Learning Practical Policies for Populations with Implementation Costs

Junlong Aaron Zhou*, Michael Gechter†, and Cyrus Samii‡

October 24, 2022

Abstract

We use a sample in which treatment assignment is unconfounded to learn a treatment-targeting policy that accounts for implementation costs in the population. The resulting "practical" treatment policy trades gains from complex targeting for reductions in costs from simplifying the targeting. We define a penalized welfare criterion that incorporates measurement and complexity costs and solve for the optimal coarsened policy. We derive a bound on welfare loss for the coarse policy. We use a revealed preference method for bounding complexity costs on a scale commensurate with a welfare measure and illustrate using a conditional cash transfer experiment in Mexico.

Keywords: *Machine learning, Statistical decisions, Policy design, Empirical welfare maximization*

*TenCent, Palo Alto, USA; email: jlzhou@nyu.edu

†Department of Economics, Pennsylvania State University, University Park, USA; email: mdg5396@psu.edu

‡(Corresponding author) Politics Department, New York University, 19 W 4th Street, New York, New York, 10002, USA; email: cds2083@nyu.edu

1 Introduction

We consider the problem of using data from a sample in which treatment is unconfounded to learn treatment-targeting policies for a population in which implementation costs are non-negligible. Our paper is motivated by practical concerns in policy design. With large-scale randomized controlled trials (RCTs), we can learn about how treatment effects vary for different types of recipients. It is natural to use these results to define optimal policies, determining who should be prioritized or eligible for receiving treatment, in the full population. However, implementing such treatment policies is never free of costs and associated constraints. Treatment policies can be expensive if they require many inputs (e.g., extensive individual-level information). Complex treatment policies, which depend on complicated combinations of inputs, may be impossible for lay persons to understand. This could undermine the public legitimacy of the policy or it could prevent practitioners from understanding the rationale behind what they are doing (Rudin, 2019). It is for these reasons that we seek ways to simplify and coarsen treatment policies.

We want the coarsening to be optimal in the sense of improving welfare for recipients while also keeping implementation costs manageable. The approach that we take brings together results on penalized welfare maximization due to Mbakop and Tabord-Meehan (2021), policy learning due to Kitagawa and Tetenov (2018), Athey and Wager (2021), and Zhou, Athey, and Wager (2022), and interpretable machine learning for practical treatment regimes (see Lakkaraju and Rudin (2017), and Rudin, Chen, Chen, Huang, Semenova, and Zhong (2022) for a review). Following Lakkaraju and Rudin (2017), we specify implementation costs as the sum of two terms: the *complexity* of the policy (e.g., the depth of an assignment tree or the number terms in a linear scoring rule) and the *measurement costs* for the inputs into the policy (e.g. the cost of a medical test). As in Athey and Wager (2021), we use a two-step approach that first constructs doubly-robust conditional treatment effect (or welfare) scores based on a high-dimensional machine learning model, and in the second step works with these scores to derive an optimally coarsened policy using an interpretable (e.g. tree-based

or linear) policy rule. Our methods and analysis covers both randomized experiments and observational studies under unconfoundedness.

The two-step approach is informed by Chernozhukov, Demirer, Duflo, and Fernandez-Val (2018), who discuss problems with using generally inconsistent machine learning predictions directly, and strategies for reliable inference with coarser estimands. The prior literature on treatment assignment starting with Manski (2004) evaluates the performance of policies in terms of regret, the deviation in welfare from that of the optimal policy at the population level. Our approach also evaluates welfare in terms of regret, while adding implementation costs. Based on the ways that welfare losses are traded for gains in terms of simplicity or measurement cost-savings, our approach arrives at an optimal coarsening. As such, the complexity of the selected policy is endogenous, in a manner that is distinct from the fixed-complexity policy classes considered by Kitagawa and Tetenov (2018) and Athey and Wager (2021). The complexity and measurement costs operate as additional penalization criteria in a welfare maximization problem already penalized to prevent overfitting. This is a variation on the penalized welfare maximization approach of Mbakop and Tabord-Meehan (2021), although one that goes beyond their concern for overfitting to also take implementation costs into account. Following Mbakop and Tabord-Meehan (2021)'s results, we derive a bound on the regret from our coarsening algorithm to show how increased information improves learning.

Using our approach requires construction of welfare scores and then specification of complexity and measurement costs in terms that are commensurate with these scores. We illustrate how to do this through an application to the experimental evaluation of the PROGRESA program in Mexico. PROGRESA aimed to alleviate current and future poverty levels through cash transfers, the largest portion of which were distributed conditional on children's school enrollment. The PROGRESA evaluation collected many covariates, allowing in principle for defining targeting strategies that could be more efficient than the program's original eligibility criteria. For the welfare scores, we use the social welfare outcome from

Gechter, Samii, Dehejia, and Pop-Eleches (2018), which has the policymaker taking into account a desire for poverty alleviation in the present as well as future earnings gains due to the program's conditionality causing children to enroll for an additional year of school. The welfare scores are on the scale of contemporary Mexican pesos. The monetary costs of alternative measurement plans are documented for this case in Skoufias, Davis, and Behrman (1999). For complexity costs, we compare the program eligibility criteria chosen by the Mexican government to other plausible alternatives of greater or lesser complexity to identify the government's implicit complexity cost in monetary terms using a revealed preference argument.

We contribute to the literature on statistical policy learning in econometrics by addressing practical cost concerns. This strand of literature (a partial list beyond those already cited includes Dehejia, 2005; Hirano and Porter, 2009; Stoye, 2009) considers optimal policy design as a statistical decision problem of maximizing population welfare. We extend this literature by addressing the practical issues a social planner faces when bringing a policy to the full population: costs of accessing the data needed to implement the policy and concerns about interpretability.

Finally, we contribute to the literature on policy learning using machine learning methods. The development of machine learning methods to characterize treatment effect heterogeneity motivates using machine learning tools to design individual-level policies (Laber and Zhao, 2015; Qian and Murphy, 2011). In bio-statistics, the literature on subgroup analysis develops methods to detect the subgroup with the largest treatment effect based on observed outcomes (Foster, Taylor, and Ruberg, 2011; Foster, Taylor, Kaciroti, and Nan, 2015; Cai, Lu, West, Mehrotra, and Huang, 2021). More recently, the connection between policy design and classification problems has been explored (Zhang, Tsiatis, Davidian, Zhang, and Laber, 2012; Kallus, 2017; Zhang, Tsiatis, Laber, and Davidian, 2012; Chen and Lee, 2018). The literature on interpretable machine learning also addresses practical cost concerns when using machine learning for policy design (Lakkaraju and Rudin, 2017; Rudin et al., 2022). However,

prior work in this literature is specific to particular machine learning algorithms which may not accord with the policymaker's needs or fit the data well, while our approach provides conditions on the first-stage doubly-robust scores which different machine learning methods could meet. In addition, the second stage of our algorithm can use any interpretable machine learning method which can be made sensitive to implementation costs. We illustrate using linear and tree-based rules.

This paper is organized as follows: in section 2 we present our inferential setting and review optimal treatment assignment without implementation costs. In section 3 we introduce our approach to introducing implementation costs. Section 4 details the estimation algorithm and section 5 states the theoretical results. Section 6 presents the results from our application to the PROGRESA evaluation.

2 Setting

We consider a sample of units indexed by $i = 1, ..., n$ drawn from some large population $\mathcal{Z}$. Unit i in the sample is assigned a binary treatment $D_i \in \{0, 1\}$, and is characterized by a p-vector of measured pre-treatment covariates $\mathbf{X}_i \in \mathbb{X} \subset \mathbb{R}^p$, and possesses potential outcomes $(Y_i(1), Y_i(0)) \in \mathbb{R}^2$ depending on whether $D_i = 1$ or 0, respectively. Agents do not observe the full schedule of potential outcomes, but rather the realized outcome $Y_i = D_i Y_i(1) + (1 - D_i)Y_i(0)$ for unit i in the sample. Unit i's treatment effect is $\tau_i = Y_i(1) - Y_i(0)$. We write the joint distribution for the sample potential outcomes, treatment, and covariates as $(Y_i(1), Y_i(0), D_i, \mathbf{X}_i)_{i=1}^n \sim P_n$, where P_n depends on population values, sampling design, and the treatment assignment mechanism. Then, in the population, potential outcomes and covariates are distributed as $(Y(1), Y(0), \mathbf{X}) \sim P$. Outcomes realized in the population depend on whether someone is assigned to receive treatment or not, and this decision is encoded in terms of the indicator π, defined below.

Throughout this paper, we will assume unconfoundedness, i.e.

Assumption 1 *(Unconfoundedness) Under P^n, the sample distribution satisfies*

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i | \mathbf{X}_i.$$

This assumption is satisfied by construction in randomized experiments, our leading case, but may also hold in observational studies. Under assumption 1, we can estimate conditional treatment effects from the sample.

The policymaker's goal is to use the data available in the sample to design a policy determining treatment assignments for all individuals in $\mathcal{Z}$ on the basis of their characteristics $\mathbf{X}_i$. We define a policy $\pi(\cdot) \in \Pi$ as a mapping $\pi : \mathbb{X} \to \{0, 1\}$. Π is a policy space, containing all candidate policies. There are many policies that a planner could consider for assigning treatments to individuals in $\mathcal{Z}$. We provide two specific examples below.

Example 2.1 *Uniform policy:* $\Pi = \{\forall k \in \mathcal{Z}, \pi_k = 1 \text{ or } \pi_k = 0\}$.

Under the uniform policy, the planner either assigns the treatment to all individuals, or to none. Such a policy may be reasonable when the planner's fairness considerations necessitate treating everyone equally, regardless of characteristics $\mathbf{X}_i$, such as whether or not to reimburse a basic healthcare expense. The uniform policy may also be optimal when the welfare benefits of a more targeted policy do not outweigh its implementation costs.

Example 2.2 *Threshold allocation:* $\Pi = \{\pi : \pi = \mathbb{1}\{x^{(j)} \leq c\} \text{ or } \mathbb{1}\{x^{(j)} \geq c\}\}$,

where the superscript (j) denotes the jth element of x, for example student GPA. Such a policy may be reasonable to reward effort: students may get a scholarship only when a GPA requirement is met.

Note that the set of candidate policies Π can contain a range of simple to complex policies. The second example has a policy space that is more complicated than the first. We formally discuss the measure of complexity below.

Prior to introducing implementation concerns, suppose the social planner wants to select a policy based on expected impact for a suitably-chosen (welfare) outcome. For such a planner,

we write the expected social welfare of a policy π as

$$W(\pi) = E[Y_i(\pi(\mathbf{X}_i)) - Y_i(0)] = E[\tau_i \pi(\mathbf{X}_i)],$$

where $E[\cdot]$ taken with respect to the population distribution P. As in our application, by appropriately defining the outcome, this specification allows for social welfare functions that take non-linear transformations of outcome measurements and covariates in the data so long as welfare remains additive across agents (as in Haushofer, Niehaus, Paramo, Miguel, and Walker (2022)). It does not cover the non-additive social welfare functions considered in, e.g., Kitagawa and Tetenov (2021).

The optimal policy within the feasible set of policies Π satisfies

$$\pi^* = \arg\max_{\pi \in \Pi} W(\pi). \tag{1}$$

A difficulty in computing this optimum is that the τ_i are not observable, because of the fundamental problem of causal inference (Holland, 1986). A standard approach in the current literature is to construct an observable score that allows one to identify the correct optimum in expectation. So as to generalize to both randomized experiments and observational studies under unconfoundedness, instead of τ_i in the expression for $W(\pi)$, we follow Athey and Wager (2021) and use the doubly robust welfare score:

$$\Gamma_i = m(\mathbf{X}_i, 1) - m(\mathbf{X}_i, 0) + \frac{D_i - e(\mathbf{X}_i)}{e(\mathbf{X}_i)(1 - e(\mathbf{X}_i))}(Y_i - m(\mathbf{X}_i, D_i))$$

where $m(x, d)$ represents the counterfactual response surface with treatment equal to d and $e(x) = P[D_i = 1 | \mathbf{X}_i = x]$, the propensity score. Using the doubly robust scores has the benefit of generalizing our approach to settings with observational or experimental data, although this generalization does imply that some of our analyses will rely on asymptotic results.

We estimate each individual's score value according to the out-of-sample procedure:

$$\hat{\Gamma}_i = \hat{m}_n^{-k(i)}(\mathbf{X}_i, 1) - \hat{m}_n^{-k(i)}(\mathbf{X}_i, 0) + \frac{D_i - \hat{e}_n^{-k(i)}(\mathbf{X}_i)}{\hat{e}_n^{-k(i)}(\mathbf{X}_i)(1 - \hat{e}_n^{-k(i)}(\mathbf{X}_i))}(Y_i - \hat{m}_n^{-k(i)}(\mathbf{X}_i, D_i)), \quad (2)$$

where the data are evenly split into K folds and $k(i)$ denotes the fold containing i-th observation. We assume we have uniformly consistent estimators of the response surface, conditional treatment effect, and propensity score such that we satisfy satisfy assumption 2 in Athey and Wager (2021), which we restate in our terms as assumption 5 in the appendix.

