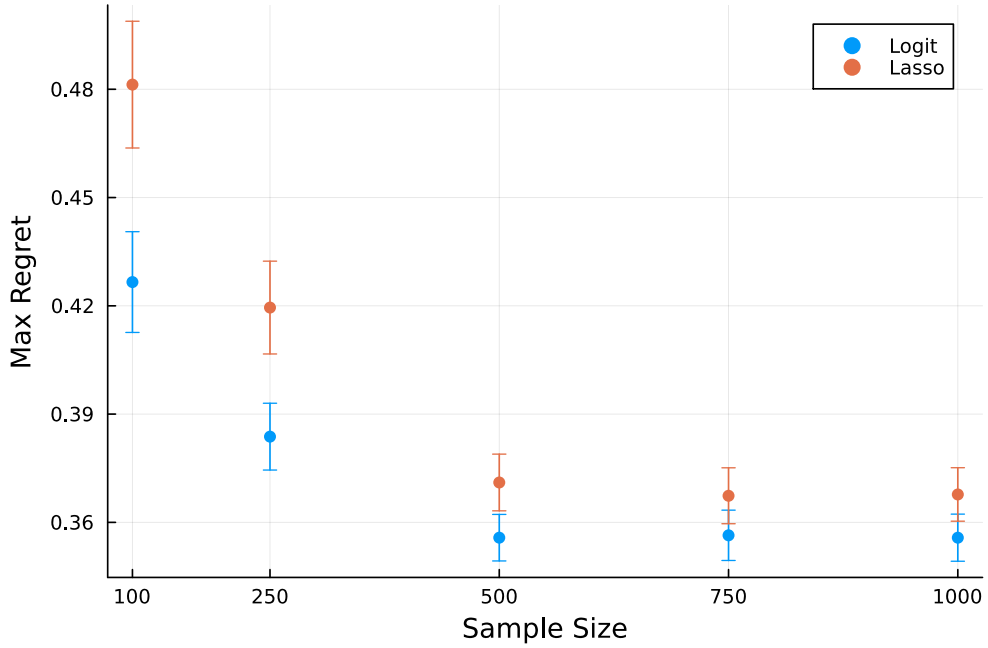


Figure 9: Maximum regret

As predicted by Theorem 4.4, the Lasso estimator achieves higher regret than the correctly specified parametric model.

D Computational details

D.1 Shape constraints

Mean takeup as a function of price, holding covariates fixed, is bounded between 0 and 1. It is assumed to be downward sloping and the subsidy is assumed to exhibit decreasing returns to scale, so that takeup is convex. These constraints can be expressed as $S\tilde{m}(.,X) \leq r$ where

$$S_1 = \begin{bmatrix} \frac{-1}{d_2-d_1} & \frac{1}{d_2-d_1} & 0 & \dots & & & \\ 0 & \frac{-1}{d_3-d_2} & \frac{1}{d_3-d_2} & 0 & \dots & & \\ \vdots & & & & & & \\ & & & & \dots & 0 & \frac{-1}{d_J-d_{J-1}} & \frac{1}{d_J-d_{J-1}} \end{bmatrix}$$

$$S_2 = \begin{bmatrix} \frac{-1}{d_2-d_1} & \frac{1}{d_2-d_1}+\frac{1}{d_3-d_2} & \frac{-1}{d_3-d_2} & 0 & \dots & & \\ 0 & \frac{-1}{d_3-d_2} & \frac{1}{d_3-d_2}+\frac{1}{d_4-d_3} & \frac{-1}{d_4-d_3} & 0 & \dots & \\ \vdots & & & & & & \\ & & & & \dots & 0 & \frac{-1}{d_{J-1}-d_{J-2}} & \frac{1}{d_{J-1}-d_{J-2}}+\frac{1}{d_J-d_{J-1}} & \frac{-1}{d_J-d_{J-1}} \end{bmatrix}$$

$$S = \begin{bmatrix} -I \\ I \\ S_1 \\ S_2 \end{bmatrix} \qquad r = \begin{bmatrix} 0_J \\ -1_J \\ 0_{J-1} \\ 0_{J-2} \end{bmatrix}$$

D.2 Dual representation of maximum regret

For each individual i, the maximum regret is obtained by

$$\hat{\Gamma}_j(X_i) = \max_{k,m} b_{jk}(X_i)'m - c_{jk}(X_i) \quad s.t. \quad Sm \leq r, \quad Fm = \hat{m}(., X_i)$$

For now, suppress the dependence on X_i and j. Let λ be the vector of Lagrange dual variables associated with the inequality constraint, and let η be the vector of Lagrange dual variables associated with the equality constraint. We can rewrite the linear program as

$$\min_{\mu} \mu \quad s.t. \quad \mu + c_k \geq \max_{m} \{b_k' m \quad s.t. \quad Sm \leq r, Fm = \hat{m}_0\} \quad \forall k$$

$$\min_{\mu} \mu \quad s.t. \quad \mu + c_k \geq \min_{\lambda,\eta} \{\lambda' r + \eta' \hat{m}_0 \quad s.t. \quad \lambda' S + \eta' F \geq b_k, \quad \lambda \geq 0\} \forall k$$

$$\min_{\mu,\lambda,\eta} \mu \quad s.t. \quad \mu + c_k \geq \lambda' r + \eta' \hat{m}_0, \quad \lambda' S + \eta' F \geq b_k, \quad \lambda \geq 0 \quad \forall k$$

Thus, computing each $\hat{\Gamma}_j(X_i)$ is a single linear program. Since the programs are independent across i, they can be solved simultaneously by summing the objective across individuals.

Since the dual formulation above is expressed as a minimization problem, it can be solved jointly with the minimization over π in (4). Alternatively, since there are finitely many $\hat{\Gamma}_j(X_i)$, we can solve these linear programs before performing the policy minimization, and plug the values into (4). In the application of Section 5, solving the problem jointly did not decrease computation time.

D.3 MILP Formulation of MMR Problem

Consider the set of policies given by (10) in the problem (4). To express this as a mixed integer-linear program, introduce the binary variables g_{ij} for $i = 1, \dots, N$ and $j = 1, \dots, J-1$ to indicate whether $X_i'\beta$ is

above cutoff j. For notational convenience, set $g_{i0} = 1$ and $g_{iJ} = 0$.

We introduce constraints to ensure that g_{ij} is one if and only if $X_i'\beta$ is above cutoff j using the "big-M" method. These constraints are

$$X_i'\beta - c_j \leq M g_{ij}$$
$$X_i'\beta - c_j \geq -M(1 - g_{ij}) + \epsilon$$

where M is a sufficiently large constant and ϵ is a sufficiently small numerical error tolerance. The first constraint ensures that $g_{ij} = 1$ if $X_i'\beta > c_j$, and the second constraint ensures that $g_{ij} = 0$ if $X_i'\beta \leq c_j + \epsilon$. The constant M must be chosen large enough to ensure that $|X_i'\beta - c_j| \leq M$ for all i and j. The tolerance ϵ is introduced to imitate a strict inequality constraint, which is not possible to impose exactly in a mixed integer-linear program. Otherwise, a solution of $\beta = 0, c_j = 0$ for all j would permit g_{ij} to be either 0 or 1 for all i and j. For any optimal policy (β, c), the policy $(t\beta, tc)$ is also optimal for $t > 0$. Thus, without loss of generality, we may set $M = 1$.

This leads to the following mixed integer-linear program:

$$\begin{aligned}
&\min_{g,\beta,c} \sum_{i=1}^{N}\sum_{j=1}^{J}(g_{ij} - g_{i,j-1})\hat{\Gamma}_j(X_i) \\
&s.t. \quad c_1 \leq c_2 \leq \cdots \leq c_{J-1} \\
&\qquad g_{ij} \geq X_i'\beta - c_j + \epsilon, \quad i = 1, \ldots, N, \quad j = 1, \ldots, J-1 \\
&\qquad g_{ij} \leq 1 + X_i'\beta - c_j, \quad i = 1, \ldots, N, \quad j = 1, \ldots, J-1 \\
&\qquad \beta_1 \geq 0
\end{aligned}$$