From the above, we can see that $W(\pi) = E[\tau_i \pi(\mathbf{X}_i)] = E[\Gamma_i \pi(\mathbf{X}_i)]$. We will use the estimated score $\hat{\Gamma}_i$ to define a sample analogue of the welfare criterion as

$$W_n(\pi) = E_n[\pi(\mathbf{X}_i)\hat{\Gamma}_i]$$

where $E_n[\cdot]$ is a consistent estimator for $E[\cdot]$ applied to the sample data. (Whereas $E[\cdot]$ defines a fixed value on $\mathcal{Z}$, $E_n[\cdot]$ defines a value that varies randomly depending on the realized sample and treatment allocation.) For given policy space Π, we can find the estimated optimal policy solving the sample analog of expression (1):

$$\hat{\pi}^* = \arg\max_{\pi \in \Pi} W_n(\pi). \quad (3)$$

3 Coarsening Policy with Implementation Costs

Having defined our concepts, we revisit the social planner's implementation problem. A social planner does not just care about participant outcomes, but also takes into account the costs of policy implementation. Kitagawa and Tetenov (2018) consider an aggregate budget constraint for (non-stochastic) treatment costs which the program may not violate.[1] For example, suppose the new treatment is costly to implement, say costing c dollars for each

[1] Sun (2021) introduces a stochastic budget constraint which must be satisfied asymptotically.

individual receiving it, and the social planner has a total budget of B dollars. This budget constraint can be directly imposed as a restriction on the policy space Π. The social planner now solves an optimization problem on the restricted policy space:

$$\max_{\substack{\pi \in \Pi \\ \sum_i c \times \pi_i \leq B}} W_n(\pi).$$

However, this is not the only cost concern a social planner may have.

First, the social planner may care about the cost of accessing the data used for treatment assignment. Experimental program evaluations often collect data on a large number of indicators, for example to detect treatment effect heterogeneity. But when the time comes to implement the policy in the full population it is more practical to collect only the minimal set of variables necessary to perform treatment assignment, with some indicators being more expensive to collect than others. For example, policymakers designing the PROGRESA program in Mexico wanted to target cash transfers to poorer households and faced a choice of how to determine eligibility in the population (Skoufias, 2005). Options included (1) measures of community deprivation (for example, the share of households with a dirt floor) available in pre-collected census data and (2) answers to survey questions about household characteristics and assets. Option 1 involves no additional cost while option 2 requires program staff to interview potential program participants.

Second, there are costs associated with complicated policies. Ultimately the PROGRESA program designers chose to combine options 1 and 2 to assign treatment in rural areas by initially classifying communities as eligible/ineligible based on the first principal component of a set of community-level correlates of poverty. Then, within eligible communities, the program determined the eligibility of households according to a linear function of 13 household characteristics (Skoufias et al., 1999; Coady, Martinelli, and Parker, 2013). Santiago Levy, chief architect of the program, attributes part of its durability across administrations to the transparency of its eligibility requirements. Thanks to the linearity of the treatment

assignment rule within eligible communities, the household-level eligibility requirements were transparent enough to be explained to citizens and members of the government as a point system, with different household assets and characteristics being worth different amounts of points (Levy, 2006).

A complicated policy can also bring the costs of requiring professionals to implement it. For example, the "assign-to-all" policy is typically easier to carry out than the threshold-allocation policy, because the latter may require professionals to carefully evaluate individuals' characteristics as inputs to a decision according to a policy manual. Medical treatments, for example, require professional expertise to check individuals' physical characteristics, comorbidity, and potential drug-drug interactions, and then make decisions based on the measured features. Simpler, interpretable, and easy-to-carry-out policy may be desirable for keeping such costs manageable, within acceptable bounds of expected treatment effectiveness.

The existing literature (e.g. Kitagawa and Tetenov, 2018) mentions these additional cost concerns and suggests that they can also be incorporated by restricting the policy space according to the budget, but this can be too narrow a view of the planner. We see a program's budget as endogenous and depending on tradeoffs between welfare gains and implementation costs. As the policy gets more complicated and granular, implementation costs increase. But the welfare benefits of better targeting may exceed these cost increases; the social planner wants a more precise policy as long as the marginal benefit exceeds its marginal cost. Mbakop and Tabord-Meehan (2021) also consider a setting without hard restrictions on the policy space, and propose to use penalized welfare maximization (PWM) to select the optimal complexity to prevent the treatment assignment rule from being overfitted to the sample. Our approach is similar in spirit, except that our penalization is based additionally on substantive features of the policymaker's objective.

To implement this idea, we follow Lakkaraju and Rudin (2017) to define the social

planner's objective (utility) as follows:

$$U(\pi) = g_1(\pi) - \lambda_2 g_2(\pi) - \lambda_3 g_3(\pi) \tag{4}$$

where

$$\begin{aligned} g_1(\pi) &= W(\pi) = E[\pi_i \Gamma_i] \\ g_2(\pi) &= \sum_{j=1}^{p} c_j \mathbb{1}\{\mathbf{X}^{(j)} \text{ used in } \pi\} \\ g_3(\pi) &= \text{Complexity}(\pi) \\ \lambda_2, \lambda_3 &> 0 \end{aligned}$$

The first term is the welfare gain from implementing the policy π, as discussed in the previous section. The second term captures the measurement cost for feature vector $\mathbf{X}^{(j)} = (X_1^{(j)}, ..., X_n^{(j)})'$, and c_j is the known cost of accessing the jth feature. The measurement cost depends on whether the policymaker's measurement plan sets $\mathbb{1}\{\mathbf{X}^{(j)} \text{ used in } \pi\} = 0$ or 1 for a given feature j.

If measurement costs can be reasonably specified in terms of a small number M of measurement plan options, then optimization in terms of measurement depends on a manageable number of discrete comparisons. For example, in our application policymakers chose between (1) simply determining eligibility at the community level using pre-collected census data, or conducting (2) minimal or (3) more detailed household surveys to determine household eligibility for the PROGRESA program. These three options, each allowing for targeting using more variables than the previous, had costs reported in Skoufias et al. (1999).

The last term reflects the cost of complexity. As the policy becomes more complex, the social planner incurs greater costs. λ_3 can be recovered from her stated or, using the strategy we pursue in our empirical application, revealed preferences. The λs determines the relative importance of the costs and benefits. The social planner wants to find the optimal π to solve

the problem in equation (4). We propose a two-step method to solve this problem.

4 Algorithm

Our two-step policy-learning algorithm fits a low-dimensional interpretable policy (step 2) using scores from a high-dimensional machine learning model fit to the sample data (step 1). This algorithm can be applied to many existing models. The high-dimensional model is used to estimate a doubly robust welfare gain score ($\hat{\Gamma}_i$) that characterizes the net welfare benefit from assigning individual i to treatment, given covariates $\mathbf{X}_i$. Our use of doubly robust scores from step 1 draws on arguments from Chernozhukov et al. (2018) and Athey and Wager (2021) and allows us to draw on existing results to bound welfare regret for the coarsened policies derived in step 2. Existing machine learning models, such as generalized random forests (Athey, Tibshirani, and Wager, 2019) or kernel-based support vector machines (Imai and Ratkovic, 2013) can serve this role in the first step. In step 2, we search for the optimal interpretable machine learning model based on the $\hat{\Gamma}_i$ values to obtain a coarsened policy to be implemented.

Specifically, we consider the following search strategy. For a sequence of policy spaces $\{\Pi_k^m\}$, where k indexes policy complexity level and m indexes a measurement plan, we search for the optimal policy $\hat{\pi}_{k,m}^*$ satisfying:

$$\hat{\pi}_{k,m}^* = \underset{\pi \in \Pi_{k,m}}{\arg\max}\ W_n(\pi) \equiv \underset{\pi \in \Pi_{k,m}}{\arg\max}\ U_n(\pi).$$

Then, we go through all Π_k^m and find the optimal policy $\hat{\pi}_n$ and k^* such that $\hat{\pi}_n = \hat{\pi}_{k^*,m^*}^*$ and $(k^*, m^*) = \underset{k,m}{\arg\max}\ U_n(\hat{\pi}_{k,m}^*) - C_n(k) - \sqrt{\frac{k}{n}}$. The terms $C_n(k)$ and $\sqrt{\frac{k}{n}}$ are model penalization terms for controlling generalization error from using sample-based estimates (Mbakop and Tabord-Meehan, 2021; Bartlett, Boucheron, and Lugosi, 2002).

The complexity cost varies depending on the social planner's choice of policy class (or, alternatively, class of interpretable machine learning models. We use the terms synonymously).

One example policy class consists of decision trees (Breiman, Friedman, Olshen, and Stone, 2017; Miller, 2019). Another consists of linear decision rules. A decision tree model maps the covariate vector $\mathbf{X}_i$ into an action via a particular path from the root node to a leaf node. Each split in the decision tree represents an "if-else" rule: "if covariate $\mathbf{X}^{(j)} \leq c$, then left-child; else, right-child." All observations in leaf l will be assigned to the same treatment arm. Following the decision tree, which is easy to visualize, one can see who is eligible to receive treatment and why. Linear decision rules determine who should be assigned to treatment on the basis of linear combinations of features. These can be interpreted as "points" associated with each feature so that there is a threshold number of points such that individuals with points above the threshold value will be treated (Kitagawa and Tetenov, 2018). As mentioned above, Levy (2006) highlights the transparency of the linear eligibility rule in the PROGRESA program as being important to its legitimacy and sustainability. In the tree-based policy case, the complexity can be defined as the number of splits in a tree; in the linear decision rule case, the complexity can be defined as the number of features incorporated.

Once the analyst has taken care to identify complexity costs in the same units as the outcome (Mexican pesos, in our application), the complexity costs in the social planner's objective can be incorporated in ways that are equivalent to the complexity penalties used for regularization in, e.g., decision trees model (Breiman et al., 2017) or regularized linear models (Hastie, Tibshirani, Friedman, and Friedman, 2009). For example in a decision tree model we may assign a cost to the number of "if-else" rules, which is the cost for an additional split. In this case, a split would only be accepted if the welfare gains from more precise targeting exceed the cost of the additional split. This is straightforward and natural: if there are more "if-else" rules, it is harder for professionals to follow the tree (as with treatment recommendations based on complex comorbitity and drug-drug interaction considerations). Therefore, if we add one more "if-else" rule to the current policy, the new policy will only be accepted if the additional rule creates a positive net benefit. Similarly, in a linear model the

number of features used serves as the complexity measure. If there is one more feature, it increases costs due to a more complex calculation and lower transparency.

The $C_n(k)$ term is a penalty which accounts for overfitting and maintains good inferential properties in generalizing from sample to population. In this paper, we follow Bartlett et al. (2002) and use the negative maximal-discrepancy estimate as our penalty term. Other penalty terms, including hold-out penalty, are also compatible with our assumptions. The maximal-discrepancy for given policy class Π_k is defined as $\sup_{\pi \in \Pi} W_n^{(1)}(\pi) - W_n^{(2)}(\pi)$, where the superscript (1) and (2) indicate the empirical welfare evaluated on the first half and the second half of the sample, respectively. Intuitively, it captures the maximal welfare difference if a policy is applied to two independent samples from the same distribution. In the appendix, we extend to our weighted classification setting Bartlett et al. (2002)'s approach to computing maximal discrepancy in binary classification by flipping the $\hat{\Gamma}_i$ for first half of the data and re-estimating a new policy within the same policy class. The new policy's empirical welfare will be one half of the maximal discrepancy. Bartlett et al. (2002) shows that the maximal discrepancy approach has better performance in terms of error bounds compared to hold-out sample estimation. The auxiliary $\sqrt{\frac{k}{n}}$ term ensures a rate at which the aggregate penalty $(C_n(k) - \sqrt{\frac{k}{n}})$ scales in k, which in turn is used when determining a bound on the generalization error from using sample values to maximize population welfare (Mbakop and Tabord-Meehan, 2021; Bartlett et al., 2002).

The algorithm is summarized as follows:

Algorithm 1: Automatic Policy Coarsening via Maximal Discrepancy

Extract the set of covariates $\mathcal{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)$ from the sample.

Estimate the doubly robust score $\hat{\Gamma}_i$ for all i in the sample using a machine learning model satisfying our assumptions.

for $\Pi_{k,m}$ *from* $\Pi_{1,1}$ *to* $\Pi_{K,M}$ **do**

- Search for the optimal policy $\hat{\pi}_{k,m} \in \Pi_{k,m}$ using the elements of $\mathbf{X}_i$ available under measurement plan m and the $\hat{\Gamma}_i$ embedded in the estimated objective value $U_n(\hat{\pi}_{k,m})$, the sample counterpart of the objective in equation (4).
- Set $\hat{\Gamma}_i^* = -\hat{\Gamma}_i$ when $i \leq \frac{n}{2}$ and $\hat{\Gamma}_i^* = \hat{\Gamma}_i$ when $i > \frac{n}{2}$.
- Search for the optimal policy $\hat{\pi}'_{k,m} \in \Pi_k$ using the elements of $\mathbf{X}_i$ available under measurement plan m, using $\hat{\Gamma}_i^*$ in place of $\hat{\Gamma}_i$ to obtain the estimated objective value $U_n^*(\hat{\pi}'_{k,m})$.
- Obtain an objective measure $\nu_{\hat{\pi},k,m} = U_n(\hat{\pi}_{k,m}) + 2U_n^*(\hat{\pi}'_{k,m}) - \sqrt{\frac{k}{n}} - \lambda_2 g_2(\hat{\pi}_{k,m}) - \lambda_3 g_3(\hat{\pi}_{k,m})$.