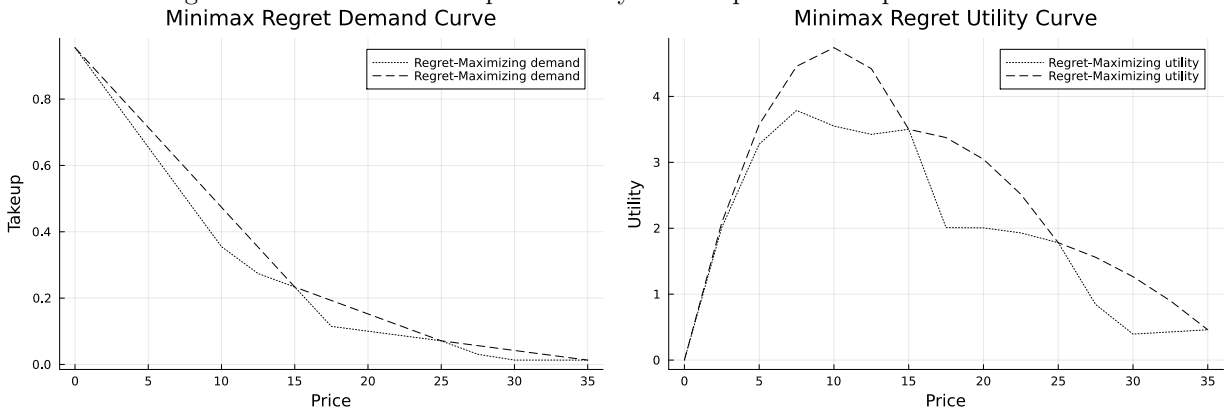
which we can solve using a standard mixed integer-linear programming solver.

E Robustness

E.1 Maximin welfare with Lipschitz constraint

I first illustrate the effect of shape restrictions on the maximin welfare policy. I give an example of shape restrictions that ensure the maximin welfare policy assigns a new treatment to the population. The maximin welfare policy without covariates is given by the price that maximizes the lower bound on welfare in Figure 2. This policy assigns a price of 15 thousand shillings to the population. Although the maximin welfare policy restricts to the original set of treatments, this is not true in general and depends on the utility function and

Figure 10: Bounds on takeup and utility at each price with Lipschitz constraint



The maximal and minimal possible expected takeup at each price, and the corresponding bounds on expected utility generated by these bounds on takeup.

the shape restrictions imposed.

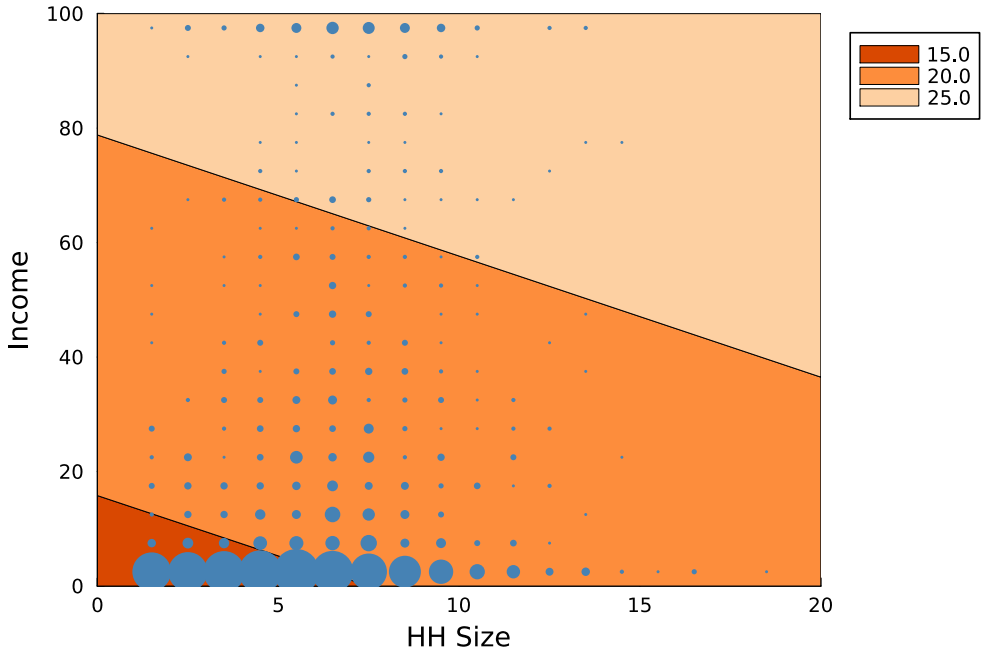
To illustrate this, I give an example in which the maximin welfare policy assigns a new treatment to the population. Specifically, in the example of Section 5, impose in addition that takeup is Lipschitz continuous, with a Lipschitz constant of 0.06. The resulting bounds on takeup and welfare are plotted in Figure 10.

From Figure 10, we see that the maximin welfare policy assigns a price of 7.5 thousand shillings to the population.

E.2 Robustness to utility parameters and treatment set

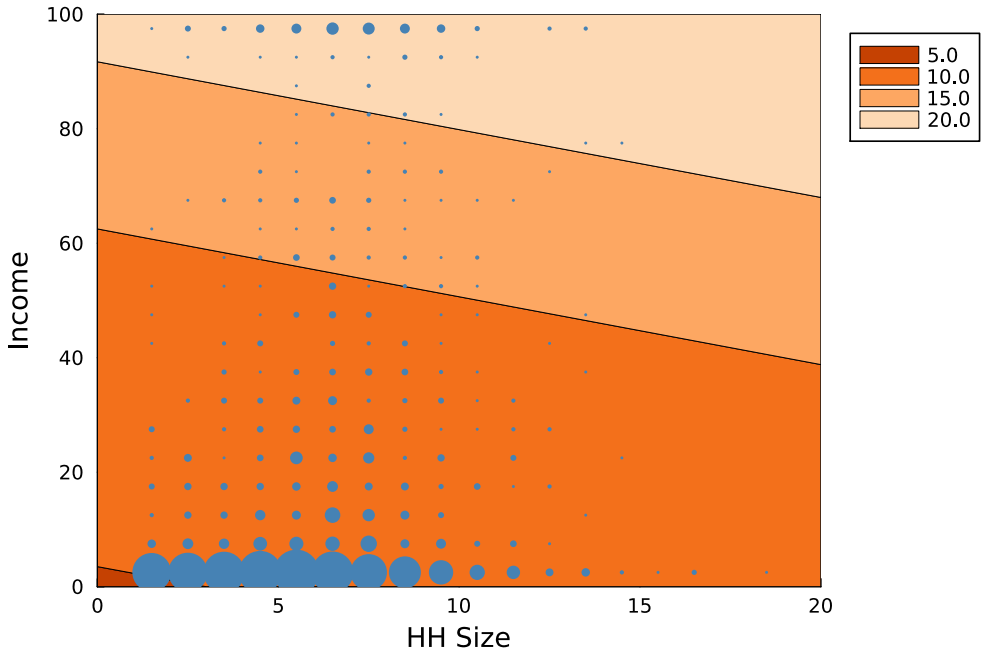
In these section I present the optimal policy under different specifications of $\mathcal{D}$ and α. I consider three specifications of $\mathcal{D}$ constructed from equally spaced prices between 0 and 35 thousand shillings, where $d_{j+1} - d_j = \Delta$ for $\Delta \in \{5, 2.5, 1\}$. I consider three values of $\alpha \in \{25, 35, 45\}$. The results are plotted in Figures 11-18, except for $\Delta = 2.5$ and $\alpha = 25$, which is shown in Section 5.