Select the model with the highest objective measure $\nu_{\hat{\pi},k,m}$.

The detailed theoretical reasoning is provided in the next section.

5 Theory

In this section, we lay out the theoretical properties of Algorithm 1, focusing on how learning improves with the experimental sample size. Our main results come in the form of regret bounds on the difference between policymaker utility under the optimal policy and the utility obtained from our algorithm. Across a variety of conditions, we find that the order of our bounds matches that of bounds on welfare regret from the penalized and policy learning literature whose algorithms do not consider implementation costs. The main message is therefore that incorporating implementation costs in the optimal policy search does not substantially degrade performance with respect to welfare regret.

To set the stage for these results, recall that the social planner's objective is given by:

$$\begin{aligned} U(\pi) &= W(\pi) - \lambda_2 g_2(\pi) - \lambda_3 g_3(\pi) \\ &= E[\pi_i \Gamma_i] - \lambda_2 g_2(\pi) - \lambda_3 g_3(\pi). \end{aligned}$$

We follow Mbakop and Tabord-Meehan (2021) and construct a sequence of nested policy spaces $\Pi_1 \subset \Pi_2 \subset \cdots \subset \Pi_k \subset \cdots \subset \Pi$, where $\pi^* \in \Pi$. For example, the sequence could be made up of decision trees with increasing depth indexed by k. For ease of exposition in this section, we hold the measurement plan fixed along this sequence and suppress the m subscript. The general case where there are multiple measurement plans under consideration makes the sequence of policy spaces non-nested. As in Mbakop and Tabord-Meehan (2021) our results can be extended to this case.

Within a given policy space Π_k we have an optimal policy π_k^* which solves

$$\sup_{\pi \in \Pi_k} E[\pi_i \Gamma_i].$$

We can decompose the welfare regret arising from choosing a policy $\hat{\pi}_k \in \Pi_k$ in place of π^* as

$$W(\pi^*) - W(\hat{\pi}_k) = \underbrace{W(\pi^*) - W(\pi_k^*)}_{approximation\ error} + \underbrace{W(\pi_k^*) - W(\hat{\pi}_k)}_{estimation\ error}.$$

Following Mbakop and Tabord-Meehan (2021) and Bartlett et al. (2002) we refer to the first term as the approximation error and the second term as estimation error. The approximation error is an irreducible error coming from the complexity of Π_k relative to π^*. The estimation error comes from a combination of the choice of estimation algorithm and sampling error. The fact that the policy spaces $\{\Pi_k\}$ are nested means the estimation error is non-decreasing with respect to k. The approximation error is non-increasing with respect to k. Mbakop and Tabord-Meehan (2021) proposes the penalized welfare maximization (PWM) method to search for the optimal k, which balances the approximation error and the estimation error.

For us, however, the social planner's utility does not only consist of the welfare from treatment, but also includes implementation costs. We perform an analogous decomposition in our setting:

$$\begin{aligned}
U(\pi^*) - U(\hat{\pi}_k) =& U(\pi^*) - U(\pi_k^*) + U(\pi_k^*) - U(\hat{\pi}_k) \\
=& W(\pi^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*) - W(\pi_k^*) + \lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) \\
&+ W(\pi_k^*) - \lambda_2 g_2(\pi_k^*) - \lambda_3 g_3(\pi_k^*) - W(\hat{\pi}_k) + \lambda_2 g_2(\hat{\pi}_k) + \lambda_3 g_3(\hat{\pi}_k) \\
=& W(\pi^*) - W(\pi_k^*) + (\lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) \\
&+ W(\pi_k^*) - W(\hat{\pi}_k). \qquad (5)
\end{aligned}$$

The last equality comes from the fact that $\pi_k, \pi_k^* \in \Pi_k$ and follow the same measurement plan so they have the same measurement and the complexity costs. Note that the practical cost of π^*, $\lambda_2 g_2(\pi^*) + \lambda_3 g_3(\pi^*)$ is a constant with respect to policy choice. The elements highlighted in the decomposition (approximation error, implementation costs, and estimation error) will play key roles in our bounding arguments.

Before we proceeding to our main results, we lay out the necessary assumptions. We maintain assumptions 2 and 4 from Athey and Wager (2021), restated in our terms as assumptions 5 and 6 in the appendix, which provide the necessary assumptions for the doubly robust welfare scores. We also assume that each policy space Π_k in the sequence has bounded VC dimension.

For a given policy space Π_k, we let

$$\hat{\pi}_k = \arg\max_{\pi \in \Pi_k} W_n(\pi) \equiv \arg\max_{\pi \in \Pi_k} U_n(\pi).$$

The equivalence is due to the fact that for a fixed policy space and fixed measurement plan, the practical costs are also fixed. We now introduce a penalization term $C_n(k)$, which is intended to account for the extent to which the in-sample welfare measure $W_n(\hat{\pi})$ is upward-biased

for the population welfare measure $W(\hat{\pi})$ as a result of $\hat{\pi}_k$ having overfitted $W_n(\cdot)$. $C_n(k)$ takes k as an argument because a policy's ability to overfit is an increasing function of its complexity. We will also refer to $C_n(k)$ as a *statistical* complexity penalty, separate from the substantive complexity cost in the planner's objective function. Finally, we refer to the difference $W_n(\hat{\pi}_k) - W(\hat{\pi}_k)$ as the *generalization error* in welfare evaluation.

We require the following condition on $C_n(k)$ which, intuitively, means that the the penalty term must provide a good approximation to the generalization error.

Assumption 2 *(Assumption 3.4 in Mbakop and Tabord-Meehan (2021)) Let $\mathcal{P}$ denote the set of distributions satisfying assumptions 1, 5, and 6. There exist positive constants c_0 and c_1 such that $C_n(k)$ satisfies the following tail inequality for every n, k and for every $\epsilon > 0$:*

$$\sup_{P \in \mathcal{P}} P_n(W_n(\hat{\pi}_k) - W(\hat{\pi}_k) - C_n(k) > \epsilon) \leq c_1 e^{-2c_0 n \epsilon^2}.$$

Drawing on Bartlett et al. (2002)'s discussion of penalty terms for model selection in classification problems, in appendix A.4 we show that our implementations of hold-out and maximal discrepancy penalties satisfy assumption 2.

We now state our first theoretical result.

Theorem 5.1 *For every P satisfying assumptions 1, 5, and 6 and penalty function $C_n(k)$ satisfying assumption 2, there exist constants Δ and c_0 such that:*

$$E_{P_n}[U(\pi^*) - U(\hat{\pi}_n)] \leq \inf_k \Bigg\{ E_{P_n}[C_n(k)] + (W(\pi^*) - W(\pi_k^*)) \\ + (\lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) + \sqrt{\frac{k}{n}} \Bigg\} + \sqrt{\frac{\log(\Delta e)}{2c_0 n}} \tag{6}$$

where $\hat{\pi}_n = \hat{\pi}_{k^*}$ and $k^* = \arg\max_k U_n(\hat{\pi}_k) - C_n(k) - \sqrt{\frac{k}{n}}$.

Proof: see appendix A.1.

$E_{P_n}[\cdot]$ in the expression above refers to the expectation of a quantity over repeated draws of

samples of size n. The infimum over k between 1 and K in the bound is the result of our algorithm's searching across the policy spaces in $\{\Pi_k\}$ and selecting the policy with the lowest error. Our algorithm thus effectively trades off the three terms in the utility decomposition in equation (5) (approximation error, estimation error, and implementation cost) against one another. The theorem is analogous to Theorem 3.1 in Mbakop and Tabord-Meehan (2021), with the addition of implementation costs, and is our first theoretical result showing that our algorithm's incorporation of implementation costs does not result in fundamentally different regret properties compared to the penalized and policy learning literature.

We now discuss several ways to obtain a uniform bound for the right hand side of equation (6). For this we require that $C_n(k)$ satisfies a modified version of Assumption 3.5 in Mbakop and Tabord-Meehan (2021):

Assumption 3 *There exists a positive constant C_1 such that for every n, $C_n(k)$ satisfies*

$$\sup_{P\in\mathcal{P}} \limsup_{n\to\infty} E_{P_n}[C_n(k)] \leq C_1\sqrt{\frac{VC(\Pi_k)}{n}},$$

where $VC(\Pi_k)$ is the VC-dimension of Π_k.

The $P \in \mathcal{P}$ refer to possible states of the world (Manski, 2004). In the appendix we show that our implementation of hold-out and maximal discrepancy penalties satisfy assumption 3.

With this result we are able invoke the following corollary, analogous to Corollary 3.2 in Mbakop and Tabord-Meehan (2021).

Corollary 1 *Suppose assumptions 1, 2, 5, and 6 hold and that $\Pi = \Pi_K$ for some finite K. Then*

$$\begin{aligned}\limsup_{n\to\infty} E_{P_n}[U(\pi^*) - U(\hat{\pi}_n)] \leq &\limsup_{n\to\infty} \inf_{1\leq k\leq K} \left\{ E_{P_n}\left[C_1\sqrt{\frac{VC(\Pi_k)}{n}} + (W(\pi^*) - W(\pi_k^*))\right]\right. \\ &\left. +(\lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) + \sqrt{\frac{k}{n}}\right\} + \sqrt{\frac{\log(\Delta e)}{2c_0 n}}.\end{aligned}$$

This result is straightforward but illustrates the link between Athey and Wager (2021)'s welfare regret bound and the bound on the difference between optimal policymaker utility and that achieved by our algorithm. When $k = K$, the approximation error is equal to zero and resulting bound is $E\left[C_1\sqrt{\frac{VC(\Pi_K)}{n}}\right]$, which appears in Athey and Wager (2021)'s welfare regret bound for policies learned from experimental and observational studies, plus additional fixed terms and terms appearing in Mbakop and Tabord-Meehan (2021) generated by the search over policy spaces. If $\pi^* \in \Pi_{k_0}$ for some known k_0, then $W(\pi^*) - W(\pi^*_{k_0}) = 0$. This gives us have another bound similar to one discussed in Mbakop and Tabord-Meehan (2021).

Corollary 2 *Suppose assumptions 1, 2, 5, and 6 hold, then*

$$\begin{aligned}
&\limsup_{n\to\infty} E_{P_n}[U(\pi^*) - U(\hat{\pi}_n)] \\
&\leq \limsup_{n\to\infty}\ C_1\sqrt{\frac{VC(\Pi_k)}{n}} + (\lambda_2 g_2(\pi^*_k) + \lambda_3 g_3(\pi^*_k) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) + \sqrt{\frac{k}{n}} + \sqrt{\frac{\log(K\Delta e)}{2c_0 n}}.
\end{aligned}$$

Finally we provide a bound which does not rely on knowledge of the approximation error $W(\pi^*) - W(\pi^*_k)$, or conditions under which it is equal to zero. To do so, we must choose $C_n(k)$ and derive a uniform bound on $W(\pi^*) - W(\pi^*_k)$. As discussed above, both the hold-out penalty and the maximal discrepancy penalty satisfy assumption 2 and are thus appropriate for our purposes. For the uniform bound on the approximation error $W(\pi^*) - W(\pi^*_k)$, we conduct a decomposition of $E_{P_n}[U(\pi^*) - U(\hat{\pi}_k)]$.

$$\begin{aligned}
E_{P_n}[U(\pi^*) - U(\hat{\pi}_k)] &= E_{P_n}[U(\pi^*) - U(\hat{\pi}^*) + U(\hat{\pi}^*) - U(\hat{\pi}_k)] \\
&= E_{P_n}[\underbrace{(W(\pi^*) - W(\hat{\pi}^*))}_{A} + \underbrace{(W(\hat{\pi}^*) - W(\hat{\pi}_k))}_{B}] \\
&+ \lambda_2 g_2(\pi_k) + \lambda_3 g_3(\pi_k) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)
\end{aligned}$$

where $\hat{\pi}^*$ is the optimal in-sample policy in Π. We further make an assumption on the ultimate policy space Π, following Athey and Wager (2021).

Assumption 4 *We assume there are constants* $0 < \beta < min(\zeta_m, \zeta_g)$ *and* $N^* \geq 1$ *such that the VC-dimension of policy class* Π *is bounded:* $VC(\Pi) \leq n^\beta$ *for all* $n \geq N^*$.

This assumption guarantees a bound for term A, the regret from using $\hat{\pi}^*$ in place of π^* which has the form $A \leq C\sqrt{\frac{VC(\Pi)}{n}}$. Because both $\hat{\pi}^*$ and π^* are elements of Π, and $\hat{\pi}^*$ is the empirical maximizer, we can invoke theorem 1 in Athey and Wager (2021) when further maintaining assumption 8. Term B is bounded by the weighted misclassification error of $\hat{\pi}_k$ with respect to $\hat{\pi}^*$. To see this, note that

$$\hat{\pi}^*(X_i)\Gamma_i - \hat{\pi}_k(X_i)\Gamma_i = \begin{cases} |\Gamma_i| & \text{if } \hat{\pi}_k = 1, \hat{\pi}^* = 0 \text{ and } \pi^* = 0 \\ |\Gamma_i| & \text{if } \hat{\pi}_k = 0, \hat{\pi}^* = 1 \text{ and } \pi^* = 1 \\ -|\Gamma_i| & \text{if } \hat{\pi}_k = 1, \hat{\pi}^* = 0 \text{ and } \pi^* = 1 \\ -|\Gamma_i| & \text{if } \hat{\pi}_k = 0, \hat{\pi}^* = 1 \text{ and } \pi^* = 0 \\ 0 & \text{otherwise} \end{cases}$$

Weighting the misclassification error by $|\Gamma_i|$, the absolute value of the doubly robust score, we have that $\mathbb{1}\{\hat{\pi}_k(X_i) \neq \hat{\pi}^*(X_i)\}|\Gamma_i| = |\Gamma_i|$ when they disagree. Therefore, we have

$$E_{P_n}[\mathbb{1}\{\hat{\pi}_k(X_i) \neq \hat{\pi}^*(X_i)\}|\Gamma_i|] \geq E_{P_n}[W(\hat{\pi}^*) - W(\hat{\pi}_k)].$$

Lemma 4 in Athey and Wager (2021) shows that we can use $\hat{\Gamma}_i$ in place of Γ_i and the difference between $E_{P_n}[\mathbb{1}\{\hat{\pi}_k(X_i) \neq \hat{\pi}^*(X_i)\}|\Gamma_i|]$ and $E_{P_n}[\mathbb{1}\{\hat{\pi}_k(X_i) \neq \hat{\pi}^*(X_i)\}|\hat{\Gamma}_i|]$ is bounded by an $O\left(\sqrt{\frac{VC(\Pi_k)n}{n^{1+\beta}}}\right)$ term.

Results from the statistical learning literature lead to a general result bounding misclassification error.

Lemma 5.1 *For binary classification algorithm class* Π_k *with finite VC-dimension* $VC(\Pi_k)$ *and the target being* π^**, for any* $\delta > 0$*, with probability at least* $1 - \delta$*, the following holds for*

all $\pi_k \in \Pi_K$*:*

$$\limsup_{n\to\infty} L(\pi_k) \leq \hat{L}(\pi_k) + 2\zeta\sqrt{\frac{VC(\Pi)\log\frac{en}{VC(\Pi)}}{n}} + \sqrt{\frac{\kappa}{n}\log\left(\frac{2}{\delta}\right)}$$

where $\hat{L}(\pi_k) = E_n[\mathbb{1}\{\pi_k(X_i) \neq \pi^*(X_i)\}|\hat{\Gamma}_i|]$*and* $L(\pi_k) = E_{P_n}[\hat{L}(\pi_k)]$ *, and* ζ *and* κ *are constants.*

Proof: see appendix A.2.

This result combined with the bound on term A provides a bound for $E[U(\pi^*) - U(\hat{\pi}_k)]$ for given k.

Theorem 5.2 *Suppose assumptions 1, 2, 4, 5, and 6 hold and that* $\Pi = \Pi_K$ *for some finite* K*, we have with probability* $1-\delta$

$$\limsup_{n\to\infty} E_{P_n}[U(\pi^*) - U(\hat{\pi}_n)] \leq \limsup_{n\to\infty} \inf_{1\leq k\leq K} \left\{ \hat{L}(\hat{\pi}_k) + 2\zeta\sqrt{\frac{VC(\Pi)\log\frac{en}{VC(\Pi)}}{n}} + \sqrt{\frac{\kappa}{n}\log\left(\frac{2}{\delta}\right)} + C\sqrt{\frac{VC(\Pi)}{n}} \right.$$
$$\left. + (\lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) + \sqrt{\frac{k}{n}} \right\}$$

where $\hat{\pi}_n$ *is given in the statement of theorem 5.1.*

Proof: see appendix A.3.

Note that when $VC(\Pi)$ is large, the value and the convergence rate of the bound is governed by $C\sqrt{\frac{VC(\Pi)}{n}}$. It shows that the regret for our practical policy has the same order as if we were searching for the optimal policy only in the largest policy space using the method introduced in Athey and Wager (2021).

Our results altogether imply that: (1) our algorithm yields a policy with the same convergence rate as existing EWM methods when the planner has information on the optimal policy class (corollaries 1 and 2); (2) without such information and when the complexity of the optimal policy does not grow too fast with the sample size, the welfare of the practical policy is still bounded and achieves regret of the same order as the most complex policy space considered.

6 Application

We now illustrate the value of our coarsened policy learning algorithm by applying it in the context of Mexico's PROGRESA conditional cash transfer program. We will show that our algorithm can appropriately balance the trade-off between policy benefits and the social planner's implementation costs.

From 1997 to 2019 PROGRESA[2] was the centerpiece of the Mexican government's policy agenda aimed at poverty alleviation and developing the human capital of poor households. PROGRESA worked to alleviate current and future poverty levels through cash transfers mainly given to mothers in households. The largest portion of the cash transfers were conditioned on children's regular school attendance. Compared to previous poverty alleviation programs in Mexico, PROGRESA was novel in targeting eligibility at the household level and delivering benefits electronically in cash rather than in-kind, with the intention of providing resources directly to the poor households who would most benefit from the program (Skoufias, 2005).

PROGRESA was rolled out as a randomized controlled trial, in which 504 villages were partitioned into treatment and control groups, with treatment communities providing benefits to eligible households satisfying the prescribed conditions. This experimental design allows for the identification of the welfare effects of the program. We use the experimental data to derive optimal targeting policies.

6.1 Welfare

We use Gechter et al. (2018)'s (henceforth GSDP) welfare outcome as our Y_i. GSDP's outcome allows for an evaluation of the redistributive benefits of PROGRESA by comparing 1) the upside to transfer recipients against 2) the downside to those taxed to pay for the transfers. To arrive at 1), the net upside for transfer recipients, GSDP's welfare outcome compares the cash benefits with the cost of complying with the condition that a child be

[2]Later renamed Oportunidades and then Prospera under subsequent central government administrations.

enrolled in school. This cost could be explicit, like when the child would otherwise work for wages, or implicit, when the child would help with care of younger siblings or on the family farm. Since the child's counterfactual use of time when enrolled is unknown, GSDP take a conservative approach to valuing program benefits and assume families who would not enroll a child without PROGRESA need the entirety of the transfer to compensate them for the child's alternative use of time. The result is that the net benefit from the cash only accrues to households who would enroll children even without the program. For 2), GSDP consider the transfer program to be financed by a uniform income tax. The per-child cost is the average size of the grant provided to children who enroll when eligible for PROGRESA, net of any savings to the government provided by the higher lifetime income tax revenue of individuals who complete an additional year of education because of PROGRESA.

Putting 1) and 2) together, the motive for redistributing from taxpayers to program recipients arises because increments/decrements to each's income are valued according to a function with constant elasticity of substitution in income across individuals in the population (Atkinson, 1970). The function has diminishing marginal returns to each individual's income (governed by the parameter σ) so the overall value of the function increases when income is transferred from the average taxpayer to the poor. GSDP identify σ by assuming the break-even point of social welfare occurs at the Mexican poverty line, as intended by the program designers (Skoufias et al., 1999). We use GSDP's estimated value of 0.33. Income taxation is an imperfect instrument for this transfer, however, since some taxpayers will reduce the amount of labor they supply to the market in response to a lower amount of takehome pay. By moving taxpayers away from their unconstrained optimal behavior, taxation imposes an additional cost on them which economists refer to as the deadweight loss of taxation. GSDP use the benchmark figure of a 30% additional cost to taxpayers beyond the average value of transfers claimed (Finkelstein and Hendren, 2020).

Taking a first order approximation to the redistribution, child i's welfare contribution if

the targeting policy assigns them to treatment (eligibility for PROGRESA) is

$$\underbrace{g(ed_i, male_i)\left(\frac{\bar{N}}{N_i}\right)^{\sigma} enroll_i(0)}_{welfare\ benefit} \tag{7}$$

$$-\underbrace{1.3\Omega\left[g(ed_i, male_i)enroll_i(1) - T(enroll_i(1), age_i, ed_i, male_i) + T(enroll_i(0), age_i, ed_i, male_i)\right]}_{cost\ net\ of\ savings}.$$

$g(ed, male)$ is the size of the grant a child having completed ed years of education is entitled to depending on whether the $male$ indicator is zero or one. $\bar{N}$ is the average per capita income earned by adults across Mexican households, N_i is i's parents' income. Ω is the average of $\left(\frac{\bar{N}}{N_i}\right)^{\sigma}$ across Mexican households (taxpayers). $T(enroll, age, ed, male)$ measures the present discounted value of i's lifetime income tax contributions based on whether she enrolls for an additional year of school given her age, years of education already completed, and gender. Finally $enroll_i(1)$ and $enroll_i(0)$ represent i's potential enrollment with and without eligibility for PROGRESA, respectively. i's welfare contribution is normalized to zero if the targeting policy does not assign her to receive treatment.

Equation (7) translates to defining the outcome of interest for the application as

$$\begin{aligned} Y_i =& 1.3\Omega D_i\left[T(enroll_i, age_i, ed_i, male_i) - g(ed_i, male_i)enroll_i\right] \\ &+ (1 - D_i)[g(ed_i, male_i)\left(\frac{\bar{N}}{N_i}\right)^{\sigma} enroll_i - 1.3\Omega T(enroll_i, age_i, ed_i, male_i)]. \end{aligned}$$

Treatment effects on this outcome can be interpreted as equivalent to per-eligible-child increases in the adult income of an average Mexican household since for such a household $\left(\frac{\bar{N}}{N_i}\right)^{\sigma} = 1$.

6.2 Identifying the Complexity Cost

The costs of complexity, while acknowledged as important in the ways stated above, were not numerically articulated by PROGRESA's designers. We therefore use a revealed preference

argument to identify λ_3 in the policymaker's objective function specified in equation (4). We model PROGRESA's designers as choosing between the following three candidate policies.

- $M1$: A simple linear treatment rule with PROGRESA's locality poverty index as the only feature. The poverty index is costless for the policymaker to obtain since it was derived from census and administrative data collected by the government for other reasons. To construct the index, PROGRESA used community information such as access to schools, share of illiterate adults, share of dwellings missing key assets, number of occupants per room, and the share of the population working in agriculture. Program designers then inputted the community information to a Principal Components Analysis (PCA) model, extracting the first principle component as the poverty index.

- $M2$: A linear treatment rule using the poverty index, the costly variables ultimately used for eligibility determination, and their interaction with the poverty index. The variables used to determine eligibility are described in Coady et al. (2013) and include household and individual level characteristics such as home amenity possession and children's ages. We list all of these variables in appendix B. The interaction mimics PROGRESA's two-stage eligibility rule, which first classified a locality as poor or non-poor according to the poverty index, then used individual characteristics to determine eligibility in poor localities.

- $M3$: A linear treatment rule using the same variables as $M2$ but including the interactions of all variables with "household head without education", which has been shown by GSDP to be associated with treatment effect heterogeneity.

The three treatment rules differ in the welfare gains they provide, their measurement cost, and their complexity cost. In terms of measurement cost, Skoufias et al. (1999) report that the variables needed to determine eligibility required administration of a minimal household survey costing an average of 60 Mexican pesos per household. We convert this to a per-child

measure by dividing by the average number of children per household in the PROGRESA experimental sample, yielding a measurement cost of 13.69 pesos for this set of costly variables.

PROGRESA's designers finally chose $M2$ as their eligibility rule so we know $M2 \succeq M1$ and $M2 \succeq M3$. This means that

$$W_n(M1) - \lambda_3 g_3(M1) \leq W_n(M2) - 13.69 - \lambda_3 g_3(M2) \tag{8}$$

$$W_n(M3) - \lambda_3 g_3(M3) \leq W_n(M2) - \lambda_3 g_3(M2). \tag{9}$$

The measurement cost features in the first line because we consider the poverty index to be costless since it was produced from independently-collected census data. Both M_2 and M_3, in contrast, involve the costly variables PROGRESA actually used to determine eligibility. By our construction, we have $g_3(M3) > g_3(M2) > g_3(M1)$. We set the $g_3(\cdot)$ function to measure the number of additive terms in the linear treatment rules. Hence, we have $g_3(M2) - g_3(M1) = 24$ and $g_3(M3) - g_3(M2) = 12$. By estimating the welfare associated with each method, the two inequalities from (8) involve only λ_3, and together bound the complexity cost.

To arrive at the welfare estimate for $M1$ we need to consider the welfare associated with treating individuals in selected communities who were not eligible for PROGRESA in the initial program evaluation. To do so, we follow Skoufias et al. (1999)'s approach assuming a) the Mexican government's objective during the design phase did not involve any treatment effects and b) that enrollment among all treated 8 - 18 year olds would be universal so that Equation (7) can be rewritten as

$$\begin{aligned} & g(ed_i, male_i) \left(\frac{\bar{N}}{N_i}\right)^{\sigma} enroll_i(0) \\ & - 1.3\Omega \left[g(ed_i, male_i) - T(1, age_i, ed_i, male_i) + T(enroll_i(0), age_i, ed_i, male_i)\right]. \end{aligned}$$

Since the authors of Skoufias et al. (1999) were involved in the planning phases of the

PROGRESA evaluation, we consider this a reasonable approximation to the PROGRESA planners' thinking prior to receiving the experimental results, justifying the revealed preference inequalities above.