Figure 11: $\Delta = 5$, $\alpha = 25$



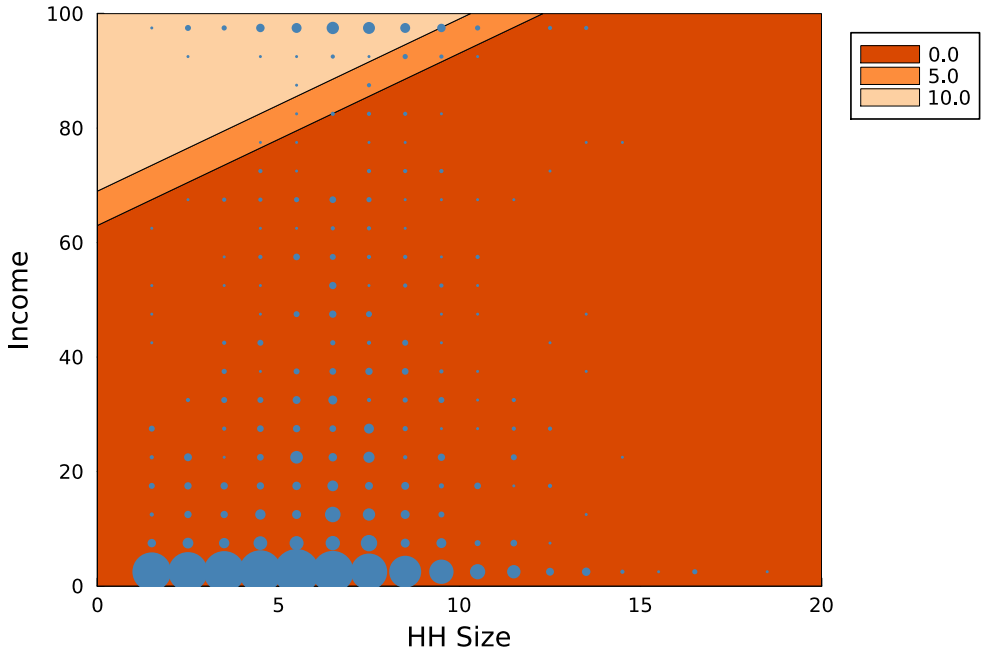
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 12: $\Delta = 5$, $\alpha = 35$



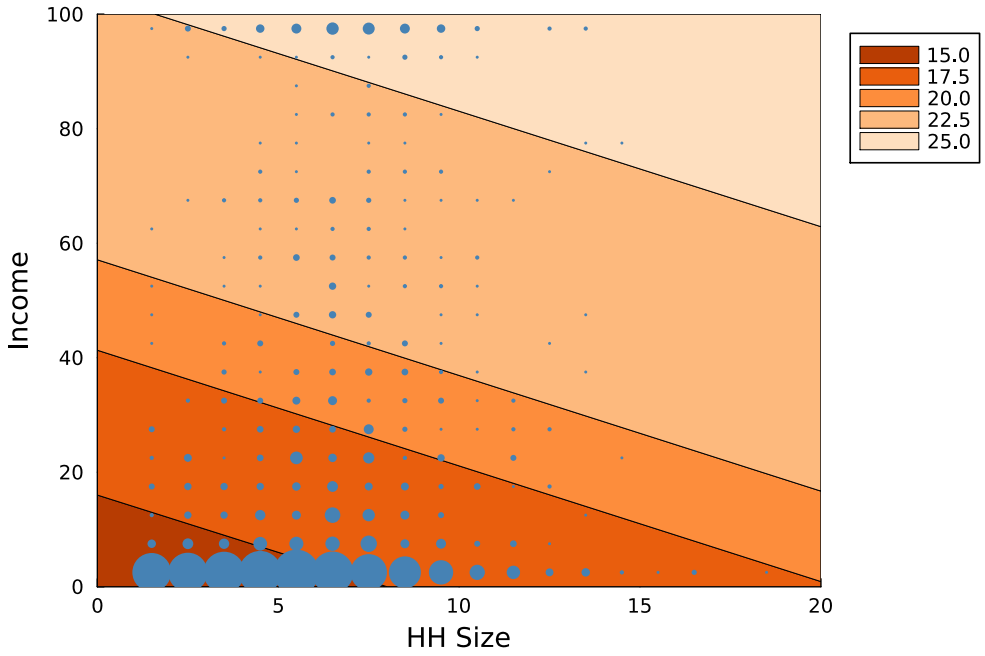
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 13: $\Delta = 5$, $\alpha = 45$