Inserting our estimates, $W_n(M1)$, $W_n(M2)$, and $W_n(M3)$ into inequalities (8), we obtain

$$\lambda_3(g_3(M2) - g_3(M1)) \leq 15.49$$
$$\lambda_3(g_3(M3) - g_3(M2)) \geq 0.02.$$

This yields our final bounds on λ_3, $0.002 \leq \lambda_3 \leq 0.645$. We take the midpoint - 0.324 pesos - as our complexity cost.

6.3 Measurement Costs

Moving to the design of *our* policies which *do* make use of the results of the PROGRESA evaluation, we consider three possible measurement plans:

1. A plan using only the costless locality poverty index,

2. A plan using the poverty index and additionally collecting the variables used by PROGRESA to determine eligibility,

3. A plan collecting the same household survey used in the PROGRESA evaluation, now in the broader population. The evaluation survey went well beyond the minimal survey needed to determine PROGRESA eligibility and gives the planner access to an additional set of variables which can be used for targeting. Skoufias et al. (1999) report that the cost of running the full household survey is 170 pesos per household on average, so an increase of 23.679 pesos relative to the eligibility-only survey when divided by average number of children per household.

We note here that only eligible individuals (according to the rule PROGRESA actually used) could receive transfers in PROGRESA's experimental evaluation. Therefore we can

only evaluate treatment effect heterogeneity in the PROGRESA experiment and perform treatment assignment for subsets of this group. This is always the case with statistical treatment assignment: the thought experiments entertained in the literature are necessarily about rendering *ineligible* some members of a target population, never expanding eligibility. So we consider any treatment assignment rule except one that withholds treatment from everyone to be using either the second or third measurement plan, since it involves selecting subsets of the original eligible children for treatment and therefore must pay the cost of collecting the variables needed to employ PROGRESA's original eligibility rule.

6.4 Results

The first step of our algorithm is to use a high dimensional model to estimate the doubly-robust welfare scores and provide an initial policy. We estimate the welfare treatment effect $\hat{\Gamma}_i$ scores using the generalized random forest algorithm (*grf*) developed by Athey et al. (2019). We include all variables in the third measurement plan in estimation. The variable importance function for the *causal_forest* function in the *grf* package shows that the main factors affecting the individual $\hat{\Gamma}_i$ scores are children's years of education, age, the locality poverty index (*indice*), and household head's age. As shown in Figure 1, we find that the poverty index can provide as much information as children's age and education level do.

To get a sense of the variation in the estimated welfare scores, we show their distribution and estimated conditional mean over children's age and the locality poverty index in the left and right panels of Figure 2. Figure 2 shows that although both age and the locality poverty index both explain treatment effect heterogeneity, age alone may not provide enough information to determine treatment assignment because the average score conditional on age always exceeds zero. In contrast, the mean welfare score conditional on locality poverty index does dip below zero for individuals in the richest communities included in the PROGRESA experiment.

For step 2, we consider a variety of linear and tree-based policy rules as our coarsening

Figure 1: Variable Importance: Main Factors Affecting PROGRESA Treatment Effects

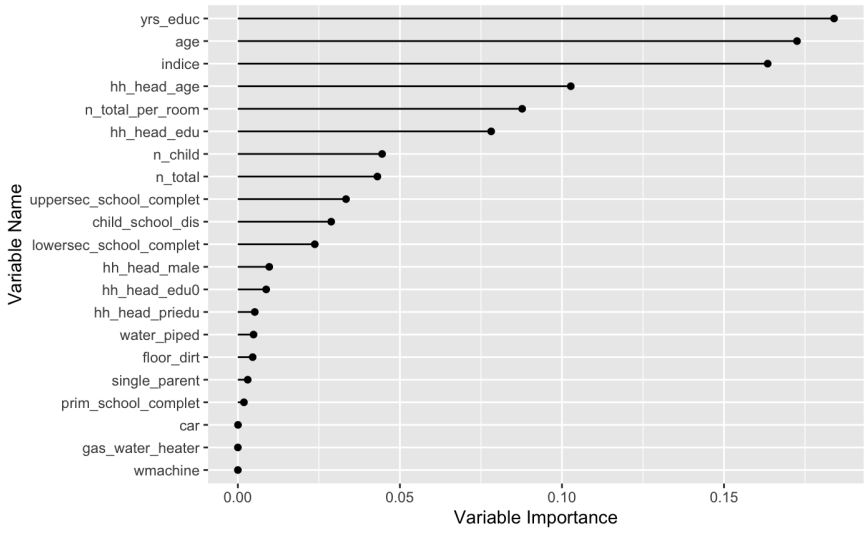
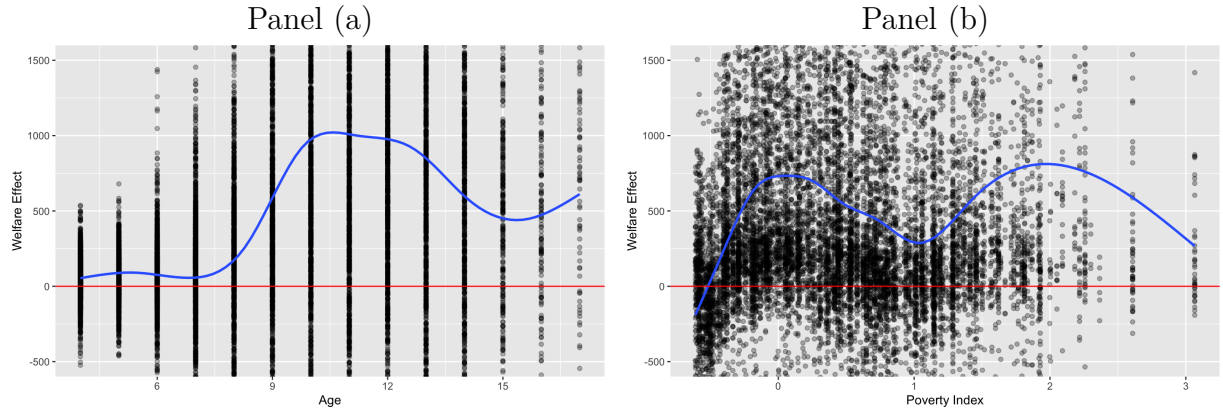


Figure 2: Treatment Effect Heterogeneity and Policy Assignment



strategy. Linear treatment rules, as per Kitagawa and Tetenov (2018), have the following form:

$$\Pi_{linear} \equiv \{\{x \in \mathbb{R}^p : \beta_0 + \mathbf{x}'_{(k)}\beta_{(k)} > 0\} : (\beta_0, \beta_{(k)}) \in \mathbb{R}^{k+1} \text{ and } k \leq p\},$$

where $\mathbf{x}_{(k)} \in \mathbb{R}^k$ is the subvector of terms in $\mathbf{X}$ used in the scoring rule. This could include any covariates in $\mathbf{X}_{costless}$ and $\mathbf{X}_{costly}$. We first consider a policy including the variables from measurement plan 3, and their interaction with the poverty index to mimic PROGRESA's original two-stage eligibility rule. This policy represents the case where the social planner makes a policy without considering measurement or complexity costs. Then, we present a

coarsened policy using only the variables available from measurement plan 2.

Table 1 details the welfare impact of each policy, the estimated proportion of individuals assigned to treatment, the practical costs incurred under each rule, as well as the social planner's utility, summing the welfare benefits and practical costs. We also report the welfare of treat-everyone and treat-no-one policies as benchmarks in the first two rows. When the social planner decides to treat everyone, retaining PROGRESA's original eligibility criteria, the per-eligible-child increment to social welfare is equivalent to a 492.376 pesos (53.87 contemporary USD) increase in the annual adult income of an average Mexican household at the time. Treating all those originally determined to be eligible for PROGRESA requires the use of measurement plan 2. Subtracting the cost of this measurement plan, the treat-all policy yields the final social planner utility of $492.376 - 13.69 = 478.686$. The treat-no-one policy, the only policy we consider which does not require a household survey, has no measurement cost but also yields zero welfare.

Row 3 in Table 1 presents the results of using a linear treatment rule taking advantage of the full list of covariates in plan 3. This policy uses all available information and therefore should target the population relatively (in the class of linear rules) precisely. It assigns treatment to 94.9% of the originally-eligible population, and has high welfare per assigned child: 497.271 pesos. However, using all the information from the full PROGRESA evaluation means incurring additional measurement costs along with additional complexity costs, which ends up with the total social planner's utility being 445.97 pesos per eligible child. In contrast, row 4 of the table shows results from a linear treatment rule using only the variables from measurement plan 2. Here the social planner may lose some precision in targeting by using only the minimal set of variables needed to determine eligibility under PROGRESA's original targeting scheme. The policy suggests that we should assign treatment to 96.6% of originally-eligible children and its per-assigned-child welfare is a lower 492.414 pesos. However, the social planner does not pay the additional measurement cost, which leads to the total utility being 470.624 pesos including complexity costs. The comparison shows that

the additional information available when using measurement plan 3 does modestly improve targeting precision but ultimately reduces the planner's objective. In fact the treat-all policy, a linear treatment rule with zero terms, achieves the highest social planner utility.

We next consider a set of classification tree-based policies:

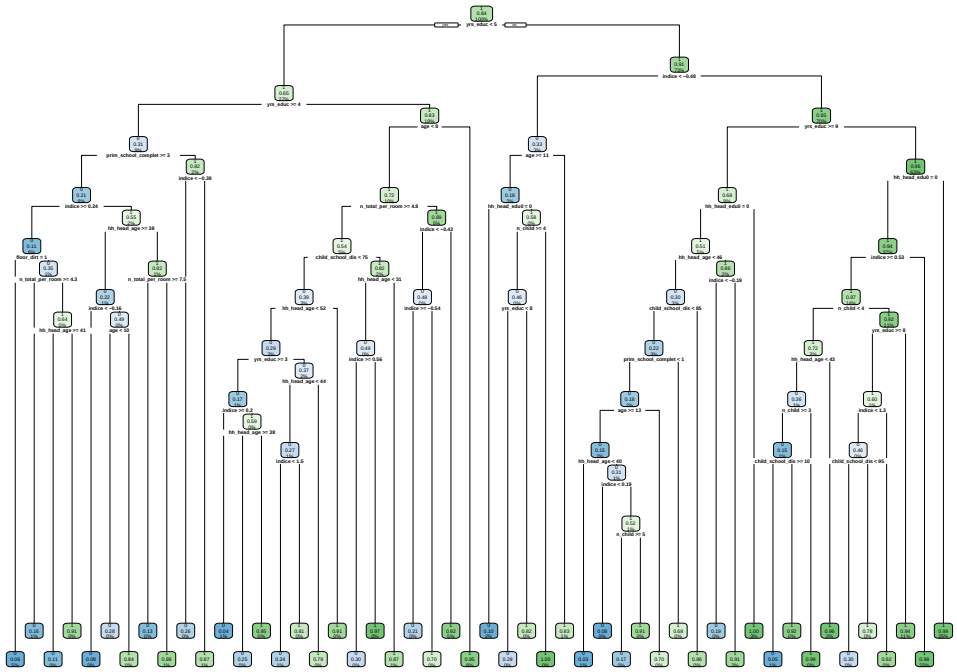
$$\Pi_{tree} \equiv \{\{x \in \mathbb{R}^p : \mu(x; \mathcal{T}_{(k)}) = 1\} : k \leq \mathcal{D}(P)\},$$

where k is the tree depth, with maximum depth $\mathcal{D}(P)$ that depends on the population covariate distribution, and $\mu(x; \mathcal{T}_{(k)})$ is the value that a k-depth classification tree $\mathcal{T}_{(k)}$ returns for input value x. First, we implement a classification-tree-based algorithm without considering measurement or complexity costs. The results are shown in rows 5 and 6 of Table 1. Row 5 uses all variables available under measurement plan 3. It allocates treatment to 78.3% of originally-eligible children and achieves the highest welfare among policies we consider: 554.257 pesos per assigned child. After subtracting measurement and complexity costs from welfare, the planner's utility under this policy is 501.336 pesos. The tree policy using only variables available under measurement plan 2, shown in row 6 of the table, has lower policy maker utility at 484.571 pesos per assigned child. In tree-based policies which do not take into account implementation costs, the benefits of undertaking the full PROGRESA evaluation survey rather than the minimal eligibility survey outweigh the costs.

The tree using variables from measurement plan 3 grown without concern for implementation costs (from Table 1 row 5) is presented in Figure 3. The policy has a visibly complex structure. We now proceed to including implementation costs in the policy search to obtain a coarsened policy, as proposed in this paper.