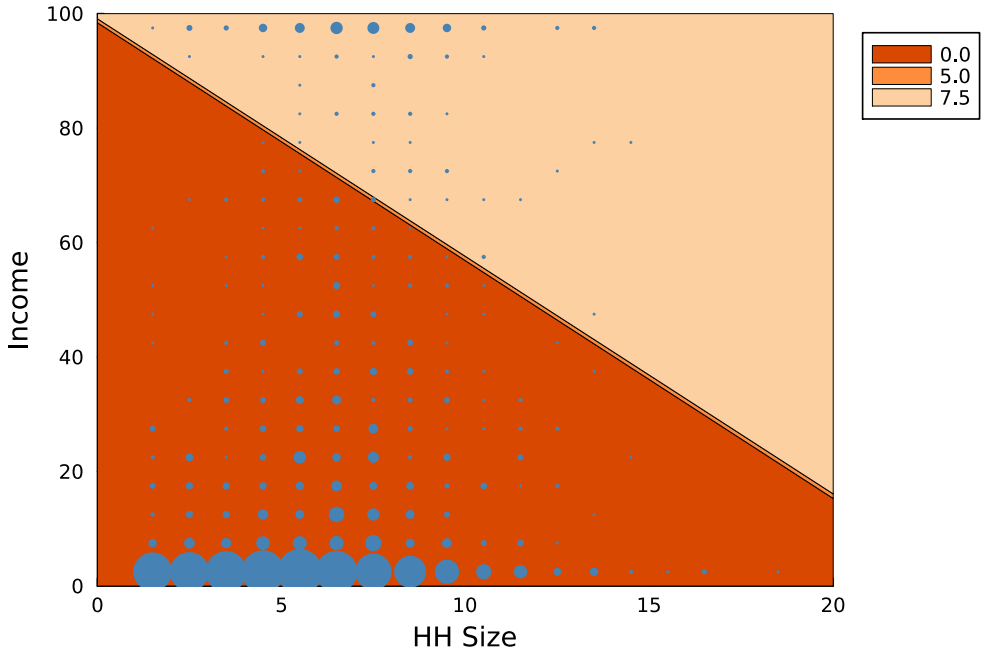
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 14: $\Delta = 2.5$, $\alpha = 25$



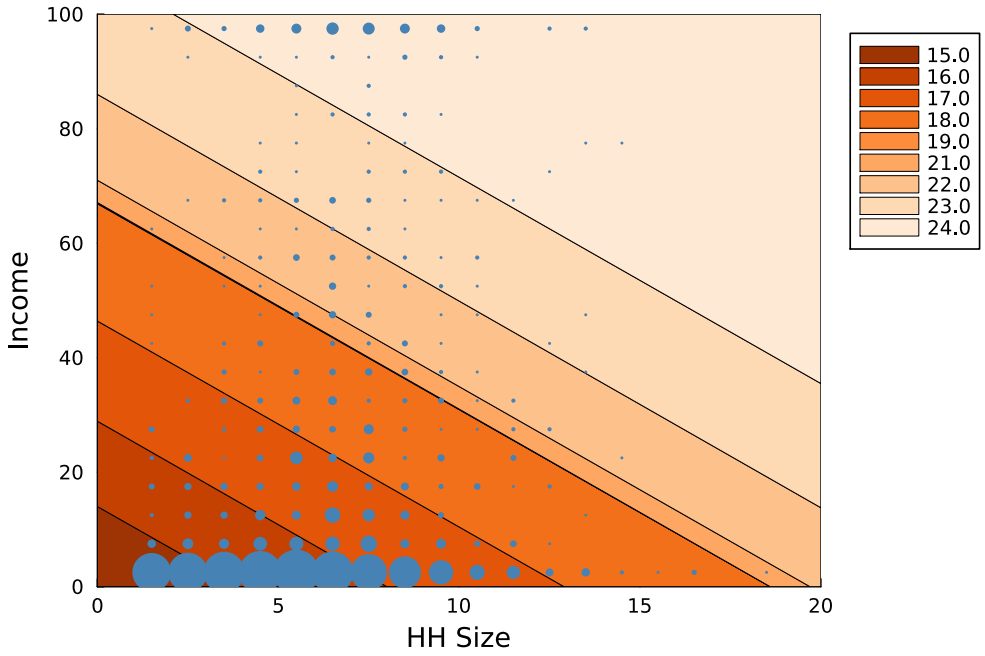
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 15: $\Delta = 2.5$, $\alpha = 45$