We incorporate complexity costs by increasing the complexity parameter of the tree-search algorithm. The complexity parameter in tree-search algorithms such as *rpart* (Therneau and Atkinson, 2022) is used to prune and restrict the size of the tree. It prevents further splitting unless the percent improvement provided by adding a split to the tree is larger than the

Figure 3: Tree-based Policy without Considering Implementation Costs

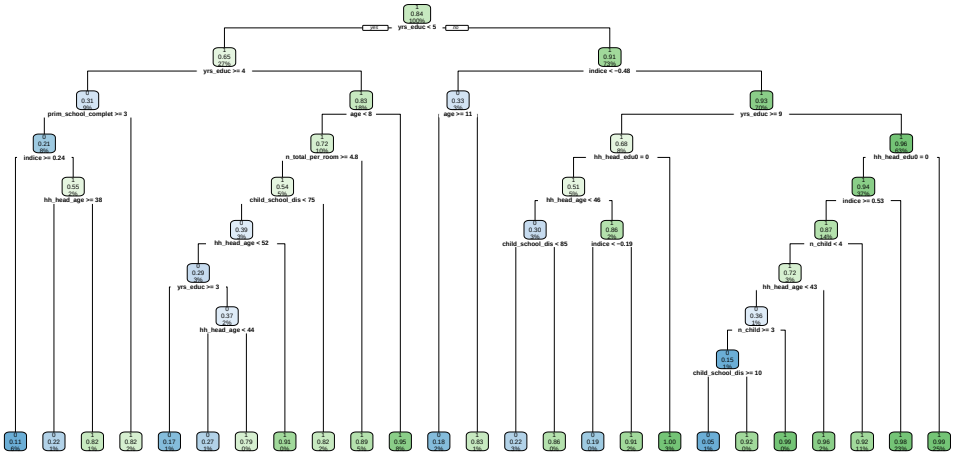


value of the complexity parameter. Because we adopt a weighted-classification algorithm, where the weight is the peso-valued individual welfare treatment effect estimate $\hat{\Gamma}_i$, the classification error improvement also has a monetary value. Therefore, we add the value of the complexity cost, 0.324 pesos, divided by 106.279, the initial weighted classification error, to arrive at the percent improvement adding a split to the policy must provide to exceed the *substantive* complexity cost. We then add this figure to the cross-validated *statistical* complexity parameter.

Figure 4 shows the coarsened tree after considering both measurement and complexity costs, with results detailed in row 6 of Table 1. The coarsened tree allocates treatment to 79.4% of originally-eligible children. The policy uses variables exclusive to measurement plan 3 and therefore incurs the maximum measurement cost. Because we use a larger complexity parameter, the tree is simpler and the planner pays a lower total complexity cost. The welfare value attained by this policy, 550.298 pesos per assigned child, is lower than that associated with the tree grown with only a statistical complexity penalty due to a loss of

targeting precision. However, our coarsened tree achieves higher total utility (505.477 pesos per assigned child), after taking into account complexity costs. We visualize the coarsened policy in Figure 4 below.

Figure 4: Coarsened Tree for Policy Assignment



As a social planner, this result indicates that at least for PROGRESA, gathering more information is helpful in the case of a tree-based policy. Although collecting information implies paying extra measurement costs, it helps for designing a more precise policy which brings more social welfare. However, such gains were not attained for a linear treatment rule.

Table 1: Estimated Welfare Gains

Policy Rule	Outcome: Adjusted Net Welfare					
	Share of population being treated	Welfare	Measurement cost	Number of selected plan 3 variables ∉ plan 2	Complexity cost	Total planner utility
Treat All Policy	1	492.376	13.69	0	0	478.686
Treat No-one Policy	0	0	0	0	0	0
Linear Treatment Rule with Full set of Variables	0.949	497.271	37.369	9	13.932	445.97
Linear Treatment Rule with Only Original Eligibility Variables	0.966	492.414	13.69	0	8.1	470.624
Tree with Full set of Variables	0.783	554.257	37.369	4	15.552	501.336
Tree with with Only Original Eligibility Variables	0.812	512.517	13.69	0	14.256	484.571
Tree with Cost Terms	0.794	550.298	37.369	4	7.452	505.477

Notes: The welfare is evaluated on a separate test set, net of measurement and complexity cost.

7 Conclusion

The expansion of large-scale randomized experiments brings opportunities for practitioners and researchers to analyze and design better policies. With individual-level data, one can design an individualized policy to maximize social welfare rather than a uniform policy for the entire population. However, any policy that tries to use targeting to increase social welfare also comes with implementation costs. The cost to a social planner consists of the cost of the treatment (e.g. the cash transfer given to individual households), cost of gathering data, and cost of complexity.

An unrestricted individualized policy can be very complex and require detailed, costly-to-acquire data. The previous literature in statistical policy design addresses the issue by constraining the policy space according to a fixed budget constraint (Kitagawa and Tetenov, 2018; Athey and Wager, 2021), or estimated budget constraint (Sun, 2021) which holds asymptotically. In this paper, we proposed a method with a flexible cost-benefit trade-off incorporating costs for measurement and complexity. Our aim in this paper has been threefold: (1) to endogenize policy complexity to trade-offs between gains to recipients versus costs of implementation, (2) to establish how learning improves with increasing experimental sample size even with this flexibility regarding the policy class, and (3) to demonstrate methods for determining implementation costs and using a cost-sensitive algorithm for learning an optimal policy.

We proposed a two-step algorithm to optimally coarsen a high-dimensional policy to account for practical costs. We showed that the algorithm provides a treatment regime with a regret bound that tightens in available experimental information. The result is a "practical" treatment policy that, by construction, trades potential treatment effect gains from highly granular targeting for gains in terms of practical implementation. We showed the policy selected by the proposed algorithm is sensitive to the policy class (e.g., linear versus tree-based policy rules) and the trade-off between welfare gains for recipients and implementation costs.

Data Availability

Upon publication, data and code to reproduce the empirical results in this paper will be available through the Harvard Dataverse under the authors' names.

Funding Statement

This work did not involve any funding from outside the authors' home institutions.

References

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

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

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory 2*(3), 244–263.

Bartlett, P. L., S. Boucheron, and G. Lugosi (2002). Model selection and error estimation. *Machine Learning 48*(1), 85–113.

Bartlett, P. L., O. Bousquet, and S. Mendelson (2005). Local rademacher complexities. *The Annals of Statistics 33*(4), 1497–1537.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone (2017). *Classification and regression trees*. Routledge.

Cai, H., W. Lu, R. M. West, D. V. Mehrotra, and L. Huang (2021). Capital: Optimal subgroup identification via constrained policy tree search. *arXiv preprint arXiv:2110.05636*.

Chen, L.-Y. and S. Lee (2018). Best subset binary prediction. *Journal of Econometrics 206*(1), 39–56.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india. Technical report, National Bureau of Economic Research.

Coady, D., C. Martinelli, and S. W. Parker (2013). Information and participation in social programs. *the world bank economic review 27*(1), 149–170.

Dehejia, R. H. (2005). Program evaluation as a decision problem. *Journal of Econometrics 125*(1-2), 141–173.

Devroye, L., L. Gyorfi, and G. Lugosi (1996). *A Probabilistic Theory of Pattern Recognition*. New York: Springer.

Finkelstein, A. and N. Hendren (2020). Welfare Analysis Meets Causal Inference. *Journal of Economic Perspectives 34*(4), 146–67.

Foster, J. C., J. M. Taylor, N. Kaciroti, and B. Nan (2015). Simple subgroup approximations to optimal treatment regimes from randomized clinical trial data. *Biostatistics 16*(2), 368–382.

Foster, J. C., J. M. Taylor, and S. J. Ruberg (2011). Subgroup identification from randomized clinical trial data. *Statistics in medicine 30*(24), 2867–2880.

Gechter, M., C. Samii, R. Dehejia, and C. Pop-Eleches (2018). Evaluating ex ante counterfactual predictions using ex post causal inference. *arXiv preprint arXiv:1806.07016*.

Hastie, T., R. Tibshirani, J. H. Friedman, and J. H. Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*, Volume 2. Springer.

Haushofer, J., P. Niehaus, C. Paramo, E. Miguel, and M. W. Walker (2022). Targeting impact versus deprivation. Technical report, National Bureau of Economic Research.

Hirano, K. and J. R. Porter (2009). Asymptotics for statistical treatment rules. *Econometrica 77*(5), 1683–1701.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association 81*(396), 945–960.

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

Kallus, N. (2017). Recursive partitioning for personalization using observational data. In *International conference on machine learning*, pp. 1789–1798. PMLR.

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

Kitagawa, T. and A. Tetenov (2021). Equality-minded treatment choice. *Journal of Business & Economic Statistics 39*(2), 561–574.

Laber, E. B. and Y.-Q. Zhao (2015). Tree-based methods for individualized treatment regimes. *Biometrika 102*(3), 501–514.

Lakkaraju, H. and C. Rudin (2017). Learning cost-effective and interpretable treatment regimes. In *Artificial intelligence and statistics*, pp. 166–175. PMLR.

Levy, S. (2006). *Progress against poverty: sustaining Mexico's Progresa-Oportunidades program*. Washington, DC: Brookings Institution Press.

Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica 72*(4), 1221–1246.

Mbakop, E. and M. Tabord-Meehan (2021). Model selection for treatment choice: Penalized welfare maximization. *Econometrica 89*(2), 825–848.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence 267*, 1–38.

Mohri, M., A. Rostamizadeh, and A. Talwalkar (2018). *Foundations of machine learning*. MIT press.

Qian, M. and S. A. Murphy (2011). Performance guarantees for individualized treatment rules. *Annals of statistics 39*(2), 1180.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence 1*(5), 206–215.

Rudin, C., C. Chen, Z. Chen, H. Huang, L. Semenova, and C. Zhong (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys 16*, 1–85.

Skoufias, E. (2005). *PROGRESA and its impacts on the welfare of rural households in Mexico*, Volume 139. Intl Food Policy Res Inst.

Skoufias, E., B. Davis, and J. Behrman (1999). An evaluation of the selection of beneficiary households in the education, health, and nutrition program (progresa) of mexico. *International Food Policy Research Institute, Washington, DC*.

Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics 151*(1), 70–81.

Sun, L. (2021). Empirical welfare maximization with constraints. *arXiv preprint arXiv:2103.15298*.

Therneau, T. and B. Atkinson (2022). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1.16.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, Volume 47. Cambridge university press.

Zhang, B., A. A. Tsiatis, M. Davidian, M. Zhang, and E. Laber (2012). Estimating optimal treatment regimes from a classification perspective. *Stat 1*(1), 103–114.

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

Zhou, Z., S. Athey, and S. Wager (2022). Offline multi-action policy learning: Generalization and optimization. *Operations Research*.

A Appendix

Technical Assumptions

Throughout our paper, we maintain several assumptions from Athey and Wager (2021). We provide the assumptions we require here.

Assumption 5 *(Assumption 2 in Athey and Wager, 2021) In the setting of equation (2), assume that second moments are controlled as* $E_n[m_n^2(\mathbf{X}_i, D_i)], E_n[(m_n(\mathbf{X}_i, 1) - m_n(\mathbf{X}_i, 0))^2] < \infty$ *and* $E_n[e_n^2(\mathbf{X}_i)] < \infty$ *for all* $n = 1, 2, \ldots,$ *and that we have access to uniformly consistent estimators of these nuisance components,*

$$\sup_{x,d}(|\hat{m}_n(x,d) - m_n(x,d)|), \sup_{x}(|(\hat{m}_n(x,1) - \hat{m}_n(x,0)) - (m_n(x,1) - m_n(x,0))|) \xrightarrow{p} 0,$$

$$\sup_{x}(|\hat{e}_n(x) - e_n(x)|) \xrightarrow{p} 0.$$

The L_2 *errors of each component decay as follows. For some* $0 < \zeta_m, \zeta_g < 1$ *with* $\zeta_m + \zeta_g \geq 1$ *and some* $a(n) \to 0$*, where* $(\mathbf{X}_i, D_i)$ *is taken to be an independent test example drawn from the same distribution as the training data*

$$E[(\hat{m}_n(\mathbf{X}_i, D_i) - m_n(\mathbf{X}_i, D_i))^2], E[((\hat{m}_n(\mathbf{X}_i, 1) - \hat{m}_n(\mathbf{X}_i, 0)) - (m_n(\mathbf{X}_i, 1) - m_n(\mathbf{X}_i, 0))^2] \leq \frac{a(n)}{n^{\zeta_m}},$$
$$E[(\hat{e}(\mathbf{X}_i) - e_n(\mathbf{X}_i)^2] \leq \frac{a(n)}{n^{\zeta_g}}.$$

The subscript n *in functions like* $m_n(x,d) = E_n[Y_i(d)|X_i = x]$ *allows problem-specific quantities to depend on the sample size.*

Assumption 6 *(Assumption 4 in Athey and Wager, 2021) There is an* $\eta > 0$ *such that* $\eta \leq e_n(x) \leq 1 - \eta$ *for all* x*,*n*.*

Assumption 7 *(Assumption in Lemma 2 of Athey and Wager, 2021) The true influence scores* Γ_i *are drawn from a sequence of uniformly sub-Gaussian distribution with variance*

bounded from below,

$$\mathbb{P}_n[|\Gamma_i| > t] \leq C_\nu e^{-\nu t^2}, \textit{for all } t > 0, Var_n[\Gamma_i | \mathbf{X}_i = x] \geq s^2$$

for some constants $C_\nu, \nu, s > 0$ *and for all* n*.*

Assumption 8 *(Assumption in theorem 1 of Athey and Wager, 2021) The irreducible noise* $\epsilon_i = Y_i - m(X_i, D_i)$ *is both uniformly sub-Gaussian conditionally on* X_i *and* D_i *and has second moments uniformly bounded from below,* $Var(\epsilon_i | X_i = x, D_i = d) \geq s^2$*.*

A.1 Proof for theorem 5.1

The proof closely follows the proof of Theorem 3.1 in Mbakop and Tabord-Meehan (2021). Define $R_{n,k}(\pi) = U_n(\pi) - C_n(k) - \sqrt{\frac{k}{n}}$. By our previous construction, we have $k^* = \arg\max_k R_{n,k}(\hat{\pi}_k)$.

For every k, we can write

$$U(\pi^*) - U(\hat{\pi}_n) = (U(\pi^*) - U(\pi_k^*)) + (U(\pi_k^*) - U(\hat{\pi}_n)).$$

We also have that

$$U(\pi_k^*) - U(\hat{\pi}_n) = (U(\pi_k^*) - R_{n,k^*}(\hat{\pi}_n)) + (R_{n,k^*}(\hat{\pi}_n) - U(\hat{\pi}_n)).$$

By the definition of R_{n,k^*}, we know that $R_{n,k^*}(\hat{\pi}) \geq R_{n,k}(\hat{\pi}_k)$ for all k. Therefore we have that

$$\begin{aligned} U(\pi_k^*) - R_{n,k^*}(\hat{\pi}_n) &\leq U(\pi_k^*) - R_{n,k}(\hat{\pi}_k) \\ &= U(\pi_k^*) - U_n(\hat{\pi}_k) + C_n(k) + \sqrt{\frac{k}{n}} \\ &= W(\pi_k^*) - W_n(\hat{\pi}_k) + C_n(k) + \sqrt{\frac{k}{n}} \end{aligned} \tag{10}$$

Fix $\delta > 0$, and choose some $\tilde{\pi}_k \in \Pi_k$ such that $W(\tilde{\pi}_k) + \delta \geq W(\pi_k^*)$. We have

$$W(\pi_k^*) - W_n(\hat{\pi}_k) + C_n(k) + \sqrt{\frac{k}{n}} \leq W(\tilde{\pi}_k) + \delta - W_n(\hat{\pi}_k) + C_n(k) + \sqrt{\frac{k}{n}}.$$

Taking expectations and letting δ go to zero gives us

$$E_{P_n}[U(\pi_k^*) - R_{n,k^*}(\hat{\pi}_n)] \leq E_{P_n}[C_n(k)] + \sqrt{\frac{k}{n}}.$$

Now we look at the $R_{n,k^*}(\hat{\pi}_n) - U(\hat{\pi}_n)$ term. Note that

$$P_n(R_{n,k^*}(\hat{\pi}_n) - U(\hat{\pi}_n)) > \epsilon) \leq P_n\left(\sup_k (R_{n,k}(\hat{\pi}_k) - U(\hat{\pi}_k)) > \epsilon\right).$$

By Boole's inequality, we have that:

$$P_n\left(\sup_k (R_{n,k}(\hat{\pi}_k) - U(\hat{\pi}_k)) > \epsilon\right) \leq \sum_k P_n(R_{n,k}(\hat{\pi}_k) - U(\hat{\pi}_k) > \epsilon).$$

Combining the definition of $R_{n,k}$ and assumption 2, we have

$$\begin{aligned}\sum_k P_n(R_{n,k}(\hat{\pi}_k) - U(\hat{\pi}_k) > \epsilon) &= \sum_k P_n\left(U_n(\hat{\pi}_k) - U(\hat{\pi}_k) - C_n(k) > \epsilon + \sqrt{\frac{k}{n}}\right) \\ &\leq \sum_k c_1 e^{-2c_0 n(\epsilon + \sqrt{\frac{k}{n}})^2} \leq e^{-2c_0 n\epsilon^2} \sum_k c_1 e^{-2kc_0}\end{aligned}$$

We therefore have $P_n(R_{n,k^*}(\hat{\pi}_n) - U(\hat{\pi}_n) > \epsilon) \leq \Delta e^{-2c_0 n\epsilon^2}$ by setting $\Delta = \sum_k c_1 e^{-2kc_0} < \infty$. Following a standard integration argument (see Problem 12.1 in Devroye, Gyorfi, and Lugosi (1996)), we have $E_{P_n}[R_{n,k^*}(\hat{\pi}_n) - U(\hat{\pi}_n)] \leq \sqrt{\frac{log(\Delta e)}{2c_0 n}}$.

Combining these bounds gives us that

$$\begin{aligned} E_{P_n}[U(\pi^*) - U(\hat{\pi}_n)] &\leq E_{P_n}[C_n(k)] + U(\pi^*) - U(\pi_k^*) + \sqrt{\frac{log(\Delta e)}{2c_0 n}} + \sqrt{\frac{k}{n}} \\ &= E_{P_n}[C_n(k)] + W(\pi^*) - W(\pi_k^*) + \sqrt{\frac{log(\Delta e)}{2c_0 n}} + \sqrt{\frac{k}{n}} \\ &+ (\lambda_2 g_2(\pi_k^*) + \lambda_3 g_3(\pi_k^*) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)) \end{aligned}$$

for every k, and the result follows.

A.2 Proof for lemma 5.1

First, we define Rademacher complexity. Similar to VC-dimension, it provides a measure of complexity of a function space.

Definition 1 *(Rademacher Complexity (Mohri, Rostamizadeh, and Talwalkar, 2018))*
Let σ_i be independent Rademacher random variables, each of which which takes a value in $\{-1, +1\}$ with $\frac{1}{2}$ probability.

We define the empirical Rademacher Complexity as $\mathcal{R}_n(\Pi) = E_\sigma \left[\sup_{\pi \in \Pi} \{ \frac{1}{n} \sum_{i=1}^n \sigma_i \mathbb{1}\{\pi_i \neq R_i\} \} \right]$, *where $R_i = \mathbb{1}\{\Gamma_i \geq 0\}$. The Rademacher Complexity is the expectation of the empirical Rademacher Complexity over all samples of size n according to the same distribution:* $\mathcal{R}(\Pi) = E_{P_n}[\mathcal{R}_n(\Pi)]$.

To see why Rademacher complexity is a natural complexity measure, note that $\mathcal{R}_n(\Pi)$ characterizes the maximum in-sample classification accuracy on randomly generated labels σ_i over classifiers $\pi \in \Pi$; thus, $\mathcal{R}_n(\Pi)$ measures how much we can overfit to random coin flips using Π.

By introducing Rademacher complexity, we are able to link the policy to the classification loss function. For any population loss function $L(\pi, R)$, we have the inequality

$$L(\pi, R) \leq \hat{L}(\pi, R) + \sup_{\pi} \{L(\pi, R) - \hat{L}(\pi, R)\}$$

where $\hat{L}$ is the loss evaluated at given sample. In our case, L is the weighted classification error. That is, we have $\hat{L}(\pi, R) = \frac{1}{n}\sum_{i=1}^{n}(\Gamma_i \mathbb{1}\{\pi(X_i) \neq R_i\})$ and $L(\pi, R) = E_{P_n}[\hat{L}(\pi, R)]$.

Note that the sup part in the function above changes by at most Γ_i/n when changing one element of our sample, and Γ_i is a sub-Gaussian variable, so the condition of theorem 2.6.2 in Vershynin (2018) satisfied. It gives us that with probability $1-\delta$

$$\sup_{\pi}\{L(\pi, R) - \hat{L}(\pi, R)\} \leq E_{P_n}\left[\sup_{\pi}(L(\pi, R) - \hat{L}(\pi, R))\right] + \sqrt{\frac{\kappa}{n}\log\left(\frac{2}{\delta}\right)}$$

where κ is a constant determined by the C_ν and ν.

The symmetrization inequality from Bartlett, Bousquet, and Mendelson (2005) shows that $E_{P_n}\left[\sup_{\pi}\{L(\pi, R) - \hat{L}(\pi, R)\}\right] \leq E_{P_n}\mathcal{R}_n(L(\Pi))$, the Rademacher complexity of the loss function. Therefore, we have the following lemma:

Lemma A.1 *For loss function L defined as weighted binary classification loss evaluated at the population weighted by scores which are sub-Gaussian; for a given sample of size n, for any $\delta > 0$, with probability $1-\delta$, and for any policy $\pi \in \Pi : \mathbb{X} \to \{0, 1\}$, we have:*

$$L(\pi, R) \leq \hat{L}(\pi, R) + E_{P_n}\mathcal{R}_n(L(\Pi)) + \sqrt{\frac{\kappa}{n}\log\left(\frac{2}{\delta}\right)}$$

where κ is a constant.

Lemma A.1 shows that the loss in the population for a given policy π is bounded by the in-sample misclassification plus the Rademacher complexity of the policy class Π. It guarantees that population level welfare loss minimization is achievable when we learn the policy from a given sample. However, because the Rademacher complexity is also growing with the sample size n, in order to provide a bound for welfare loss, we need to show that the Rademacher complexity does not grow too fast.

We now introduce the growth function. For any sample with size n, let $S = \{x_1, \ldots, x_n\}$. For any classification algorithm $\pi \in \Pi$, we know π will map S to a 0/1 vector: $\pi(x_1, \ldots, x_n) \in$

$\{0,1\}^n$. The total number of dichotomies given by Π is limited by 2^n. We define the growth function $\prod_{\Pi}(n)$ as the total number of dichotomies which can be provided by Π when sample size is n, formally:

Definition 2 *(Growth function, Definition 3.6 in Mohri et al. (2018)) The growth function of a classification candidate set* Π, $\prod_{\Pi} : \mathbb{N} \to \mathbb{N}$ *is defined as:*

$$\forall n \in \mathbb{N}, \prod_{\Pi}(n) = \max_{\{x_1,\ldots,x_n\} \subset \mathcal{X}} |\{(\pi(x_1), \ldots, \pi(x_n)) : \pi \in \Pi\}|.$$

Massart's lemma links the growth function to Rademacher complexity:

Lemma A.2 *(Massart's Lemma) Let* $A \subset \mathbb{R}^n$ *be a finite set, with* $r = \max_{x \in A} ||x||_2$ *and Rademacher variables* σ_i*, we have* $\underset{\sigma}{E}\left[\frac{1}{n}\sup_{x \in A} \sum_{i=1}^n \sigma_i x_i\right] \leq \frac{r\sqrt{2\log|A|}}{n}$.

Note for any classification algorithm Π, the $\pi(S)$ set is a dichotomy set, so the size is $\prod_{\Pi}(n)$. Because we consider the weighted classification, we have $r \leq ||\Gamma||_2$. Because Γ_i is sub-Gaussian, Khintchine's inequality (as in exercise 2.6.5 in Vershynin, 2018) shows that $r \leq \zeta\sqrt{2n}$ where ζ is a constant under assumption 7. Therefore we have $\mathcal{R}_n(\Pi) \leq 2\zeta\sqrt{\frac{\log \prod_{\Pi}(n)}{n}}$ (see Corollary 3.8 in Mohri et al. (2018)).

Finally, we use Sauer's lemma to link the growth function with VC-dimension.

Lemma A.3 *(Sauer's Lemma, Theorem 3.17 in Mohri et al. (2018)) Let* Π *be a policy space with* $VC(\Pi) = d$*, then for all* $n \geq d$*, we have* $\prod_{\Pi}(n) \leq \sum_{i=0}^{d} \binom{n}{i}$

Furthermore,

$$\begin{aligned}\prod_{\Pi}(n) &\leq \sum_{i=0}^{d}\binom{n}{i} \\ &\leq \sum_{i=0}^{d}\binom{n}{i}\left(\frac{n}{d}\right)^{d-i} \\ &\leq \sum_{i=0}^{n}\binom{n}{i}\left(\frac{n}{d}\right)^{d-i} \\ &= \left(\frac{n}{d}\right)^{d}\sum_{i=0}^{n}\binom{n}{i}\left(\frac{d}{n}\right)^{i} \\ &= \left(\frac{n}{d}\right)^{d}\left(1+\frac{d}{n}\right)^{n} \leq \left(\frac{en}{d}\right)^{d},\end{aligned}$$

where e denotes Euler's number .

This yields the following corollary

Corollary 3 *For policy $\pi \in \Pi$ with VC-dimension $VC(\Pi)$, the following holds:*

$$\mathcal{R}_n(\pi) \leq 2\zeta\sqrt{\frac{VC(\Pi)\log\frac{en}{VC(\Pi)}}{n}}.$$

which shows that for any policy with limited complexity (VC-dimension), the Rademacher complexity is bounded.

Combining lemma A.1 and corollary 3, the result follows.

A.3 Proof for theorem 5.2

By definition, $\hat{\pi}_n = \hat{\pi}_{k^*}$ and $k^* = \arg\max_k \left\{U_n(\hat{\pi}_k) - C_n(k) - \sqrt{\frac{k}{n}}\right\}$. For each k,

$$U_n(\pi^*) - U_n(\hat{\pi}_n) \leq U_n(\pi^*) - U_n(\hat{\pi}_k).$$

Note we also have that $E_{P_n}[\mathbb{1}\{\hat{\pi}_k \neq \hat{\pi}^*\}|\Gamma_i|] \geq E_{P_n}[W(\hat{\pi}^*) - W(\hat{\pi}_k)]$.