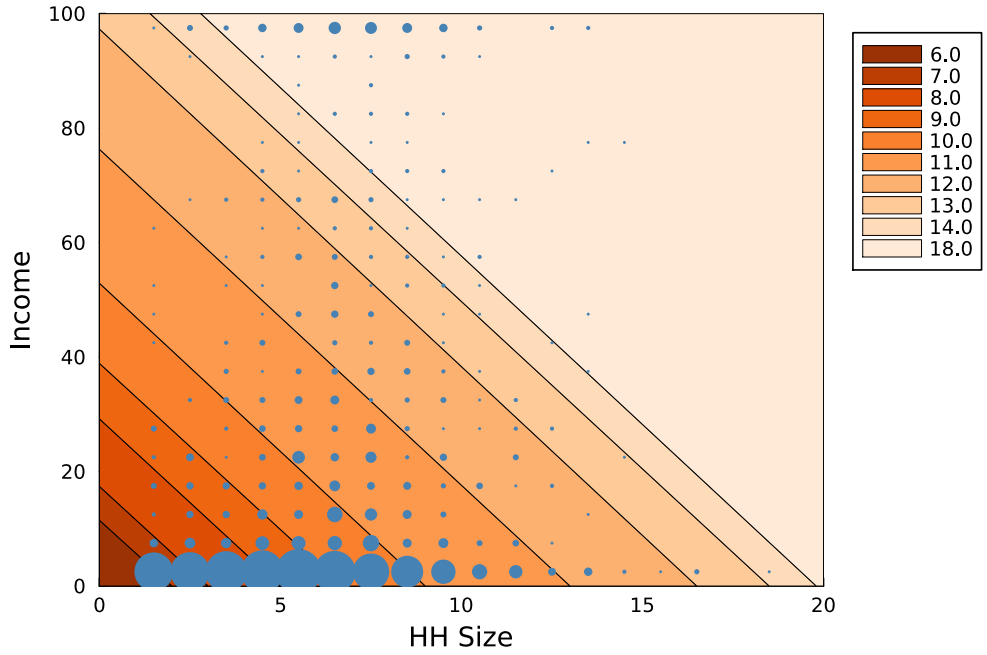
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 16: $\Delta = 1$, $\alpha = 25$



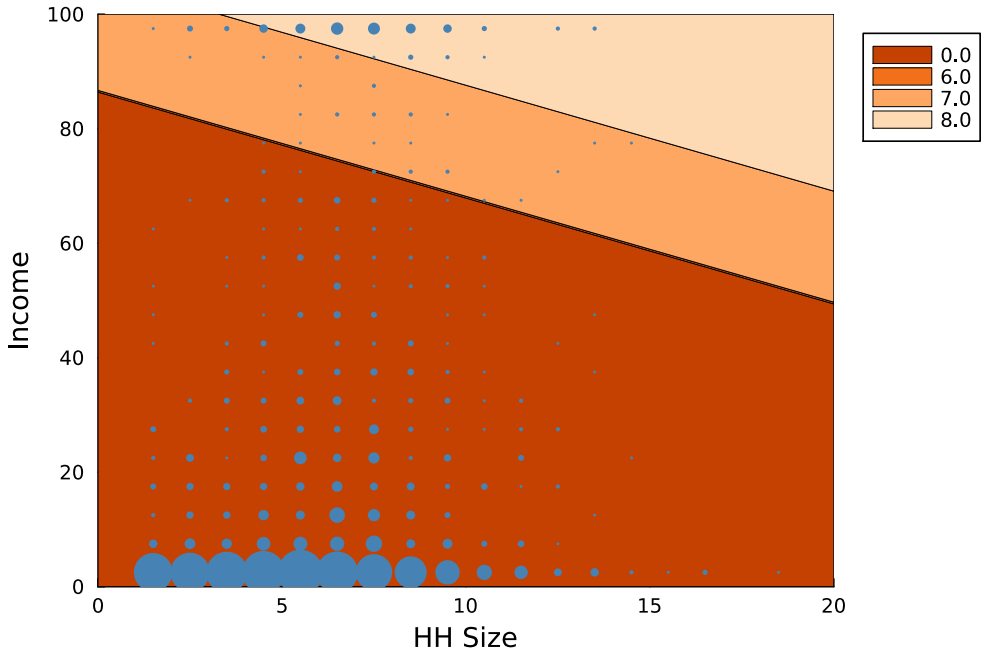
The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 17: $\Delta = 1$, $\alpha = 35$



The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.

Figure 18: $\Delta = 1$, $\alpha = 45$



The estimated optimal treatment allocation as a function of household size and earnings. The size of the dots is proportional to the number of people at each value of covariates. The shaded regions indicate which covariate values are assigned to each treatment.