By lemma 5.1, we have that $E_{P_n}[\mathbb{1}\{\hat{\pi}_k \neq \hat{\pi}^*\}|\Gamma_i|] \leq \hat{L}(\hat{\pi}_k) + 2\zeta\sqrt{\frac{VC(\Pi)\log\frac{en}{VC(\Pi)}}{n}} + \sqrt{\frac{\kappa}{n}\log\left(\frac{2}{\delta}\right)}$

with probability $1-\delta$ for every k. Because $E_{P_n}[U(\pi^*) - U(\hat{\pi}_k)] = E_{P_n}[(\underbrace{W(\pi^*) - W(\hat{\pi}^*)}_{A}) + (\underbrace{W(\hat{\pi}^*) - W(\hat{\pi}_k)}_{B})] + \lambda_2 g_2(\hat{\pi}_k) + \lambda_3 g_3(\hat{\pi}_k) - \lambda_2 g_2(\pi^*) - \lambda_3 g_3(\pi^*)$, our result follows directly.

A.4 Choice of Penalty Terms

Bartlett et al. (2002) shows that several penalty terms $C_n(k)$ can satisfy assumption 2 in the classification setting. We provide the some of their suggestions and related results here.

Hold-out Penalty

We split the sample into a training set (S_{train}) and a test set (S_{test}). We define the sample size of S_{train} as n. For every k, we estimate $\hat{\pi}_k$ using S_{train}. We can get the following empirical welfare:

$$U_{train}(\hat{\pi}_k) = W_n^{S_{train}}(\hat{\pi}_k) - \lambda_2 g_2(\hat{\pi}_k) - \lambda_3 g_3(\hat{\pi}_k)$$
$$U_{test}(\hat{\pi}_k) = W_n^{S_{test}}(\hat{\pi}_k) - \lambda_2 g_2(\hat{\pi}_k) - \lambda_3 g_3(\hat{\pi}_k).$$

We define the hold-out penalty as

$$C_n(k) = U_{train}(\hat{\pi}_k) - U_{test}(\hat{\pi}_k).$$

Recall that the optimal policy is selected by

$$\hat{\pi}_n = \arg\max_k U_{train}(\hat{\pi}_k) - C_n(k) - \sqrt{\frac{k}{n}}$$

which simplifies to

$$\hat{\pi}_n = \arg\max_k U_{test}(\hat{\pi}_k) - \sqrt{\frac{k}{n}}.$$

We want to show that the $C_n(k)$ formulated in this way satisfies assumption 2. We can see

that

$$P_n(W_n(\hat{\pi}_k) - W(\hat{\pi}_k) - C_n(k) > \epsilon) = P_n(W_n^{S_{test}}(\hat{\pi}_k) - W(\hat{\pi}_k) > \epsilon)$$

We further define $\tilde{W}_n(\pi) = \frac{1}{n}\sum \pi(X_i)\Gamma_i$, where the true score is used. Note that $W_n(\hat{\pi}_k) - W(\hat{\pi}_k) = W_n(\hat{\pi}_k) - \tilde{W}_n(\hat{\pi}_k) + \tilde{W}_n(\hat{\pi}_k) - W(\hat{\pi}_k)$ and we have

$$P_n(W_n(\hat{\pi}_k) - W(\hat{\pi}_k) - C_n(k) > \epsilon) = P_n\left(W_n^{S_{test}}(\hat{\pi}_k) - \tilde{W}_n^{S_{test}}(\hat{\pi}_k) + \tilde{W}_n^{S_{test}}(\hat{\pi}_k) - W(\hat{\pi}_k) > \epsilon\right).$$

The term $W_n^{S_{test}}(\hat{\pi}_k) - \tilde{W}_n^{S_{test}}(\hat{\pi}_k)$ is bounded following the proof of lemma 4 in Athey and Wager (2021), which shows that $W_n(\hat{\pi}_k) - \tilde{W}_n(\hat{\pi}_k)$ is sub-Gaussian conditional on n with mean zero. Because the welfare is constructed using doubly robust scores, which are sub-Gaussian by our assumption 7, the second term $\tilde{W}_n^{S_{test}}(\hat{\pi}_k) - W(\hat{\pi}_k)$ is also a sum of sub-Gaussian with mean zero. Following from the general Hoeffding inequality, the validity of assumption 2 follows.

We also check assumption 3. Note that

$$\begin{aligned} E_{P_n}[C_n(k)] &= E_{P_n}[U_{train}(\hat{\pi}_k) - U_{test}(\hat{\pi}_k)] \\ &= E_{P_n}[W_{train}(\hat{\pi}_k) - W(\hat{\pi}_k) + W(\hat{\pi}_k) - W_{test}(\hat{\pi}_k)] \end{aligned}$$

By the law of iterated expectations, we have $E_{P_n}[W(\hat{\pi}_k) - W_{test}(\hat{\pi}_k)] = 0$. The first term is bounded from lemma 4 in Athey and Wager (2021).

Maximal Discrepancy

For simplicity assume that the sample size n is even and randomly order the data and split them into two equally sized sets. Define the empirical welfare for the two parts as

$$\begin{aligned} U_n^{(1)}(\pi) &= W_n^{(1)}(\pi) - \lambda_2 g(\pi) - \lambda_3 g(\pi) \\ U_n^{(2)}(\pi) &= W_n^{(2)}(\pi) - \lambda_2 g(\pi) - \lambda_3 g(\pi) \end{aligned}$$

where the superscript denotes the evaluation sample, the first half (1) or the second half (2).

We denote the penalty term based on the negative maximal discrepancy as

$$C_n(k) = -\max_{\pi_k \in \Pi_k} U_n^{(1)}(\pi_k) - U_n^{(2)}(\pi_k) = -\max_{\pi_k \in \Pi_k} W_n^{(1)}(\pi_k) - W_n^{(2)}(\pi_k).$$

Following Bartlett et al. (2002), we propose a strategy to estimate $C_n(k)$. Denote our usual data as $D = \{(X_1, \Gamma_1), \ldots, (X_n, \Gamma_n)\}$, we construct a modified dataset $D' = \{(X_1', \Gamma_1'), \ldots, (X_n', \Gamma_n')\}$, where $(X_i', \Gamma_i') = (X_i, -\Gamma_i)$ for $i \leq \frac{n}{2}$ and $(X_i', \Gamma_i') = (X_i, \Gamma_i)$ for $i > \frac{n}{2}$.

Note that

$$\begin{aligned} W_n'(\pi) &= \frac{1}{n}\sum_{i=1}^{n} \pi(X_i)\hat{\Gamma}_i' \\ &= -\frac{1}{n}\sum_{i=1}^{n/2} \pi(X_i)\hat{\Gamma}_i + \frac{1}{n}\sum_{i=n/2+1}^{n} \pi(X_i)\hat{\Gamma}_i \\ &= \frac{1}{2}(W_n^{(2)}(\pi) - W_n^{(1)}(\pi)) \end{aligned}$$

Therefore, the π maximizing W_n' also solves the maximal discrepancy.

To see why maximal discrepancy works, consider a ghost sample D'' independent from the sample and from the same population. For each k, we have

$$\begin{aligned} & E\left[\max_{\pi_k \in \Pi_k} \{W_n''(\pi_k) - W_n(\pi_k)\}\right] \\ =& \frac{1}{n} E\left[\max_{\pi_k \in \Pi_k} (\sum_{i=1}^{n} (\pi(X_i'')\hat{\Gamma}_i'' - \pi(X_i)\hat{\Gamma}_i))\right] \\ \leq& \frac{1}{n} E\left[\max_{\pi_k \in \Pi_k} (\sum_{i=1}^{n/2} (\pi(X_i'')\hat{\Gamma}_i'' - \pi(X_i)\hat{\Gamma}_i)\right] + \max_{\pi_k \in \Pi_k} \left(\sum_{i=n/2+1}^{n} (\pi(X_i'')\hat{\Gamma}_i'' - \pi(X_i)\hat{\Gamma}_i))\right) \\ =& \frac{2}{n} E\left[\max_{\pi_k \in \Pi_k} (\sum_{i=1}^{n/2} (\pi(X_i'')\hat{\Gamma}_i'' - \pi(X_i)\hat{\Gamma}_i)\right] \\ =& E\left[\max_{\pi_k \in \Pi_k} (W_n^{(1)}(\pi_k) - W_n^{(2)}(\pi_k))\right]. \end{aligned}$$

We now check assumption 2. As shown above, following the proof of lemma 4 in Athey and Wager (2021), we only need to focus on $P_n(\Lambda + \tilde{W}_n(\hat{\pi}_k) - W(\hat{\pi}_k) - C_n(k) > \epsilon)$ where we know $\Lambda = W_n(\hat{\pi}_k) - \tilde{W}_n(\hat{\pi}_k)$ is sum of mean zero sub-Gaussian random variables. Note that

$$
\begin{aligned}
&P_n(\Lambda + \tilde{W}_n(\hat{\pi}_k) - W(\hat{\pi}_k) - C_n(k) > \epsilon) \\
=&P_n\left(\Lambda + \tilde{W}_n(\hat{\pi}_k) - W(\hat{\pi}_k) + \max_{\pi_k \in \Pi_k}\{W_n^{(1)}(\pi_k) - W_n^{(2)}(\pi_k)\} > \epsilon\right) \\
\leq&P_n\left(\Lambda + \max_{\pi_k \in \Pi_k}\{\Lambda^{(1)} - \Lambda^{(2)} + \tilde{W}_n^{(1)}(\pi_k) - \tilde{W}_n^{(2)}(\pi_k)\} - \inf_{\pi_k \in \Pi_k}\{\tilde{W}_n(\pi_k) - W(\pi_k)\} > \epsilon\right) \\
\leq&P_n\left(\Lambda + \max_{\pi_k \in \Pi_k}\{\Lambda^{(1)} - \Lambda^{(2)} + \tilde{W}_n^{(1)}(\pi_k) - \tilde{W}_n^{(2)}(\pi_k)\} - \inf_{\pi_k \in \Pi_k}\{\tilde{W}_n(\pi_k) - W(\pi_k)\}\right. \\
&\quad \left. > E\left[\max_{\pi_k \in \Pi_k}\{\Lambda^{(1)} - \Lambda^{(2)} + \tilde{W}_n^{(1)}(\pi_k) - \tilde{W}_n^{(2)}(\pi_k)\} - \inf_{\pi_k \in \Pi_k}\{\tilde{W}_n(\pi_k) - W(\pi_k)\}\right] + \epsilon\right)
\end{aligned}
$$

where the last inequality comes from the fact that the *max* term is greater or equal than zero, and the *inf* term is less or equal to zero. Note that for any sample $\{(X_i, \Gamma_i)\}$, if we change the value of Γ_i while fixing the other Γs, $\max_{\pi_k \in \Pi_k}\{\Lambda^{(1)} - \Lambda^{(2)} + \tilde{W}_n^{(1)}(\pi_k) - \tilde{W}_n^{(2)}(\pi_k)\} - \inf_{\pi_k \in \Pi_k}\{\tilde{W}_n(\pi_k) - W(\pi_k)\}$ will change at most $\frac{\Gamma_i}{n}$. We also have that Γ_i is a sub-Gaussian variable. Therefore, $\Lambda + \sup_{\pi_k \in \Pi_k}(\tilde{W}_n(\pi_k) - W(\pi_k)) - \max_{\pi_k \in \Pi_k}(W_n^{(1)}(\pi_k) - W_n^{(2)}(\pi_k))$ satisfies the condition of theorem 2.6.2 in Vershynin (2018), and the result follows.

We finally check assumption 3. Again, note that

$$
\begin{aligned}
E_{P_n}[C_n(k)] &= E_{P_n}\left[\max_{\pi_k \in \Pi_k}\{W_n^{(1)}(\pi_k) - W_n^{(2)}(\pi_k)\}\right] \\
&= E_{P_n}\left[\max_{\pi_k \in \Pi_k}\{W_n^{(1)}(\pi_k) - W(\pi_k) + W(\pi_k) - W_n^{(2)}(\pi_k)\}\right] \\
&\leq 2E_{P_n}\left[\max_{\pi_k \in \Pi_k}\{W(\pi_k) - W_n^{(1)}(\pi_k)\}\right]
\end{aligned}
$$

Apply lemma 4 and corollary 3 in Athey and Wager (2021) and the result follows.

B Variables Used to Assign Treatment in the PROGRESA Application

The variables originally used by PROGRESA to determine eligibility, as listed in Coady et al. (2013), are the following:

- Household head without education
- Household head with only primary education
- Female household head
- No refrigerator
- No washing machine
- No vehicle (car or truck)
- No toilet connected to running water
- Age of household head (years)
- Unpaved floor
- No gas heating
- Family size/rooms in the house
- Children 0 to 11 years old
- Children/working age adults

For our analysis, we focus at the individual level rather than household level. We include children's age, and exclude the "Children 0 to 11 years old" and "Children/working age adults".

At the additional measurement cost of conducting a full household survey, we assume the policy maker can access the following measures derived from the PROGRESA impact evaluation's household questionnaire.

- Child's distance from the school where they would complete their next grade
- Whether the child's household is single parent
- How many years of education the child has completed at the time of surveying
- Whether the child has completed primary school
- Whether the child has completed lower secondary school
- Whether the child has completed upper secondary school
- Household size
- Whether the household has piped water